A Review on Polyp Detection and Segmentation in Colonoscopy Images using Deep Learning

Krishnendu S. M. Tech Image Processing College of Engineering Chengannur College of Engineering Chengannur College of Engineering Chengannur Kerala, India

Geetha S. Asst. Professor in CSE Kerala, India

Gopakumar G. Associate Professor in CSE Kerala, India

Abstract— Colorectal cancer is one of the main causes of cancer death worldwide. Polyp growth on the inner lining of the colon or rectum can change into cancer, over time. Early detection and diagnosis of polyp are key to ensure the survival of the patient. Detection and segmentation of polyps in colonoscopy is a challenging task due to the variations in size, shape, the texture of polyps, and variations between different types of hard mimics. Colonoscopy is the most efficient method for polyp screening and detection; however, it is operator dependent, timeconsuming, and error-prone. Sometimes polyps are missed or hard to detect in colonoscopy. In recent years many efficient deep learning techniques are introduced for polyp segmentation. This paper reviews several works that deal with deep learning techniques for polyp segmentation.

Keywords— Colorectal cancer, colonoscopy. polyp, segmentation, deep learning

INTRODUCTION

Colorectal cancer is the third most commonly occurring cancer in men and second most commonly occurring cancer in women [1]. Most of the colon cancer starts with adenomatous polyps which are small and benign cell groups. But not all polyps become cancer. The chance of a polyp turning into cancer depends on the type of polyp it is. Early detection and removal of a polyp are essential for the prevention and timely diagnosis of colorectal cancer. Colonoscopy is a successful preventative procedure for the screening and detection of polyps. But colonoscopy analysis is time-consuming and suffers a high miss rate of polyps. Moreover, manual screening is not only time consuming and laborious but also heavily relies on clinical experience. Miss detection and false polyps can lead to the death of the patient. Hence automatic detection of polyps is highly desirable. However automatic detection is a very challenging task. Several approaches have been used for tackling these challenges in the automatic detection of polyps.

Most of the recent studies are based on deep learning methods. Deep learning has been successfully applied to many areas of science and technology, especially in medical image analysis. Deep learning now becomes the dominant technique with considerable capabilities in image processing tasks. The deep learning models such as convolutional neural networks have been applied to the task of automatic polyp detection and segmentation with promising results.

This paper provides a comprehensive review and insight on a different aspect of these methods, which includes the preprocessing, choice of network architectures, training strategies, and training data. Fig.1 shows an example of a polyp in a colonoscopy image and its location.

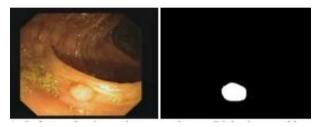


Fig. 1. polyp in a colonoscopy image and its location.

The rest of this paper is organized as follows, section II reviews some methodologies of polyp detection and segmentation. Section III presents a discussion on these methods.

II. **METHODS**

Most of the recent studies regarding polyp detection and segmentation are based on various deep learning techniques. Among these CNN, FCN, R-CNN, U-NET, and Mask R-CNN are popular. We make a review of the most recent papers of polyp segmentation in colonoscopy images.

Yu et al. [2] adopted a novel offline and online threedimensional deep learning integration framework to automatically detect polyps from colonoscopy videos by leveraging 3D-FCN's. Offline 3D-FCN is first developed and exploited for learning Spatio-temporal features from the training samples extracted from colonoscopy videos. Here explore 3D-CNN to learn spatiotemporal features from colonoscopy videos for automated polyp detection. This is the first approach that employs 3D-CNN for endoscopic video analysis. Fig.2 shows the flow chart of the 3D deep learning framework. A typical 3D-CNN consists of 3D convolutional layers, fully connected layers, and softmax layers.3D convolution and 3D pooling operations are performed in spatial and temporal dimensions. Also, the outputs of 3D convolution

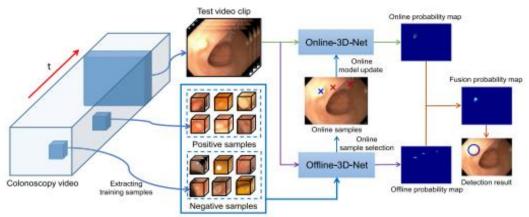


Fig. 2. Flow chart of the online and offline 3-D deep learning framework [2]

and pooling is 3D features when the input is a video clip. On the other-hand 3D-CNN can sufficiently preserve temporal information of colonoscopy videos when extracting hierarchical features, hence it can effectively distinguish polyps from hard mimics such as specular spots and air bubbles by using spatiotemporal information. But 3D-CNN is quite computationally expensive and So 3D-CNN is converted to 3D-FCN for fast detection by adopting the fully convolutional concept in [15]. 3D FCN needs only to feed the whole video clip into the 3D-FCN and get the probability map of the whole video clip within a forward propagation directly. So, these methods can reduce the redundant computations and accelerate detection compared to the traditional sliding window approach.

The architecture of these 3D-FCN for offline representation learning consists of 6 conventional convolutional layers and each of them is followed by a rectified linear unit (RELU) [15] as an activation function. Here also add four maxpooling layers between these convolutional layers to increase the receptive field and reduce the feature volume size. After each pooling layer, double the number of feature volume to preserve the necessary information. Offline 3D-FCN is trained using the cropped subvolumes and finetuned with the backpropagation method using the training subvolumes. The online model is also implemented based on 3D-FCN and adopts the same architecture as the offline network. Weights from offline 3D-Net are used to initialize each online 3D-Net weight and update its weights with back-propagation using online training samples extracted from the video. Fig.3 shows the extracted online training samples. The main purpose of the online network is to remove the specific polyp like FP's detected by offline network and further improve the detection performance. To do this, the next step is to combine the outputs of the offline network and online network to get the final polyp detection result. Through the online sample selection and offline training, the influence of hard mimics can reduce to an extent.

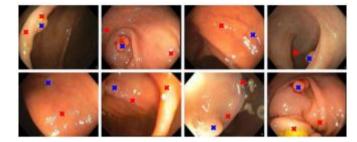


Fig. 3. Extracted online training samples.[2]

Zhang et al. [3] presented another novel hybrid classification method for automatic polyp segmentation in colonoscopy video frames. The segmentation method is composed of two main stages that are, the region proposal stage using the FCN and region refinement stage using texton-based patch representation followed by a random forest (RF) classifier. FCN provides initial polyp candidates and texton based patch representation which further discriminate polyp from non-polyp regions. Data-driven and hand-designed features are taken for segmentation. Moreover, the hierarchical features of polyps are learned by FCN, while the context information related to the polyp boundaries is modeled by texton patch representation.

FCN-8's architecture with VGG16(CNN classification net) is adopted as a region proposal net. The VGG16 net is composed of 5 stacks followed by 3 transformed convolution layers, where each block contains several convolutional layers, a ReLu layer, and max-pooling layers. The FCN-8's produces more detailed segmentation compared to the FCN-16's. However, some false positives may present due to the lack of spatial regularization for FCN. Here FCN-8's was trained with two classes for polyp and non-polyps(background) given by the ground truth images. FCN8's are trained using MatConvNet which is commonly used in a deep learning framework.

Bardi et.al. [4] address colon polyp detection using a convolutional neural network (CNN) combined with autoencoder, but there is no image processing applied here. The tensor flow library is used for training the convolutional encoder-decoder model. In the encoder part, here used three similar modules, each consisting of a convolution layer with stride 2 and a non-overlapping max pool with kernel 2. In

the decoder section, each layer in the encoder contains its counter-part. The network dimension is equal to the input dimension.

Xiao et al. [5] attempted to use the existing deep neural network called Deep Lab-V3 to detect polyps in colonoscopy images and for the semantic segmentation of polyps and to transmit it effectively, a long short-term memory is combined with Deep Lab-V3 to augment the signal of the location of the polyp. DeepLab_V3 is used to learn and extract features of polyps. DeepLab v3 has three sub-frameworks, ResNet, multi-grid methods, and atrous spatial pyramid pooling (ASPP) in cascade. The long shortterm memory network (LSTMs) network preserves information for long periods. The memory cell is changing with the input gate, forget gate, and output gate. The input gate decides what information to be thrown away from the cell state. Forget gate decides what new information to be stored in the cell state. The output gate decides what to be output. The LSTMs can remove or add information to horizontal path, which ensures keeping important information for a longer period. In DeepLab_v3, the ResNet, multi-grid methods, and the ASPP acquire different information.

Urban et al. [6] designed and trained a deep CNN to detect polyps using a diverse and representative set of handlabeled images from screening colonoscopies collected from more than 2000 patients. They trained different CNN architectures in this study. All trained CNN consists of the same fundamental building blocks, including convolutional layer, fully connected layer, maximum or average pooling, activation function, batch normalization operations, and skip connections. Here each layer is followed by a rectified linear (ReLu) activation function. The last hidden layer is connected to the output unit and optimized the loss with linear output units for localization problems. Softmax output units and optimized kull back-Leibler divergence are used for classification. Localization is implemented by predicting the size and location of a bounding box that tightly enclosed any identified polyps. This allowed building CNNs that could operate in real-time.

Shin et al. [7] applied a region-based object detection scheme for polyp detection. Here adopted the region proposal network (RPN) which was introduced in a faster R-CNN method [24] to obtain a polyp candidate region in polyp frames. Then applied a proper augmentation strategy such as rotating, scaling, shearing, blurring, and brightening. Then apply two post-learning schemes: false-positive learning and offline learning. In the FP learning scheme, post-training the detector system with automatically selected negative detection outputs (FP's) which are detected from normal colonoscopy videos. This is effective to reduce many of the polyp-like false positives. The offline scheme further, improved the detection performance by using the video-specific reliable polyp detection and post-training procedure. Fig.4 shows an example of polyp augmentation.

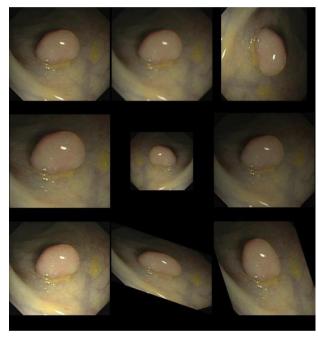


Fig. 4. An example of different augmentation of polyps. (a) original polyp image frame, (b) blurred image with 1.0 of standard deviation, (c) 90-degree rotated image, (d) zoom-in image, (e) zoom-out image, (f) dark image, (g) bright image, (h) sheared image by y-axis, (i) sheared image by x-axis.[7].

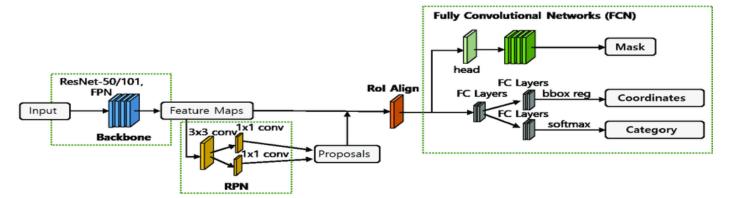


Fig 5. The framework of Mask R-CNN [8]

Kang et al. [8] employed a Mask R-CNN network to identify and segment polyps. Mask R-CNN in this model consists of different backbone structures that are ResNet50 and ResNet101. Then use an ensemble method to combine the output of two Mask R-CNN networks. The bitwise combination is used as the ensemble method. Some data augmentation is used here as a preprocessing process. Mask R-CNN first detects targets in the image and produces a high-quality segmentation result for each target i.e., it provides an instance segmentation of polyps. Compare to other networks discussed in these studies Mask R-CNN is very fast and a little more efficient for polyp segmentation. The detailed framework of Mask R-CNN is shown in Fig.5.

Zheng et al [9] proposed an algorithm for automatic polyp detection and localization in colonoscopy video. An efficient on-the-fly trained CNN has been deployed. To overcome tracking failure caused by motion effects, here also use object detection or segmentation network such as U-Net. It utilizes optical flow to track polyps and fuse temporal information. A CNN model is first trained to detect and segment polyp in each video frame. Once a polyp is detected, the center of the polyp is computed and traced through the following frames until stopping criteria are met. During tracing, optical flow is utilized to trace easier cases and CNN is used to process harder ones. If a frame doesn't contain any polyp center seed, the frame will be regarded as a negative frame. If there are multiple polyp seeds in a frame, a spatial voting algorithm is run and the most confident center is kept as the detection while others are eliminated. The overview of the proposed method is shown in Fig.6.

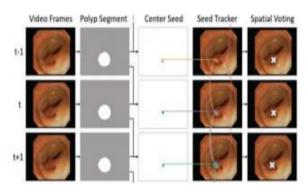


Fig. 6. Overview of the proposed method [9].

Tashk et al. [10] proposed a network, which has a novel U-Net architecture. This paper adopted a novel approach for fully automatic polyp detection. This includes three main steps: first, a preprocessing step is applied to the dataset images. The preprocess comprises 3 distinct color transformations known as La*b*, CMYK, and gray-level. In the second step, the U-Net is proposed for segmentation and the final step is post-processing for improving the pixel-wise

classification outcomes. The architecture of the proposed network includes

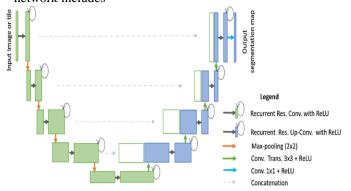


Fig. 7. The architecture of U-Net.

fully 3D layers that enable the network to be fed with multi or hyperspectral images or even video streams. Moreover, there is a dice prediction output layer. The architecture of U-Net is shown in Fig.7.

Sun et al. [11] design a U-Net with dilation convolution, which is a novel end to end deep learning framework for the colorectal polyp segmentation. The model consists of an encoder to extract multi-scale semantic features of polyps and a decoder to expand the feature map to a polyp segmentation map. The dilated convolution is added to the encoder part of the network to learn high-level semantic features without resolution reduction which improves feature representation ability. The architecture of the model consists of an end-toend convolutional neural network which includes a construction part on the left and an expensive part on the right. The model takes a single colonoscopy image as the input and outputs a binary mask segmentation of polyps that has the same size as the input image on the last layer. To improve display effectiveness during colonoscopy, several post-processing operations are applied, such as smoothening, drop small objects, and combine nearby objects.

Feng et al. [12] develop a stair-shape network (SSN) for real-time polyp segmentation in colonoscopy images. The SSN can well balance the inference time and segmentation accuracy. The lightweight backbone with four specific residual blocks and simplified upsampled operation allows fast inference time. For the backbone network, designed an FCN to extract diverse features on different levels. Besides, some intestinal folds in colonoscopy images are likely to be taken mistakenly as polyps. To address these issues, a specific dual attention module is applied to refine the output feature of each residual block in the backbone. Then designed a multiscale fusion module (MFM) for fully fusing features of different layers. Fig.8 shows an overview of the SSN model.

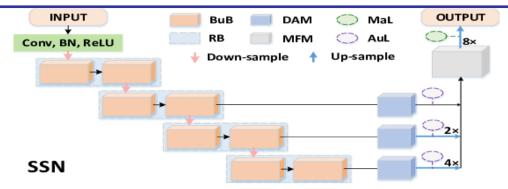


Fig. 8. An overview of our SSN model.[12]

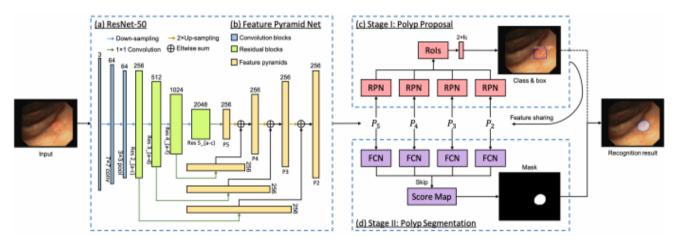


Fig. 9. The overview of PLP-Net.[13]

Jia et al. [13] introduced a two-stage approach called "polyp for automated pixel-accurate polyp recognition in colonoscopy images" (PLP-Net) for automated polyp recognition in colonoscopy images, using deep convolutional neural network. The PLP-Net improves the performance of polyp segmentation by using a two-stage learning strategy. The PLP- Net comprises two stages, that are the polyp proposal stage and the polyp segmentation stage. The learning process would be complicated by the complex colonic wall if pixel-wise training is performed directly on the CNN model. Therefore, a two-stage framework is proposed, where the polyp proposal stage is constructed as a region-level polyp detector, aiming to accurately segment the area of the polyp that occupies in the image. In addition to this, here apply the very deep ResNet-50 as well as a pyramid component to seek deeper and richer semantics from each frame. Feature sharing and skim schemes are adopted to perform multiscale transfer learning between stage 1 and stage 2. The overview of PLP-Net is shown in Fig.9.

Tan et al. [14] proposed a three-dimensional GLCM based CNN for 3D polyp classification. This proposed model contains three steps. The first step is to convert the original Hounsfield unit CT value of the 3D polyp into gray-level value based on CT value. Here performs a gray level scaling on the original CT image pixel values to an appropriate value range.

The second step is to generate multiple 3-dimensional gray-level values. A multi-channel CNN model is used to perform the classification of polyps using GLCM feature images. The CNN model used in this method consists of seven layers that are three convolutional layers, two maxpooling layers, and two fully connected layers. In each convolution layer, batch normalization and activation functions are performed. The model uses the ReLu as the activation function, the crossing entropy loss as the training loss, and softmax function at the last fully connected layer. The proposed GLCM method can enrich the information of the small polyp without bringing in any artificial information. The proposed GLCM feature-based model can naturally overcome the problem of the various lesion sizes existing in all the CNN based models, which may shift the current CNN-raw image paradigm to a CNN-texture image (GLCM) paradigm. The workflow of the 3D-GLCM CNN model is shown in Fig.10.

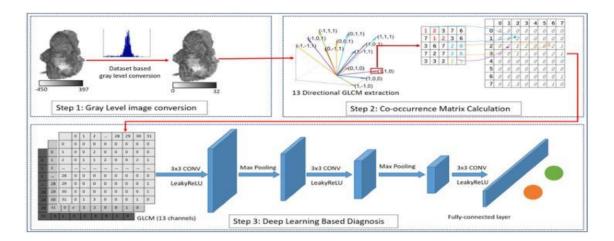


Fig. 10. 3D-GLCM CNN model.[14]

TABLE 1. SUMMARY OF DISCUSSED MODELS

Authors	Methods	Datasets	Quantitative measures (%)	Limits
Yu et al [2]	Novel offline and online 3D deep learning integration framework	ASU-Mayo	Prec=78.7; Rec=53.8; F1=63.9; F2= 98.7	High processing time
Zhang et al. [3]	FCN with novel hybrid classification	CVC-ColonDB	Acc=97.54; sens=75.66; spec=98.81	High processing time
Bardi et al. [4]	Convolutional encoder- decoder model	CVC-ColonDB, CVC- ClinicDB, ETIS-LARIB	Acc=96.7	No promising result
Xia et al [5]	Deep lab and LSTM network	CVC-ClinicDB	mIoU=93.21	High training time and prone to overfitting
Urban et al [6]	Deep CNN with Resnet50	proprietary	Acc=99.0; sens=96.8	Missed polyps
Shin et al [7]	Region-based deep CNN and post-learning	CVC-ClinicDB, ETIS- LARIB	Prec=91.4; Rec=71.2; F1=80; F2= 74.5	High processing time
Kang et al [8]	Instance segmentation using Mask R-CNN	CVC-ColonDB, CVC- ClinicDB, ETIS-LARIB	Prec=73.84; Rec=74.37	Limited segmentation
Zheng et al. [9]	Optical flow with an on- the-fly trained CNN	CVC-videoClinicDB	Prec=84.58; Rec=97.29; F1=90.49; F2=94.45	No balance between FP and FN
Tashk et al [10]	U-Net and morphological post-process	CVC-ColonDB, CVC- ClinicDB, ETIS-LARIB	Acc=99.02; Prec=70.2; Rec=82.7; F1=76.0;	High training time
Sun et al [11]	U-Net with dilation convolution	CVC-ColonDB, CVC- ClinicDB, ETIS-LARIB	Prec=96.71; Rec=95.51; F1=96.11	High training time
Feng et al [12]	Novel stair shape network (SSN)	CVC-ColonDB, CVC- ClinicDB, Endo Scene	Prec=92.85; Rec=94.83	Not completely reduce miss polyp rates
Jia et al [13]	PLP-Net with two-stage pyramidal feature prediction	CVC-612 test set, ETIS- LARIB	Prec=85.9; Rec=59.4	High computational cost
Tan et al [14]	3D GLCM based CNN model	proprietary	Acc=91; sens=90; spec=71	Model is complex

Most of the models discussed in this paper, used data augmentation as a preprocessing procedure. One of the challenges in training the polyp segmentation model is the insufficient number of data for training. obtaining a large number of polyp images with the corresponding ground truth of a polyp mask is generally quite difficult because access data is limited due to privacy concerns. Endoscopy procedures associated with moving camera control, and the color-setting is not consistent. So, the appearance of available endoscopy images changes across different laboratories. The data augmentation steps bring endoscopy images into an extended space that can cover all their variances. Moreover, by augmenting the training data, can reduce the problem of overfitting. Table 1 shows a summary of the discussed approaches.

III. DISCUSSIONS AND CONCLUSIONS

The polyp detection and segmentation from colonoscopy images are still a challenging task in the medical field. Many studies are employed in this case, among these deep learning has shown an efficient performance over other techniques. Each model discussed in this paper has its advantages and limits. All of the models have achieved impressive performance in various image segmentation tasks also. In this paper, we provided a comprehensive review of some recent works for the detection and segmentation of colon polyps.

ACKNOWLEDGEMENT

We would like to thank our Director of IHRD and the Principal of our institution for providing us the facilities and support for our work. We express our heartfelt gratitude to Jyothi R.L., Asst. Professor in dept. Computer science of our institution for providing timely advice and valuable suggestions to complete our work.

REFERENCES

- F. Bray, A. Jemal, R.A. Smith et al, "Global cancer transactions according to the human development index (2008-2030): a population-based study", Lancet Oncol, Vol.13, pp.790—801, 2012.
- [2] Yu L, H. Chen, Q. Dou, J. Qin, and P.A. Heng," Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos," IEEE Journal of Biomedical and Health Informatics, vol.21, pp.65–75, January 2017.
- [3] L. Zhang, S. Dolwani, and X. Ye, "Automated Polyp Segmentation in Colonoscopy Frames Using Fully Convolutional Neural Network and Textons," Springer international publishing, pp.707–717,2017.
- [4] O. Bardi, D.S.Sosa, B.G.Zapirain and A.Elmaghrby, "Automatic colon polyp detection using convolutional encoder-decoder model, "2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),pp.445–448, 2017
- [5] W. T.Xiao, L.J. Chang, and W.M. Liu, "Semantic segmentation of colorectal polyps with deep lab and LSTM networks, "2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW),pp.1–2, 2018
- [6] G. Urban, P. Tripathi, T. Alkayali, M. Mittal, F. Jalali, W. Karnes, and P. Baldi, "Deep learning localizes and identifies polyps in real-time with 96% accuracy in screening colonoscopy," Gastroenterology, pp.1069–1078,2018
- [7] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasing-ham, "Automatic colon polyp detection using region-based deep CNN and post-learning approaches," IEEE Access, vol. 6, pp. 40950–40962, 20185
- [8] J. Kang and J. Gwak, "Ensemble of instance segmentation models for polyp segmentation in colonoscopy images" IEEE Access, vol.7, pp.26440–26447, February 2019.

- [9] H. Zheng, H. Chen, J. Huang, X. Li, X. Han and J. Yao "Polyp tracking in video coloscopy using optical flow with an on-the-fly trained CNN," IEEE international symposium on biomedical imaging, pp.79–82, April 2019
- [10] A. Tashk, J. Herp, and E. Nadimi, "Fully automatic polyp detection based on a novel U-Net architecture and morphological postprocess," 2019 International Conference on Control, Artificial Intelligence, Robotics and Optimization (ICCAIRO), pp.37– 41,2019.
- [11] X. Sun, P. Zhang, D. wang, Y. Cao, and B. Liu, "Colorectal polyp segmentation by U-Net with dilation convolution, "2019 18th IEEE International conference on machine learning and applications, pp.851–858, 2019
- [12] R. Feng, B. Lei, W. Wang, T. Chen, J. Chen, D.Z. Chen and J. Wu, "SSN: A stair-shape network for real-time polyp segmentation in colonoscopy images, "IEEE 17th international symposium on biomedical imaging, pp.225–229, April 2020.
- [13] X. Jia, X. mai, Y. Cui, Y. Yuan, X. Xing, H. Seo, L. Xing, and M.Q.H. Meng, "Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction," IEEE Transactions on automation science and engineering, pp.1–15,2020.
- [14] J. Tan, Y. Gao, Z. Liang, W. Cao, M. Pomeroy, Y. Huo, L. Li, M.A. Barish, A.F. Abbasi, and P. J. Pickhardt, "3D-GLCM CNN: A 3-dimensional gray- level co-occurrence matrix-based CNN model for polyp classification via CT colonography," Transaction on medical imaging, vol. 39, pp. 2013–2024, June 2020.
- [15] J. Long, F. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conference of Computer Vision and Pattern Recognition, 2015, pp. 3431—3440.
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097—1105.
- [17] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," IEEE Trans.Med. Imag., vol. 35, no. 2, pp. 630— 644, Feb. 2016.
- [18] N. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: Full training or fine-tuning?" IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1299—1312, May 2016.
- [19] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of the convolutional neural networks," in Proc. 2015 IEEE 12th Int. Symp. Biomed. Imag., 2015, pp. 79—83.
- [20] J. Bernal, J. Sachez, F. Vilariño, "Towards Automatic Polyp Detection with a Polyp Appearance Model," Elsevier Pattern Recognition, vol. 45, no. 9, pp. 3166–82, September 2012.
- [21] J. Bernal, J. Sachez, G. Fernadez-Esparrach, D. Gil, C. Rodríguez and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," Elsevier Computerized Medical Imaging and Graphics, vol. 43, pp. 99–111, July 2015.
- [22] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer," International Journal of Computer Assisted Radiology and Surgery, vol. 9, no. 2, pp. 283–293, 2014.
- [23] M. Ganz, X. Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic Narrow-band imaging data," IEEE Transactions on biomedical engineering, vol.59, no.8, pp.2144– 2151, August 2012.
- [24] R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis., Santiago, Chile, pp. 1440–1448, December 2015