# WSPD:Weakly Supervised Network for Polyp Detection

**Xiuquan Du, Jiajia Chen, Xuejun Zhang, Hancheng Wu** and **Lou Sun**

School of Computer Science and Technology, Anhui University, Hefei, China

dxqllp@ahu.edu.cn

# 1.Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second most common cause of cancer death.The incidence of colorectal cancer continues to rise in low - and middle-income countries and remains high in most developed countries, and the threat to health of colorectal cancer will be further aggravated[19].Colorectal polyps are an early feature of colorectal cancer and are a driver of colorectal cancer incidence and mortality.Studies have shown that colonoscopy polyp screening is an effective technique for colorectal cancer prevention by providing information on the location and appearance of colorectal polyps, enabling physicians to remove them before they develop into colorectal cancer, reducing CRC incidence by 30%[1].However, the accuracy of colonoscopy depends on the experience of endoscopists. Clinical studies have shown that 22%-28% of polyps are missed in patients undergoing colonoscopy [20], with the omission rate of 8.65% by skilled physicians and 57.14% by unskilled physicians [23].Missed polyps can lead to late diagnosis of colon cancer, with a survival rate as low as 10%[21].Therefore, accurate and comprehensive polyp detection is very important in clinical diagnosis.

However, this is a challenging task for doctors for the following reasons.First of all, with the increasing number of colorectal cancer patients, the number of colonoscopies continues to rise, but there is a serious shortage of experienced endoscopists, which makes many doctors tired work, resulting in a high number of missed polyps.In addition, the colon wall has mucous membranes, which form a large number of folds, some of which are similar to colon polyps, making polyps difficult to identify;Some colon polyps are very small, or even in the initial bulge stage, and are extremely difficult to detect.This also leads to inaccurate and missed polyp detection.Therefore, a computer-aided technique (CAD) that can detect all possible polyps at an early stage is of great significance for polyp detection.

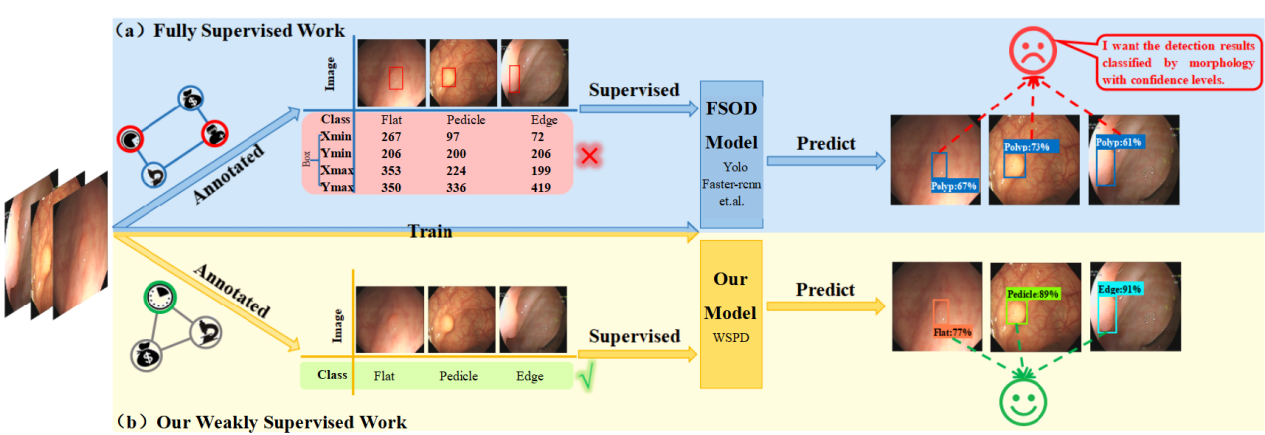
Previous polyp detection methods have relied on classifiers to distinguish polyps from their surroundings, using a combination of features such as colour wavelet analysis, local binary pattern (LBP), texture, Haar, directional gradient histogram (HoG) and other features based on colour, texture and shape. These methods are crucial for the identification of abnormal growths in the mucosa. **It should be noted that the use of technical terminology will be explained in the introduction.** However, polyps often resemble surrounding features and manual feature representation is limited, resulting in higher miss rates in such models [15].

Fig. 1. (a) Traditional fully supervised polyp detection (FSPD) uses the instance-level annotations as supervision. (b) Our weakly supervised polyp detection (WSPD)uses the image-level annotations as supervision.

**Deep learning has transformed the approach to manual feature learning, resulting in significantly improved performance in a range of medical tasks [8], including analysing medical images to diagnose and treat diseases, including detecting lung nodules [9], lung cancer [10] [11], classifying and identifying dermatological diseases [18], and diagnosing breast cancer [13]**. Recently, promising results have been achieved with fully supervised detection algorithms such as Faster-Rcnn [16] and YOLO [17]. The detector is trained through a fully supervised network, using CNN as the feature extractor and bounding box as the labeling, which results in the identification and confidence level of the colon polyp present in the image.

While fully supervised methods have shown promising results, colon polyp diagnosis poses significant challenges to the extension of fully supervised deep learning methods. This is due to the need for a large amount of accurately annotated object-level datasets, with the process of performing such annotation consuming considerable human resources. In particular, the high requirements for annotators in colon polyp object-level annotation has resulted in only experienced clinicians being tasked with such tasks, which means that colon polyp images are not annotated during clinical diagnosis [14]. As a result, clinicians face an increased data annotation burden and must spend additional time annotating data. The current detection network can be divided into two groups based on detection results: One group detects the presence of polyps without distinguishing between categories, as shown in Fig2(a), while the other group detects polyps and classifies them as adenoma/hyperplasia(cancer), as shown in Fig2(b). Our discussions with partner hospitals revealed that

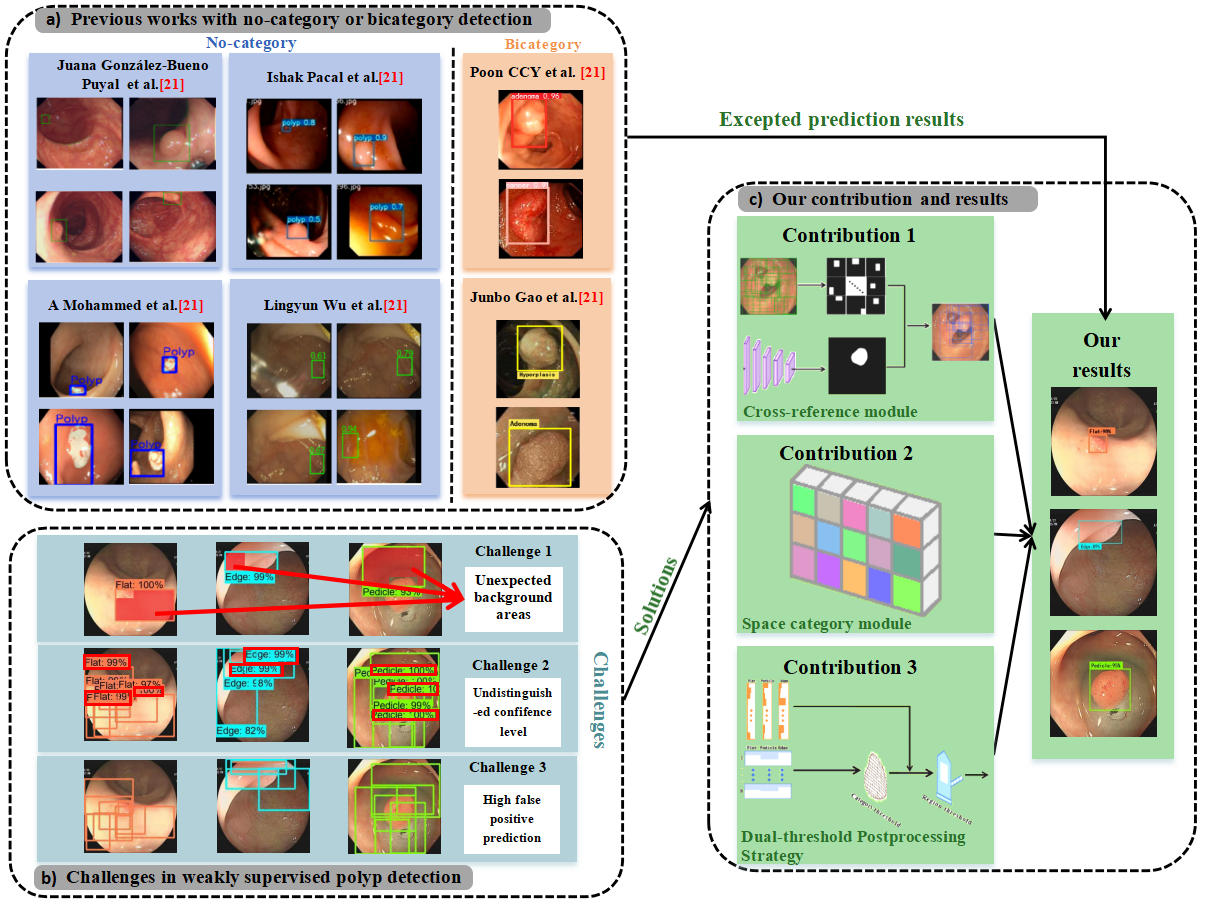


Fig. 2. Detection results from established methods and challenges within our framework. (a) is previous works with no-category or bicategory detection.(b) is challenges in weakly supervised polyp detection. (c) is our methods with the excepted prediction.

the networks can only identify polyps when the physician pinpoints their location with the colonoscope during an actual colonoscopy, so assessing network performance for clinical use is not meaningful.It should be noted that polyp shape is related to physician miss rates in clinical diagnosis.(加入过渡语句) The research shows that the missed diagnosis rate of flat polyps is up to 23.52%, while the missed diagnosis rate of raised polyps is 8.95%[23]. Therefore, polyp shape should be given priority in the training and evaluation of polyp detection networks.

Considering the FSOD problems we mentioned in the previous section, in this paper, we design the weakly supervised colon polyp detection method with image-level labeling,as shown in Fig. 1(b). Our method annotates the image with image-level polyp morphology, feeds the image into the proposed weakly supervised polyp detection network (WSPD) for training, and outputs the detection frame and polyp category confidence in the image.

Earlier works have addressed the problem of WSOD, as demonstrated by Bilen et al [22], whose approach is to use a pre-trained CNN to describe image regions, which is then combined with MIL to build an end-to-end weakly supervised detection network. Developed from Selective Search Window (SSW) [24], which uses pre-trained CNNs to describe image regions, understand object categories, and select positive samples by combining classification and detection results. . However, as shown in Fig2(c) coarse prediction box, we observe that the network's detection structure tends to select candidate boxes that contain a significant number of background regions, while only containing polyps in a localised area. We also observe that the classification structure of the network does not learn the most discriminating parts of the different categories at the beginning of training, but gradually overfits each candidate box as the number of training sessions increases, as shown in Fig2(c) similarity confidence level. Furthermore, observing the candidate boxes of the same image also shows that there is a large amount of spatial redundancy in the different candidate boxes generated by SSW, and the category scores of the redundant candidate boxes are extremely similar, so the network's post-processing is unable to accurately filter them, and there are a large number of false positives, as shown in Fig. 2(c) redundant predicted boxes.

Given the above observations, we propose a Cross-domain reference module(CRM) that pre-trains the polyp segmentation network using a publicly available colon polyp dataset, then decentralises it to a private dataset to generate pseudo-labels, and finally uses the obtained pseudo-labels as a reference to filter candidate frames in the data pre-processing stage. In this way, the network can screen potential frames before starting training, minimizing the impact of imprecise candidate frames on the network's detection structure and reducing computation during detection.

To further improve the discriminative abilities of the suggested network, we suggest implementing the Spatial Category Module (SCM) into the backbone network. Our belief is that categorisation by global pooling will drastically reduce the network parameters. Consequently, the network will put more learning pressure on the feature extraction stage. Consequently, the feature extraction stage will need to learn not only high-level classification features, but also generalised low-level features such as position, shape and size. Thus, we sign a channel-guided global attention module to properly address the global features of the target, thus increasing the ability of the network to perform accurate discrimination and detection.

To reduce the number of false positives in the network, we propose a two-threshold postprocessing strategy, which is motivated by the notion that the confidence in the box region of densely packed items is higher than the confidence in the box region of loosely packed items. In the first step, the candidate boxes are filtered by the category threshold to eliminate candidate boxes with low category scores, and then the region threshold is implemented to eliminate candidate boxes with low region scores. This post-processing mechanism identifies and eliminates redundant candidate boxs during the secondary filtering stage, which can reduce the rate of false positive detection.

Indeed,it is important to note that the joint learning of classification and box regression has been proven beneficial for fully supervised polyp detection.But it remains a challenging task for weakly supervised polyp detection and requires innovative ideas and insight.Our contributions can be briefly summarised as follows:

* We designed a Cross-domain reference module to reduce network localization interferences and improve network localization accuracy.
* We designed a spatial category module to enhance the ability to learn features, and the localization and categorization ability made significant progress.
* We designed a dual-threshold post-processing strategy to refine the filtering conditions, and the category average accuracy was significantly improved.
* We are the first work on weakly supervised detection guided by colon morphology categories, with significant results on public/private datasets.

# **2.Related Work**

## 2.1. Polyp Segmentation Network

Ronnebergeretal et al. proposed the medical segmentation network Unet[27], which uses an end-to-end architecture with a coding-decoding structure.Because of its accuracy, generality to the task, and easiness of implementation, subsequent work followed the standard example. For example, Unet++ [28] uses a branch-cut strategy to speed up inference and address unknown depths.ResUnet++ [29] fuses residual strategies to solve the gradient vanishing, includes SE block (squeeze and excitation block) to enhance the sensitivity of relevant features and suppress interference, and ASPP module to integrate different sizes of contextual information.

However, these methods do not handle polyp boundaries well.Subsequently, PraNet [26], PsiNet [30], MSNet [32] and SFANet [33], LODNet [31] force the model to learn the polyp boundary differences, which greatly enhances the model's ability to discriminate the polyp boundaries, and achieves satisfactory results. In addition, ACSNet [34], HRENet [35], and CCBANet [36] obtained high confidence predictions by focusing on contextual information, aggregating multi-scale context, and reducing local feature ambiguities.ICGNet [37] used an inverse contour guiding module to aggregate the underlying edge information, constrain the inverse region, and used an adaptive context module to extract the current layer's local global information and complementary information of the previous layer, expanding the dense features to solve the problems of size difference, irregular shape, and boundary ambiguity.

## 2.2. Weak supervision object detection

The problem to be solved for weakly supervised target detection coincides with the research goal of multiple example learning (MIL) [38] in weakly supervised learning, and thus weakly supervised target detection is usually treated as a multiple example learning problem.Bilen [22] firstly proposed a framework for weakly supervised target detection based on multiple example learning. The important challenge of mapping instance-level candidate box scores to image-level category labelling is addressed.The complete framework contains three main parts: 1) Candidate box generator. Selective search windows [24] or Edge boxes [39] algorithms are used to generate a large number of target candidate boxes on the input image; 2) Feature extraction. CNN is used to extract features from the input image, and then spatial pyramid pooling [40] or region of interest pooling [41] is used to generate a fixed-size candidate box feature matrix; 3) Detector. The candidate frame features are mapped to the image category labels, and the multi-example learning loss function is computed to complete the localisation and classification of target objects in the image.

# 3.Method

In this section, we present the structure of the weakly supervised polyp detection network and the dual-threshold post-processing strategy. The structure of the network consists of three main components: a Cross-domain reference module(CRM), a spatial category module (SCM), and a multiple instance branch.

The overall architecture of the proposed network is shown in Fig. 3.Given the image data, the image features are extracted by the pre-trained ConvNet on ImageNet [42] and the candidate frames are generated by the cross domain control module. The candidate frames and image features are sent to the ROI pooling layer for region feature mapping, and the results are sent to the spatial category module and the multi-instance detection module, respectively. The multi-instance detection module gives the location score and the spatial category module gives the category score. The dot product score, clamped column sums are used as prediction results. The regression training of the network is guided by the categorical labels y = [y1, y2，···，yC]∈{0,1}C, where yc = 1 or 0 denotes the presence or absence of an object class c. The rest of this section discusses these three modules in detail, as well as the dual-thresholding post-processing strategy.

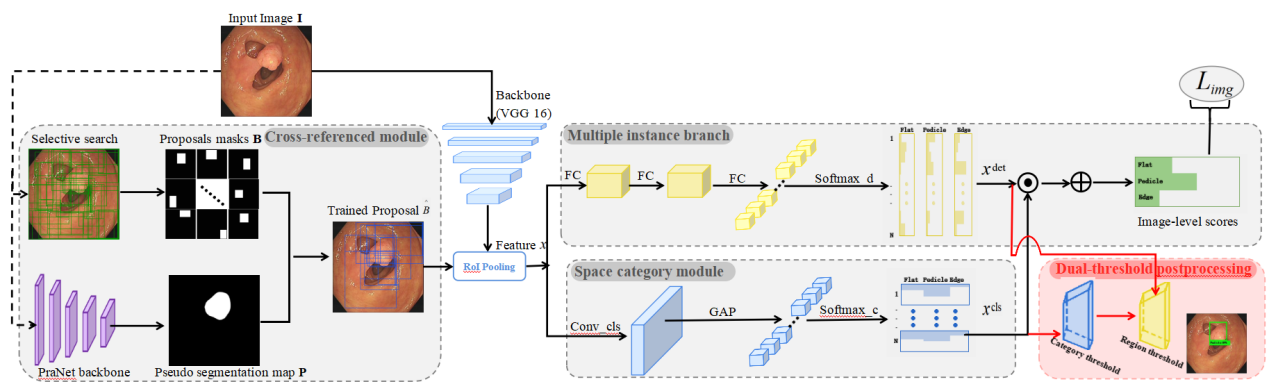


Fig. 3.Overview of the proposed WSPD.The WSPD consists of end-to-end architecture , including three parts for train with cross-reference module (CRM), multiple instance branch , space category module (SCM) and an extra dual-threshold post-processing for test.The solid ellipses denote the image-level loss functions . The black and red arrows are denoted as train data streams and post-processing operations , respectively.

## 3.1. Cross-domain reference module

We generated polyp candidate proposals using SSW algorithm to locate polyps. However, the generated proposals have a large number of errors and spatially redundant proposals , which bring noise to the detection. To solve the above problems, we propose a Cross-domain reference module(CRM) to generate proposals with less noise interference.

Specifically, we pre-trained a segmentation network heterocentric to our data on the publicly available colon polyp dataset, after which we fed our data into the network to produce segmentation results (P). At the same time, an SSW algorithm is used to generate initial candidate frames for our data and transform them into binary masks (B). Intuitively, if if B differs significantly from P, then B may be a hard or mislabelled sample, as the pre-trained segmentation model must have learnt some generic features of the polyp. In this case, this bounding box will be filtered out from model training. As a result, the problem shown in Fig. 1(d) will be alleviated.

In practice,we divide the dataset CVC-ClinicDB [25] according to the ratio of 8:2 and select PraNet [26] as the segmentation network (S) for training. 80% of the data is used to train the network, 20% of the data is used to test the performance of the network, and the weight parameters when the network performs best are saved. Given our private dataset image I, we use it as an input segmentation network S, load the saved weighting parameters and get the prediction result P with the following formulation:



where I denotes the input image and w is the weight of the pre-trained segmentation network S.P is a binary segmentation map for two class with a size of h×w.

At the same time, the SSW algorithm is applied to the image and the binary mask is converted to obtain the coarse candidate box B. The Jaccard Coefficient jc=  is chosen to measure the difference between B and P, and only  the candidate boxes are selected, and the final output is the refined candidate box ={b1,b2,...,bN}(where N is the number of proposals), the formula is as follows:





wheredenote is filtering threshold and SSW is the Selective search windows, Details are referred to[24].For each image,is a shortlist of candidate object proposals .Each element B(i,:) indicates the coordinate value of the proposals and the form is (xmin，ymin)、(xmax，ymax).

After optimization according to Eq.(3), potential noise interference can be minimized, and fine candidate boxes are fed into the network for subsequent training.

## **3.2. Spatial Category Module**

Generally, the pattern of training multiclassification network is to extract features from convolutional network, flatten the features to one-dimensional vectors by fully connected layer, and give classification to different categories by softmax. However, the ways of directly flattening to make fully connected retains a lot of spatial information,since larger region with more variation are more likely to have high category consistency across the whole training set.

To ensure the accuracy of the classification, we design the spatial category module(SCM). Formally, the spatial category module contains a convolutional layer, a global average pooling layer and a Flatten layeras follows:



where  is a flatten operation and  is a Global Average Pooling (GAP) layer to get the classification score map. w and b are the parameters of the spatial category module, which is a 3 × 3 convolutional layer.Then, it is delivered to the softmax operator.The classification scores can be calculated by:



The classification scores is a matrice with the same shape as the detection scores,which is described in detail in the next section. In this case,we expect it to learn more foreground/background information to help categorize objects.

## **3.3. Mutiple Instance** Branch

We only have image-level labels to indicate whether an object category is present or not.In order to train a standard object detector with regression, it is necessary to tap into instance-level supervision. The multi-instance detection module performs detection by scoring regions relative to each other. Given a region feature mapping graphextracted from a ConvNet,the multiple instance detection module takes it as input and outputs a detection scores matrice 。Specifically, the feature mapping map is fed into three tandem fully connected layers FC1~3 to obtain the candidate proposals stretching vector. This is then passed to the softmax operator, this time defined as follows:





Here, has  channels,which differents from the input feature map channels, and the former is the result of the fully connected layer processing. Each elementindicates the probability of the ith proposal bi belonging to the jth category.

In the case, in fact, the softmax operator compares, for each region independently, class scores。In this method, In this method, the module can evaluate which region holds a better information image fragment.

After that, the scores of all proposals are generated by element-wise product⊙.During the training stage, the loss function can be formulated as follows:



## **3.4. Dual-threshold Postprocessing Strategy**

Dual thresholds of different nature are designed to filter the detection results to obtain more accurate predictions that are neither too large to cover too much of the background nor too small to degrade into object parts.

**Category filtering strategy** Although categories may not be able to distinguish candidate frames with high spatial coverage in terms of categories, they are effective in distinguishing the target from the background. To filter out untrustworthy candidates, we set a category threshold, and only candidates with category confidence greater than the threshold are retained.After that，non-maxima suppression (with the threshold) is applied to the proposals.Finally, proposals with scores higher thanand overlap lighter than are held for the further filtering process.

**Region filtering strategy** As mentioned,filtering only on categories leads to a large number of false positive (FP) predictions. This is due to the fact that classification is unable to select the most valuable regions, and it prefers candidate frames that contain a lot of background as predictions because they have more categorical information. In contrast to this region confidence is more concerned with whether the candidate frames are tightly surrounding the object.So naturally, we further set the region thresholds which are dynamically changing unlike the fixed category thresholds. Consider the worst situation, where the candidate frames input to the network have a high spatial overlap and tightly frame different parts of the object. In this case, the region scores of the candidate frames are averaged across the candidate frames. Therefore, we set the region threshold to the worst-case mean normalisation with the following formula:



where N denotes the number of proposals in an image, and the result is between 0 and 1.

1. Haggar, F.A., Boushey, R.P.: Colorectal cancer epidemiology: incidence, mortality,survival, and risk factors. Clinics in colon and rectal surgery 22(04), 191–197 (2009)
2. S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, “Computer-aided tumor detection in endoscopic video using color wavelet features,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 141–152, Sep. 2003.
3. S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. De Groen, “Polyp detection in colonoscopy video using elliptical shape feature,” in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 2, pp. II-465–II-468.
4. L. A. Alexandre, N. Nobre, and J. Casteleiro, “Color and position versus texture features for endoscopic polyp detection,” in *Proc. Int. Conf. BioMed. Eng. Informat.*, 2008, vol. 2, pp. 38–42.
5. S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, “Texturebased polyp detection in colonoscopy,” in *Proc. Bildverarbeitung fur die Medizin*, 2009, pp. 346–350.
6. J. Bernal, J. Sanchez, and F. Vilarino, “Towards automatic polyp detection with a polyp appearance model,” *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, 2012.
7. S. Park, D. Sargent, I. Spofford, K. G. Vosburgh, and Y. A-Rahim, “A colon video analysis framework for polyp detection,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1408–1418, May 2012.
8. Qadir H A , Balasingham I , Solhusvik J ,et al.Improving Automatic Polyp Detection Using CNN by Exploiting Temporal Dependency in Colonoscopy Video[J].IEEE Journal of Biomedical and Health Informatics, 2020, 24(1):180-193.DOI:10.1109/JBHI.2019.2907434.
9. Jiang B, Li N, Shi X, Zhang S, Li J, de Bock GH, Vliegenthart R, Xie X. Deep Learning Reconstruction Shows Better Lung Nodule Detection for Ultra-Low-Dose Chest CT. Radiology. 2022 Apr;303(1):202-212. doi: 10.1148/radiol.210551. Epub 2022 Jan 18. PMID: 35040674.
10. Wang L. Deep Learning Techniques to Diagnose Lung Cancer. Cancers (Basel). 2022 Nov 13;14(22):5569. doi: 10.3390/cancers14225569. PMID: 36428662; PMCID: PMC9688236.
11. Feng J, Jiang J. Deep Learning-Based Chest CT Image Features in Diagnosis of Lung Cancer. Comput Math Methods Med. 2022 Jan 19;2022:4153211. doi: 10.1155/2022/4153211. PMID: 35096129; PMCID: PMC8791752.
12. Nguyen VD, Bui ND, Do HK. Skin Lesion Classification on Imbalanced Data Using Deep Learning with Soft Attention. Sensors (Basel). 2022 Oct 4;22(19):7530. doi: 10.3390/s22197530. PMID: 36236628; PMCID: PMC9572097.
13. Balkenende L, Teuwen J, Mann RM. Application of Deep Learning in Breast Cancer Imaging. Semin Nucl Med. 2022 Sep;52(5):584-596. doi: 10.1053/j.semnuclmed.2022.02.003. Epub 2022 Mar 24. PMID: 35339259.
14. A N R , B M W S , A J G ,et al.Automated Detection of Acute Respiratory Distress Syndrome from Chest X-Rays Using Directionality Measure and Deep Learning Features[J].Computers in Biology and Medicine, 2021.DOI:10.1016/j.compbiomed.2021.104463.
15. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Integrating online and offline threedimensional deep learning for automated polyp detection in colonoscopy videos.IEEE JBHI 21(1), 65–75 (2016)
16. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun.Faster R-CNN: Towards real-time object detection with region proposals networks. In NIPS, 2015.
17. Redmon J , Divvala S , Girshick R ,et al.You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition.IEEE, 2016.DOI:10.1109/CVPR.2016.91.
18. Naqvi M, Gilani SQ, Syed T, Marques O, Kim HC. Skin Cancer Detection Using Deep Learning-A Review. Diagnostics (Basel). 2023 May 30;13(11):1911. doi: 10.3390/diagnostics13111911. PMID: 37296763; PMCID: PMC10252190.
19. Zheng Ying, Wang Zezhou. Interpretation of global colorectal cancer statistics. Chinese Journal of Epidemiology, 2021, 42(1): 149-152.
20. A. M. Leufkens, M. G. H. van Oijen, F. P. Vleggaar, and P. D. Siersema, “Factors influencing the miss rate of polyps in a back-to-back colonoscopy study,” Endoscopy, vol. 44, no. 5, pp. 470–475, 2012.
21. L. Rabeneck, J. Souchek, and H. B. El-Serag, “Survival of colorectal cancer patients hospitalized in the veterans affairs health care system,” Am. J. Gastroenterology, vol. 98, no. 5, pp. 1186–1192, 2003.
22. H. Bilen and A. Vedaldi. Weakly supervised deep detection networks. In CVPR, pages 2846–2854, 2016. 1, 2, 3, 4, 7, 8
23. Chen Y L. Analysis of missed diagnosis rate and related factors of colorectal polyps in colonoscopy [J]. 2020. (in Chinese)
24. Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. IJCV, 2013.
25. Bernal, J., S´anchez, F.J., Fern´andez-Esparrach, G., Gil, D., Rodr´ıguez, C., Vilari˜no, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. CMIG 43, 99–111 (2015)
26. Deng-Ping Fan, Ge-Peng Ji, Tao Zhou,Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao.Pranet: Parallel reverse attention network for polyp segmentation. MICCAI, pages 263–273, 2020.
27. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015).
28. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018).
29. ha, D., et al.: Resunet++: an advanced architecture for medical image segmentation. In: 2019 IEEE International Symposium on Multimedia (ISM), pp. 225–2255. IEEE (2019)
30. Murugesan, B., Sarveswaran, K., Shankaranarayana, S.M., Ram, K., Joseph, J., Sivaprakasam, M.: Psi-Net: shape and boundary aware joint multi-task deep network for medical image segmentation. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7223–7226 (2019)
31. Cheng, M., Kong, Z., Song, G., Tian, Y., Liang, Y., Chen, J.: Learnable orientedderivative network for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 720–730. Springer, Cham (2021).
32. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 120–130. Springer, Cham (2021)
33. Fang, Y., Chen, C., Yuan, Y., Tong, K.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 302–310. Springer, Cham (2019).
34. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 253–262. Springer, Cham (2020).
35. Shen, Y., Jia, X., Meng, M.Q.-H.: HRENet: a hard region enhancement network for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 559–568. Springer, Cham (2021).
36. Nguyen, T.-C., Nguyen, T.-P., Diep, G.-H., Tran-Dinh, A.-H., Nguyen, T.V., Tran, M.-T.: CCBANet: cascading context and balancing attention for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 633–643. Springer, Cham (2021)
37. Xiuquan Du,Xuebin Xu,and Kunpeng Ma.:ICGNet: Integration Context-Based Reverse-Contour Guidance Network for Polyp Segmentation. IJCAI 2022.
38. Dietterich T G, Lathrop R H and Lozano-pérez T.1997.Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89(1-2): 31 - 71. [DOI: 10.1016/S0004-3702(96)00034-3]
39. C. L. Zitnick and P. Dollar. Edge boxes: Locating object 候选框 from edges. In Proc. ECCV, pages 391–405, 2014.
40. K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In Proc. ECCV, pages 346–361, 2014.
41. Girshick R.2015.Fast r-cnn// Proceeding of the International Conference on Computer Vision[C]. Santiago, Chile: IEEE, 1440 - 1448.[DOI: 10.1109/ICCV.2015.169]
42. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy,Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. IJCV, 2015