

Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation

Hanyang Chi¹ Jian Pang¹ Bingfeng Zhang^{1*} Weifeng Liu¹

¹China University of Petroleum (East China)

{chihanyang, jianpang}@s.upc.edu.cn, {bingfeng.zhang, liuwf}@upc.edu.cn

1. More Analysis

1.1. Different Displacement Strategies

The original Tab. 4 shows the various displacement strategies in the ABD-R module. The “Random” yield the worst results due to the uncertainty of this random displacement greatly increases the uncontrollability. While “Reliable” exhibits a 0.2% improvement in DSC compared to the “Same” strategy, it theoretically effectively removes uncontrolled regions and displaces them with semantically similar information. Here, uncontrolled regions denote areas with relatively low confidence in predictions. We use predictions to evaluate whether the region is suitable to learn for the model. Although there are limited differences among the various displacement strategies in Tab. 4, our primary objective in the ABD-R module is to suppress uncontrollable regions to enhance model stability. It is only when the ABD-R and ABD-I modules are combined that their full potential can be realized.

2. More Experimental Results

2.1. The Influence of Mixed Perturbation

Fig. 1 indicates that involving additional input perturbation to Cross Pseudo Supervision (CPS) [2] may lead to misclassifications, including erroneously recognizing background as foreground and incorrectly identifying foreground as background.

2.2. Detailed Comparison on ACDC

In Tab. 1, we show the detailed comparison between our approach and Cross teaching on ACDC dataset. It can be seen that our approach, *i.e.*, **Ours**-ABD (Cross Teaching [3]) significantly improves the performance of the baseline (Cross Teaching). After introducing our approach to the more powerful approach BCP [1], it reaches a new state-of-the-art result.

*Corresponding author.

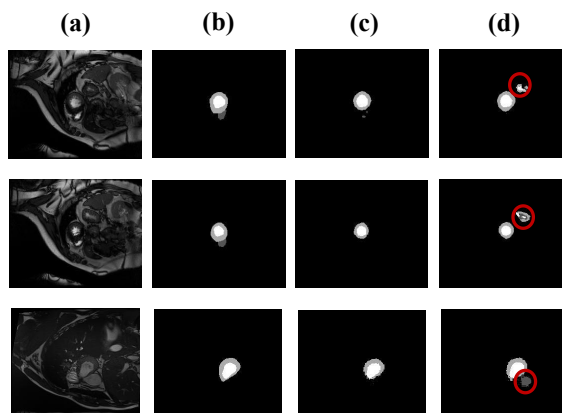


Figure 1. (a) Inputs; (b) Labels; (c) CPS prediction results; (d) Prediction results after adding input perturbation. The first two lines represent the prediction results of $f(\theta_1)$, while the last line represents the prediction results of $f(\theta_2)$. The areas highlighted by the red circles indicate regions of misclassification.

2.3. Visualization of Adaptive Bidirectional Displacement

Fig. 2 shows the generated samples of the Adaptive Bidirectional Displacement with Reliable Confidence (ABD-R) under different iterations. It can be observed that, in different iterations, ABD-R replaces the lowest confident regions with the final regions selected with high confidence sourced from another augmented image. As a result, the generated samples have compensated semantics, ensuring the reliability of consistency learning. Fig. 3 visualizes the generated samples of the Adaptive Bidirectional Displacement with Inverse Confidence (ABD-I) under different iterations. Fig. 3 indicates that the newly generated samples contain more low confident semantics information compared to the original augmented image, those samples can empower the model to effectively learn from the potentially uncontrollable regions.

Table 1. Detailed Comparison on ACDC dataset

Method	Scans used		RV		Myo		LV		Avg	
	Labeled	Unlabeled	DSC \uparrow	95HD \downarrow	DSC \uparrow	95HD \downarrow	DSC \uparrow	95HD \downarrow	DSC \uparrow	95HD \downarrow
Cross Teaching			57.70	21.4	62.80	11.5	76.30	15.7	65.60	16.2
Ours-ABD (Cross Teaching)	3(5%)	67(95%)	84.67	2.67	83.65	3.59	90.73	6.11	86.35	4.12
Ours-ABD (BCP)			88.69	1.46	85.78	1.99	92.42	1.27	88.96	1.57
Cross Teaching			84.80	7.80	84.40	6.90	90.10	11.20	86.40	8.60
Ours-ABD (Cross Teaching)	7(10%)	63(90%)	87.87	5.42	85.92	3.12	91.83	6.64	88.52	5.06
Ours-ABD (BCP)			89.47	1.27	86.70	1.22	93.25	1.88	89.81	1.46

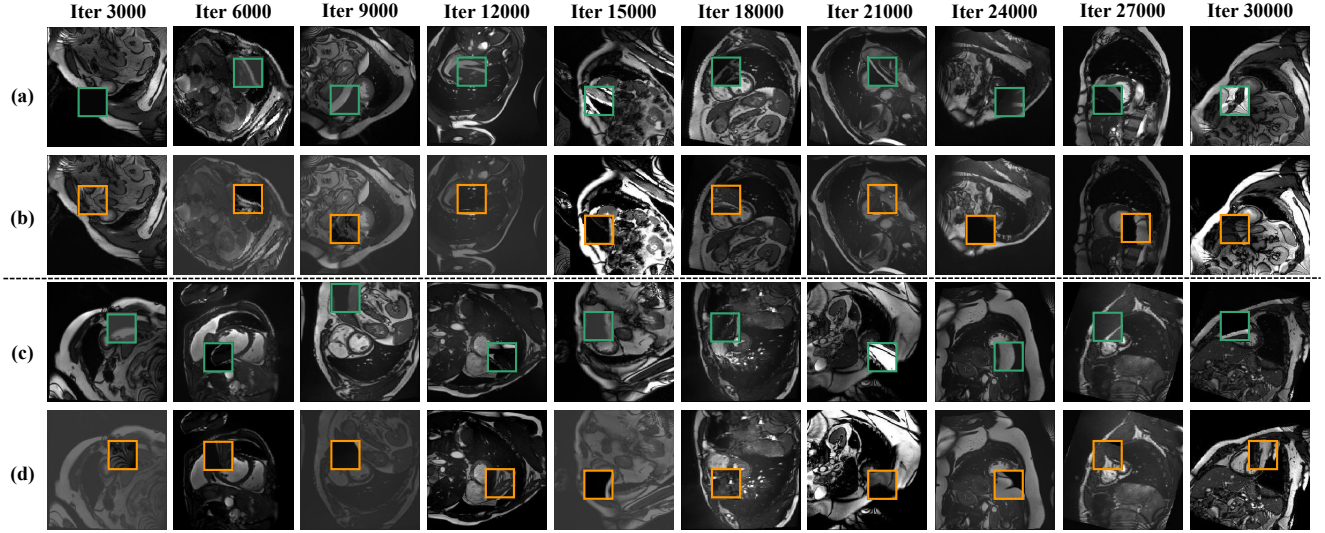


Figure 2. Visualization of generated samples from ABD-R with unlabeled data. The generated sample pairs (a) and (b) are specifically derived from weakly and strongly augmented images. The green box in (a) represents the final selected semantic derived from (b). The orange box in (b) represents the final selected semantic derived from (a). Samples in (c) and (d) are generated in the same way as above.

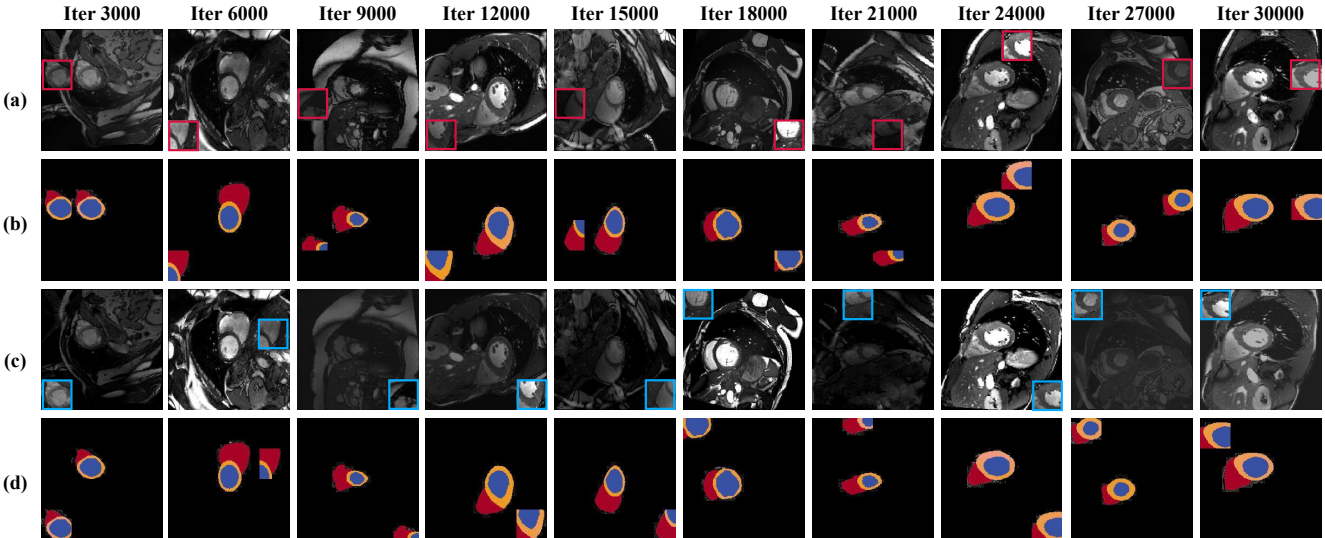


Figure 3. Visualization of generated samples and labels from ABD-I with labeled data. (a) and (c) are the generated samples based on weakly and strongly augmented images. The red box in (a) represents the lowest confidence semantic derived from (c). (b) The generated labels correspond to the samples from (a). The blue box in (c) represents the lowest confidence semantic derived from (a). (d) The generated labels correspond to the samples from (c)

Table 2. Ablation study of patch number K with 5% labeled data.

K	DSC \uparrow	Jaccard \uparrow	95HD \downarrow	ASD \downarrow
4	85.31	75.35	4.32	1.47
16	86.35	76.73	4.12	1.22
64	85.56	75.64	4.50	1.35

Table 3. Ablation study of patch number K on PROMISE12.

K	DSC \uparrow	Jaccard \uparrow	95HD \downarrow	ASD \downarrow
4	78.20	64.58	3.81	1.63
16	82.06	69.94	3.02	1.33
64	78.85	65.45	5.59	2.25

Table 4. Ablation study of top- n highest confidence patches with 10% labeled data.

n	DSC \uparrow	Jaccard \uparrow	95HD \downarrow	ASD \downarrow
2	87.90	79.13	5.53	1.70
4	88.52	79.97	5.06	1.43
8	87.94	79.10	6.32	1.80

2.4. Selection of Patch Number on ACDC and PROMISE12

Tab. 2 demonstrates how different patch numbers impact the performance of the model on ACDC when only 5% labeled data is used. We also add the ablation study of the patch number on PROMISE12 dataset in Tab. 3. K represents the number of patches in the image. It can be observed that selecting $K = 16$ is the most appropriate choice in both cases, as both excessively large and small values of K do not contribute significantly to learn meaningful semantics. If K is too small, e.g., K is 2, suppose the 2nd region will be replaced by the 1st region from the other image, it will produce a new sample formed by two “1st region”, totally lost information in the 2nd region. The opposite situation occurs if K is too large.

2.5. Selection of Top- n Highest Confidence Patches on ACDC

Tab. 4 demonstrates how different numbers of top highest confidence patches impact the performance of the model when only 10% labeled data is used. n represents the top n highest confidence levels in the image. It can be observed that $n = 4$ generates better performance than others. When n is too small, the final regions selected with high confidence may not closely align with the original semantics. Conversely, if n is too large, it may result in the final selection of regions with relatively low confidence, leading to uncontrollable outcomes.

References

- [1] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11514–11524, 2023. 1
- [2] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 1
- [3] Xiangde Luo, Minhao Hu, Tao Song, Guotai Wang, and Shaoting Zhang. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In *International Conference on Medical Imaging with Deep Learning*, pages 820–833. PMLR, 2022. 1