MOST: Multi-Formation Soft Masking for Semi-Supervised Medical Image Segmentation (Supplementary Materials)

Xinyu Liu[1,*], Zhen Chen[2,*], and Yixuan Yuan[1,✉]

1 Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong SAR
2 Centre for Artificial Intelligence and Robotics (CAIR), Hong Kong Institute of Science and Innovation, Chinese Academy of Sciences, Beijing, China
∗ Equal Contribution
yxyuan@ee.cuhk.edu.hk

A Detailed Results on BraTS 2019

In Sec. 3.2 in the main paper, we present the Dice of MOST on BraTS 2019. In this section, we present the detailed performance comparison in Table 1, MOST gives the best Dice (84.17%), Jaccard (74.24%), and 95HD (11.02), indicating the potential of our proposed method on various data modalities.

B Sensitivity to the Pseudo Label Threshold

The value τ determines the threshold at which a segmentation result is considered confident enough to be used as a pseudo label for training. We conduct experiments on LA dataset using a comprehensive range of thresholds and list the results in Table 2. As the threshold increases from 0.60 to 0.75, there is a consistent improvement. This indicates that our method performs better when selecting confident predictions as pseudo labels. When the value increases to larger than 0.75, the performance starts to decline. Therefore, MOST is robust to the choice of τ, and we select τ=0.75 to maximize the overall performance while avoiding excessive false positives or negatives in the pseudo labels.

Table 1. Detailed results on BraTS 2019 with 10% data labeled.

Method	Metrics			
	Dice↑	Jac↑	95HD↓	ASD↓
Supervised	83.84	74.79	8.32	2.13
DTC [1]	81.75	71.63	15.73	2.56
URPC [2]	82.59	72.11	13.88	3.72
CPCL [6]	83.36	73.23	11.74	1.99
AC-MT [5]	83.77	73.96	11.37	**1.93**
MOST	**84.17**	**74.24**	**11.02**	2.83

Table 2. Sensitivity to the pseudo label threshold τ.

τ	Dice↑	Jac↑	95HD↓	ASD↓
0.60	90.64	83.06	5.73	1.71
0.65	91.04	83.65	5.15	**1.54**
0.70	90.73	83.11	5.51	1.64
0.75	**91.17**	**83.85**	5.63	1.76
0.80	90.84	83.31	5.32	1.64
0.85	90.59	82.98	5.41	1.71
0.90	91.12	83.85	**4.88**	1.63
0.95	90.94	83.48	5.28	1.65

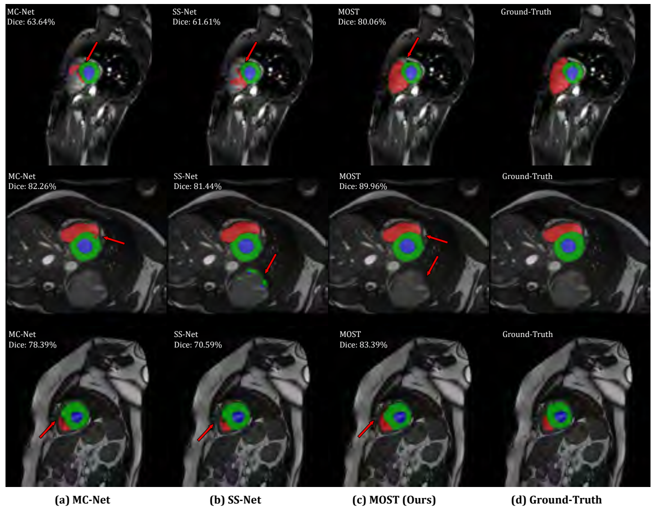

Fig. 1. Segmentation performance comparison on 2D ACDC dataset among (a) MC-Net [3], (b) SS-Net [4], (c) the proposed MOST, (d) ground-truth.

C Qualitative Comparison on 2D Dataset ACDC

Fig. 1 shows some qualitative results on the multi-class 2D dataset ACDC. MOST not only identifies the region of each class precisely (3^{rd} row), but also differentiates the boundary between classes accurately (1^{st} and 2^{nd} rows), which further demonstrate the effectiveness of MOST in multi-class segmentation.

References

1. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: AAAI. vol. 35, pp. 8801–8809 (2021)
2. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Med. Image Anal. **80**, 102517 (2022)
3. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. Med. Image Anal. **81**, 102530 (2022)
4. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: MICCAI. pp. 34–43. Springer (2022)
5. Xu, Z., Wang, Y., Lu, D., Luo, X., Yan, J., Zheng, Y., Tong, R.K.y.: Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. Med. Image Anal. **88**, 102880 (2023)
6. Xu, Z., Wang, Y., Lu, D., Yu, L., Yan, J., Luo, J., Ma, K., Zheng, Y., Tong, R.K.y.: All-around real label supervision: Cyclic prototype consistency learning for semi-supervised medical image segmentation. IEEE J. Biomed Health Inform. **26**(7), 3174–3184 (2022)