Thanks for your careful and valuable reviews. We will explain your questions point by point:

Reviewer #1:

**Q1:What does hetero-centric segmentation network, tandem feeding, L\_{BCE} mean?**

A1:

1. hetero-centric segmentation network：对于分割网络而言用于训练参数的数据只来源于一个中心的部分数据，而测试来源于多中心数据集，多中心数据的相关解释请参看文献[1]
2. tandem feeding意思是将region mapping feeding给FC1结果送给FC2，FC2的结果再送给FC3类似于物理上串联的电阻。
3. L\_{BCE}指的是标准的二分类交叉熵损失。

[1]Ali S, Jha D, Ghatwary N, Realdon S, Cannizzaro R, Salem OE, Lamarque D, Daul C, Riegler MA, Anonsen KV, Petlund A, Halvorsen P, Rittscher J, de Lange T, East JE. A multi-centre polyp detection and segmentation dataset for generalisability assessment. Sci Data. 2023 Feb 6;10(1):75. doi: 10.1038/s41597-023-01981-y. PMID: 36746950; PMCID: PMC9902556.

**Q2:In formula 9, can N be zero?**

A2:N不会为0，因为N指的是整张图像中经由CRM产生的proposals的个数，为了保证整张图像proposals的存在性我们在过滤后对剩余proposals的数量进行了判断，如果为0我们会在初始的候选框中随机选取几个候选框作为proposals送入后续网络。

**Q3:formula questions**

A3:非常感谢您指出的错误和受益匪浅的修改意见。我们已经重新检查并修正了文章相关公式的定义，这些修正将展示在文章的最终版中

Reviewer #2:

**Q1:How does the method address limitations depicted in Fig.2(A) if segmentation results are incorrect ?**

A1:我们非常同意您的担忧：当分割的结果不正确网络将会再次面对图2（A）所表述的限制。从本质上来讲，这个模块的意义是将proposals聚合到息肉所在位置的附近，所以在文中我们除了CRM模块以外还设计了SCM、DPS模块。其中SCM模块可以使网络更好的学习到息肉的形态类别，同时使网络更好地关注息肉所在的位置，因为这是对息肉进行形态分类的前提。这个模块会将网络的注意力聚焦在息肉上起到聚合息肉的作用，此时那些明显不同的proposals的预测得分将会降低，再经过DPS进行后处理这些限制也是能够克服的只是效果会有所下降。如我们在消融实验表2第7、8行所展示的。

The SCM module enables the network to focus better on the location of the polyps, reducing the prediction scores of significantly different proposals. Then, DPS module will filter them.

**Q2:What is the rationale behind the correlation between coordinates and polyp categories?**

A2:坐标与类别之间没有直接的关系，但是roi的坐标表示息肉可能所在的位置，这是对息肉进行分类的前提。因为如果想要对息肉进行形态分类那么首先要找到息肉所在的位置，而在息肉图像中息肉的位置是不固定的，因此需要坐标指出图像中息肉可能出现的位置。

**Q3:The multiple instance branch should be evaluated in the ablation study.**

A3:您的评论非常正确，MIL在弱监督乃至半监督训练中至关重要，然而对于我们的方法而言去掉MIL进行的消融实验是不可取的。这是因为如果去掉MIL此时proposals的学习模块（CRM+SCM）将会仅剩下SCM，这个模块会根据proposals对类别打分，也就是说一个proposal的所有类别概率加和将为1，这使得网络无法学习哪个proposal最有可能包含物体。而弱监督息肉检测不仅需要学习proposals里包含的是什么类型的息肉还需要在众多proposals找出哪个proposal最有可能包含息肉。

仅用MIL：这是因为对于每个类别MIL会按照proposals打分，简单的讲就是一张图像中每类所有proposals概率加和等于1，这导致网络无法进行优化学习，因为没有具体的注释指导它对同一类中不同proposals打分的正确性。

所以我们很抱歉您所提的消融实验在弱监督检测中是无法完成的。

**Q4:The weakly-supervised methods compared are outdated.It is recommended to**

**include more recent weakly-supervised detection methods.**

A4:感谢您指出的我们工作中的不足，我们已经在调研最近的弱监督检测相关工作，我们将会在最终版的论文中加入这些新的对比方法。

Reviewer #3:

**Q1: What is the evidence that pre-trained segmentation model must have learned some generic features of the polyp? Does the pre-training dataset contain any polyp images?**

A1:感谢您提出的具有建设性的意见。预训练模型学习到息肉通用的特征这个证据我们认为泛化性实验可以提现。因为模型只在ClinicDB数据集上训练，其他两个数据集对于他来说是全新且不可见的。因此模型无法事先学习这两个数据集的特征。但我们的实验表明模型在这种情况下依然有不错的性能，所以网络一定学习到了不同数据集中息肉的通用特征，例如边缘纹理等。

**Q2: Why do we not use the pretrained model directly?Could a better pretrained model solve this problem entirely?**

A2:感谢您提出的疑问和建议。但是直接使用预训练模型是不合适的，因为分割结果并不是完全准确的，就像Reviewer #2 Q1所指出的那样。如果出现预测结果不正确时，仅使用预训练的分割模型将无法纠正这个错误。因为分割结果是proposals框的唯一来源。此外，训练更好的预训练模型需要逐点标记的像素级标签，这与弱监督降低标注量的出发点相悖。这也是全监督息肉分割如何提高泛化性能的研究内容。但就我们目前所知，这个问题在这个领域目前还没有被很好的解决。

**Q3：Could the paper provide some independent validation of pre-trained model on the polyp detection task?**

A3:您指出的验证实验很有必要，我们将在最终版的论文中进行补充。这个实验将主要从两个方面展开：1）预训练模型在不同iou阈值下的正确定位率2）直接使用预训练模型进行弱监督息肉检测时的mAP

Reviewer #4:

**Q1: The proposed model mostly shows less competitative accuracy than DiffusionDet50 or DiffusionDet500.**

A1:我们很抱歉方法的性能不够显著。但我们认为这种比较是不公平的。因为检测的性能同时受到定位和分类准确度的影响。全监督方法可以同时提供定位和分类的学习引导但图像级弱注释目标检测没有明确的定位引导。因此弱监督要比全监督更为困难。所以我们认为这种比较是不合适的。此外，less competitative accuracy在自然图像领域上也有发生，例如文章[2]所提到的。

[2]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

**Q2:The training of the model needs the morphological category labels. Do these labels increase**

**the amount of annotation effort of physicians?**

A2:我们想解释：我们的模型只需要形态标签不需要病理标签。获取病理标签需要医生先将息肉切除然后送至化验室分析病理，之后才能进行标注，这需要大量的时间也增加了医生的工作量。而形态标签仅需要医生在发现息肉时通过肉眼就可以判断，这并不会增加医生工作量。

**Q3:Better explanations for advantages of Cross Reference module and Spatial Category module.**

A3:我们想补充说明一下，这些模块以及DPS模块的优势在补充材料中我们已经进行了分析，请参见补充材料图Figure S1、Figure S2、Figure S3.

**Q4:Why the performance Of YOLO or Faster RCNN appears not so competitive for polyps detection.**

A4:You have raised an important point,通过我们对数据集的分析和文献[3]所展示的结果，我们认为这可能是因为这两种法是基于Anchor的，他需要预先设置大量的参数例如scale和宽高比。这些参数是针对自然图像而设计的，运用到息肉上是不合适的。此外，与自然图像相比，息肉图像目标与背景的差异较小，分类更为困难准确率会因此有所降低。mAP不仅和定位准确度相关同时也和分类准确度相关因此他们的性能appears not so competitive。

typos and errors,paper editing error,formula error