基于合作的图像级注释弱监督息肉检测网络

# CBIA-WSPD:Cooperation Based Image-level Annotation for Weakly Supervised Polyp Detection

## Introduction

结直肠癌癌症（CRC）是癌症中诊断频率第三高的癌症，其中80%以上源于息肉[1]。未经治疗的息肉可能会变成恶性的，并可能危及生命的癌症[2]，尽管它们在早期是良性的。在临床实践中，借助结肠镜手术检测和切除息肉被视为预防结肠癌发展的黄金标准。然而，这种方法在很大程度上取决于医生的经验和熟练程度，由于内镜医生在长时间工作后的不规范操作和疏忽，失误率很高（高达30%）[3]。幸运的是，**卷积神经网络（CNNs）和变压器已被证明是自动息肉检测的最先进的计算方法**，并将误诊风险降至最低[4-5]。

Colorectal cancer (CRC) ranks the third most frequently diagnosed cancer, which more than 80% originates from polyps[1]. Untreated polyps might become malignant and potentially life-threatening cancer [2] although they are benign in the early stage. In clinical practice, the detection and removal of polyps with the help of colonoscopy procedure is regarded as the golden standard to prevent the development of colon cancer. However, the method highly depends on the experience and proficiency of the medical practitioner that suffers a high miss rate (as much as 30%) due to the irregular operation and negligence of endoscopists after long duty[3]. Fortunately, the Convolutional Neural Networks (CNNs) and Transformers have been proven to be the most rominent computational approach for automatic polyp detection and minimize the risk of misdiagnosis[4-5].

在过去的几年里，计算机辅助检测（CAD）取得了重大进展，在检测和分割息肉区域方面表现出了令人印象深刻的能力[6-12]。但它们需要与具有实例级边界框注释的大规模数据集相结合，由于结肠图像的复杂性（图2a上）和息肉的多样性（图2a下）,这是非常费力和具有挑战性的。相比之下，弱监督目标检测网络(WSOD)只需要更容易获得的图像级注释就可以进行训练。因此，研究将注意力转移到了依赖粗标记的WSOD上。基于这一事实，**在本文中，我们专注于仅使用图像级监督训练息肉检测器，即弱监督息肉检测(WSPD)**。

During the last few years, computer-aided detection(CAD) have made significant advancements, showcasing impressive capabilities in detection and segmenting polyp regions[6-12]. But they require in conjunction with large scale datasets with instance level bounding box annotations, which is very laborious and challenging due to the complexity of colonic images and the diversity of polyps as shown in Fig. 2 (a). By contrast, weakly supervised object detection networks(WSOD) can be trained with only needs image-level annotations that are much easier to acquire.Hence, studies have shifted their attention to WSOD that rely on coarse markers. Motivated by this fact, in this paper we focus on training polyp detectors using only image-level supervisions, i.e., weakly supervised polyp detection (WSPD).

由于实例级监督的不可用，普遍的工作将其视为一个多示例学习(MIL)问题来处理，这需要预先生成候选框。在众多研究中，选择性搜索[14]和边缘盒[15]是创建候选提案的标准方法。然而，后来的研究认为他们大多数是负例，这不仅影响了检测效果，而且降低了预测速度[13]。考虑到这一点，他们设计了区域提议网络(RPN)[16]. 确实，与传统方法相比，它们在速度上有了显著的提高。不幸的是，这种策略需要大量的先验参数，如尺度和比例。更糟糕的是，这些参数的有效性在很大程度上控制了建议的质量，进而影响检测结果。我们统计了两个公开的以及一个私人息肉数据集的相关信息。如图2 (b)-(c)所示，不同数据集的息肉大小和box宽高比的范围变化很大。这些差异对参数的设置提出了严峻的挑战，如果只设置高频参数会导致召回率降低造成漏检；如果设置所有参数，则会导致大量冗余的候选框和更复杂的计算。此外，这些参数是动态的，因为它们与结肠镜和息肉之间的距离密切相关。进一步地，为了确保RPN的有效性，网络训练需要实例级注释，这不仅偏离了弱监督的要求，更无法应用于WSPD。**考虑到这些，我们认为传统方法可能比RPN更适合WSPD。加入长尾数据（背景框远多于息肉框）**

Since instance-level supervision is not available, prevailing work regards it as a multiple instance learning (MIL) problem to handle, whereby pre-generated proposals are required. Among various studies Selective search [14] and Edge Boxes [15] is the standard methods to create candidate proposals. However, later studies believe that most of them are negative cases, which not only affect the detection effect but also reduce the prediction speed[13]. With that in mind, they designed the region proposal network (RPN)[16]. Indeed, it showed a significant improvement in speed compared to traditional methods. Unfortuantely, this strategy calls for a large number of priori parameters such as scales and ratios. Even worse, the effectiveness of these parameters heavily controls the quality of the proposals and thereby impacts the detection results. We statistically correlated information from two publicly available along with a private polyp dataset. As Fig. 2 (b)-(c) shows the range of polyp scales and box ratios in different datasets is extremely variable. These variations in differentiation pose a severe challenge to parameters setting, as setting only the high-frequency parameter will cause a lower recall rate resulting in missed detections, while result in a large number of redundant candidate proposals and more complex calculations if all the parameters are set. In addition, these numbers are constantly dynamic because they are closely linked to the distance between the colonoscope and polyps. Furthermore, to ensure the high performance of RPN, instance-level annotations are required to train the network, which not only deviates from the requirement of weak supervision (WS) but also fails to deliver on WSPD.In consideration of these we argue that rather than RPN probably traditional methods are more adequate for WSPD.

尽管有几项很有前途的工作将MIL与深度学习[17-19]相结合，突破了自然图像的界限将弱监督成功地应用于医学领域。然而，与自然图像领域类似，它们也很容易过度拟合对象部分（如图1a left），因为它们本质上都高度依赖于区域分类器，而最具辨别力的分类证据可能来自整个对象区域，但也可能来自关键部分。幸运的是，在弱监督分割（WSS）中，检测区域的完整性更容易确保。考虑到他们内在的联系，一些结合分割指导检测的努力[20-21]已经被做出。然而，WSS需要检测热图作为伪标签，但其过于粗糙不足以作为指导标记。得益于迁移学习，最近的视觉大模型方面的进展如（SAM）使这一想法成为现实。糟糕的是，直接的息肉掩码预测是失败的（如图1a right），而高质量掩码需要输入提示（例如点、框）[22]。具体来说，边界框是最有效的点仅次于它。然而对于WSPD，点是唯一可用的选项。考虑到前面提到的事实，即分类器更倾向于关注对象中最具鉴别力的部分，我们认为它是自动获取输入提示的最强大的助手 (详见第3.1节)。**总之，WSPD和SAM不应该独立工作，而是天生的合作精神，应该共同努力克服自己的内在弱点**（如图1b）。

Despite several promising works combining MIL with deep learning [17-19] have greatly pushed the boundaries of natural images to successfully apply WS to the medical field. However, similar to the natural image domain, they also easily overfit the object parts as shown in Fig. 1 (a) left, because they both highly rely on regional classifiers in essence, while the most discriminating classification evidence may come from the entire object region, but may also from the crucial parts. Fortunately, the completeness of a detected region is easier to ensure in weakly supervised segmentation (WSS). For the inherent relations, some efforts [20-21] have been done to combine segmentation in order to guided the detection task. Nevertheless, WSS demands the detection heatmap as pseudo-labels, which is too coarse to be sufficient as guidemark. Benefiting from transfer learning, the recent advances in large vision models such as segment anything model (SAM) brings this assumption to reality. Badly, direct polyp mask prediction is failed as shown in Fig. 1 (a) right, and input prompts (e.g., points, boxes) are desired for high-quality masks of the specified location [22]. Specifically, the bounding box is the most effective, with the point next to it. However, for WSPD, point is the only available option. Considering the fact mentioned previously that the classifiers are more tend to focus on the most discriminative parts of the object, we believe that it is the most powerful assistant to automatically get the input prompts (detailed in section 3.1). In a word, instead of working independently, WSPD and SAM are naturally cooperative and should work together to overcome their intrinsic weaknesses as shown in Fig. 1 (b).

尽管在弱注释检测模型方面取得了一些显著进展，但准确可靠的息肉检测很容易被小而平坦的息肉所欺骗。正如我们在表1中的分析，在图像中，与背景相比，息肉区域相对较小，这使得前景（息肉）和背景（结肠直肠壁）之间的严重区域不平衡，导致息肉被大背景淹没，造成模型过度拟合不相关信息。而后，扁平息肉通常缺乏清晰可见的边界，并且表现出与周围结直肠组织相似的方面，这使得传统的网络难以准确区分前景息肉和外来背景。这背后的原因在于，与低分辨率深层特征相比，高分辨率浅层特征可以很好地学习细节特征和小结构，但检测会不断进行下采样以获得细粒度的表示，这会损害浅层特征，削弱对小而平坦目标的识别[23]。为了解决这一问题，我们引入了特征语义流融合模块（FSFF）来对齐多尺度特征之间的空间信息以丰富特征信息(详见第3.2节)。

Although some notable progress made in weak annotation detection models, accurate and reliable polyp detection can be easily fooled by small and flat polyps. As our analysis in Table 1, in the image, the polyp regions are relatively small when compare to the background leads to the severely regional imbalance between the foreground (polyp) and background (colorectal wall), which results in polyps being overwhelmed by the large background causing the model to overfit irrelevant information. Subsequently, typically the flat polyp lacks clearly visible borders and exhibits a similar aspect to the tissue of the surrounding colorectum, which makes conventional networks struggle to accurately distinguish foreground polyps from extraneous backgrounds. The reason behind this is that high-resolution shallow features work well to learn detailed features as well as small structures compared to low-resolution deep features, but the detection is constantly downsampled to obtain a fine-grained representation, which damages the shallow features and weakens recognition of small, flat targets[23]. To address this, we introduce feature semantic flow fusion (FSFF) module to align spatial information between multi-scale features to achieve a more efficient transfer of semantic information from shallow to deep layers to enrich feature information(detailed in section 3.2).

总之，我们的主要贡献有四个方面：

* 在提案生成阶段，我们通过对主流数据集的多维分析，证实了RPN在弱监督息肉检测任务中的不适用性。因此，我们使用传统方法为WSPD生成提案。
* 我们提出了CBIA-WSPD，一个基于互补协作机制的框架，在只有图像级注释的弱监督环境中增强对彼此的帮助。
* 为了更好地捕捉扁平和小息肉，我们设计了特征语义流融合模块，以从不同层获取更高质量的特征信息，通过聚合多尺度特征层的空间信息。
* 我们的方法在三个数据集（即ClinicDB、Kvasir、private）上进行，这些数据集在WSPD中获得了最先进的性能，并在某些方面与完全监督的检测具有竞争力。

In summary, our main contributions are listed four-fold:

* In the proposals generation stage, we confirmed the inapplicability of RPN in the task of weakly supervised polyp detection by performing multi-dimensional analyses on popular datasets. Therefore, we retain the traditional method to generate proposals for WSPD.
* We propose the CBIA-WSPD, a framework that employ complementary collaborative mechanism to enhance valuable assistance to each other in the weakly supervised setting with only image-level annotation.
* To better capture the flat and small polyps, we design the feature semantic flow fusion module to get higher quality feature information from different level by aggregating spatial information from multi-scales feature layers.
* Our method is conducted on three datasets(i.e. ClinicDB, Kvasir, private) , which obtains state-of-the art performance in WSPD and are competitive with fully supervised detection in certain aspects.

## **Related Work**

Region Proposal Generation.

Segment Anything.

# MedSAM在包含超过一百万对医学图像-掩模对的前所未有的数据集上微调，以提升SAM在医学图像分割领域的适用性。由于SAM本质上是一种可提示的分割方法，需要使用点或边界框来指定分割目标，所以MedSAM为每张训练数据都提供了bounding box作为输入提示。然而结果表明，当边界框提示不明确时，分割将会出现困难（例如眼底血管分割）。

**Method**

**Experiment**

**Conclusion**

1. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,”CA: a cancer journal for clinicians, vol. 71, no. 3, pp. 209–249, 2021.
2. Alrushaid N, Khan FA, Al-Suhaimi E, Elaissari A. Progress and Perspectives in Colon Cancer Pathology, Diagnosis, and Treatments. Diseases. 2023 Oct 24;11(4):148. doi: 10.3390/diseases11040148. PMID: 37987259; PMCID: PMC10660546.
3. P. Wang, P.X. Liu, J.R.G. Brown, T.M. Berzin, G.Y. Zhou, S. Lei, X.G. Liu, L.P. Li, X. Xiao, Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine WhiteLight Colonoscopy in a Prospective Tandem Study, Gastroenterology, 159 (2020) 1252-+.https://doi.org/10.1053/j.gastro.2020.06.023
4. Jiaxin Mei and Tao Zhou and Kaiwen Huang and Yizhe Zhang and Yi Zhou and Ye Wu and Huazhu Fu:A Survey on Deep Learning for Polyp Segmentation: Techniques, Challenges and Future Trends. arXiv preprint arXiv:2311.18373(2023)
5. Hassan, C., Spadaccini, M., Iannone, A., Maselli, R., Jovani, M., Chandrasekar, V. T., Antonelli, G., Yu, H., Areia, M., Dinis-Ribeiro, M., et al., “Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis,” Gastrointestinal endoscopy 93(1), 77–85 (2021).
6. Yuncheng Jiang, Zixun Zhang, Yiwen Hu, Guanbin Li, Xiang Wan, Song Wu:ECC-PolypDet: Enhanced CenterNet with Contrastive Learning for Automatic Polyp Detection. arXiv preprint arXiv:2401.04961(2024)
7. Q. Chang, D. Ahmad, J. Toth, R. Bascom, and W. E. Higgins, “Esfpnet: efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video,” in Medical Imaging: Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 12468. SPIE, 2023, p. 1246803.
8. M. M. Rahman and R. Marculescu, “Medical image segmentation via cascaded attention decoding,” in IEEE/CVF WACVW, 2023, pp. 6222–6231.
9. Y. Li, M. Hu, and X. Yang, “Polyp-sam: Transfer sam for polyp segmentation,” arXiv preprint arXiv:2305.00293, 2023.
10. Hemin Ali Qadir, Younghak Shin, Jacob Bergsland, Ilangko Balasingham:Accurate Real-time Polyp Detection in Videos from Concatenation of Latent Features Extracted from Consecutive Frames. arXiv preprint arXiv:2303.05871(2023).
11. N. K. Tomar, D. Jha, U. Bagci, and S. Ali, “Tganet: Text-guided attention for improved polyp segmentation,” in MICCAI. Springer, 2022, pp. 151–160.
12. J. Wang, Q. Huang, F. Tang, J. Meng, J. Su, and S. Song, “Stepwise feature fusion: Local guides global,” in MICCAI. Springer, 2022, pp.110–120.
13. Chen Z Y，Wang Z D and Gong C. 2023. Image-level labeled weakly supervised object detection：a survey. Journal of Image and Graphics，28（09）：2644-2660
14. Uijlings, J.R., van de Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. IJCV 104(2), 154–171 (2013)
15. Zitnick, C.L., Doll´ar, P.: Edge boxes: Locating object proposals from edges. In:ECCV. pp. 391–405 (2014)
16. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. TPAMI 39(6), 1137–1149 (2017)
17. Philip Müller and Felix Meissen and Johannes Brandt and Georgios Kaissis and Daniel Rueckert.Anatomy-Driven Pathology Detection on Chest X-rays.MICCAI.2023
18. Su Z, Tavolara TE, Carreno-Galeano G, Lee SJ, Gurcan MN, Niazi MKK. Attention2majority: Weak multiple instance learning for regenerative kidney grading on whole slide images. Med Image Anal. 2022 Jul;79:102462. doi: 10.1016/j.media.2022.102462. Epub 2022 Apr 17. PMID: 35512532; PMCID: PMC10382794.
19. Qu J, Wei X, Qian X. Generalized pancreatic cancer diagnosis via multiple instance learning and anatomically-guided shape normalization. Med Image Anal. 2023 May;86:102774. doi: 10.1016/j.media.2023.102774. Epub 2023 Feb 21. PMID: 36842410.
20. Xu Y Q， Zhou C L， Yu X， Xiao B and Yang Y. 2021. Pyramidal multiple instance detection network with mask guided self-correction for weakly supervised object detection. IEEE Transactions on Image Processing，30：3029-3040［DOI：10.1109/TIP.2021.3056887］
21. Gao W， Wan F， Yue J， Xu S C and Ye Q X. 2022. Discrepant multiple instance learning for weakly supervised object detection. Pattern Recognition， 122： #108233 ［DOI： 10.1016/j. patcog. 2021.108233］
22. Kirillov, A., et al., Segment anything. arXiv preprint arXiv:2304.02643, 2023.
23. X. Li, A. You, Z. Zhu, H. Zhao, M. Yang, K. Yang, S. Tan, and Y. Tong,“Semantic flow for fast and accurate scene parsing,” in ECCV. Springer,2020, pp. 775–793