Effective Polyp Segmentation in Medical Imaging: A Semi-Supervised Approach with Progressive Supervision

Shunlong Ye[1*] and Zexuan Ji[1]

[1] Dept. of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

yeshunlong@njust.edu.cn, jizexuan@njust.edu.cn

Abstract. Medical imaging plays a pivotal role in diagnosing and monitoring various diseases, and accurate polyp segmentation is essential for early disease detection and treatment planning. However, the scarcity of well-annotated medical imaging data hampers the development of effective segmentation models. Semi-supervised learning stands out as a critical approach to address this issue. Traditional semi-supervised medical image segmentation models leverage strong-weak perturbation consistency losses to provide supervision signals. While these losses are effective in promoting consistency, they often exhibit a drawback, that is the dominance of strong perturbations during early training stages. This dominance hinders the model's ability to extract meaningful features, ultimately impacting its overall performance. In response to these challenges, we introduce a novel training method applied to polyp segmentation. We employ a progressive approach, gradually introducing strong-weak perturbation supervision signals. This gradual guidance empowers the model to focus on learning relevant features from the beginning of training, mitigating the issue of early dominance and enhancing its performance. Our method is evaluated on two widely-used polyp segmentation datasets and surpasses state-of-the-art methods, demonstrating substantial improvements in segmentation accuracy.

Keywords- polyp segmentation**,** semi-supervised learning**,** progressive supervision

1. Introduction

Medical imaging assumes a crucial role in diagnosing and managing a wide array of medical conditions, equipping healthcare providers with the knowledge needed for well-informed decision-making [1][2]. Within the realm of medical imaging, accurate polyp segmentation is of paramount importance as it aids in the early detection of colorectal diseases and facilitates treatment planning [3]. However, the development of robust polyp segmentation models is challenged by the scarcity of meticulously annotated medical imaging data. To address this data scarcity, semi-supervised learning has emerged as a promising avenue [4]. This approach leverages both labeled and unlabeled data, effectively expanding the available dataset for training segmentation models [5].

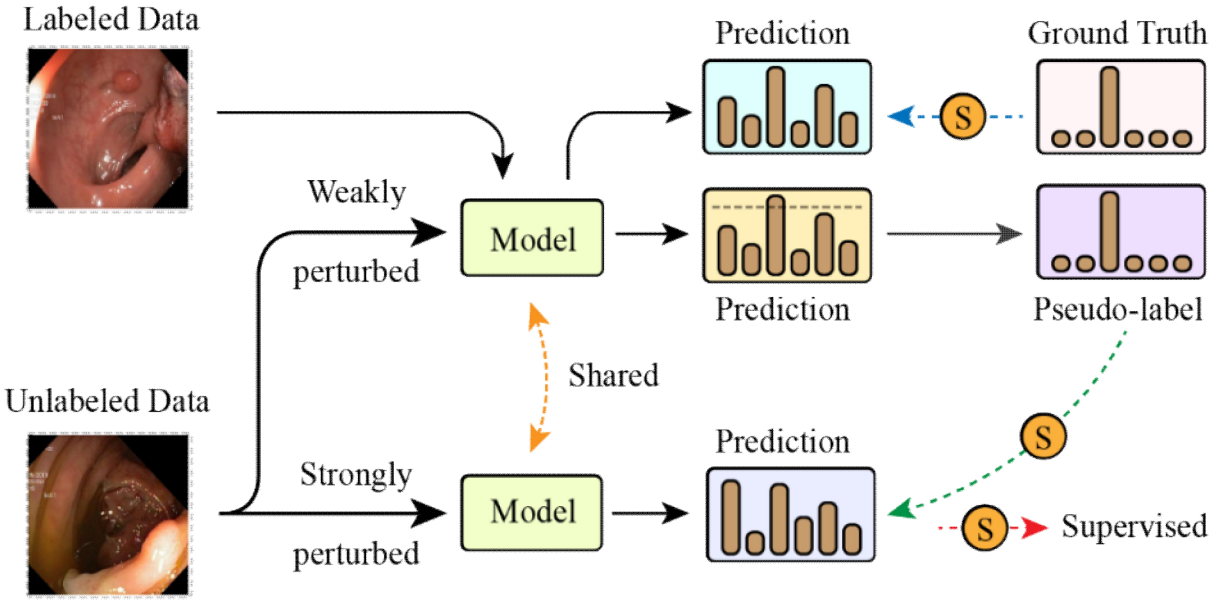


Figure 1. The core concept of the FixMatch algorithm.

Nevertheless, traditional semi-supervised medical image segmentation models often confront a significant issue. Take FixMatch [6] in Figure 1 as an example, one branch receives a strong perturbation while the other receives a weak perturbation. Intuitively, the model is expected to provide more accurate predictions for the branch subjected to weaker perturbations. Consequently, these predictions are employed to generate pseudo-labels, striking a balance between ensuring pseudo-label accuracy and leveraging the benefits of stronger perturbations for enhanced model performance, these models typically rely on strong-weak perturbation consistency losses to provide supervision signals. Regrettably, during the initial training stages, strong perturbations tend to dominate the learning process, posing challenges for the model to effectively capture salient features and impeding its overall performance [7].

In this paper, we introduce a novel training approach, applied to polyp segmentation, addressing the aforementioned issue inspired by DART [8]. Our approach involves the progressive provision of strong-weak perturbation supervision signals, breaking the training process into distinct stages. In the initial stage, only weak perturbation supervision signals are provided. Subsequently, in the second stage, a limited number of strong perturbation supervision signals are introduced, followed by an increase in perturbation signals in the third stage. This design enables the model to better utilize the supervision signals obtained from pseudo-labeling while avoiding the excessive influence of strong perturbations, ultimately enhancing the learning of meaningful features.

In Summary, this paper contributes to the field with three primary innovations. Firstly, we discuss the limitations of classic strong-weak perturbation supervision signals. Building upon this insight, we introduce a novel semi-supervised training method that facilitates early feature learning while enabling effective utilization of strong perturbation supervision signals in the later stages. Finally, our approach is evaluated on two widely used polyp segmentation datasets, demonstrating substantial performance improvements that surpass the current state-of-the-art methods.

2. Method

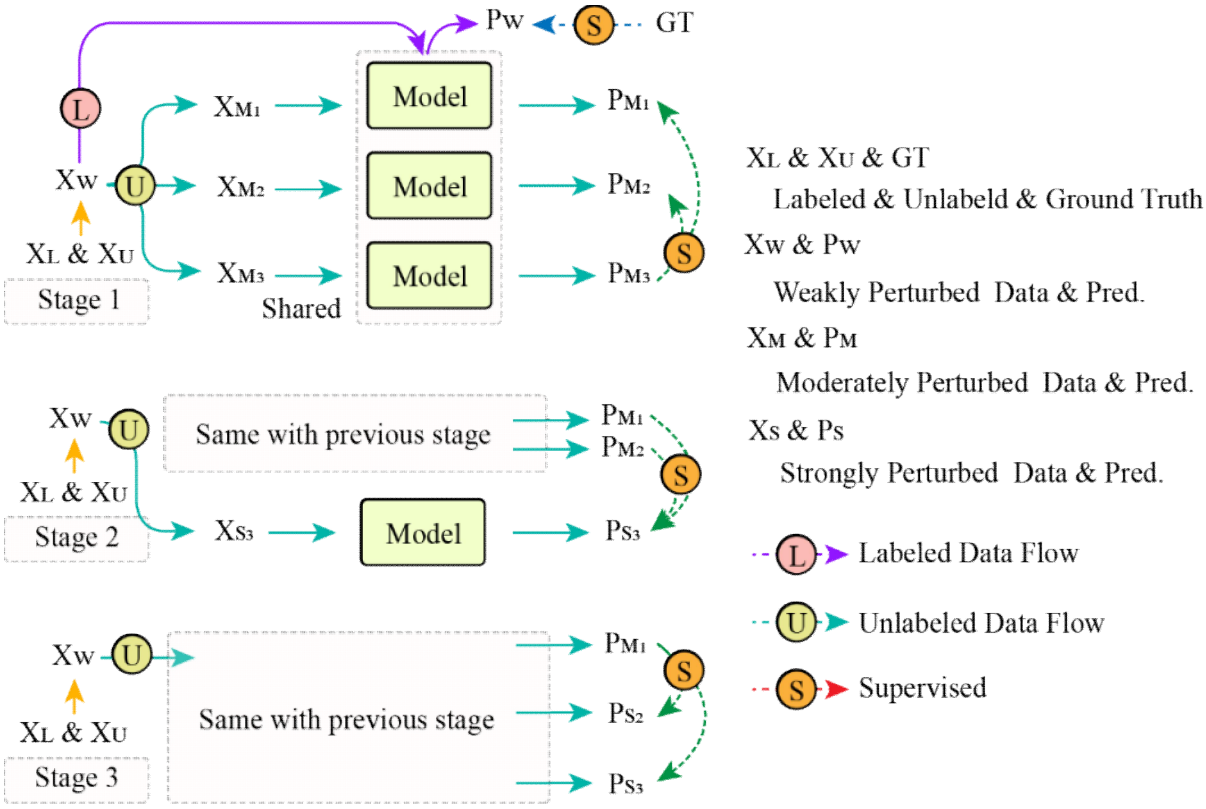
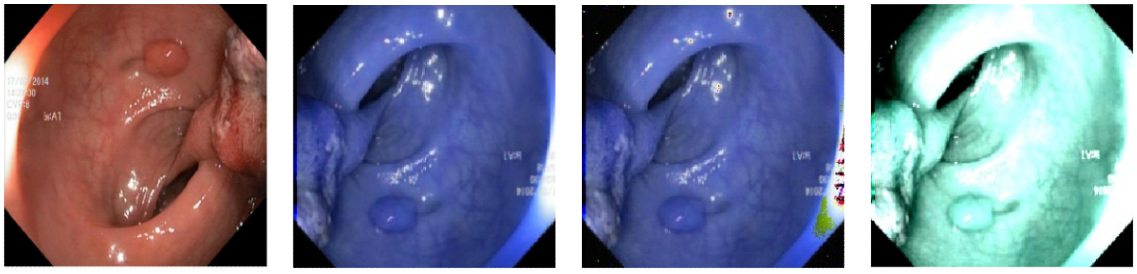


Figure 2. Model training overview with progressive perturbations.

Figure 2 illustrates the structure and training process of our model. The data flow begins with basic data augmentation, termed as "weakly perturbations", applied to all data. Subsequently, labeled data directly undergo training, while unlabeled data undergo augmentations with three distinct random parameter settings, resulting in moderate or strong perturbations. The predictions of data subjected to moderate perturbations are treated as pseudo-labels, guiding the training of data subjected to strong perturbations. During this process, all models share parameters to optimize memory efficiency and maintain computational efficiency comparable to conventional methods.

2.1. Perturbation Strategies

To introduce perturbations, we employ three techniques. Firstly, the original data undergo weak perturbations through random cropping and flipping. Subsequently, moderate perturbations are introduced through the addition of random Gaussian noise, followed by the application of strong perturbations by random alterations in color saturation and contrast. The effects of these perturbations are demonstrated in Figure 3.



(a) image (b) weakly perturbed (c) moderately perturbed (d) strongly perturbed

Figure 3. Perturbation effects on data augmentation.

2.2. Loss Computation

Our loss computation consists of two integral components: the supervised loss and the unsupervised loss, each serving a distinct purpose in our semi-supervised learning framework.

The supervised loss, denoted as l_s is responsible for quantifying the dissimilarity between the model's predictions (p_i) for labeled data and their corresponding ground-truth labels ($p_m(y|x_b)$). In this equation:

$$l_s = \frac{1}{\mathrm{Batch}} \sum_{i=1}^{\mathrm{Batch}} \mathrm{CE}\left(p_i, p_m(y|x_b)\right) \quad (1)$$

Here, CE represents the cross-entropy loss, and Batch signifies the number of samples within a batch.

On the other hand, the unsupervised loss, denoted as l_u, is designed to evaluate the model's predictions for unlabeled data in relation to their pseudo-labels. These pseudo-labels are derived from the predictions of data subjected to moderate perturbations. The equation for l_u is as follows:

$$l_u = \frac{1}{w*\mathrm{Batch}} \sum_{i=1}^{w*\mathrm{Batch}} [\max(p_m(y|u_b) \geq \tau] \mathrm{CE}(\mathrm{argmax}(p_m(y|u_b)), p_m(y|u_b)) \quad (2)$$

In this equation, w represents the ratio of unlabeled data, and τ is a threshold for pseudo-label assignment. l_u calculates the cross-entropy loss based on the most probable class prediction (argmax) for the unlabeled data and their respective pseudo-labels.

The final loss is obtained as the sum of l_s and l_u with a scaling factor consistency_weight applied to l_u to balance the contributions of both components:

$$\mathrm{Loss} = l_s + \mathrm{consistency\_weight} * l_u \quad (3)$$

This hybrid loss function serves as the foundation of our semi-supervised learning approach, enabling model to effectively leverage both labeled and unlabeled data for improved model performance.

3. Experiments

We conducted a rigorous evaluation on two widely-used colonic polyp segmentation datasets, Kavsir [9] and CVC-ClinicDB [10]. We maintained consistency with the data preprocessing and dataset partitioning standards defined by the baseline models specific to each dataset. This alignment ensured a fair comparison with recent state-of-the-art models and established a robust foundation for our assessments. Our primary evaluation metric was the mean Intersection over Union (mIoU), which calculates the average IOU across all classes. We experimented with varying annotation data ratios of 10% and 20%.

Table 1. Performance on Kavsir and CVC-CLinicDB Datasets. Bold entries in the table indicate the best-performing results for each metric.

Dataset	Kavsir		CVC-CLinicDB	
Labeled Ratio	10%	20%	10%	20%
U-Net [11]	0.3560	0.4354	0.5780	0.6874
PSPNet [11]	0.3259	0.4158	0.5600	0.6210
LinkNet [11]	0.3897	0.4698	0.4950	0.5614
MT [12]	0.5741	0.6698	0.7125	0.7695
GAN [5]	0.5632	0.6599	0.68994	0.7595
CCT [4]	0.6058	0.6629	0.6925	0.75201
FixMatch [6]	0.6647	0.6816	0.6955	0.7483
SemiSegPolyp [2]	0.6523	0.7058	0.7045	0.7799
Ours	0.7074	0.7114	0.7452	0.7998

The results, as depicted in the Table 1, underscore the impressive performance of our model across both datasets and different annotation data ratios. Furthermore, the consistently higher mIoU values achieved by our proposed method underscore its effectiveness in addressing the complexities of colonic polyp segmentation tasks. Across both the Kavsir and CVC-ClinicDB datasets, at both 10% and 20% labeled ratios, our approach demonstrated a notable improvement in capturing the intricate details of colonic polyps compared to established baseline models. The superior performance of our

method is particularly evident when juxtaposed with various state-of-the-art techniques, such as U-Net, PSPNet, LinkNet, MT, GAN, CCT, FixMatch, and SemiSegPolyp. Notably, at a labeled ratio of 20%, our method achieved an impressive mIoU of 0.7998 on the CVC-ClinicDB dataset, outperforming all other models. This signifies the robustness of our approach in scenarios where a higher percentage of labeled data is available. The observed trend of our method consistently outshining its counterparts suggests its adaptability to different annotation data ratios and datasets. This adaptability is crucial in real-world scenarios where obtaining a large labeled dataset might be challenging. The ability of our method to maintain high segmentation accuracy even with a limited amount of labeled data highlights its potential for practical deployment in medical image analysis, specifically in colonic polyp segmentation applications. Overall, these results affirm the promising capabilities of our proposed approach and position it as a compelling candidate for advancing the state of the art in colonic polyp segmentation.

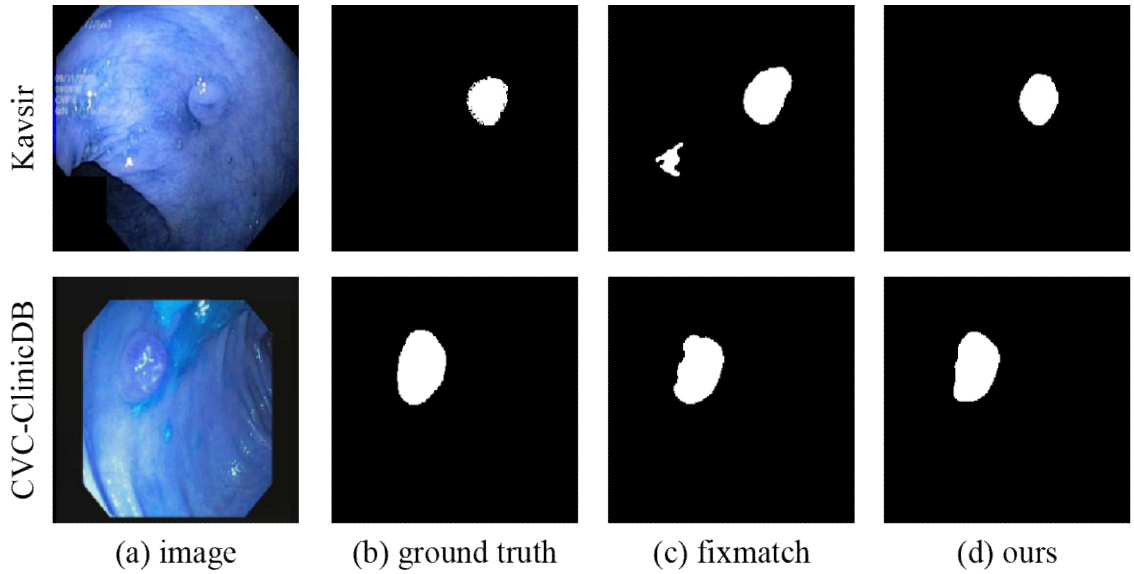


Figure 4. Enhanced polyp boundary segmentation compared to classic methods.

4. Conclusions

In this paper, we conducted a comprehensive analysis of traditional semi-supervised medical image segmentation models, revealing a critical limitation prevalent during early training stages. To address this challenge, we introduced a novel progressive semi-supervised training approach. By gradually providing supervision signals with varying perturbation strengths, our method empowers the model to more efficiently capture valuable features during early training while maintaining the capacity to exploit the benefits of stronger perturbations in the later stages. Our experiments on two widely-adopted polyp segmentation datasets demonstrated the superiority of our approach, surpassing the state-of-the-art results. Future directions encompass extending the application of our approach to diverse medical image segmentation tasks and exploring its potential in other domains requiring semi-supervised learning techniques.

5. References

[1] Ren G, Lazarou M, Yuan J, et al. Towards Automated Polyp Segmentation Using Weakly-and Semi-Supervised Learning and Deformable Transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 4354-4363.

[2] El Rai M C, Darweesh M, Al-Saad M, et al. SemiSegPolyp: Semi-Supervised Polyp Segmentation using Graph Signals[C]//2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS). IEEE, 2022: 1-4.
[3] Wu H, Chen G, Wen Z, et al. Collaborative and adversarial learning of focused and dispersive representations for semi-supervised polyp segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 3489-3498.
[4] Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12674-12684.
[5] Souly N, Spampinato C, Shah M. Semi supervised semantic segmentation using generative adversarial network[C]//Proceedings of the IEEE international conference on computer vision. 2017: 5688-5696.
[6] Sohn K, Berthelot D, Carlini N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence[J]. Advances in neural information processing systems, 2020, 33: 596-608.
[7] Yang L, Qi L, Feng L, et al. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7236-7246.
[8] Jain S, Addepalli S, Sahu P K, et al. DART: Diversify-Aggregate-Repeat Training Improves Generalization of Neural Networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16048-16059.
[9] Jha D, Smedsrud P H, Riegler M A, et al. Kvasir-seg: A segmented polyp dataset[C]//MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26. Springer International Publishing, 2020: 451-462.
[10] Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized medical imaging and graphics, 2015, 43: 99-111.
[11] Minaee S, Boykov Y, Porikli F, et al. Image segmentation using deep learning: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(7): 3523-3542.
[12] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.