Computers, Materials & Continua CMC, vol.63, no.3, pp.1263-1272, 2020

A Polyp Detection Method Based on FBnet

Jingjing Wan[1], Taiyue Chen[2, *], Bolun Chen[2, 3, *], Yongtao Yu[2], Yiyun Sheng[2] and Xinggang Ma[1]

Abstract: The incidence of colorectal cancer (CRC) in China has increased in recent years. The mortality rate of CRC has become one of the highest among all cancers; CRC increasingly affects the health and quality of people's lives. However, due to the insufficiency of medical resources in China, the workload on medical doctors has further increased. In the past few decades, the adult CRC mortality and morbidity rate dropped sharply, mainly because of CRC screening and removal of adenomatous polyps. However, due to the differences in polyp itself and the skills of endoscopists, the detection rate of polyps varies greatly. In this paper, we adopt an anchor-free mechanism and introduce a better method to factorize the process of bounding box regression. Firstly, we regress the shape of object by the variant of Faster RCNN. Secondly, we re-define the target function of the location of object. The experimental result shows that our method achieves a mAP of 55.8%, which outperforms other state-of-the-art methods by at least 11.9%. This will greatly help to reduce the missed diagnosis of clinicians during endoscopy and treatment, and provide effective help for early diagnosis, early treatment and prevention of CRC.

Keywords: Colorectal cancer, polyp detection, anchor free, two step decomposition.

1 Introduction

Colorectal cancer (CRC) is one of the common malignant tumors in China. With the continuous development of people's living standards and dietary habits, the incidence and mortality of colorectal cancer keep rising in recent years [Society and Society (2018)], which seriously endangers the health and living quality of people. CRC has become a major public health problem due to its high morbidity and high mortality.

According to statistics, CRC is the second and third leading cause of death in men and women, respectively [Society and Society (2018)]. In addition, a recent study reported a significant increase in the annual percentage of CRC incidence among young people [Bailey, Hu, You et al. (2015)]. In clinical diagnosis, colonoscopy plays an important role in the screening of CRC [Rex, Boland, Dominitz et al. (2017)]. The use of colonoscopy

[1] Department of Gastroenterology, Affiliated Huaian Hospital of Xuzhou Medical University, The Second People's Hospital of Huaian, Huaian, 223002, China.

[2] College of Computer Engineering, Huaiyin Institute of Technology, Huaian, 223003, China.

[3] University of Fribourg, Fribourg, 1700, Switzerland.

[*] Corresponding Authors: Bolun Chen. Email: chenbolun1986@163.com;

Taiyue Chen, Email: taiyuechen@foxmail.com.

Received: 11 February 2020; Accepted: 24 February 2020.

to reduce CRC mortality and incidence is mainly due to the ability to detect polyps/adenomas [Brenner, Chang-Claude, Jansen et al. (2014)] and remove them by resection [Doubeni, Corley, Quinn et al. (2018); Brenner, Chang-Claude, Jansen et al. (2014)]. In addition, there is evidence that for every 1.0% increase in adenoma detection rate (ADR), the risk of interphase CRC is reduced by 3.0% [Corley, Jensen, Marks et al. (2014); Kaminski, Regula, Kraszewska et al. (2010)]. In the past few decades, the adult CRC mortality and morbidity rate dropped sharply (reduced by 51% and 32%, respectively), mainly because of CRC screening and removal of adenomatous polyps [Burke, Kaul and Pohl (2017)]. However, due to the differences in polyp itself and the skills of endoscopists, the detection rate of polyps varies greatly [Shaukat, Oancea, Bond et al. (2009)], and, in some cases, polyps may be missed by the diagnosis, and the rate of missed diagnosis is as high as 27% [Mahmud, Cohen, Tsourides et al. (2015); Ahn, Han, Bae et al. (2012)]. Thus, unrecognizable polyps in the field of view during colonoscopy are an important issue [Mahmud, Cohen, Tsourides et al. (2015)]. Some studies have shown that the second observer's observations increase the polyp detection rate (PDR), but such strategies are still controversial in improving adenoma detection rate (ADR) [Aslanian, Shieh, Chan et al. (2013); Buchner, Shahid, Heckman et al. (2011)].

At present, the medical industry has incorporated more high-technology such as computer sciences and sensor technology, making medical services become more intelligent and precise. With the latest breakthroughs in artificial intelligence, especially the development of deep learning (DL), computer-aided diagnosis (CADx) of polyps during colonoscopy has attracted wide attention [Chen, Lin, Lai et al. (2018); Byrne, Chapados, Soudan et al. (2019); Fang, Cai, Sun et al. (2018)]. Deep belief network studied by Wan et al. [Wan, Chen, Kong et al. (2019)] is adopted to help doctors to detect the early intestinal cancer.

The ultimate goal of a real-time automatic polyp detection system is to assist endoscopic detection of polyp lesions. Although several automated polyp detection systems have been developed over the past decade [Tajbakhsh, Gurudu and Liang (2015); Misa-wa, Kudo, Mori et al. (2018)], there is a lack of the ability of this technique to locate and track polyps in clinical practice during on-site colonoscopy.

This paper proposes an anchor-free and two-steps-decomposition method to improve the detection rate of polyps/adenomas. This will greatly help to reduce the missed diagnosis of clinicians during endoscopy and treatment, and provide effective help for early diagnosis, early treatment and prevention of CRC.

2 Description of the problem

In this paper, to model the correlation between polyp and bounding box in a small dataset, we propose a novel method to enormously reduce the difficulty of learning and risk of overfitting. The whole process of polyp detection is shown in Fig. 1.

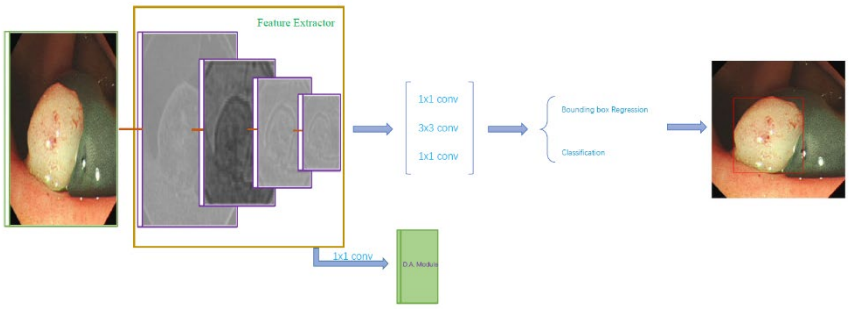

Figure 1: The whole polyp detection architecture. Keyword N×N Conv represents standard convolution operation with NxN kernel

3 Detailed process of the algorithm

3.1 Anchor-free mechanism

In recent years, mainstream literatures usually use anchor-based methods to regress bounding box (i.e., researchers regress the deltas between an anchor box and a ground-truth box instead of directly regressing the ground-truth box.). Although anchor-based methods reduce the difficulty of regression, it introduces some drawbacks. Firstly, in order to improve recall as much as possible, lots of anchor boxes are defined to guarantee to capture all ground-truth boxes in an image. It not only causes that most of anchor boxes have few IOU with the ground-truth boxes but also brings in a huge class imbalance between positive and negative anchor boxes. Secondly, researchers must use prior knowledge to design anchor boxes, including scales and aspect ratios of anchor boxes. Obviously, it is hard to choose a set of appropriate anchor boxes in a dataset.

In our experiment, due to drastic variations in the sizes and aspect ratios of ground-truth boxes, we have no choice but to define more anchor boxes. Unfortunately, it greatly increases the number of parameters and finally leads to over-fitting. To alleviate it, we employ an anchor-free mechanism which means that we must directly regress ground-truth boxes without the help of anchor boxes. However, as you see in YOLO-v1 [Brenner, Chang-Claude, Jansen et al. (2016)], it performs so badly when it tries to directly regress ground-truth boxes. To fix this problem, we, in the next section, introduce two steps for bounding box regression to reduce the difficulty of learning without increasing the number of parameters.

3.2 Two steps of bounding box regression

In consideration of the questions mentioned above, we factorize the bounding box regression into two steps. First, we regress the shape of object by the variant of bounding box regression in Faster RCNN. Second, we re-define the target function of locating objects.

3.2.1 Object shape regression

To reduce the learning difficulty, we take the following variants of bounding box regression in RCNN [Ren, He, Girshick et al. (2015)]:

$$\mathrm{G_w} = \lambda \mathrm{e}^{\mathrm{P_w}} \tag{1}$$

$$\mathrm{G_h} = \lambda \mathrm{e}^{\mathrm{P_h}} \tag{2}$$

Let G_w, G_h, P_w, P_h, respectively, denote the width and height of the ground-truth and the predicted boxes, and λ is a normalized factor.

3.2.2 Object location regression

Notably, it is bad to couple object shape (i.e., h, w) with object location (i.e., x, y) like RCNN [Ren, He, Girshick et al. (2015)]. If the object shape has huge variation, object location also has the same nasty properties. To address this issue, we decompose it into two steps. First, we predict each position to identify whether or not it contains the center point. Second, we predict the location offsets relative to the center point.

We convert a location (x, y) in an image to (x/s, y/s) in a feature map where s is the down-sample factor. However, x/s and y/s are not exactly integers. Hence, we have two choices.

In the first choice, we simply map (x/s, y/s) to (⌊x/s⌋, ⌊y/s⌋). Then, we predict the offset relative to (x/s, y/s) as follows:

$$offset_x = \frac{x}{s} - \left\lfloor \frac{x}{s} \right\rfloor, offset_y = \frac{y}{s} - \left\lfloor \frac{y}{s} \right\rfloor \tag{3}$$

where (x, y) is the center point of an object in the heatmap.

In the second choice, we map (x/s, y/s) to (⌊x/s+0.5⌋, ⌊y/s+0.5⌋). Then, we predict the offset relative to (x/s, y/s) as follows:

$$offset_x = \left\lfloor \frac{x}{s} + 0.5 \right\rfloor - \frac{x}{s}, offset_y = \left\lfloor \frac{y}{s} + 0.5 \right\rfloor - \frac{y}{s} \tag{4}$$

3.3 Loss function

The loss function of FBnet contains two parts (i.e., the loss in two steps of bounding box regression.). In our experiment, we simply view the two parts as regression problems. For the first step, we assign Smooth-L1 loss to object shape as in RCNN [Ren, He, Girshick et al. (2015)] for being robust to outliers. For the second step, we adopt MSE loss for both object categories and object locations. Consequently, the total loss is formulated as follows:

$$L = \frac{1}{N_{obj}} \sum_i 1_i^{obj} \{L_{cls} + L_{offset} + L_{loc}\} + \frac{1}{N_{noobj}} \sum_j 1_j^{noobj} \left(O_j^{gt} - O_j^{pred}\right)^2 \tag{5}$$

$$L_{cls} = \left(O_i^{gt} - O_i^{pred}\right)^2 \tag{6}$$

$$L_{offset} = \left(offset_i^{gt} - offset_i^{pred}\right)^2 \tag{7}$$

$$L_{loc} = SmoothL1\left(\{w_i^{pred}, h_i^{pred}\}, \{w_i^{gt}, h_i^{gt}\}\right) \tag{8}$$

where 1_i^{obj} denotes whether an object appears in a cell, and N_{noobj} and N_{obj} are two zoom factors.

3.4 Training details

In this paper, we take VGG-16 [Simonyan and Zisserman (2014)] as our feature extractor.

3.4.1 Iterating training

We take two steps training to optimize our networks: (1) freeze all layers in VGG-16 [Simonyan and Zisserman (2014)] and train customed layers by the total loss, (2) freeze the front layers in VGG-16 [Simonyan and Zisserman (2014)] and fine-tune them.

3.4.2 Super-parameter setting

In YOLO-v2 Redmon et al. [Redmon and Farhadi (2017)], they explored that a low-resolution classifier cannot extract robust features from high resolution images. For the sake of convenience, the input resolution in all experiment is resized as 224×224.

In the first step, we freeze all layers in VGG-16 Simonyan et al. [Simonyan and Zisserman (2014)] and add detection head whose parameters are randomly initialized. We use Adam to train the customed layers for 20 epochs with an initial learning rate of 5e-3 which is divided by 10 separately at 5 and 17 epochs.

In the second step, we use the weights from Step 1 to initialize all layers and take Adam to train it for 3 epochs with a very tiny learning rate of 1e-5.

3.4.3 Data augmentation

All data augmentation methods are merely random cropping and horizontal flip-ping. In this experiment, we leverage many kinds of data augmentation approaches. However, most of them have negative effects on mAP. It even causes misconvergence. When we check the feature map, we found that it is so bad and hard to find a potential pattern in it. As we know, human body environment is simple and structured. For example, almost all the colors of colonoscopy pictures are red. When we use color jittering, it may generate blue or other color images to train, intro-duce extraneous noise and finally make the network confused of this noise and hard to convergence.

3.5 Inference details

3.5.1 Inference post-procession

(1) Filter all boxes whose confidence are lower than α.

(2) Filter all boxes whose area are lower than β.

(3) Run non-maximum suppression with threshold γ.

3.5.2 Evaluation metrics

The results of our methods are measured with Mean Average Precision (mAP) as in Faster RCNN [Ren, He, Girshick et al. (2015)].

4 Experimental results and analysis

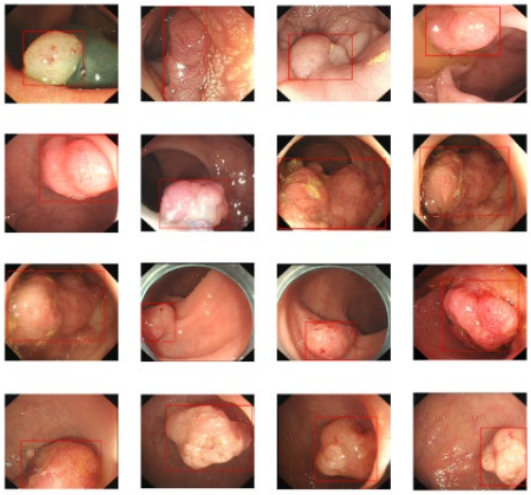

Figure 2: Some samples from our polyp dataset

4.1 Dataset

In this polyp dataset, it contains 201 colonoscopy pictures. As shown in Fig. 2, Each picture has one to three objects. There are 150 for training and 51 for testing. For preventing data leaking, all images from one patient belong to either training set or testing set.

4.2 Super-parameter sensitive experiment

First, we verify the effectiveness of Step 1 with different hyper-parameters λ. All experimental results are shown in Tab. 1. Especially, the scores are very close to each other when the hyper-parameter λ is set as 3, 4, 5, 6, and 7.

Table 1: The effects of different λ in the first step

λ	1	2	3	4	5	6	7	8
mAP	0.431	0.454	0.456	**0.466**	0.463	0.455	0.457	0.443

For verifying the effect of λ in the first step, all models in Tab. 1 share the same architecture and use the same regression (i.e., $G_x = G_w * P_x$, $G_y = G_h * P_x$, $G_w = \lambda * \mathrm{e}^{\mathrm{Pw}}$, $G_h = \lambda * e^{P_h}$) without the participation of the second step. Notably, when λ=1, it is similar to the target function of Faster RCNN [Ren, He, Girshick et al. (2015)]. As the increment of λ, mAP is slightly advancing. When λ=4, FBnet achieves the top result with the mAP of 0.466. It proves that λ reduces the difficulty of learning and ultimately improves the performances.

Then we test our Eqs. (3) or (4) under different settings of λ. All results are shown in Tab. 2.

Table 2: The effects of combining two steps under different settings of **λ**

λ	1	2	3	4	5	6	7	8
Eq. (3)	0.452	0.489	0.513	**0.541**	0.522	0.537	0.522	0.511
Eq. (4)	0.455	0.488	0.523	**0.558**	0.530	0.542	0.539	0.528

For checking the effect of combining two steps, all models in Tab. 2 take the same regression (i.e., $G_x = x + offset_x, G_y = y + offset_y, G_w = \lambda * \mathrm{e}^{\mathrm{P_w}}, G_h = \lambda * e^{P_h}$). In Eq. (3), it converts all points in a rectangle whose left-top point locates in $(\left\lfloor\frac{x}{s}\right\rfloor * s,\ \left\lfloor\frac{y}{s}\right\rfloor * s)$ and right-bottom point locates in $(\left\lfloor\frac{x}{s}\right\rfloor * s + s,\ \left\lfloor\frac{y}{s}\right\rfloor * s + s)$ to $(\left\lfloor\frac{x}{s}\right\rfloor,\ \left\lfloor\frac{y}{s}\right\rfloor)$. However, Eq. (4) maps all points in a rectangle whose left-top point locates in $(\left\lfloor\frac{x}{s}\right\rfloor * s - 0.5 * s,\ \left\lfloor\frac{y}{s}\right\rfloor * -0.5 * s)$ and right-bottom point locates in $(\left\lfloor\frac{x}{s}\right\rfloor * s + 0.5 * s,\ \left\lfloor\frac{y}{s}\right\rfloor * s + 0.5 * s)$ to $(\left\lfloor\frac{x}{s}\right\rfloor, \left\lfloor\frac{y}{s}\right\rfloor)$. In Eq. (3), a point $(\left\lfloor\frac{x}{s}\right\rfloor * s - 1,\ \left\lfloor\frac{y}{s}\right\rfloor * s - 1)$ must be mapped to $(\left\lfloor\frac{x}{s}\right\rfloor - 1,\ \left\lfloor\frac{y}{s}\right\rfloor - 1)$. Nevertheless, it's still converted to $(\left\lfloor\frac{x}{s}\right\rfloor,\ \left\lfloor\frac{y}{s}\right\rfloor)$. Even though the point changes a little, it has absolutely different meaning in Eqs. (3) and (4). As shown in Tab. 2, Eq. (4) performs better than Eq. (3).

Tab. 2 shows that FBnet improves the mAP by 12.7%. In addition, as we have seen in the experiment, such decoupling operation makes our network converge more quickly and locate more exactly.

Actually, if we remove the operation of decoupling, it is equivalent to YOLO-v1 [Brenner, Chang-Claude, Jansen et al. (2016)]. In YOLO-v1, it was creative to introduce anchor-free mechanism. However, it lacks effective techniques to stabilize the process of training. For these problems, we introduce the two steps factorizations. As shown in Tab. 3, compared with YOLO-v1, it boosts great improvements.

4.4 Comparisons with state-of-the-art methods

Finally, we compare our method with some classic methods and all result are shown in Tab. 3.

Table 3: FBnet *vs.* other state-of-the-art two-stage or one-stage detectors

Methods	Faster-RCNN	Yolo-v2	SSD	RetinaNet	Yolo-v1	Ours
mAP	0.231	0.307	0.221	0.449	0.178	0.558

For the sake of a fair comparison with the state-of-the-art counterparts and demonstrate the feasibility of our network including Faster-RCNN [Ren, He, Girshick et al. (2015)], Yolo-v2 [Redmon and Farhadi (2017)], SSD [Liu, Anguelov, Erhan et al. (2016)], and RetinaNet [Lin, Goyal, Girshick et al. (2017)], all super-parameters of the state-of-the-art methods are fine-tuned to the best. All details are shown in Appendix A. As shown in Tab. 3, with VGG-16 as the backend, our FBnet is superior to the other state-of-the-art methods.

5 Conclusion

In this paper, we adopt an anchor-free mechanism and introduce a better method to factorize the process of bounding box regression. Firstly, we regress the shape of object by the variant of Faster RCNN [Ren, He, Girshick et al. (2015)]. Secondly, we re-define the target function of the location of object. The experimental result shows that our method achieves a mAP of 55.8%, which outperforms other state-of-the-art methods by at least 11.9%. This will greatly

help to reduce the missed diagnosis of clinicians during endoscopy and treatment, and provide effective help for early diagnosis, early treatment and prevention of CRC.

Funding Statement: This research was supported in part by the National Natural Science Foundation of China under grants No. 61602202 and 61603146, the Natural Science Foundation of Jiangsu Province under contracts BK20160428 and BK20160427, the Six talent peaks project in Jiangsu Province under contract XYDXX-034, the Natural Science Foundation of Huaian under contract HAB201934 and the project in Jiangsu Association for science and technology.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Ahn, S. B.; Han, D. S.; Bae, J. H.; Byun, T. J.; Kim, J. P. et al. (2012): The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. *Gut and Liver*, vol. 6, no. 1, pp. 64.

Aslanian, H. R.; Shieh, F. K.; Chan, F. W.; Ciarleglio, M. M.; Deng, Y. et al. (2013): Nurse observation during colonoscopy increases polyp detection: a randomized prospective study. *American Journal of Gastroenterology*, vol. 108, no. 2, pp. 166-172.

Bailey, C. E.; Hu, C. Y.; You, Y. N.; Bednarski, B. K.; Rodriguez-Bigas, M. A. et al. (2015): Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surgery*, vol. 150, no. 1, pp. 17-22.

Brenner, H.; Chang-Claude, J.; Jansen, L.; Knebel, P.; Stock, C. et al. (2014): Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. *Gastroenterology*, vol. 146, no. 3, pp. 709-717.

Buchner, A. M.; Shahid, M. W.; Heckman, M. G.; Diehl, N. N.; McNeil, R. B. et al. (2011): Trainee participation is associated with increased small adenoma detection. *Gastrointestinal Endoscopy*, vol. 73, no. 6, pp. 1223-1231.

Burke, C.; Kaul, V.; Pohl, H. (2017): Polyp resection and removal procedures: insights from the 2017 digestive disease week. *Gastroenterology & Hepatology*, vol. 13, no. 19, pp. 1.

Byrne, M. F.; Chapados, N.; Soudan, F.; Oertel, C.; Pérez, L. et al. (2019): Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*, vol. 68, no. 1, pp. 94-100.

Chen, P. J.; Lin, M. C.; Lai, M. J.; Lin, J. C.; Lu, H. H. S. et al. (2018): Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology*, vol. 154, no. 3, pp. 568-575.

Corley, D. A.; Jensen, C. D.; Marks, A. R.; Zhao, W. K.; Lee, J. K. et al. (2014): Adenoma detection rate and risk of colorectal cancer and death. *New England Journal of Medicine*, vol. 370, no. 14, pp. 1298-1306.

Doubeni, C. A.; Corley, D. A.; Quinn, V. P.; Jensen, C. D.; Zauber, A. G. et al. (2016): Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. *Gut*, vol. 67, no. 2, pp. 291-298.

Fang, S. Q.; Cai, Z. P.; Sun, W. C.; Liu, A. F.; Liu, F. et al. (2018): Feature selection method based on class discriminative degree for intelligent medical diagnosis. *Computers, Materials & Continua*, vol. 55, no. 3, pp. 419-433

Kaminski, M. F.; Regula, J.; Kraszewska, E.; Polkowski, M.; Wojciechowska, U. et al. (2010): Quality indicators for colonoscopy and the risk of interval cancer. *New England Journal of Medicine*, vol. 362, no. 19, pp. 1795-1803.

Lin, T. Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. (2017): Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980-2988.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. et al. (2016): SSD: Single shot multibox detector. *European Conference on Computer Vision*. pp. 21-37.

Mahmud, N.; Cohen, J.; Tsourides, K.; Berzin, T. M. (2015): Computer vision and augmented reality in gastrointestinal endoscopy. *Gastroenterology Report*, vol. 3, no. 3, pp. 179-184.

Misawa, M.; Kudo, S. E.; Mori, Y.; Cho, T.; Kataoka, S. et al. (2018): Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology*, vol. 154, no. 8, pp. 2027-2029.

Redmon, J.; Farhadi, A. (2017): YOLO9000: better, faster, stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263-7271.

Ren, S.; He, K.; Girshick, R.; Sun, J. (2015): Faster R-CNN: towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, pp. 91-99.

Rex, D. K.; Boland, C. R.; Dominitz, J. A.; Giardiello, F. M.; Johnson, D. A. et al. (2017): Colorectal cancer screening: recommendations for physicians and patients from the USA multi-society task force on colorectal cancer. *Gastroenterology*, vol. 153, no. 1, pp. 307-323.

Shaukat, A.; Oancea, C.; Bond, J. H.; Church, T. R.; Allen, J. I. (2009): Variation in detection of adenomas and polyps by colonoscopy and change over time with a performance improvement program. *Clinical Gastroenterology and Hepatology*, vol. 7, no. 12, pp. 1335-1340.

Simonyan, K.; Zisserman, A. (2014): Very deep convolutional networks for large-scale image recognition. ArXiv:1409-1556.

Society, A. C.; Society, A. C. T. (2018): American cancer society: cancer facts and figures. *American Cancer Society*.

Tajbakhsh, N.; Gurudu, S. R.; Liang, J. (2015): Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630-644.

Wan, J. J.; Chen, B. L.; Kong, Y. X.; Ma, X. G.; Yu, Y. T. (2019): An early intestinal cancer prediction algorithm based on deep belief network. *Scientific Reports*, vol. 9, no. 1, pp. 1-13.

Appendix A

All models have taken optimal training epochs, learning rate and optimizer. Meanwhile, different models have different decisive parameters. All results are shown in Tabs. 4-6. In this paper, we also fine-tune them.

In Faster RCNN, we mainly fine-tune those fateful parameters including prior anchor, train-time region proposals, test-time region proposals, input resolution, data augmentation and so on. For prior anchor, we adopt the principle in YOLO-v2 [Redmon and Farhadi (2017)]. In this paper, we run the k-means algorithm in training set to cluster k (k>0) prior anchors without human intervention.

Table 4: The setting of prior anchor box

k	2	3	4	5	6	7	8
mAP	0.132	0.159	0.162	0.188	0.184	0.189	0.172

For the train/test-time region proposals, we fix train-time proposals and find best test-time proposals.

Table 5: The setting of train/test-time region proposals

Train proposals	100	300	500	800	1000	1500	2000
mAP	0.166	0.194	0.208	0.144	0.139	0.095	0.097

At the same time, we also change other parameters including backend network.

In Yolo-v2 and SSD, we mainly focus on the setting of prior anchor, backend network, input resolution, data augmentation and so on.

In RetinaNet, we pay attention to the setting of the parameters of focal loss, backend network, input resolution, data augmentation and so on.

Table 6: The setting of parameters of focal loss

λ	α	mAP
0	0.75	0.359
0.1	0.75	0.367
0.2	0.75	0.394
0.5	0.50	0.395
1.0	0.25	0.407
2.0	0.25	0.388
5.0	0.25	0.343

By fine tuning other parameters, the best results of all models are shown in Tab. 3.