Minding Fuzzy Regions: A Data-driven Alternating Learning Paradigm for Stable Lesion Segmentation

Supplementary Material

1. Details of Compared Methods

- BECO [4] learns all possible noisy data in a collaborative training manner. The experiments we conducted retain only non-fuzzy labels, creating pseudo-labels for fuzzy regions as [4].
- GCE [6] leverages a robust loss function that combines cross-entropy (CE) loss with mean absolute error (MAE) loss to handle label noise.
- MW-Net [5] employs an MLP network to learn a weighting function for noisy data.
- CIRL [1] utilizes a clustering-inspired loss function that transforms regions with label noise from supervised learning to unsupervised clustering.

2. Supplementary Experimental Results

2.1. Experiments on Organ Segmentation

We demonstrate DALE's performance on the organ segmentation task in the table below. It shows that DALE brings significant improvements to organ segmentation for various models, which DALE's applicability. We demonstrate DALE's performance on the organ segmentation task in the table below. The results show that DALE significantly enhances organ segmentation across various models, achieving an average Dice score improvement of 2.27%. This illustrates DALE's effectiveness and broad applicability in medical image segmentation.

Method	Backbone	Synapse										Growth Rate
		Dice↑	95HD↓	Aorta↑	GB↑	KL↑	KR↑	Liver↑	PC↑	SP↑	SM↑	
EMCAD	CNN-based	0.7507	46.0207	0.8403	0.5988	0.6852	0.7729	0.9358	0.5588	0.8603	0.7537	+2.02%
***EMCAD**		**0.7659**	**38.0397**	**0.8663**	**0.6066**	**0.7804**	**0.7244**	**0.9403**	**0.5955**	**0.8516**	**0.7622**	
Convformer	CNN-Transformer hybrid	0.7666	38.1309	0.8714	0.6158	0.7961	0.7629	0.9430	0.5231	0.8574	0.7628	+1.43%
***Convformer**		**0.7775**	**31.1414**	**0.8696**	**0.6382**	**0.8255**	**0.7735**	**0.9349**	**0.5705**	**0.8477**	**0.7600**	
TransFuse		0.7748	31.6900	0.8723	0.6313	0.8187	0.7702	0.9408	0.5586	0.8508	0.7562	+2.40%
***TransFuse**		**0.7934**	**29.9120**	**0.8852**	**0.6845**	**0.7990**	**0.7864**	**0.9498**	**0.5998**	**0.8700**	**0.7723**	
Xboundformer		0.7649	34.7228	0.8622	0.6262	0.7561	0.7114	0.9345	0.5695	0.8676	0.7919	+2.81%
***Xboundformer**		**0.7864**	**30.7898**	**0.8589**	**0.6289**	**0.8242**	**0.7690**	**0.9386**	**0.6038**	**0.8679**	**0.7998**	
SwinUmamba	Mamba	0.7485	42.1225	0.8513	0.5903	0.7625	0.7006	**0.9257**	0.5538	0.8363	**0.7675**	+2.70%
***SwinUmamba**		**0.7687**	**32.7155**	**0.8699**	**0.6180**	**0.8032**	**0.7532**	0.9218	**0.5727**	**0.8581**	0.7527	
Average Growth Rate		+2.27%										

Table 1. Evaluation of the proposed DALE on various advanced models using Synapse datasets.

2.2. Experiments on SAM-Based Models

We conducted experiments on SAM-based models, including SAM2 [3], MedSAM-2 [7], and InstaSAM [2], using skin disease datasets. As shown in Table 2, DALE significantly improves the segmentation performance of these models, achieving up to a 4.5% increase in accuracy for skin disease segmentation.

Model	Model Type	ISIC2016&PH2				Growth Rate
		Dice↑	mIoU↑	95HD↓	ASD↓	
SAM2	SAM-based	0.8514	0.7563	9.6231	1.7151	+4.50%
***SAM2**		**0.8897**	**0.8091**	**5.7227**	**0.8856**	
MedSAM-2		0.8856	0.8031	5.3054	0.8379	+1.52%
***MedSAM-2**		**0.8991**	**0.8248**	**4.2772**	**0.7085**	
InstaSAM		0.8961	0.8195	5.0288	0.8121	+2.68%
***InstaSAM**		**0.9201**	**0.8566**	**2.9673**	**0.4623**	

Table 2. DALE on ISIC2016 & PH2 (*: Trained with DALE).

References

[1] Zhuangzhuang Chen, Zhuonan Lai, Jie Chen, and Jianqiang Li. Mind marginal non-crack regions: Clustering-inspired representation learning for crack segmentation. In *CVPR*, pages 12698–12708, 2024. 1

[2] Siwoo Nam, Hyun Namgung, Jaehoon Jeong, et al. Instasam: Instance-aware segment any nuclei model with point annotations. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 232–242, 2024. 1

[3] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, et al. Sam 2: Segment anything in images and videos, 2024. arXiv preprint arXiv:2408.00714. 1

[4] Shenghai Rong, Bohai Tu, Zilei Wang, and Junjie Li. Boundary-enhanced co-training for weakly supervised semantic segmentation. In *CVPR*, pages 19574–19584, 2023. 1

[5] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. *NeurIPS*, 32, 2019. 1

[6] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *NeurIPS*, 31, 2018. 1

[7] Jiayuan Zhu, Yunli Qi, and Junde Wu. Medical sam 2: Segment medical images as video via segment anything model 2, 2024. arXiv preprint arXiv:2408.00874. 1