

# CSC-PA: Cross-image Semantic Correlation via Prototype Attentions for Single-network Semi-supervised Breast Tumor Segmentation

## Supplementary Material

### 1. Confidence-based Augmentation Strategy

For unlabeled data, we design a confidence-based augmentation strategy to obtain masked images  $x^m$ , which selectively masks simple regions based on the confidence of the predictions  $f(x^u)$ . Specifically, we first divide the unlabeled images into patches of size  $N \times N$  ( $N = 57$  in our implementation) and compute the average entropy of all foreground pixels within each patch to assess the difficulty of corresponding patch  $G$ . The difficulty is calculated as follows:

$$\mathcal{I}_{i,j} = - \sum_{w=0}^{W-1} f^w(x_{i,j}^u) \log(f^w(x_{i,j}^u)), \quad (1)$$

$$\mathcal{I}_G = \frac{\sum_{(i,j) \in G} \mathcal{I}_{i,j} 1[\arg \max_w (f(x_{i,j}^u)) = 1]}{\sum_{(i,j) \in G} 1[\arg \max_w (f(x_{i,j}^u)) = 1]}, \quad (2)$$

where  $f^w(x_{i,j}^u)$  represent the classification probability of pixel  $(i, j)$  for class  $w$ , and  $W$  denotes the number of classes. Then, we randomly mask  $K_m$  patches, while excluding the  $K_h$  patches with the highest difficulty  $\mathcal{I}_G$ . The values of  $K_m$  and  $K_h$  are determined by the lesion size and shape of the input image, defined as follows:

$$K_h = \beta \frac{T-t}{T} \sum_w 1[\sum_{(i,j) \in G} (\arg \max_w (f(x_{i,j}^u))) > 0], \quad (3)$$

$$K_m = \lambda \left( \frac{HW}{N^2} - K_h \right), \quad (4)$$

where  $t$  and  $T$  denote the current epoch and the total epoch, respectively.  $H$  and  $W$  represent the length and width of the input images, respectively.  $\beta$  and  $\lambda$ , both set to 0.7, are hyper-parameters that control the number of difficult patches and masked patches, respectively. By inferring overall predictions from complex regions and limited simpler regions, the model can capture additional mutual information from challenging areas, leading to improved segmentation performance, particularly at lesion boundaries.

### 2. Additional Ablation Studies

**Impact of Selection Number.** In the edge container update process of edge prototype attention, we select  $n$  edge pixels to form the current edge set  $E_h$ , which is used to update the proposed adaptive edge container via a linear layer. As depicted in Tab. 1, we perform an ablation study on the selection number of edge features  $n$  using the UDIAT dataset

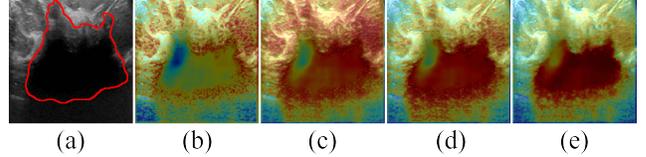


Figure 1. Visualization of partial self-correlation maps. (a) Image and GT. (b)-(e) Self-correlation maps of different channels.

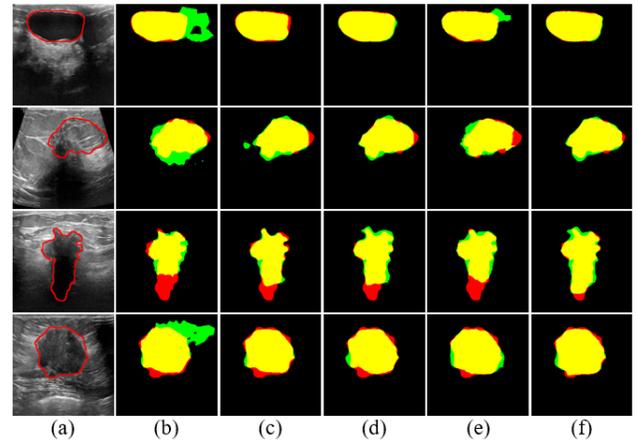


Figure 2. Visualization of component ablation on the BUSI and UDIAT test sets under 1/4 partition protocol. (a) Image and ground truth. (b) SupOnly. (c) w/FPA. (d) w/(FPA+EPA). (e) w/PAL. (f) w/(FPA+EPA+PAL). Red lines represent ground truth. Red, green and yellow represent the GT, prediction and their overlapping regions, respectively.

under the 1/4 partition protocol. The results indicate that our method achieves the highest Dice when  $n$  is set to 256. Besides, selecting an excessively large or small number of edge features impedes the construction of an effective edge container. In the former case, the insufficient number of lesion edges results in inadequate updates to the container. In the latter case, the lack of diversity in the selected features prevents the container from generating a representative edge prototype.

**Impact of Threshold  $\tau$ .** In the correlation modeling of pixel affinity loss, the threshold  $\tau$  is employed to filter out pixel features with low confidence, thereby improving the accuracy of the generated pixel affinity matrices. Tab. 2 presents the ablation study of the threshold  $\tau$ , demonstrating that our method achieves optimal performance when  $\tau = 0.80$ . In addition, a high  $\tau$  may exclude more correct pixel features during the correlation modeling process,

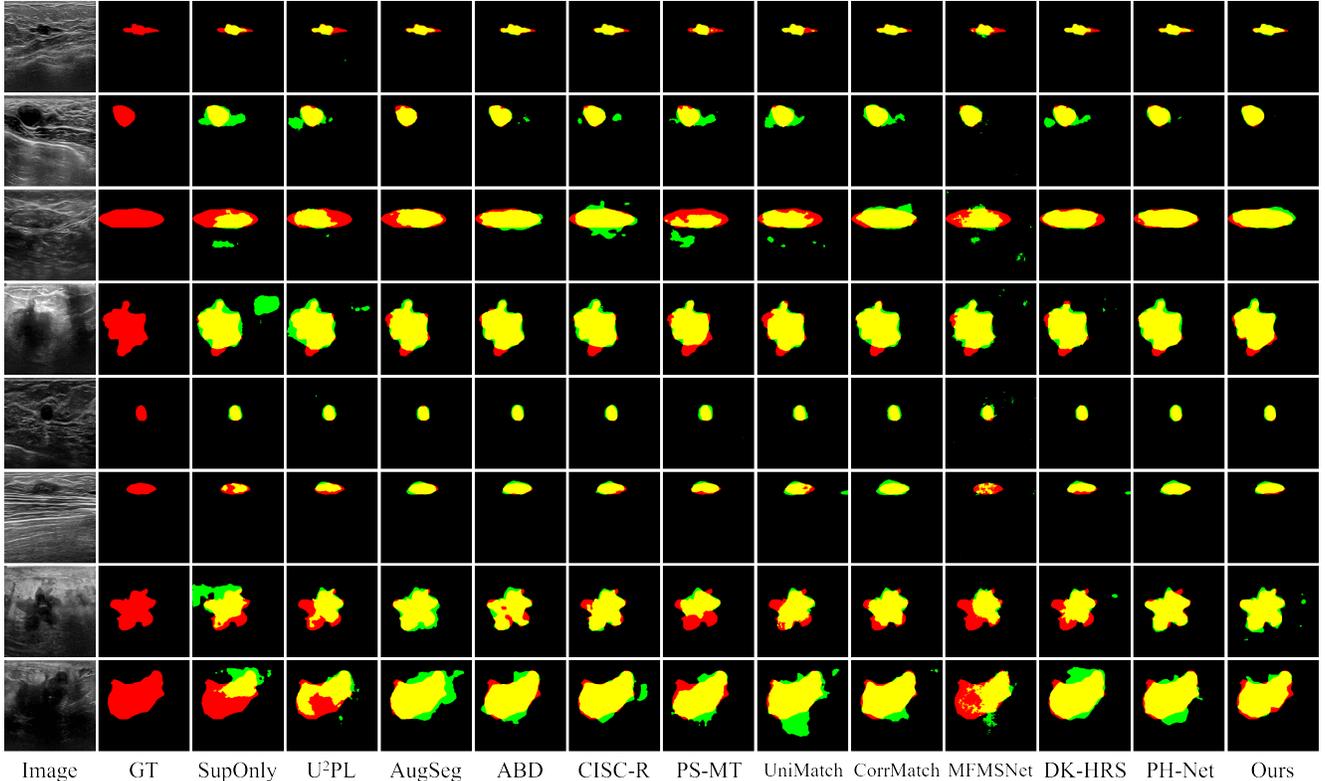


Figure 3. Additional visual comparison with different state-of-the-art methods on the BUSI and UDIAT test sets under 1/2 and 1/8 partition protocols, where the first four rows under the 1/2 split and the last four rows under the 1/8 split. Red, green and yellow represent the ground truth, prediction and their overlapping regions, respectively.

$n$	Dice $\pm$ std(%)	IoU $\pm$ std(%)
4	79.47 $\pm$ 2.91	69.22 $\pm$ 3.20
16	79.84 $\pm$ 3.04	69.74 $\pm$ 2.81
64	80.32 $\pm$ 1.81	70.21 $\pm$ 2.08
256	<b>81.26<math>\pm</math>1.75</b>	<b>71.65<math>\pm</math>1.89</b>
1024	79.96 $\pm$ 2.58	70.34 $\pm$ 2.47

Table 1. Ablation study of different selection number  $n$  on UDIAT test set under 1/4 partition protocol.

leading to an affinity matrix that inadequately captures the semantic correlations between pixels and their surrounding pixels. Conversely, a low  $\tau$  may cause the affinity matrix to be influenced by incorrect pixel features, resulting in degraded performance of the segmentation model.

### 3. More Experimental Results

**More Visualization Results.** In Fig. 1, we present partial self-correlation maps to provide a more intuitive understanding. In addition, as shown in Fig. 2, to further demonstrate the effectiveness of each component in our proposed CSC-PA framework, we visualize the prediction results of component ablation on the BUSI [1] dataset with 129 (1/4)

$\tau$	Dice $\pm$ std(%)	IoU $\pm$ std(%)
0.75	78.97 $\pm$ 2.27	68.87 $\pm$ 2.92
0.80	<b>81.26<math>\pm</math>1.75</b>	<b>71.65<math>\pm</math>1.89</b>
0.85	80.25 $\pm$ 1.62	70.48 $\pm$ 2.30
0.90	79.57 $\pm$ 1.76	69.61 $\pm$ 1.69

Table 2. Ablation study of different threshold  $\tau$  on UDIAT test set under 1/4 partition protocol.

labels and the UDIAT [11] dataset with 33 (1/4) labels.

**More Visual Comparisons.** We provide additional visual comparisons with SupOnly and other ten state-of-the-art methods under 1/2 and 1/8 partition protocols. Among them, U<sup>2</sup>PL [6], AugSeg [12] and ABD [2] are implemented on the mean-teacher (MT) framework. CISC-R [7] is a cross-image method based on contrastive learning. PS-MT [4] introduces an auxiliary teacher on the MT architecture, while UniMatch [10] and CorrMatch [5] are both single network. In addition, MFMSNet [8] is a fully-supervised method for BUS images. DK-HRS [9] and PH-Net [3] are semi-supervised methods for BUS data. As demonstrate in Fig. 3, the first four rows represent the predictions under the 1/2 split, while the last four rows show the predictions under the 1/8 split.

## References

- [1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020. [2](#)
- [2] Hanyang Chi, Jian Pang, Bingfeng Zhang, and Weifeng Liu. Adaptive bidirectional displacement for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4070–4080, 2024. [2](#)
- [3] Siyao Jiang, Huisi Wu, Junyang Chen, Qin Zhang, and Jing Qin. Ph-net: Semi-supervised breast lesion segmentation via patch-wise hardness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11418–11427, 2024. [2](#)
- [4] Yuyuan Liu, Yu Tian, Yuanhong Chen, Fengbei Liu, Vasileios Belagiannis, and Gustavo Carneiro. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4258–4267, 2022. [2](#)
- [5] Boyuan Sun, Yuqi Yang, Le Zhang, Ming-Ming Cheng, and Qibin Hou. Corrmatch: Label propagation via correlation matching for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3097–3107, 2024. [2](#)
- [6] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4248–4257, 2022. [2](#)
- [7] Linshan Wu, Leyuan Fang, Xingxin He, Min He, Jiayi Ma, and Zhun Zhong. Querying labeled for unlabeled: Cross-image semantic consistency guided semi-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8827–8844, 2023. [2](#)
- [8] Ruichao Wu, Xiangyu Lu, Zihuan Yao, and Yide Ma. Mfm-snet: A multi-frequency and multi-scale interactive cnn-transformer hybrid network for breast ultrasound image segmentation. *Computers in Biology and Medicine*, 177: 108616, 2024. [2](#)
- [9] Xiaozheng Xie, Jianwei Niu, Xuefeng Liu, Yong Wang, Qingfeng Li, and Shaojie Tang. A domain knowledge powered hybrid regularization strategy for semi-supervised breast cancer diagnosis. *Expert Systems with Applications*, 243:122897, 2024. [2](#)
- [10] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7236–7246, 2023. [2](#)
- [11] Moi Hoon Yap, Manu Goyal, Fatima Osman, Robert Martí, Erika Denton, Arne Juette, and Reyer Zwiggelaar. Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine*, 107:101880, 2020. [2](#)
- [12] Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, and Jingdong Wang. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11350–11359, 2023. [2](#)