

- Chairs have requested users to enter domain conflicts. Please [click here](#) to enter your conflict domains.

View Reviews

Paper ID

361

Paper Title

MDNet: Morphology-Driven Weakly Supervised Polyp Detection

Track Name

IJCAI2024 Main Track

Reviewer #1

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).

This abstract discusses the challenges in accurately detecting polyps, which are crucial indicators of potentially lethal colorectal cancer, in clinical settings due to morphological differences. Existing methods either require annotated bounding boxes, overlook unprepared polyp proposals, or provide incomplete predictions. Moreover, these methods primarily focus on the detection rate of polyps based on pathological classification. To address these issues, the authors propose the Morphology-Driven network (MDNet), which detects polyps using only image-level supervision. The MDNet incorporates a Cross-Domain Reference Module (CRM) to mitigate the negative effects of uncertain proposals and a Spatial Category Module (SCM) to enhance discrimination among similar polyps with different morphologies. Additionally, a dual-threshold post-processing strategy (DPS) integrates class and region scores to improve detection accuracy. Experimental results on three datasets demonstrate that MDNet offers better robustness and performance compared to existing methods. The authors have made the code available at an anonymous GitHub repository.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).

The performance in the weakly supervised domain.

The thorough experimental evaluation.

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.

The paper's clarity can be improved as it needs to be proofread to correct typos and errors. The notation is poor and can be improved to better explain the proposal. Some examples below:

If P is binary segmentation map, should it be defined as $P \in \{0,1\}^{h \times w}$

On line 238-239, I guess you select all the boxes that has a jc greater than the threshold. If so, the sentence is not correct.

Formula 3 is not correct, parameters are missing useful to compute the outcome. Moreover on line 240, it is stated that \hat{B} is a set of candidate boxes, while in formula 3 it is described as a function. Please fix it. Also on lines from 243 to 246 the definitions are wrong. What is \hat{B} on line 243? A set of candidate boxes or a single number? Use $B(b_i, :)$ with $b_i \in \hat{B}$ instead of i . Change $xmin$ (and the other coordinate) with $xmin_i$ (and the other coordinates as well).

Minor issues:

Figure 2 caption refers to letter from a) to d), but these are not available in the figure, please fix it.

On page 3, Given an Image $I \in \mathbb{R}^{C \times H \times W}$, it is better so specify what C , H , and W are.

On page 4, I don't know what is a hetero-centric segmentation network, can you use a citation for it?

Line 250, "pattern" -> "goal"

Line 253, "direct flatten will loss a lot of spatial information" -> "direct flattening will result in the loss of a lot of spatial information"

Line 278, what do you mean for tandem feeding? 串联喂食的意思是什么

Line 283, "which different from the input feature map channels" -> "which differs from the input feature map channels."

What is L_{BCE} in formula 8? 公式8中的 L_{BCE} 是什么

In formula 9, can N be zero?

Line 417, "an visualization" -> "a visualization"

Line 371, "it demonstrates the each" -> "it demonstrates that each"

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly influence your evaluation of the paper and 2) do not require new results.

can you briefly describe what is a hetero-centric segmentation network?

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines <https://ijcai24.org/reproducibility/>).

CREDIBLE: I believe that the obtained results can, in principle, be reproduced. Even though key resources (e.g., proofs, code, data) are unavailable at this point, the key details (e.g., proof

sketches, experimental setup) are sufficiently well described for an expert to confidently reproduce the main results, if given access to the missing resources.

Reviewer #2

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).

The paper presents a semi-supervised polyp detection method that relies solely on image-level annotations. The method incorporates a cross-domain reference module (CRM), a spatial category module (SCM), and a dual-threshold post-processing strategy (DPS). The proposed method outperforms several semi-supervised methods in terms of detection performance.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).

The proposed CRM and DPS can filter out some false positives with the assistance of a pre-trained segmentation model and threshold filtering.

The experiments demonstrate superior detection and generalization performance compared to several semi-supervised methods.

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.

1. CRM依赖于预先训练的分割模型的分割结果来消除明显不同的提议。然而，分割结果也可能是不正确的，因为它们面临着与图2（A）所示相同的限制。该方法如何解决这个问题？

1. The CRM relies on the segmentation results of a pre-trained segmentation model to eliminate significantly different proposals. However, the segmentation results could also be incorrect, as they face the same limitations depicted in Fig.2(A). How does the method address this issue?

2. The proposed MIB translates ROI coordinates into probability scores across different categories. What is the rationale behind the correlation between coordinates and polyp categories?

3. The SCM, being a common operation in the detection field, is difficult to consider as a contribution. Instead, the multiple instance branch, which is crucial for semi-supervised training, should be evaluated in the ablation study. SCM作为检测领域的一种常见操作，很难被认为是一种贡献。相反，对于半监督训练至关重要的多实例分支应在消融研究中进行评估。

4. The weakly-supervised methods compared are outdated (prior to 2019). It is recommended to include more recent weakly-supervised detection methods. 所比较的弱监督方法已经过时（2019年之前）。建议包括更新的弱监督检测方法。

5. The process of converting a proposal feature with D channels into a classification score with C channels using a single softmax operator (Eq.4&5, Eq.6&7) is unclear. It appears that an operation is missing. 使用单个softmax算子（等式4和5，等式6和7）将具有D个通道的提议特征转换为具有C个通道的分类分数的过程尚不清楚。似乎缺少一个操作

6. The paper contains several errors, such as:

- The detection visualizations in Fig.1(b) under two supervision are identical. 在两个监督下，图1（b）中的检测可视化是相同的

- Red arrows, which should denote post-processing operations mentioned in the caption, are missing in Fig.3. 图3中缺少红色箭头，该箭头应表示标题中提到的后处理操作

- Missing k in the denominator and unclear B in Eq.7.

- 'WSPD' in Fig.5 is incorrect.

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly

2. 所提出的MIB将ROI坐标转换为不同类别的概率得分。坐标和息肉类别之间的相关性背后的原理是什么？

influence your evaluation of the paper and 2) do not require new results.

Please refer to the weaknesses.

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines <https://ijcai24.org/reproducibility/>).

CONVINCING : I am convinced that the obtained results can be reproduced, possibly with some effort. Key resources (e.g., proofs, code, data) are already available, will be made available upon acceptance, or good reasons as to why they are not (e.g., proprietary data or code) are reported in the paper. Key details (e.g., proofs, experimental setup) are sufficiently well described but their exact recovery may require some work.

Reviewer #3

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).

The paper studies detecting polyps, one of the early signs of colorectal cancers, from endoscopic images. Detecting polyps can lower the rate of cancer, and hence reliable detection is an important task.

The paper's main contributions are:

1. Proposing MDNet, a morphology driven new polyp detection mechanism.
2. Various components of MDNet, including a cross-domain reference module and spatial category module, which can have impact outside the immediate field of polyp detection.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).

1. The paper provides detailed study of a complex and important problem — polyp detection.
2. Proposes multiple independent mechanisms for improving polyp detection, namely cross-domain reference module, spatial category module, multiple instance branch and dual threshold post-processing strategy. I suspect these methods can be insightful for other medical image tasks as well.
3. MDNet achieves very strong empirical performance with minimal supervision, performing much better than the weakly supervised baselines, and only being surpassed by diffusion model based baselines.
4. The paper provides thorough ablation studies on the various components of MDNet, showing their relative contribution to the performance.

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.

The paper has a lot of typos, for example, only to name a few:

1. Line 72, various morphologies polyp

2. Line 121, unaccurate
3. Line 144, proposaled
4. Line 166, we employs

Moreover:

Cross-domain reference module:

文章提到，预先训练的分割模型必须学习息肉的一些通用特征。有什么证据可以证明这一点？预训练数据集是否包含任何息肉图像？该论文能否为该模型在息肉检测任务中提供一些独立的验证？

1. The paper mentions that the pre-trained segmentation model must have learned some generic features of the polyp. What is some evidence for this? Does the pre-training dataset contain any polyp images? Could the paper provide some independent validation of this model on the polyp detection task?
2. I don't understand why we need a separate model for predicting binary masks B, if we only use masks B that match with the prediction from the pretrained model, P, to a high degree (by thresholding on the Jaccard coefficient). Why do we not use the pretrained model directly? How much performance boost do we get by using a separate model over simply using this pretrained model? Could a better pretrained model solve this problem entirely?

我不明白为什么我们需要一个单独的模型来预测二进制掩码B。如果我们只使用在很大程度上与来自预训练模型P的预测匹配的掩码B（通过对Jaccard系数进行阈值处理）。为什么不直接使用预训练的模型？与简单地使用这个预训练的模型相比，使用单独的模型可以提高多少性能？一个更好的预训练模型能完全解决这个问题吗？

Spatial Category Module:

这似乎只是在平坦层之前添加一个GAP层。这已经在先前的工作[1]中完成了，特别是为了找出类激活映射（CAM），并且不是新颖的。

1. This seems to be just adding a GAP layer before a flattening layer. This has been done in prior work [1], specifically to find out class-activation maps (CAM), and is not novel.
2. Why does FC represent a flatten layer, and not a fully-connected layer? The terminology needs to be fixed. 为什么FC代表扁平层，而不是完全连接的层？术语需要固定

[1] Learning Deep Features for Discriminative Localization, <https://arxiv.org/abs/1512.04150>

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly influence your evaluation of the paper and 2) do not require new results.

Please check weaknesses listed above.

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines <https://ijcai24.org/reproducibility/>).

CONVINCING : I am convinced that the obtained results can be reproduced, possibly with some effort. Key resources (e.g., proofs, code, data) are already available, will be made available upon acceptance, or good reasons as to why they are not (e.g., proprietary data or code) are reported in the paper. Key details (e.g., proofs, experimental setup) are sufficiently well described but their exact recovery may require some work.

Reviewer #4

Questions

1. Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).

This paper presents a model to detect polyps based on morphology-driven weakly supervised learning.

It considers some weaknesses of existing models and uses special some strategies to over the weaknesses.

2. What are the main strengths of the paper? Please focus on novelty (how novel are the concepts, problems addressed, or methods introduced in the paper), soundness (is the

paper technically sound), clarity (is the paper well-organized and clearly written), and significance (comment on the likely impact the paper may have on the AI research community as a whole or on its own sub-field).

Strengths:

- + It proposes a morphology-driven weakly supervised model as opposed to pathology-driven ones in the literature
- + AP on different morphology is used to evaluate the performance
- + The designed method considers the cross-domain reference and spatial category

3. What are the weaknesses of the paper? Please focus on novelty, soundness, clarity, and significance.

Weaknesses:

- The performance appears to be not so impressive, e.g., judging from Table 1. The proposed model, MDNet, mostly shows less competitive accuracy than supervised models with Box, such as DiffusionDet50 or DiffusionDet500, though it shows improvement over models without using Box or Faster R-CNN and YOLOv3. 性能似乎并不那么令人印象深刻

- For a conceptual point of view, the training of the model needs the morphological category labels. So instead of frame-level pathological labels, morphological labels are also needed. These labels appear to increase the amount of labors for annotations. 从概念的角度来看，模型的训练需要形态学类别标签。因此，还需要形态学标签来代替帧级的病理标签。这些标签似乎增加了注释的工作量。

- The presentation of the paper still has a lot of issues:

It still contains a lot of grammatical or editing errors, which hinder readers from smooth reading and understanding.

For example, line 18: "... (SCM) is designed, who enhances the ability..."

"who" should be "which".

For another example, in Fig 1 sub-figure (b), the left plot (pathological supervision) and the right plot (morphological supervision) appear to be exactly the same. This appears to be an editing error. 例如，在图1的子图（b）中，左侧图（病理学监督）和右侧图（形态学监督）看起来完全相同。这似乎是一个编辑错误

- Figure 2 is pretty complex. Though the figure caption mentions about parts (1)-(d); I could find Figure 2 (a-c), but I couldn't find (d) part for the contributions. Do I miss that part or do you miss that part? 图2相当复杂。尽管图中提到了第（1）-（d）部分：我可以找到图2（a-c），但我找不到贡献的（d）部分。我错过了那个部分还是你错过了那个？

- A minor suggestion: Though I believe that the claimed novelties regarding Cross Reference module and Spatial Category module are okay, it is not so clear regarding what advantages they may produce and why they were not used in the literature yet. Better explanations may help readers appreciate the novelties or the contributions. 尽管我认为所声称的关于交叉参考模块和空间类别模块的新颖性是可以的，但它们可能产生什么优势以及为什么还没有在文献中使用还不太清楚。更好的解释可以帮助读者欣赏这些新颖之处或贡献。

4. Please carefully list the questions that you would like the authors to answer during the author-feedback period. List only questions about specific issues that 1) could directly influence your evaluation of the paper and 2) do not require new results.

1. Briefly speaking, why do you think that morphology can help with detection of polyps? YOLO or Faster-R-CNN can detect objects of different shapes/morphologies but their performance appears not so competitive for polyps detection. 简单地说，为什么你认为形态学可以帮助检测息肉？YOLO或Faster-R-CNN可以检测不同形状/形态的物体，但它们的性能在息肉检测方面似乎没有那么竞争力

2. The training of the model needs the morphological category labels. Do these labels increase the amount of annotation effort of physicians? 模型的训练需要形态学类别标签。这些标签是否增加了医生的注释工作量？

5. Are the results of this paper easily reproducible? (Please refer to our reproducibility guidelines <https://ijcai24.org/reproducibility/>).

CREDIBLE: I believe that the obtained results can, in principle, be reproduced. Even though key resources (e.g., proofs, code, data) are unavailable at this point, the key details (e.g., proof sketches, experimental setup) are sufficiently well described for an expert to confidently reproduce the main results, if given access to the missing resources.