An Innovative Polyp Detection Method from Colon Capsule Endoscopy Images Based on A Novel Combination of RCNN and DRLSE

Ashkan Tashk[1] and Esmaeil Nadimi[2]
Mærsk McKinney-Muller Institute (MMMI)
Syddansk Universitet (SDU)
Odense, Denmark
[1]email : asta@mmmi.sdu.dk
[2]email : esi@mmmi.sdu.dk

***Abstract*—Background: Direct detection of polyps from colon capsule endoscopy (CCE) videos is an ultimate goal not only for physicians but also for biomedical engineers who are working on automatic internal lesions like polyps. There is also a great enthusiasm among biomedical professionals to make advanced systems for aiding doctors to have a faster and accurate diagnosis by the means of polyp detection from CCE acquired video streams. Such systems must be able to localize polyps correctly and extract the whole lesions from the video frames completely.**

Material and Methods: In this paper, a new approach toward object-wise polyp detection from CCE frames in a video stream is proposed. The proposed method employs modified region proposal CNNs to localize the existing polyps from CCE acquired video frames and after that a level-set method known as Distance Regularized Level Set Evolution (DRLSE) is employed for automatic model-based segmentation of localized polyps. The pixel-wise detection of polyps is necessary for polyp classification and will help gastroenterologists to determine appropriate prognosis and treatment for the patients.

Results and conclusion: The proposed method is trained by the means of an CCE still image dataset which includes manually annotated polyps. The trained network is then applied to CCE video images. The results demonstrate that the proposed method is able to localize and detect polyps both region-wise and pixel-wise with a good rate of accuracy.

Keywords— colon capsule endoscopy (CCE), Distance Regularized Level Set Evolution (DRLSE), Faster Region Proposal Convolutional Neural Network (Faster R-CNN).

I. INTRODUCTION

Polyp is a kind of colorectal lesions which are not dangerous in their early stages but can cause cancerous malignancy if they progress to the late stages. Fortunately, they could be cured if they are diagnosed soon and are treated correctly. In figure 1, a polyp progression from benign to malignancy stages is depicted. The traditional screening of colorectal cancer (CRC) is basically done by biomedical imagery like optical colonoscopy (OC), which is an invasive and inconvenient method for patients. Moreover, this method provides polypectomy and biopsy ability which are essential for later histopathologic studies.

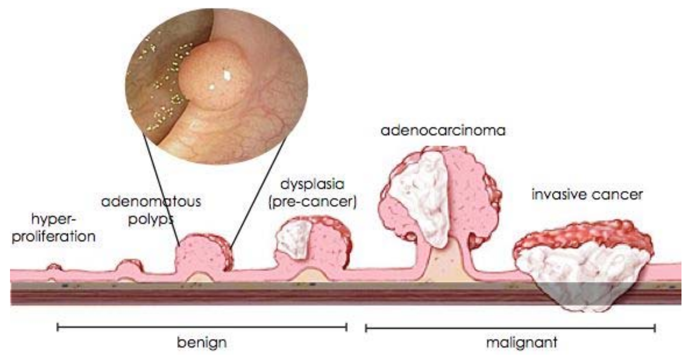


Fig. 1. Benign to Malignant progression of Colorectal Polyps

The main issue in front of colorectal studying based on conventional imagery techniques like optical colonoscopy was their quality and patient inconvenience, since in most cases, the patients do not consent and like to have such invasive imagery. Moreover, the acquired OC imagery can not cover all the parts of colon for detecting probably existing polyps. So, it is very an important issue in front of gastroenterologists to improve the uptake admissibility of OC technique among the patients imposed to such imageries [1].

In the ending years of twentieth century, a more minimally invasive and convenient CRC screening technique, in compare to OC, has been introduced. This colorectal cancer screening technology is named colon capsule endoscopy (CCE); since, its imagery device is like a capsule including miniature cameras and advance electronic boards. The marketing name of this capsules is PilCams. Accordingly, a new horizon was open in front of physicians and gastroenterologists to employ PilCams as a more popular and patient friendly technique for colorectal polyp screening [2]. PilCams were approved by US Food and Drug Administration (FDA) to be used for investigation of small bowel in 2001 [3]. In figure 2, some sample PilCams which are the main tools for CCE imagery are shown. As it is obvious, their dimensionality is like the conventional pills prescribed by physicians for medical treatment.

Fig. 2. Sample PilCams and their superficial manifestation

However, the first PilCams were proposed to be used for small bowel and esophageal screening, their application was extended to colon in 2006. In recent years, there are new versions of PilCams like PilCam colon 1 & 2 which are used for outpatient colon polyp screening programs [4].

Besides the resolution and quality differences between CCE and OC imageries, the accuracy and perspective of CCE for detecting colorectal polyps is significantly higher. The differences between OC and CCE acquired images for the same colorectal polyp is shown in figure 3.

(a) (b)

Fig. 3. A same Polyp manifestation for two different imaging modalities: (a) OC, and (b) CCE

CCE acquired video signals can be converted and viewed in softwares like rapidreader which are provided by the PilCam producing company. Therefore, the most common technique for investigating CCE acquired images for detecting polyps and analyzing them is employment of software tools. Such softwares provide easier reading and studying opportunities for gastroenterologists to detect polyp in shorter time [5]. The proposed studies demonstrate that such software could help even non-expert observers to detect colorectal polyps with good accuracy. However, such investigations do not lead to reliable polyp risk rate diagnosis or classification which are necessary for treatment purposes. In [6], the authors addressed a new strategy for detecting colorectal polyps from both OC and CCE screening using machine learning strategies in an unmanned and automatic manner. The results of this study are promising for automatic polyp localization from CCE still images. In this paper, we presented an innovative method based on artificial intelligence methods to localize polyps and extract them pixel-wise.

The organization of the rest of this paper is organized as follows: In section II, the strategies for providing materials and applying them to proposed polyp detection method is presented. Section III assigns to evaluation and assessment results. Finally, the paper is concluded in section IV.

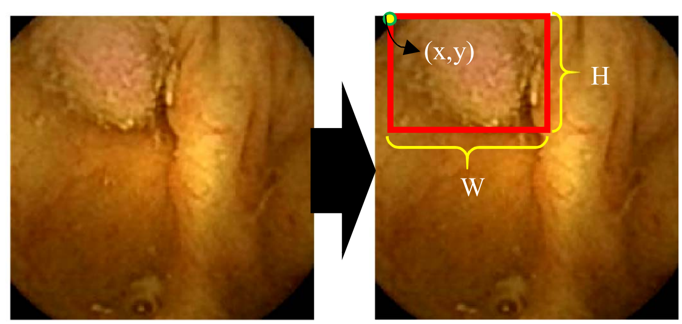
II. Material and Methods

This section comprises of two subsections. In the first subsection, the procedures for providing polyp dataset images including their region-wise annotations are explained. The second subsection is assigned to the proposed polyp detection and localization method. The proposed method comprises of two main steps. In the first step, a region proposal deep convolutional neural network in a transfer learning manner is employed for extracting the coordinates of polyps which localize the region of interest in the form a rectangle. In the second step, the localized polyps by a rectangle are segmented based on a modified level set approach. More descriptions are provided in the following subsections.

A. Material

The principal material for polyp detection from CCE images is to provide an acceptable number of CCE images and video streams which include polyps as the main studying colorectal (gastrointestinal) lesions. Therefore, the CCE images are acquired from a determined number of patients participating in a clinical study and was conducted by experts at Odense University hospital (OUH). We also used some of the data for our current research with their permission.

There are some still images extracted out of the acquired CCE video frames which include polyps. The location of polyps in each frame was not mentioned or annotated by the physicians at the early stages. So, the coordinates bounding boxes confining polyps are extracted and saved as the mentioned region-wise ground truths for training and evaluating purposes. Figure 4 illustrates one of the CCE still images which include a polyp. The identifying features for confining bounding box around each polyp are the coordinates of the bounding box around each polyp and the height and the width of the rectangle around it.



(a) (b)

Fig. 4. Sample CCE frame including polyp: frame without regional polyp annotation (a) , and (b) frame with regional polyp annotation

For the training purposes of classification layer of the pre-determined region proposal deep CNN network, a set of CCE frames including at least a polyp are employed. The polyps' annotation is done in the format of rectangular bounding boxes. In total, about 812 CCE frames are annotated for their existing polyps.

B. Automatic Segmentation of Detected Polyps

The segmentation stage in the proposed method contains two steps. In the first step, a region proposal network (RPN) based on a pre-defined deep convolutional neural network is employed. There are several approach toward proposing RPNs for object detection in image processing applications. These approaches have same Classification Procedure as follows:

1) *Applying a selective search for desired objectives.*
2) *Creating different RPs from one image.*
3) *Creating a feature vector from those proposals using a predefined CNN.*
4) *Classifying each feature vector with an appropriate classifier like the SVM for each object class.*

Consequently, the throughput of the aforementioned procedures forms the following outputs:

1) *Object-wise classification.*
2) *Bounding Box Refinement:* Rejecting a regional image proposal with a high IoU compared to a higher scoring region.
3) *Bounding Box Regression:* input is the center, width and height in pixels of the RP and the label is the Ground Truth (GT) bounding box.

There are several state-of-the-art proposed RPN techniques based on CNNs for object detection and localization in images such as R-CNN [7], Fast R-CNN [8], Faster R-CNN [9], Single shot MultiBox Detector (SSD) [10] and You Only Look Once (YOLO) [11]. Among these methods, we employed faster R-CNN as the main object detection technique. The block diagram of a faster R-CNN network is illustrated in figure 5. So, it is possible to apply the prepared CCE still images with their polyp annotation GTs to the mentioned faster R-CNN. In this way, there is a main issue and it is necessary that an appropriate pre-trained CNN is chosen to be embedded into faster R-CNN network as its conducting deep convolutional network for feature extraction and region-based classification.

TABLE I. EVALUATION RESULTS FOR SEVERAL CNNS EMPLOYED AS THE CORE NETWORK IN FASTER R-CNN APPROACH

RPN Methods \ Criteria	Assessment Metrics (%)			
	Precision	***Sensitivity***	***F_1-Score***	***AP***
Faster RPN-Alexnet	83.92	34.98	45.35	50.90
Faster RPN-VGG16	85.07	41.34	50.80	59.16
Faster RPN-VGG19	89.29	41.54	52.94	64.16
Faster RPN-GoogleNet	65.51	12.24	19.86	15.41
Faster RPN-squeezenet	31.18	18.18	20.27	10.39
Faster RPN-resnet18	81.20	39.76	48.52	53.57
Faster RPN-resnet50	86.83	33.37	44.43	51.66
Faster RPN-resnet101	88.81	31.55	43.26	50.25
Faster RPN-inceptionv3	91.05	32.32	44.23	51.43
Faster RPN-mobilenetv2	62.96	28.79	36.52	31.88

Figure 6 illustrates some sample results for predicting the location of polyps in the form bounding boxes around them.

The results show that VGG19 as a series convolutional neural network provides the most higher value as F1-score for polyp

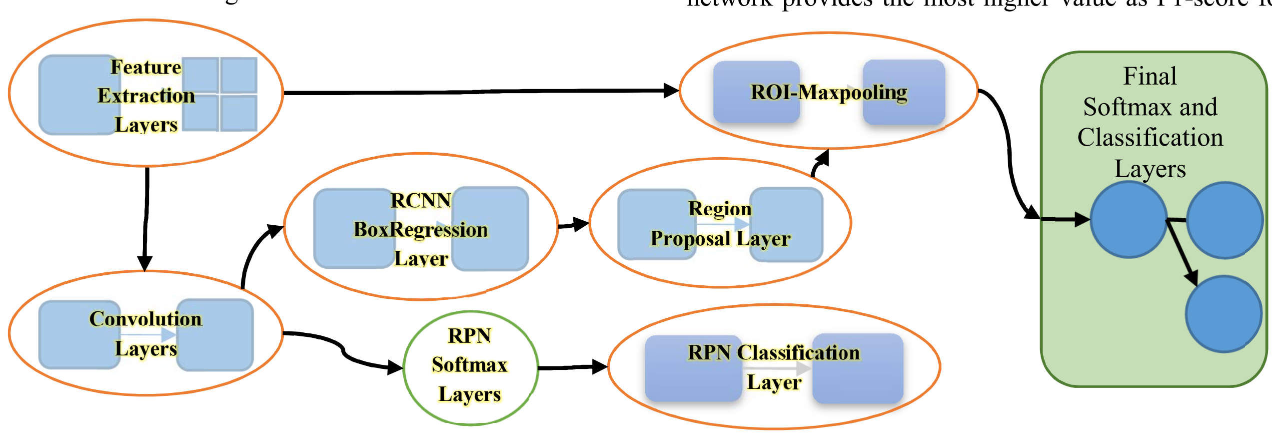
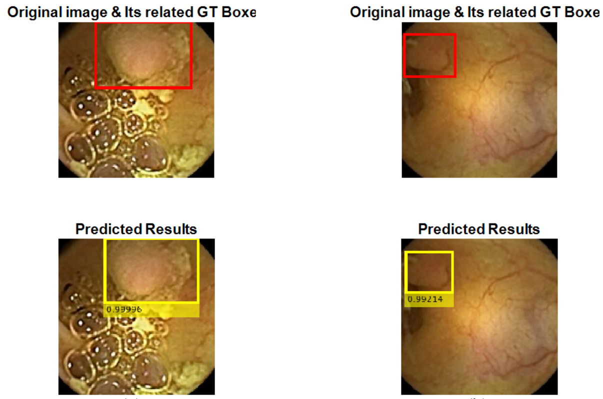


Fig. 5. Block diagram of faster region proposal deep convolutional network (Faster R-CNN)

To find a suitable core CNN for polyp localization purpose, we applied the provided CCE images with polyp region-wise GTs to a list of most popular pre-trained deep convolutional neural networks. By this strategy, we could find those CNNs which provide higher accuracy along with a better harmonic rate between precision and sensitivity metrics. According to this goal, the evaluation results shown in table 1 are achieved.

localization based on faster R-CNN approach. However, applying the faster R-CNN with VGG19 core network to test CCE video frames show that this network produces more False negative (FN) polyp objectives (figure 7) while the other networks with good performance compensate these shortcoming but suffers from producing higher false positive (FP) polyp candidates (figure 8).



(a) (b)

Fig. 6. Sample Results for Polyp Localization based on faster R-CNN with two different CNN core network: (a) AlexNet as a series CNN, (b) Resnet18 as a directed acyclic graph (DAG) CNN

To overcome this issue, we used a conditional combination of both faster R-CNN methods to improve polyp localization based on the following conditional statement:

1- Polyp Localization procedure:

F.R-CNN-Seriesnet: Localization output based on faster R-CNN with a Series CNN (like VGG19) core network as the coordinates of bounding boxes

F.R-CNN-DAGnet: Localization output based on faster R-CNN with a DAG (squeezenet or inceptionv3) core network as coordinates of bounding boxes

2- Applying the trained F.R-CNNs to test datasets

3- *For reducing the No. of FNs:*

IF (F.R-CNN-VGG19 ∪ F.R-CNN-DAGnet) ≠ ∅ THEN

4- *For decreasing the No. of FPs:*

IF (F.R-CNN-VGG19 ∩ F.R-CNN-DAGnet) ≠ ∅ THEN

Bounding Box = (F.R-CNN-VGG19 ∩ F.R-CNN-DAGnet);

ELSE

No Polyp is Detected.

As the logics in the above box state, it is possible to improve the accuracy and performance of polyp localization algorithm by the means of logical combination of the object detection CNNs with each other which have been trained based on the provided CCE images annotated for the polyps.

In the second step of the proposed polyp segmentation method, an automatic segmentation strategy based on a modern and popular model-based approach known as distance regularized leve-set evolution (DRLSE) [12]. This method is able to employ the bounding box around the polyp objectives detected by the means of aforementioned faster R-CNN approach as an initial contour for automatic segmentation of the polyps in a pixel-wise manner.

The DRLSE segmentation strategy comprises of three active contour energy functions: (a) external energy Eext, (b) internal energy Eint and (c) constraint energy Econs. The combinational relationship between Econs and Eint terms is named as level set regularization term and the term Eext depends strongly on the content of images as input data.

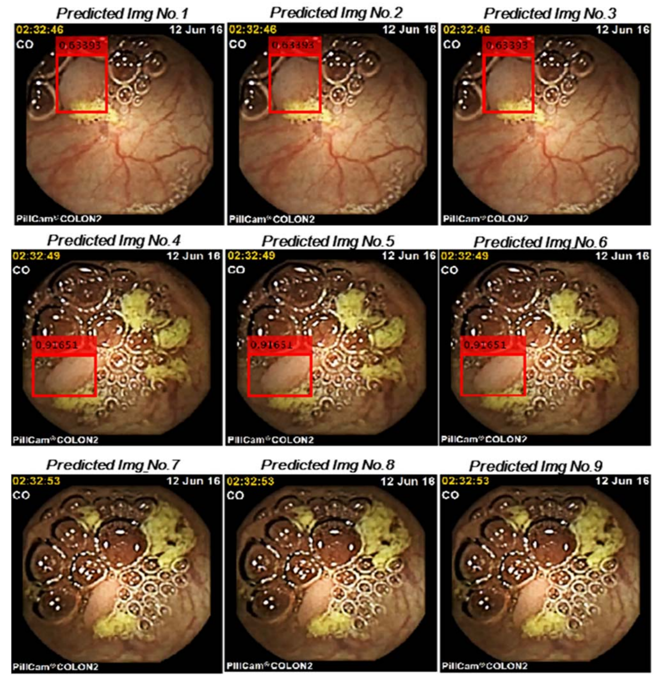


Fig. 7. TP polyps (frames in rows 1 and 2) vs. FN polyps (frames in row 3) detected or missed by a faster RCNN with VGG19 CNN core

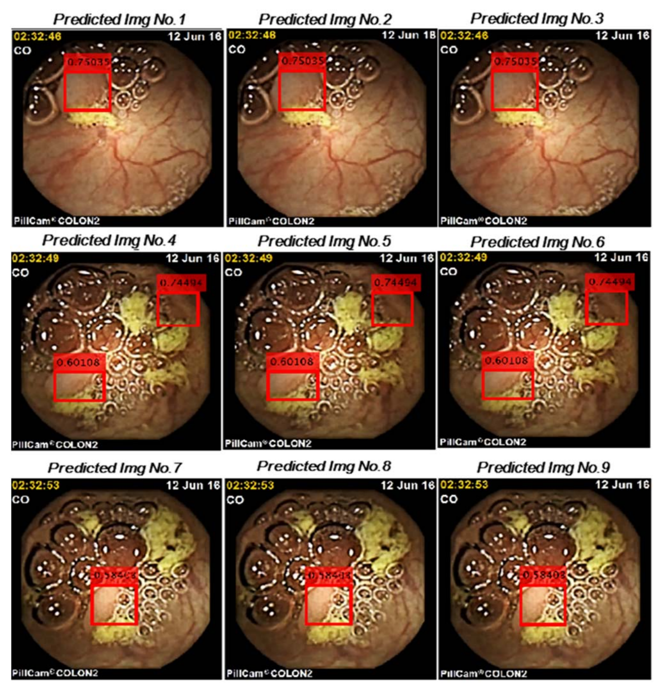


Fig. 8. TP detected polyps (frames in rows 1 and 3) vs. FP polyps (frames in row 2) detected by a faster RCNN with squeezenet CNN core

So, the total energy function in DRLSE is formulated as follows:

$$E_{total} = \eta E_{lvreg}(\phi) + E_{ext}(\phi) \tag{1}$$

where $E_{lvreg.}$ is the combination of E_{cons} and E_{int} known as level-set regularized term. Moreover, η and ϕ stand for a constant parameter and the contour variable, respectively. The initial value of ϕ variable can be defined as the following piece-wise function

$$\phi_0(x) = \begin{cases} -a_0, & if\ x \in R_0 \\ a_0, & otherwise \end{cases} \tag{2}$$

where a_0 is a positive constant value and R_0 is an initial area in the range of real numbers.

The main issue in front of a robust and reliable segmentation based on the mentioned DRLSE approach is that it is a parameter-oriented approach and the adjustment of the parameters for different modalities of the employed images need to be set according to the content of those images. To solve this problem, we employed the parameter adjustment similar to the ones proposed in [13].

The extracted bounding boxes around polyp objectives in CCE images by employed faster R-CNN are used as the initial zero level set contour for DRLSE minimization procedure during shrinkage procedure. The formula for initial contour extraction from the detected bounding boxes G_σ in an image I can be considered as follows

$$g = \frac{1}{1+\left|\nabla G_\sigma \times I\right|^2} \tag{3}$$

where $\nabla(.)$ stands for gradient function which in this paper is applied to the detected bounding box around polyps as the initial contour of DRLSE segmentation technique.

A sample of object-wise segmentation of a polyp based on the extracted bounding box around it using DRLSE algorithm is depicted in figure 9.

(a) (b) (c)

Fig. 9. Sample polyp localization and pixel-wise segmentation based on DRLSE: (a) CCE image with detected bounding box around polyp, (b) binary mask extracted using ROI rectangle, and (c) level-set segmented mask

A summary of the proposed polyp semantic segmentation method applied to CCE images is as shown in the following box:

- **Stage1:** providing polyp bounding box annotated database from CCE frames.
- **Stage2:** Automatic Semantic segmentation of polyps
    - **Step 1:** Training Faster R-CNN networks with several series and DAG CNNs.
    - **Step 2:** Employing extracted bounding boxes around polyps from test CCE databases as initial contours for pixel-wise segmentation of polyps.

III. Implementation Results

In this section, the whole procedures for the implementation of the proposed method from both hardware and software associated aspects are described. The former is mainly based on the employed hardware platform and the latter is according to the soft results achieved as the throughput of hardware system.

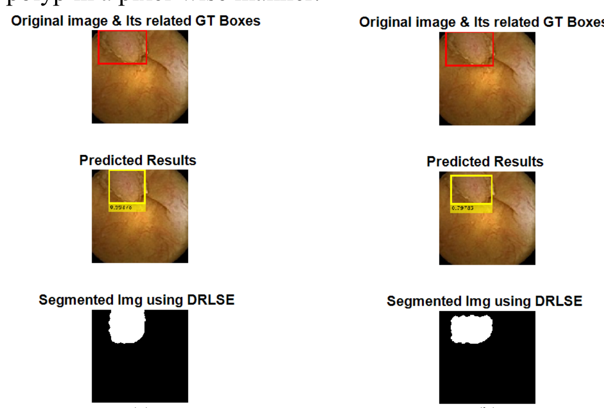
A. *Hardware specification*

The polyp detection based on faster R-CNN approach and its segmentation based on DRLSE technique are implemented on a system with Matlab2019b and a hardware specification of core i7 CPU 2.80 GHz and 16 GB RAM.

B. *Software implementation resutls*

The implementation of the proposed method to CCE frames video files leads to visual results shown in figures10 and 11.

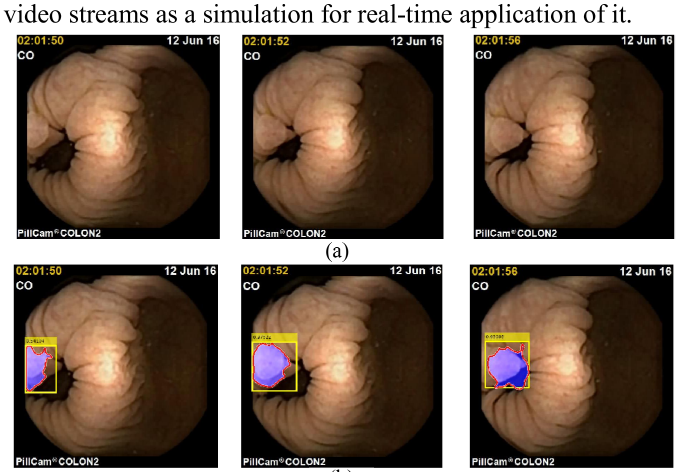
The images in figure 10, illustrate the whole procedures of proposed method for two different region proposal networks. The comparison between these two results shown in figure 10 demonstrates that the more precise the employed faster R-CNN localizes the polyp, the more accurate the DRLSE will segment the polyp in a pixel-wise manner.



(a) (b)

Fig. 10. Sample results for polyp semantic segmentation based on the implementation of proposed DRLSE method to different faster R-CNNs: (a) VGG19, and (b) inceptionv3.

Figure 11 shows some CCE video frames in a sequential order before and after proposed polyp segmentation method. These results are achieved by applying the proposed method to CCE video streams as a simulation for real-time application of it.



(b)

Fig. 11. Visual results of the proposed method's implementation: (a) frames with no detection, (b) frames after polyp localization and segmentation

Moreover, the statistical evaluation results of the proposed method are shown in figure 12.

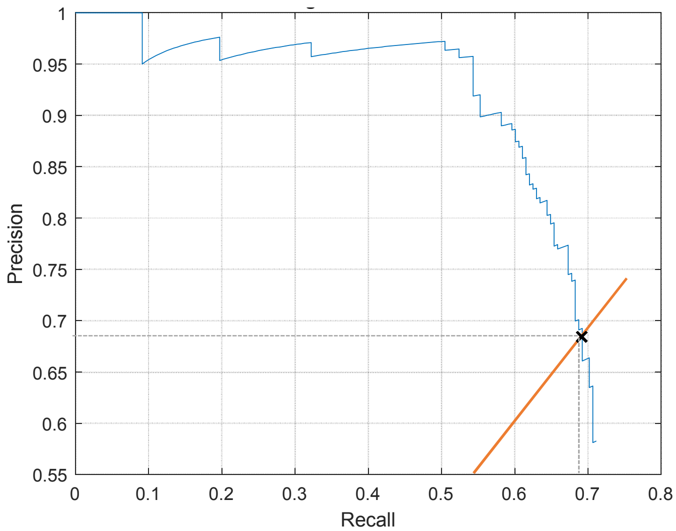


Fig. 12. Precision vs. Recall curve for polyp detection in CCE frames

According to the graph shown in figure 12, the Precision and Recall values are at an acceptable rate and the highest harmonic mean of them is reached at around 0.68 which is equal to maximum Dice or F_1-Score. The mathematical relationship between Dice or F_1-Score and Precision-Recall metrics is defined as follows:

$$F_1\text{-Score (Dice)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The important note about Dice Coefficient is that it is equal to the two times multiplication of the Area of Overlap divided by the total number of pixels in ground truth and predicted images. In fact, the Dice coefficient is very similar to the intersection over union (IoU) metric.

IV. Conclusion

Polyp detection from CCE acquired images is a challenging issue in front of physicians and experts of gastroenterology since studying CCE acquired video streams using software applications is a time-consuming and boring activity and due to these facts may lead to polyp missing. In addition, the object-wise extraction of polyps enables medical experts to diagnose their grade and predict their appropriate malignancy risk scores for further observation and better treatment.

In this paper, a new polyp segmentation method from colorectal CCE images based on an innovative artificial intelligence approach is proposed. The proposed polyp detection method not only provides a reliable polyp detection from CCE still images but also can provide semantic segmentation of them for further polyp studies. The proposed method is so fast that can be applied to Live CCE videos and it is possible to find polyps in a real-time manner.

References

[1] R. S. Kerrison, L. M. McGregor, S. Marshall, J. Isitt, N. Counsell et al, “Improving uptake of flexible sigmoidoscopy screening: a randomized trial of nonparticipant reminders in the English Screening Programme,” Endoscopy, Vol. 49, No. 1, pp. 35–43, 2017.

[2] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, “Wireless capsule endoscopy,” *Nature*, Vol. 405, pp. 417–417, 2000.

[3] B. S. Lewis, and P. Swain, “Capsule endoscopy in the evaluation of patients with suspected small intestinal bleeding: Results of a pilot study,” Elsevier Journal of Gastrointestinal Endoscopy, vol. 56, No. 3, pp. 349–353, 2002.

[4] S. N. Adler, and Y. C. Metzger, “PillCam COLON capsule endoscopy: recent advances and new insights,” Therapeutic Advances in Gastroenterology, Vol. 4, No. 4, pp. 265–8, 2011. doi:10.1177/1756283X11401645.

[5] J. Hausmann, J. P. Linke, J. G. Albert, J. Masseli, A. Tal, A. Kubesch, N. Filmann, M. Philipper, and M. Farnbacher, “Time-saving polyp detection in colon capsule endoscopy: evaluation of a novel software algorithm,” International Journal of Colorectal Disease, Vol. 34, pp. 1857-1863, 2019. https://doi.org/10.1007/s00384-019-03393-0.

[6] V. Blanes-Vidal, G. Baatrup, and E.il S. Nadimi, “Addressing priority challenges in the detection and assessment of colorectal polyps from capsule endoscopy and colonoscopy in colorectal cancer screening using machine learning,” Acta Oncologica, Vol. 58, No. S1, S29-S36, 2019. DOI: 10.1080/0284186X.2019.1584404.

[7] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR, 2014.

[8] R. Girshick, “Fast R-CNN,” IEEE International Conference on Computer Vision (ICCV2015), pp. 1440–1448, Dec. 2015. https://doi.org/10.1109/ICCV.2015.169

[9] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” 28th International Conference on Neural Information Processing Systems, Vol. 1, pp. 91–99, December 2015.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, “SSD: Single Shot MultiBox Detector,” 14th European Conference on Computer Vision (ECCV2016), Amsterdam, The Netherlands, pp. 21-37, Oct. 2016. DOI: 10.1007/978-3-319-46448-0_2

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016), Jun. 2016, DOI: 10.1109/CVPR.2016.91

[12] C. Li, C. Xu, C. Gui and M. D. Fox, "Distance Regularized Level Set Evolution and Its Application to Image Segmentation," in IEEE Transactions on Image Processing, vol. 19, no. 12, pp. 3243-3254, Dec. 2010. doi: 10.1109/TIP.2010.2069690

[13] A. Tashk, T. Hopp and N. V. Ruiters, “An Innovative Practical Automatic Segmentation of UltraSound Computer Tomography Images Acquired from USCT Systems,” Springer Journal of Iranian Journal of Science and Technology Transactions of Electrical Engineering (ISTE), Jul. 2018, DOI: 10.1007/s40998-018-0098-9.