MDNet: Morphology-Driven Weakly Supervised Polyp Detection

Supplementary Material

******

1 Ablation experiments

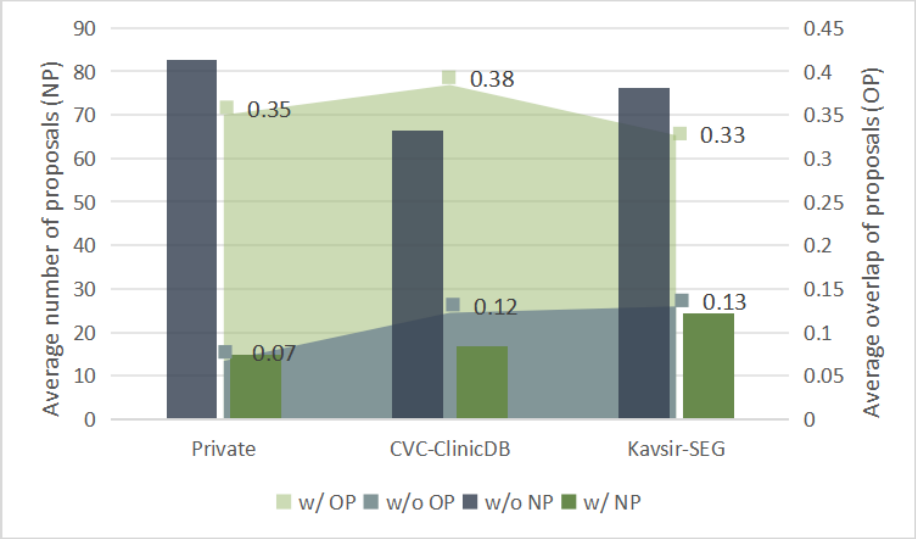

Fig S1: Visualization qualitative results of ablation study in CRM.

Due to the fact that weak supervised networks lack the supervision of instance bounding boxes, the bounding boxes do not have clear guidance directions,so the boxes in the predicted results only come from the proposal generator. However, in practical, the number of negative examples generated by the proposal generator far exceeds that of positive cases, which makes it difficult for the network to accurately learn the features of positive examples among numerous negative examples, thereby reducing the ability of polyp detection. Thus, it is necessary to previously filter before feeding them to network for training. To evaluate the effectiveness of modules, we compared the average number of proposals (NP) with or without CRM, as well as their average overlap (OP, intersection over union) with ground truth. As shown in the Fig S1, when we add CRM, the number of generated proposals decreased from 83 to 23 on private dataset, while the overlap increased by $\Delta + 4.22$. In addition, we also compared CVC-ClinicDB and Kvasir-SEG (in the last two columns of Fig S1), NP decreased from 66 to 17 (CVC-ClinicDB), from 76 to 24 (Kvasir-SEG), while OP increased by $\Delta + 2.14$ and $\Delta + 1.51$, respectively.

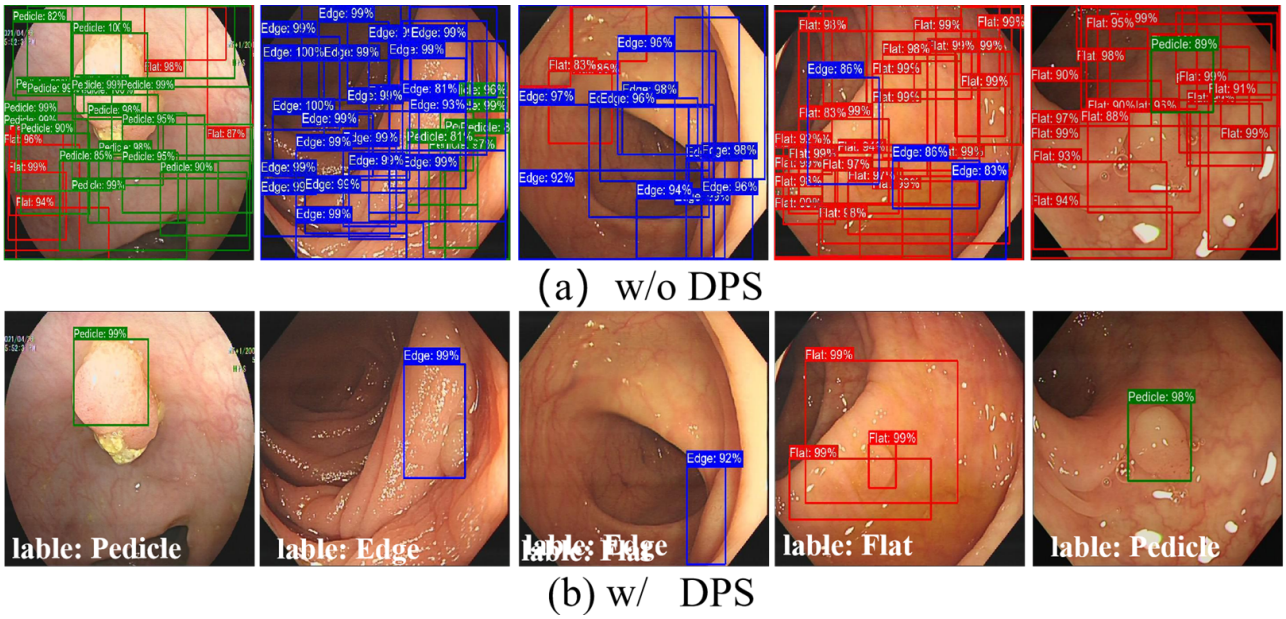

Fig S2: Visualization qualitative results of ablation study in DPS.

Through our carefully designed SCM and CRM, the proposals used for training will cover at least part of the polyp, which makes the category scores of all candidate boxes very similar, so it will be difficult to determine the final prediction box only by the category threshold. Even more unfortunately, boxes containing a certain background (such as 55%) that contain more category information than boxes that tightly surround the polyp, thereby tending to get higher category scores [4]. In brief, if we rely solely on the category threshold to determine prediction results, it will cause a large number of false positives false positives (as shown in Fig S2 (a)). Notely, unlike scores for the focus category, the region scores focus more on the clinginess and integrity of the target within the box. In other words, boxes with local object or more background will get a lower score. Thus, there is a natural complementarity between regional scores and category scores. Based on this, we put forward the DPS, first of all, threshold filter by category contains specific categories box, and then regional score will choose the box that tightly surround the polyp. As shown in Fig S2 (b), when we adopt the DPS strategy for post-processing, not only the false positives are greatly reduced, but also the contribution of DPS is proved.

2 Generalization capability

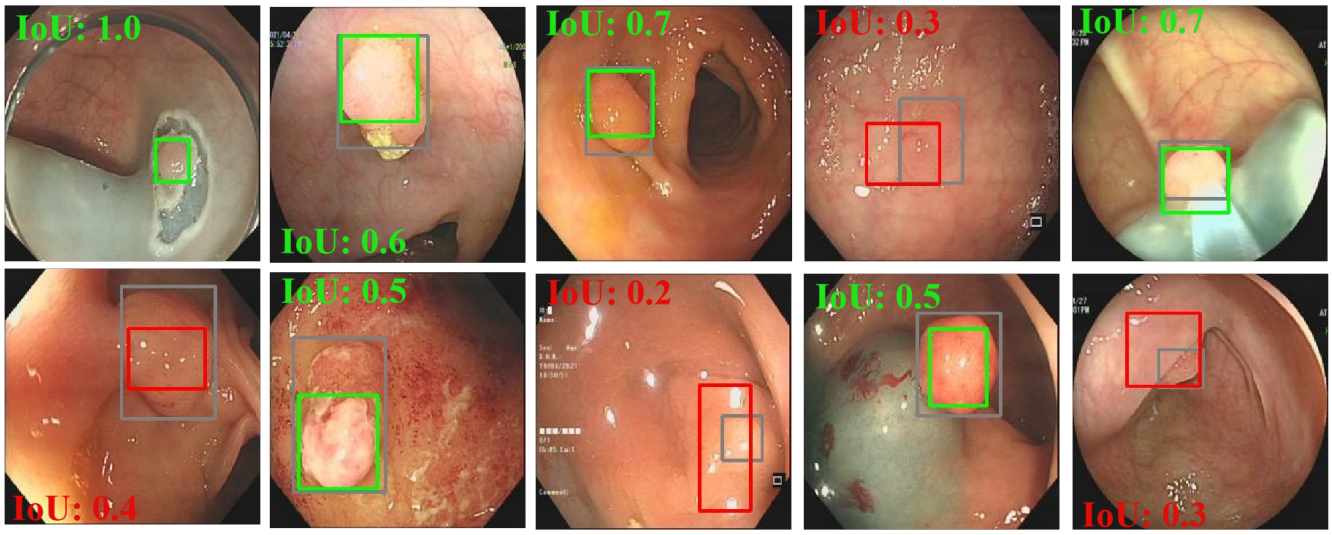

Fig S3: Some detection results for different polyp morphologies. Green rectangle indicates success cases (IoU >0.5), and red rectangle indicates failure cases (IoU <0.5).

Some success and failure detection results by MDNet, as shown in Fig S3. We can observe that, MDNet is robust to the size and aspect of polyps. The main failures are always due to overlarge boxes that not only contain objects, but also include their adjacent similar background. For non-rigid polyps, they typically exhibit considerable deformation and are more deformation of their most representative parts, so our detector unable to find these parts. An ideal solution is yet in pursuit because there exists potential for further refinement.

3 Learning Ability Supplementary Material

Methods	Supervision		IoU@10			IoU@30			IoU@50		
	Box	Class	Flat	Pedicle	Edge	Flat	Pedicle	Edge	Flat	Pedicle	Edge
Faster Rcnn [6]	✓	✓	–	77.50	46.15	–	77.50	46.15	–	76.20	46.15
YOLO [5]	✓	✓	–	82.47	39.14	–	82.47	38.41	–	81.54	35.36
DiffusionDet50 [2]	✓	✓	–	98.05	92.19	–	98.05	92.19	–	96.49	81.77
DiffusionDet500 [2]	✓	✓	–	94.26	91.39	–	94.25	91.39	–	94.25	82.49
WSDDN [3]	–	✓	14.42	23.04	12.97	6.79	16.27	4.93	5.65	11.45	2.99
OICR [8]	–	✓	–	5.20	7.69	–	1.10	–	–	0.37	–
WSOD2 [9]	–	✓	26.97	51.57	34.04	4.72	38.63	34.04	2.97	**38.63**	**34.04**
Grad-CAM [7]	–	✓	32.06	74.49	19.78	3.74	34.99	–	0.12	11.70	–
Grad-CAM++ [1]	–	✓	26.70	73.17	45.74	3.94	30.47	4.62	0.05	10.92	2.37
Ours	–	✓	**77.01**	**68.10**	**76.92**	**63.96**	**49.48**	**51.05**	**55.92**	15.21	30.91

Table S1: Average precision (in %) for different methods on CVC-ClinicDB test set. The best results of weakly supervised methods are marked in bold.

Methods	Supervision		IoU@10			IoU@30			IoU@50		
	Box	Class	Flat	Pedicle	Edge	Flat	Pedicle	Edge	Flat	Pedicle	Edge
Faster Rcnn [6]	✓	✓	–	79.69	–	–	78.67	–	–	77.07	–
YOLO [5]	✓	✓	–	75.78	7.04	–	74.62	7.04	–	70.94	6.53
DiffusionDet50 [2]	✓	✓	–	74.23	11.53	–	72.28	11.53	–	70.19	9.13
DiffusionDet500 [2]	✓	✓	–	70.73	10.79	–	69.71	10.79	–	66.90	3.66
WSDDN [3]	–	✓	5.30	17.07	1.39	**3.08**	14.28	–	**2.79**	11.36	–
OICR [8]	–	✓	–	–	–	–	–	–	–	–	–
WSOD2 [9]	–	✓	–	4.80	–	–	4.61	–	–	4.55	–
Grad-CAM [7]	–	✓	0.71	36.53	–	0.08	10.81	–	–	2.25	–
Grad-CAM++ [1]	–	✓	0.36	36.85	–	0.10	10.90	–	0.01	2.78	–
Ours	–	✓	**4.57**	**60.69**	**12.50**	1.00	**41.27**	**12.50**	0.50	**19.23**	**12.50**

Table S2: Average precision (in %) for different methods on Kvasir-SEG test set The best results of weakly supervised methods are marked in bold.

Methods	Supervision		IoU@10			IoU@30			IoU@50		
	Box	Class	Flat	Pedicle	Edge	Flat	Pedicle	Edge	Flat	Pedicle	Edge
Faster Rcnn [6]	✓	✓	–	82.77	–	–	82.77	–	–	82.77	–
YOLO [5]	✓	✓	–	73.58	17.12	–	73.58	15.32	–	73.58	11.69
DiffusionDet50 [2]	✓	✓	–	80.16	33.30	–	80.12	32.64	–	78.73	31.61
DiffusionDet500 [2]	✓	✓	–	83.77	36.42	–	83.57	34.25	–	81.45	20.97
WSDDN [3]	–	✓	8.33	29.90	5.74	4.53	27.45	2.85	3.02	25.78	2.74
OICR [8]	–	✓	3.71	–	–	3.71	–	–	3.71	–	–
WSOD2 [9]	–	✓	–	3.03	–	–	3.03	–	–	3.03	–
Grad-CAM [7]	–	✓	–	68.60	2.78	–	39.87	–	–	31.34	–
Grad-CAM++ [1]	–	✓	–	62.03	2.92	–	39.44	–	–	29.89	–
Ours	–	✓	**35.81**	**68.65**	**31.67**	**20.83**	**68.65**	**18.89**	**15.49**	**55.58**	**18.89**

Table S3: Average precision (in %) for different methods on internal test set The best results of weakly supervised methods are marked in bold.

All comparison methods are evaluated on CVC-ClinicDB, Kvasir-SEG and internal dataset as shown in Table S1, Table S2, Table S3 in terms of average precision for each category. Among weakly supervised methods, our method outperforms others on the most categories under different IoU thresholds. Especially, our method performs much better than the state-of-the-arts on ”Flat” and ”Edge”, as our approach has stronger classification ability while maintaining localization ability, though in most cases instances of these categories are extremely obvious or camouflage. Besides, the performance of our weakly supervised method is even comparable with the fully supervised methods in some aspects (the AP of flat on three datasets under different IoU thresholds), illustrating the effectiveness of the proposed MDNet.

References

[1] Aditya, C., Anirban, S., Prantik, H., N, B.V.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: IEEE Winter Conference on Applications of Computer Vision. pp. 839–847 (2018)

[2] Chen, S., Sun, P., Song, Y., Luo, P.: Diffusiondet: Diffusion model for object detection. arXiv preprint arXiv:2211.09788 (2022)

[3] Hakan, B., Andrea, V.: Weakly supervised deep detection networks. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2846–2854 (2016)

[4] Patel, K., Li, K., Tao, K., Wang, Q., Bansal, A., Rastogi, A., Wang, G.: A comparative study on polyp classification using convolutional neural networks. PLoS ONE **15**(7) (2020)

[5] Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

[6] Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(6), 1137–1149 (2017)

[7] Selvaraju, R.R., Michael, C., Abhishek, D., Ramakrishna, V., Devi, P., Dhruv, B.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision. pp. 618–626 (2017)

[8] Tang, P., Wang, X., Bai, X., Liu, W.: Multiple instance detection network with online instance classifier refinement. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3059–3067 (2017)

[9] Zeng, Z., Liu, B., Fu, J., Chao, H., Zhang, L.: Wsod2: Learning bottom-up and top-down objectness distillation for weakly-supervised object detection. In: IEEE International Conference on Computer Vision. pp. 8291–8299 (2019)