S^{2}ME: Spatial-Spectral Mutual Teaching and Ensemble Learning for Scribble-Supervised Polyp Segmentation

An Wang[1], Mengya Xu[2], Yang Zhang[1,3], Mobarakol Islam[4], and Hongliang Ren[1,2(✉)]

[1] Department of Electronic Engineering, Shun Hing Institute of Advanced Engineering (SHIAE), The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China
wa09@link.cuhk.edu.hk, yzhangcst@hbut.edu.cn, hlren@ee.cuhk.edu.hk

[2] Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore
mengya@u.nus.edu

[3] School of Mechanical Engineering, Hubei University of Technology, Wuhan, China

[4] Department of Medical Physics and Biomedical Engineering, Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London, London, UK
mobarakol.islam@ucl.ac.uk

Abstract. Fully-supervised polyp segmentation has accomplished significant triumphs over the years in advancing the early diagnosis of colorectal cancer. However, label-efficient solutions from weak supervision like scribbles are rarely explored yet primarily meaningful and demanding in medical practice due to the expensiveness and scarcity of densely-annotated polyp data. Besides, various deployment issues, including data shifts and corruption, put forward further requests for model generalization and robustness. To address these concerns, we design a framework of **S**patial-**S**pectral Dual-branch **M**utual Teaching and Entropy-guided Pseudo Label **E**nsemble Learning (S^{2}ME). Concretely, for the first time in weakly-supervised medical image segmentation, we promote the dual-branch co-teaching framework by leveraging the intrinsic complementarity of features extracted from the spatial and spectral domains and encouraging cross-space consistency through collaborative optimization. Furthermore, to produce reliable mixed pseudo labels, which enhance the effectiveness of ensemble learning, we introduce a novel adaptive pixel-wise fusion technique based on the entropy guidance from the spatial and spectral branches. Our strategy efficiently mitigates the deleterious effects of uncertainty and noise present in pseudo labels and surpasses previous alternatives in terms of efficacy. Ultimately, we formulate a holistic optimization objective to learn from the hybrid supervision of

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_4.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 35–45, 2023.
https://doi.org/10.1007/978-3-031-43907-0_4

scribbles and pseudo labels. Extensive experiments and evaluation on four public datasets demonstrate the superiority of our method regarding in-distribution accuracy, out-of-distribution generalization, and robustness, highlighting its promising clinical significance. Our code is available at https://github.com/lofrienger/S2ME.

Keywords: Polyp Image Segmentation · Weakly-supervised Learning · Spatial-Spectral Dual Branches · Mutual Teaching · Ensemble Learning

1 Introduction

Colorectal cancer is a leading cause of cancer-related deaths worldwide [1]. Early detection and efficient diagnosis of polyps, which are precursors to colorectal cancer, is crucial for effective treatment. Recently, deep learning has emerged as a powerful tool in medical image analysis, prompting extensive research into its potential for polyp segmentation. The effectiveness of deep learning models in medical applications is usually based on large, well-annotated datasets, which in turn necessitates a time-consuming and expertise-driven annotation process. This has prompted the emergence of approaches for annotation-efficient weakly-supervised learning in the medical domain with limited annotations like points [8], bounding boxes [12], and scribbles [15]. Compared with other sparse labeling methods, scribbles allow the annotator to annotate arbitrary shapes, making them more flexible than points or boxes [13]. Besides, scribbles provide a more robust supervision signal, which can be prone to noise and outliers [5]. Hence, this work investigates the feasibility of conducting polyp segmentation using scribble annotation as supervision. The effectiveness of medical applications during in-site deployment depends on their ability to generalize to unseen data and remain robust against data corruption. Improving these factors is crucial to enhance the accuracy and reliability of medical diagnoses in real-world scenarios [22,27,28]. Therefore, we comprehensively evaluate our approach on multiple datasets from various medical sites to showcase its viability and effectiveness across different contexts.

Dual-branch learning has been widely adopted in annotation-efficient learning to encourage mutual consistency through co-teaching. While existing approaches are typically designed for learning in the spatial domain [21,25,29,30], a novel spatial-spectral dual-branch structure is introduced to efficiently leverage domain-specific complementary knowledge with synergistic mutual teaching. Furthermore, the outputs from the spatial-spectral branches are aggregated to produce mixed pseudo labels as supplementary supervision. Different from previous methods, which generally adopt the handcrafted fusion strategies [15], we design to aggregate the outputs from spatial-spectral dual branches with an entropy-guided adaptive mixing ratio for each pixel. Consequently, our incorporated tactic of pseudo-label fusion aptly assesses the pixel-level ambiguity emerging from both spatial and frequency domains based on their entropy maps, thereby allocating substantially assured categorical labels to individual pixels and facilitating effective pseudo label ensemble learning.

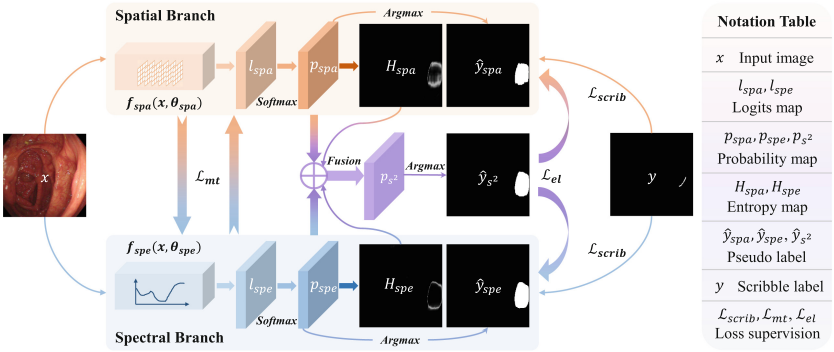

Fig. 1. Overview of our **S**patial-**S**pectral Dual-branch **M**utual Teaching and Pixel-level Entropy-guided Pseudo Label **E**nsemble Learning (S²ME) for scribble-supervised polyp segmentation. Spatial-spectral cross-domain consistency is encouraged through mutual teaching. High-quality mixed pseudo labels are generated with pixel-level guidance from the dual-space entropy maps, ensuring more reliable supervision for ensemble learning.

Contributions. Overall, the contributions of this work are threefold: First, we devise a spatial-spectral dual-branch structure to leverage cross-space knowledge and foster collaborative mutual teaching. To our best knowledge, this is the first attempt to explore the complementary relations of the spatial-spectral dual branch in boosting weakly-supervised medical image analysis. Second, we introduce the pixel-level entropy-guided fusion strategy to generate mixed pseudo labels with reduced noise and increased confidence, thus enhancing ensemble learning. Lastly, our proposed hybrid loss optimization, comprising scribbles-supervised loss, mutual training loss with domain-specific pseudo labels, and ensemble learning loss with fused-domain pseudo labels, facilitates obtaining a generalizable and robust model for polyp image segmentation. An extensive assessment of our approach through the examination of four publicly accessible datasets establishes its superiority and clinical significance.

2 Methodology

2.1 Preliminaries

Spectral-domain learning [26] has gained increasing popularity in medical image analysis [23] for its ability to identify subtle frequency patterns that may not be well detected by the pure spatial-domain network like UNet [20]. For instance, a recent dual-encoder network, YNet [6], incorporates a spectral encoder with Fast Fourier Convolution (FFC) [4] to disentangle global patterns across varying frequency components and derives hybrid feature representation. In addition, spectrum learning also exhibits advantageous robustness and generalization against adversarial attacks, data corruption, and distribution shifts [19]. In label-efficient learning, some preliminary works have been proposed to encourage

mutual consistency between outputs from two networks [3], two decoders [25], and teacher-student models [14], yet only in the spatial domain. As far as we know, spatial-spectral cross-domain consistency has never been investigated to promote learning with sparse annotations of medical data. This has motivated us to develop the cross-domain cooperative mutual teaching scheme to leverage the favorable properties when learning in the spectral space.

Besides consistency constraints, utilizing pseudo labels as supplementary supervision is another principle in label-efficient learning [11,24]. To prevent the model from being influenced by noise and inaccuracies within the pseudo labels, numerous studies have endeavored to enhance their quality, including averaging the model predictions from several iterations [11], filtering out unreliable pixels [24], and mixing dual-branch outputs [15] following

$$p_{mix} = \alpha \times p_1 + (1 - \alpha) \times p_2, \alpha = random(0, 1), \tag{1}$$

where α is the random mixing ratio. p_1, p_2, and p_{mix} denote the probability maps from the two spatial decoders and their mixture. These approaches only operate in the spatial domain, regardless of single or dual branches, while we consider both spatial and spectral domains and propose to adaptively merge dual-branch outputs with respective pixel-wise entropy guidance.

2.2 S^{2}ME: Spatial-Spectral Mutual Teaching and Ensemble Learning

Spatial-Spectral Cross-domain Mutual Teaching. In contrast to prior weakly-supervised learning methods that have merely emphasized spatial considerations, our approach designs a dual-branch structure consisting of a spatial branch $f_{spa}(x, \theta_{spa})$ and a spectral branch $f_{spe}(x, \theta_{spe})$, with x and θ being the input image and randomly initialized model parameters. As illustrated in Fig. 1, the spatial and spectral branches take the same training image as the input and extract domain-specific patterns. The raw model outputs, *i.e.*, the logits l_{spa} and l_{spe}, will be converted to probability maps p_{spa} and p_{spe} with *Softmax* normalization, and further to respective pseudo labels $\hat{y}_{spa}$ and $\hat{y}_{spe}$ by $\hat{y} = \arg\max p$. The spatial and spectral pseudo labels supervise the other branch collaboratively during mutual teaching and can be expressed as

$$\hat{y}_{spa} \rightarrow f_{spe} \text{ and } \hat{y}_{spe} \rightarrow f_{spa}, \tag{2}$$

where "$\rightarrow$" denotes supervision[1]. Through cross-domain engagement, these two branches complement each other, with each providing valuable domain-specific insights and feedback to the other. Consequently, such a scheme can lead to better feature extraction, more meaningful data representation, and domain-specific knowledge transmission, thus boosting model generalization and robustness.

Entropy-Guided Pseudo Label Ensemble Learning. In addition to mutual teaching, we consider aggregating the pseudo labels from the spatial and spectral branches in ensemble learning, aiming to take advantage of the distinctive

[1] For convenience, we omit the input x and model parameters θ.

yet complementary properties of the cross-domain features. As we know, a pixel characterized by a higher entropy value indicates elevated uncertainty in terms of its corresponding prediction. We can observe from the entropy maps $\mathcal{H}_{spa}$ and $\mathcal{H}_{spe}$ in Fig. 1 that the pixels of the polyp boundary exhibit greater difficulties in accurate segmentation, presenting with higher entropy values (the white contours). Considering such property, we propose a novel adaptive strategy to automatically adjust the mixing ratio for each pixel based on the entropy of its categorical probability distribution. Hence, the mixed pseudo labels are more reliable and beneficial for ensemble learning. Concretely, with the spatial and spectral probability maps p_{spa} and p_{spe}, the corresponding entropy maps $\mathcal{H}_{spa}$ and $\mathcal{H}_{spe}$ can be computed with

$$\mathcal{H} = -\sum_{c=0}^{C-1} p(c) \times \log p(c), \tag{3}$$

where C is the number of classes that equals 2 in our task. Unlike previous image-level fixed-ratio mixing or random mixing as Eq. (1), we can update the mixing ratio between the two probability maps p_{spa} and p_{spe} with the weighted entropy guidance at each pixel location by

$$p_{s^2} = \frac{\mathcal{H}_{spe}}{\mathcal{H}_{spa} + \mathcal{H}_{spe}} \otimes p_{spa} + \frac{\mathcal{H}_{spa}}{\mathcal{H}_{spa} + \mathcal{H}_{spe}} \otimes p_{spe}, \tag{4}$$

where "$\otimes$" denotes pixel-wise multiplication. p_{s^2} is the merged probability map and can be further converted to the pseudo label by $\hat{y}_{s^2} = \arg\max p_{s^2}$ to supervise the spatial and spectral branch in the context of ensemble learning following

$$\hat{y}_{s^2} \rightarrow f_{spa} \text{ and } \hat{y}_{s^2} \rightarrow f_{spe}. \tag{5}$$

By absorbing strengths from the spatial and spectral branches, ensemble learning from the mixed pseudo labels facilitates model optimization with reduced overfitting, increased stability, and improved generalization and robustness.

Hybrid Loss Supervision from Scribbles and Pseudo Labels. Besides the scribble annotations for partial pixels, the aforementioned three types of pseudo labels $\hat{y}_{spa}$, $\hat{y}_{spe}$, and $\hat{y}_{s^2}$ can offer complementary supervision for every pixel, with different learning regimes. Overall, our hybrid loss supervision is based on Cross Entropy loss ℓ_{CE} and Dice loss ℓ_{Dice}. Specifically, we employ the partial Cross Entropy loss [13] ℓ_{pCE}, which only calculates the loss on the labeled pixels, for learning from scribbles following

$$\mathcal{L}_{scrib} = \ell_{pCE}(l_{spa}, y) + \ell_{pCE}(l_{spe}, y), \tag{6}$$

where y denotes the scribble annotations. Furthermore, the mutual teaching loss with supervision from domain-specific pseudo labels is

$$\mathcal{L}_{mt} = \underbrace{\{\ell_{CE}(l_{spa}, \hat{y}_{spe}) + \ell_{Dice}(p_{spa}, \hat{y}_{spe})\}}_{\hat{y}_{spe} \rightarrow f_{spa}} + \underbrace{\{\ell_{CE}(l_{spe}, \hat{y}_{spa}) + \ell_{Dice}(p_{spe}, \hat{y}_{spa})\}}_{\hat{y}_{spa} \rightarrow f_{spe}}. \tag{7}$$

Likewise, the ensemble learning loss with supervision from the enhanced mixed pseudo labels can be formulated as

$$\mathcal{L}_{el} = \underbrace{\{\ell_{CE}(l_{spa}, \hat{y}_{s^2}) + \ell_{Dice}(p_{spa}, \hat{y}_{s^2})\}}_{\hat{y}_{s^2} \rightarrow f_{spa}} + \underbrace{\{\ell_{CE}(l_{spe}, \hat{y}_{s^2}) + \ell_{Dice}(p_{spe}, \hat{y}_{s^2})\}}_{\hat{y}_{s^2} \rightarrow f_{spe}}. \quad (8)$$

Holistically, our hybrid loss supervision can be stated as

$$\mathcal{L}_{hybrid} = \mathcal{L}_{scrib} + \lambda_{mt} \times \mathcal{L}_{mt} + \lambda_{el} \times \mathcal{L}_{el}, \quad (9)$$

where λ_{mt} and λ_{el} serve as weighting coefficients that regulate the relative significance of various modes of supervision. The hybrid loss considers all possible supervision signals in the spatial-spectral dual-branch network and exceeds partial combinations of its constituent elements, as evidenced in the ablation study.

3 Experiments

3.1 Experimental Setup

Datasets. We employ the SUN-SEG [10] dataset with scribble annotations for training and assessing the in-distribution performance. This dataset is based on the SUN database [16], which contains 100 different polyp video cases. To reduce data redundancy and memory consumption, we choose the first of every five consecutive frames in each case. We then randomly split the data into 70, 10, and 20 cases for training, validation, and testing, leaving 6677, 1240, and 1993 frames in the respective split. For out-of-distribution evaluation, we utilize three public datasets, namely Kvasir-SEG [9], CVC-ClinicDB [2], and PolypGen [1] with 1000, 612, and 1537 polyp frames, respectively. These datasets are collected from diversified patients in multiple medical centers with various data acquisition systems. Varying data shifts and corruption like motion blur and specular reflections[2] pose significant challenges to model generalization and robustness.

Implementation Details. We implement our method with PyTorch [18] and run the experiments on a single NVIDIA RTX3090 GPU. The SGD optimizer is utilized for training 30k iterations with a momentum of 0.9, a weight decay of 0.0001, and a batch size of 16. The execution time for each experiment is approximately 4 h. The initial learning rate is 0.03 and updated with the poly-scheduling policy [15]. The loss weighting coefficients λ_{mt} and λ_{el} are empirically set the same and exponentially ramped up [3] from 0 to 5 in 25k iterations. All the images are randomly cropped at the border with maximally 7 pixels and resized to 224×224 in width and height. Besides, random horizontal and vertical flipping are applied with a probability of 0.5, respectively.

We utilize UNet [20] and YNet [6] as the respective segmentation model in the spatial and spectral branches. The performance of the scribble-supervised model with partial Cross Entropy [13] loss (Scrib-pCE) and the fully-supervised

[2] Some exemplary polyp frames are presented in the supplementary materials.

model with Cross Entropy loss (Fully-CE) are treated as the lower and upper bound, respectively. Five classical and relevant methods, including EntMin [7], GCRF [17], USTM [14], CPS [3], and DMPLS [15] are employed as the comparative baselines and implemented with UNet [20] as the segmentation backbone referring to the WSL4MIS[3] repository. For a fair comparison, the output from the spatial branch is taken as the final prediction and utilized in evaluation without post-processing. In addition, statistical evaluations are conducted with multiple seeds, and the mean and standard deviations of the results are reported.

3.2 Results and Analysis

Table 1. Quantitative comparison of the in-distribution segmentation performance. The shaded grey and blue rows are the lower and upper bound. The best results of the scribble-supervised methods are in bold.

Method	SUN-SEG [10]			
	DSC ↑	IoU ↑	Prec ↑	HD ↓
Scrib-pCE [13]	0.633±0.010	0.511±0.012	0.636±0.021	5.587±0.149
EntMin [7]	0.642±0.012	0.519±0.013	0.666±0.016	5.277±0.063
GCRF [17]	0.656±0.019	0.541±0.022	0.690±0.017	4.983±0.089
USTM [14]	0.654±0.008	0.533±0.009	0.663±0.011	5.207±0.138
CPS [3]	0.658±0.004	0.539±0.005	0.676±0.005	5.092±0.063
DMPLS [15]	0.656±0.006	0.539±0.005	0.659±0.011	5.208±0.061
S²ME (Ours)	**0.674±0.003**	**0.565±0.001**	**0.719±0.003**	**4.583±0.014**
Fully-CE	0.713±0.021	0.617±0.023	0.746±0.027	4.405±0.119

The performance of weakly-supervised methods is assessed with four metrics,*i.e.*, Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision (Prec), and a distance-based measure of Hausdorff Distance (HD). As shown in Table 1 and Fig. 2, our S²ME achieves superior in-distribution performance quantitatively and qualitatively compared with other baselines on the SUN-SEG [10] dataset. Regarding generalization and robustness, as indicated in Table 2, our method outperforms other weakly-supervised methods by a significant margin on three unseen datasets, and even exceeds the fully-supervised upper bound on two of them[4]. These results suggest the efficacy and reliability of the proposed solution S²ME in fulfilling polyp segmentation tasks with only scribble annotations. Notably, the encouraging performance on unseen datasets exhibits promising clinical implications in deploying our method to real-world scenarios.

[3] https://github.com/HiLab-git/WSL4MIS.

[4] Complete results of all four metrics are present in the supplementary materials.

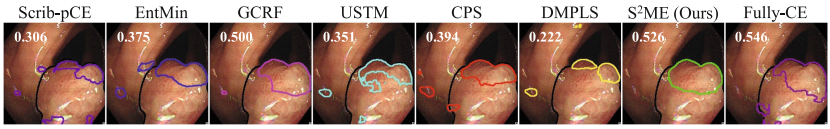

Fig. 2. Qualitative performance comparison of one camouflaged polyp image with DSC values on the left top. The contour of the ground-truth mask is displayed in black, in comparison with that of each method shown in different colors.

Table 2. Generalization comparison on three unseen datasets. The underlined results surpass the upper bound.

Method	Kvasir-SEG [9]		CVC-ClinicDB [2]		PolypGen [1]	
	DSC ↑	HD ↓	DSC ↑	HD ↓	DSC ↑	HD ↓
Scrib-pCE [13]	0.679±0.010	6.565±0.173	0.573±0.016	6.497±0.156	0.524±0.012	6.084±0.189
EntMin [7]	0.684±0.004	6.383±0.110	0.578±0.016	6.308±0.254	0.542±0.003	5.887±0.063
GCRF [17]	0.702±0.004	6.024±0.014	0.558±0.008	6.192±0.290	0.530±0.006	5.714±0.133
USTM [14]	0.693±0.005	6.398±0.138	0.587±0.019	5.950±0.107	0.538±0.007	5.874±0.068
CPS [3]	0.703±0.011	6.323±0.062	0.591±0.017	6.161±0.074	0.546±0.013	5.844±0.065
DMPLS [15]	0.707±0.006	6.297±0.077	0.593±0.013	6.194±0.028	0.547±0.007	5.897±0.045
S²ME (Ours)	**0.750±0.003**	**5.449±0.150**	**0.632±0.010**	**5.633±0.008**	**0.571±0.002**	**5.247±0.107**
Fully-CE	0.758±0.013	5.414±0.097	0.631±0.026	6.017±0.349	0.569±0.016	5.252±0.128

3.3 Ablation Studies

Network Structures. We first conduct the ablation analysis on the network components. As shown in Table 3, the spatial-spectral configuration of our S²ME yields superior performance compared to single-domain counterparts with ME, confirming the significance of utilizing cross-domain features.

Table 3. Ablation comparison of dual-branch network architectures. Results are from outputs of Model-1 on the SUN-SEG [10] dataset.

Model-1	Model-2	Method	DSC ↑	IoU ↑	Prec ↑	HD ↓
UNet [20]	UNet [20]	ME (Ours)	0.666 ± 0.002	0.557 ± 0.002	0.715 ± 0.008	4.684 ± 0.034
YNet [6]	YNet [6]		0.648 ± 0.004	0.538 ± 0.005	0.695 ± 0.004	4.743 ± 0.006
UNet [20]	YNet [6]	S²ME (Ours)	$\mathbf{0.674 \pm 0.003}$	$\mathbf{0.565 \pm 0.001}$	$\mathbf{0.719 \pm 0.003}$	$\mathbf{4.583 \pm 0.014}$

Pseudo Label Fusion Strategies. To ensure the reliability of the mixed pseudo labels for ensemble learning, we present the pixel-level adaptive fusion strategy according to entropy maps of dual predictions to balance the strengths and weaknesses of spatial and spectral branches. As demonstrated in Table 4, our method achieves improved performance compared to two image-level fusion strategies, *i.e.*, random [15] and equal mixing.

Hybrid Loss Supervision. We decompose the proposed hybrid loss $\mathcal{L}_{hybrid}$ in Eq. (9) to demonstrate the effectiveness of holistic supervision from scribbles,

Table 4. Ablation on the pseudo label fusion strategies on the SUN-SEG [10] dataset.

Fusion		Metrics	
Strategy	Level	DSC ↑	HD ↓
Random [15]	Image	0.665 ± 0.008	4.750 ± 0.169
Equal (0.5)	Image	0.667 ± 0.001	4.602 ± 0.013
Entropy (Ours)	Pixel	$\mathbf{0.674 \pm 0.003}$	$\mathbf{4.583 \pm 0.014}$

Table 5. Ablation study on the loss components on the SUN-SEG [10] dataset.

Loss			Metrics	
$\mathcal{L}_{scrib}$	$\mathcal{L}_{mt}$	$\mathcal{L}_{el}$	DSC ↑	HD ↓
✓	✗	✗	0.627 ± 0.004	5.580 ± 0.112
✓	✓	✗	0.668 ± 0.007	4.782 ± 0.020
✓	✗	✓	0.662 ± 0.004	4.797 ± 0.146
✓	✓	✓	$\mathbf{0.674 \pm 0.003}$	$\mathbf{4.583 \pm 0.014}$

mutual teaching, and ensemble learning. As shown in Table 5, our proposed hybrid loss, involving $\mathcal{L}_{scrib}$, $\mathcal{L}_{mt}$, and $\mathcal{L}_{el}$, achieves the optimal results.

4 Conclusion

To our best knowledge, we propose the first spatial-spectral dual-branch network structure for weakly-supervised medical image segmentation that efficiently leverages cross-domain patterns with collaborative mutual teaching and ensemble learning. Our pixel-level entropy-guided fusion strategy advances the reliability of the aggregated pseudo labels, which provides valuable supplementary supervision signals. Moreover, we optimize the segmentation model with the hybrid mode of loss supervision from scribbles and pseudo labels in a holistic manner and witness improved outcomes. With extensive in-domain and out-of-domain evaluation on four public datasets, our method shows superior accuracy, generalization, and robustness, indicating its clinical significance in alleviating data-related issues such as data shift and corruption which are commonly encountered in the medical field. Future efforts can be paid to apply our approach to other annotation-efficient learning contexts like semi-supervised learning, other sparse annotations like points, and more medical applications.

Acknowledgements. This work was supported by Hong Kong Research Grants Council (RGC) Collaborative Research Fund (CRF C4063-18G), the Shun Hing Institute of Advanced Engineering (SHIAE project BME-p1-21) at the Chinese University of Hong Kong (CUHK), General Research Fund (GRF 14203323), Shenzhen-Hong Kong-Macau Technology Research Programme (Type C) STIC Grant SGDX20210823103535014 (202108233000303), and (GRS) #3110167.

References

1. Ali, S., et al.: a multi-centre polyp detection and segmentation dataset for generalisability assessment. Sci. Data **10**(1), 75 (2023)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput. Med. Imaging Graph. **43**, 99–111 (2015)

3. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2613–2622 (2021)
4. Chi, L., Jiang, B., Mu, Y.: Fast Fourier convolution. Adv. Neural. Inf. Process. Syst. **33**, 4479–4488 (2020)
5. Cinbis, R.G., Verbeek, J., Schmid, C.: Weakly supervised object localization with multi-fold multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell. **39**(1), 189–203 (2016)
6. Farshad, A., Yeganeh, Y., Gehlbach, P., Navab, N.: Y-net: a spatiospectral dual-encoder network for medical image segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention - MICCAI 2022. Lecture Notes in Computer Science, vol. 13432, pp. 582–592. Springer, Cham (2022)
7. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: Advances in Neural Information Processing Systems, vol. 17 (2004)
8. He, X., Fang, L., Tan, M., Chen, X.: Intra-and inter-slice contrastive learning for point supervised oct fluid segmentation. IEEE Trans. Image Process. **31**, 1870–1881 (2022)
9. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. In: Ro, Y.M., et al. (eds.) MMM 2020. LNCS, vol. 11962, pp. 451–462. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37734-2_37
10. Ji, G.P., et al.: Video polyp segmentation: a deep learning perspective. Mach. Intell. Res. 1–19 (2022)
11. Lee, H., Jeong, W.-K.: Scribble2Label: scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 14–23. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_2
12. Li, Y., Xue, Y., Li, L., Zhang, X., Qian, X.: Domain adaptive box-supervised instance segmentation network for mitosis detection. IEEE Trans. Med. Imaging **41**(9), 2469–2485 (2022)
13. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3159–3167 (2016)
14. Liu, X., et al.: Weakly supervised segmentation of covid19 infection with scribble annotation on CT images. Pattern Recogn. **122**, 108341 (2022)
15. Luo, X., et al.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13431, pp. 528–538. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16431-6_50
16. Misawa, M., et al.: Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). Gastrointest. Endosc. **93**(4), 960–967 (2021)
17. Obukhov, A., Georgoulis, S., Dai, D., Van Gool, L.: Gated CRF loss for weakly supervised semantic image segmentation. arXiv preprint arXiv:1906.04651 (2019)
18. Paszke, A., et al.: Automatic differentiation in pyTorch. In: NIPS-W (2017)
19. Rao, Y., Zhao, W., Zhu, Z., Lu, J., Zhou, J.: Global filter networks for image classification. Adv. Neural. Inf. Process. Syst. **34**, 980–993 (2021)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation (2015)

21. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Trans. Med. Imaging **40**(8), 1990–2001 (2021)
22. Wang, A., Islam, M., Xu, M., Ren, H.: Rethinking surgical instrument segmentation: a background image can be all you need. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 355–364. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_34
23. Wang, K.N., et al.: Ffcnet: Fourier transform-based frequency learning and complex convolutional network for colon disease classification. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 78–87. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_8
24. Wang, Y., et al.: Freematch: self-adaptive thresholding for semi-supervised learning. arXiv preprint arXiv:2205.07246 (2022)
25. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 297–306. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_28
26. Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.K., Ren, F.: Learning in the frequency domain. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1740–1749 (2020)
27. Xu, M., Islam, M., Lim, C.M., Ren, H.: Class-incremental domain adaptation with smoothing and calibration for surgical report generation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 269–278. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87202-1_26
28. Xu, M., Islam, M., Lim, C.M., Ren, H.: Learning domain adaptation with model calibration for surgical report generation in robotic surgery. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 12350–12356. IEEE (2021)
29. Zhang, K., Zhuang, X.: Cyclemix: a holistic strategy for medical image segmentation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11656–11665 (2022)
30. Zhang, K., Zhuang, X.: ShapePU: a new PU learning framework regularized by global consistency for scribble supervised cardiac segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13438, pp. 162–172. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16452-1_16