DualRel: Semi-Supervised Mitochondria Segmentation from A Prototype Perspective

Supplementary Material

Huayu Mai[1*] Rui Sun[1*] Tianzhu Zhang[1,2,3†] Zhiwei Xiong[1,2] Feng Wu[1,2]
[1]University of Science and Technology of China
[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
[3]Deep Space Exploration Lab
{mai556, issunrui}@mail.ustc.edu.cn, {tzzhang, zwxiong, fengwu}@ustc.edu.cn

In the supplementary material, we first introduce more details about the calculation of confusion density (Sec. 1) and the evaluation metrics (Sec. 4.1). Then we clarify how to select reference points for referential correlation in the reliable pixel aggregation module (Sec. 3.2). Furthermore, we perform cross-domain transfer experiments to quantify the effectiveness of DualRel. Finally, we show more qualitative results including the reliability of prototypes (Sec. 4.5) and activation maps.

1. More on Confusion Density and Metrics

In this section, we provide more details about the calculation of confusion density (Fig. 1 in Sec. 1) and the evaluation metrics (Sec. 4.1) including DSC and JAC.

1.1. Calculation of Confusion Density

To further demonstrate the gap between mitochondrial images and natural images as shown in Fig. 1, we separately compute the confusion density of these two types of images derived from the same method (*i.e.*, CPS [2]), which employs pixel-level consistency regularization. In specific, we denote the predicted segmentation mask as $\widetilde{\mathbf{Y}}$, where each position (i, j) contains a pair of foreground-background probability, that is, $\widetilde{\mathbf{Y}}_{i,j} = (p_f, p_b)$. We define confusion ($cf$) as the inverse confidence of the prediction with respect to the ground truth $\mathbf{Y} \in \{0, 1\}^{H \times W}$,

$$cf_{i,j} = \begin{cases} 1 - p_b, & \text{if } \mathbf{Y}_{i,j} = 0 \\ 1 - p_f, & \text{if } \mathbf{Y}_{i,j} = 1 \end{cases}, \tag{1}$$

where $i = 1, 2, ..., H, j = 1, 2, ..., W$. In this way, we can obtain the confusion density ρ_{cf} which represents the expected confusion per pixel, formulated as

$$\rho_{cf} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} cf_{i,j}. \tag{2}$$

1.2. Calculation of Evaluation Metrics

To evaluate the accuracy of segmentation in our experiments, we adopt Dice similarity coefficient (DSC) and Jaccard-index coefficient (JAC), as described in Sec 4.1. Metric formulations are as follows:

$$\text{JAC} = \frac{|\widehat{\mathbf{Y}} \cap \mathbf{Y}|}{|\widehat{\mathbf{Y}} \cup \mathbf{Y}|} \times 100\%, \tag{3}$$

$$\text{DSC} = \frac{2 \times |\widehat{\mathbf{Y}} \cap \mathbf{Y}|}{|\widehat{\mathbf{Y}}| + |\mathbf{Y}|} \times 100\%, \tag{4}$$

where $\widehat{\mathbf{Y}}$ is the hard prediction of the network given an image, $\mathbf{Y}$ is the corresponding ground truth.

2. Selection of Reference Points

In this section, we describe in detail how to obtain reference points in Sec. 3.2, aiming to establish the referential correlation for rectifying the direct pairwise correlation. Inspired by human behavior which moves frequently accessed or studied patterns to a reliable reference event store for comparison with newly seen things, we pick reliable reference points from feature sequence $\widetilde{\mathbf{X}} \in \mathbb{R}^{hw \times C}$ with high *usage*. In specific, we define and calculate the total contribution u_i of each pixel with regard to all the K prototypes, formulated as

$$u_i = \sum_{k=1}^{K} s_{k,i}, \tag{5}$$

where $s_{k,i}$ denotes the pairwise correlation between the k-th prototype and the i-th pixel. Then we select the top-N

*Equal contribution
†Corresponding author

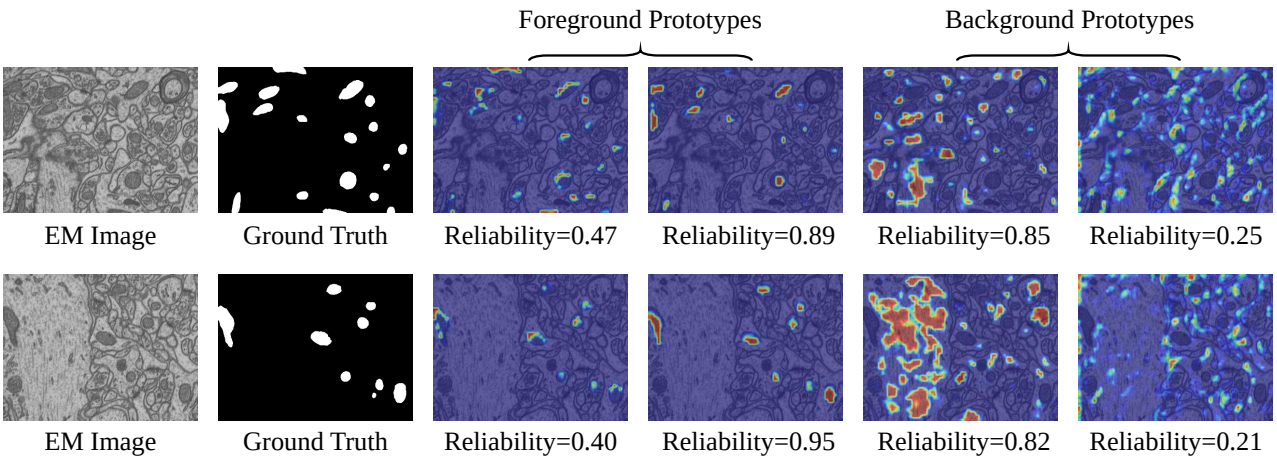

Figure 7. The visualization of the reliability of prototypes in constructing prototype-level consistency regularization. Larger weights are assigned to the more reliable prototypes.

Table 6. Cross-domain quantitative results of state-of-the-art methods **trained on Mito-R** [7] but tested on Lucchi [4] and Lucchi++ [1] dataset under different partition protocols. The fractions denote the percentage of labeled data used for training, followed by the actual number of mitochondrial images.

	1/32 (12)		1/16 (25)		1/8 (50)		1/2 (200)	
Method	JAC	DSC	JAC	DSC	JAC	DSC	JAC	DSC
Lucchi								
MT[NIPS17] [6]	41.37	57.21	42.84	58.66	49.44	65.35	58.25	71.87
CCT[CVPR20] [5]	43.81	60.01	45.54	62.40	51.72	67.95	64.97	78.12
GCT[ECCV20] [3]	46.53	62.62	48.41	65.52	54.97	71.31	68.04	81.92
CPS[CVPR21] [2]	49.62	64.89	51.59	67.48	59.52	73.78	73.32	84.40
DualRel	**56.62**	**70.59**	**58.87**	**73.41**	**67.91**	**80.26**	**79.37**	**88.29**
Lucchi++								
MT[NIPS17] [6]	38.44	53.15	39.98	54.46	46.91	60.72	55.16	66.78
CCT[CVPR20] [5]	40.00	54.84	42.82	56.97	48.31	62.04	61.22	71.92
GCT[ECCV20] [3]	43.85	59.01	45.83	61.84	51.25	67.49	65.77	77.20
CPS[CVPR21] [2]	46.38	61.84	47.32	63.88	56.47	71.53	67.05	80.07
DualRel	**53.92**	**68.11**	**55.01**	**70.36**	**65.64**	**78.78**	**72.93**	**84.16**

Table 7. Cross-domain quantitative results of state-of-the-art methods **trained on Mito-H** [7] but tested on Lucchi [4] and Lucchi++ [1] dataset under different partition protocols. The fractions denote the percentage of labeled data used for training, followed by the actual number of mitochondrial images.

	1/32 (12)		1/16 (25)		1/8 (50)		1/2 (200)	
Method	JAC	DSC	JAC	DSC	JAC	DSC	JAC	DSC
Lucchi								
MT[NIPS17] [6]	29.94	46.30	30.52	46.99	33.44	49.73	43.59	60.32
CCT[CVPR20] [5]	32.39	48.72	33.10	49.51	36.12	52.32	45.60	62.16
GCT[ECCV20] [3]	42.08	58.62	43.14	59.57	47.07	62.94	55.10	66.57
CPS[CVPR21] [2]	44.87	60.74	46.00	61.73	50.20	65.23	58.75	72.62
DualRel	**51.27**	**67.43**	**52.57**	**68.53**	**57.36**	**72.41**	**62.13**	**76.36**
Lucchi++								
MT[NIPS17] [6]	35.85	52.92	37.52	53.46	38.34	54.67	48.87	65.13
CCT[CVPR20] [5]	35.30	52.96	36.95	53.84	37.75	54.71	49.10	65.48
GCT[ECCV20] [3]	43.97	60.45	45.18	61.55	45.97	62.39	49.91	65.78
CPS[CVPR21] [2]	48.65	65.66	50.44	66.59	52.14	67.12	59.49	72.18
DualRel	**55.93**	**71.33**	**57.48**	**72.63**	**58.26**	**73.15**	**63.74**	**77.56**

pixels with the largest contribution (*i.e.*, usage) as reference points $\widetilde{\mathbf{X}}^R \in \mathbb{R}^{N \times C}$.

3. Cross-Domain Transfer Experiments

In this section, we perform cross-domain transfer experiments where the model is trained on Mito-R or Mito-H [7] dataset, but is tested on Lucchi [4] and Lucchi++ [1] dataset. This experimental setting enables greater challenge with a large domain gap to avoid evaluation overfitting caused by similar slices of the same domain (*i.e.*, the same dataset), and enables better measurement of the knowledge transfer ability of different methods.

As shown in Tab. 6, our DualRel demonstrates the superiority over other methods and shows absolute performance gains of 8.39%/6.48% in JAC/DSC under 1/8 partition, and 6.05%/3.89% in JAC/DSC under 1/2 partition over the best method (*i.e.*, CPS [2]). Similarly, Tab. 7 shows that our method consistently surpasses all the competitors under all partition protocols (*e.g.*, 7.54%/6.27% in JAC/DSC under 1/32 partition). This indicates that our method, compared to other methods directly employing pixel-level supervision, can effectively absorb reliable pixels which provide more transferable semantic information across different domains by constructing robust prototype-level consistency regularization. Furthermore, the reliability-aware consistency loss in reliable prototype selection module allows to customize the reliability of prototypes for any image from the new domain. Therefore, our model can alleviate the domain gap and have well cross-domain generalization.

4. More Qualitative Results

In this section, we show more qualitative results including the reliability of the prototypes corresponding to Fig. 5, and the diverse activation maps.

4.1. More Visualization of Reliability

To evaluate the effect of the RPrS, we visualize the reliability of more prototypes as a complement to Fig. 5. We can see that the different prototypes focus on significant distinct areas. Those prototypes whose activation regions are concentrated at the boundary have a lower reliability, which are in line with the core intention of our design. Besides, the high consistency between the reliability of prototypes and the corresponding activation regions indicates the effectiveness of our paradigm of implicitly learn the reliability in a data-driven way.

4.2. Visualization of Activation Maps

Fig. 8 visualizes the diverse activation maps, which are successfully partitioned by prototypes into different semantic patterns in an adaptive manner. For example, the 2^{nd} activation map highlighted by the *background* prototype (in the second row) mainly focuses on the boundaries while the 3^{rd} activation map mainly concentrates on the parts of the background. This proves that our diversity loss can prevent the prototypes from focusing on similar local semantic clues. In this way, diverse prototypes can capture mitochondria variations to further evaluate the reliability of prototypes in constructing prototype-level consistency regularization, and fuse with each other for more precise segmentation.

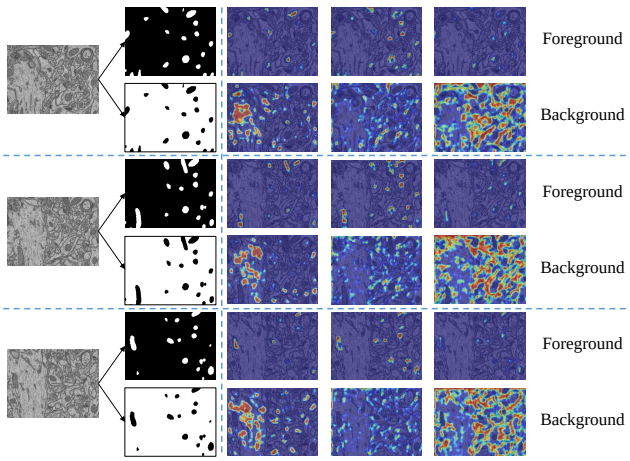

Figure 8. Visualization of the diverse activation maps highlighted by prototypes. As we can see, these prototypes mainly focus on specific mitochondrial semantic cues, such as boundaries, parts of the foreground, and parts of the background.

References

[1] Vincent Casser, Kai Kang, Hanspeter Pfister, and Daniel Haehn. Fast mitochondria detection for connectomics. In *Medical Imaging with Deep Learning*, pages 111–120. PMLR, 2020. 2

[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, 2

[3] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise semi-supervised learning. In *European conference on computer vision*, pages 429–445. Springer, 2020. 2

[4] Aurélien Lucchi, Kevin Smith, Radhakrishna Achanta, Graham Knott, and Pascal Fua. Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. *IEEE transactions on medical imaging*, 31(2):474–486, 2011. 2

[5] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 2

[6] 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*, 30, 2017. 2

[7] Donglai Wei, Zudi Lin, Daniel Franco-Barranco, Nils Wendt, Xingyu Liu, Wenjie Yin, Xin Huang, Aarush Gupta, Won-Dong Jang, Xueying Wang, et al. Mitoem dataset: large-scale 3d mitochondria instance segmentation from em images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 66–76. Springer, 2020. 2