Automatic Segmentation of Colorectal Polyps based on a Novel and Innovative Convolutional Neural Network Approach

Ashkan Tashk[1,†], Jürgen Herp[2,†] and Esmaeil Nadimi[3,†]
[†]Mærsk and Mc-kinney Muller Institute (MMMI)
Suddansk Universitet (SDU)
Campusvej 55, Odense M
Denmark
Emails: [1]asta@mmmi.sdu.dk, [2]herp@mmmi.sdu.dk, [3]esi@mmmi.sdu.dk (https://ml.sdu.dk/)

Abstract: Polyp is the name of a colorectal lesion which is created by cells clumping on the lining of the colon. The colorectal polyps can lead to severe illnesses like colon cancer if they are not treated at the early stage of their development. In current days, there are very many different polyp detection strategies based on biomedical imageries such colon capsule endoscopy (CCE) and optical colonoscopy (OC). The CCE imagery is non-invasive but the quality and resolution of acquired images are low. Moreover, it costs more than OC. So, today OC is the most desired method for detecting colorectal polyps and other lesions besides of its invasiveness. To assist physicians in detecting polyps more accurately and faster, machine learning with biomedical image processing aspect emerges. One of the most the state-of-the-art strategies for polyp detection based on artificial intelligence approach are deep learning (DL) convolutional neural networks (CNNs). As the categorization and grading of polyps need significant information about their specular highlights like their exact shape, size, texture and in general heir morphological features, therefore it is very demanded to employ semantic segmentation strategies for detecting polyps and discriminating them from the background. According to this fact, a novel and innovative method for polyp detection based on their semantic segmentation is proposed in this paper. The proposed segmentation classifier is in fact a modified CNN network named as U-Net. The proposed U-Net provides an advanced and developed semantic segmentation ability for polyp detection from OC images. For evaluating the proposed network, accredited and well-known OC image databases with polyps annotated by professional gastroenterologists known as CVC-ClinicDB, CVC-ColonDB and ETIS-Larib, are employed. The results of implementation demonstrate that the proposed method can outperform the other competitive methods for polyp detection from OC images up to an accuracy of 99% which means that the life lasting hopes could be increased to a considerable ratio.

Key-Words: Optical colonoscopy (OC); Polyp semantic segmentation; modified U-Net; Validation Criteria.

1 Introduction

The statistics revealed by the world health organization shows that the colorectal cancer is one of the third incidence and death causes among all types of cancers worldwide. For instance, according to the Cancer data from the Nordic countries' statistics [1], colorectal cancer is the first cause of death among Danish males and it is at the second rank after breast cancer for Danish females [2]. In addition, the early diagnosis of colon cancer will increase the survival rate up to 95% in the first stages [3]; hence the importance of detecting it on its early stages by using imaging techniques is very demanding and will lead to promising results [4].

Today, there are very many different clinical solutions employed for colorectal cancer diagnosis such as imagery screening comprising of colon capsule endoscopy (CCE) and CT colonography; however, these solutions suffer from lower accuracy for small polyps' detection and do not allow to obtain histological confirmation. Due to this redundancy, optical colonoscopy is the most preferable solution suggested for colorectal disease and lesions screening [5].

For accurate analysis and study of polyps as the most occurring colon lesion, the digital image processing strategies are very desired and appropriate for polyp detection [6].

One of the most not only diagnostic but also treatment strategies for colorectal cancer diagnosis is optical colonoscopy. The role of colonoscopy is both observing colon tissue in vivo and provides removing visible colorectal lesions known as polyps due to expert's diagnosis as a treatment strategy. Polyps have different types and shapes. Therefore, their

detection and accurate localization will aid colonists to make appropriate decision about them.

There are very many different computer-aided systems proposed for automatic polyp detection from colonoscopy acquired images [5-10]. Indeed, these methods are based on pattern recognition and machine learning strategies with the same procedure for stationary or video records. Such systems comprise of stages shown in Fig. 1.

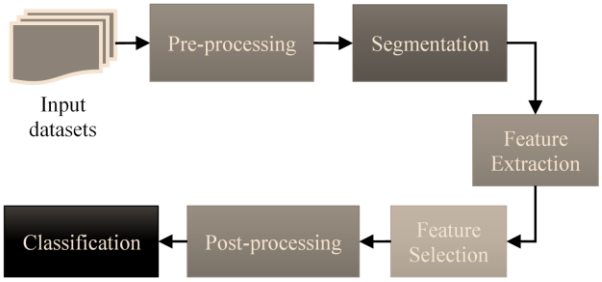


Fig. 1. Block diagram for a typical computer-aided polyp detection system.

Generally, it is possible to imply that the proposed automatic polyp detection strategies have contributions in one or more stages depicted on figure 1. The contributions can be summarized as the following aspects

- Strategies contributed on Segmentation proposals.
- Strategies with proposals and contributions about extracting features from either texture or color context or even morphology of ground truth polyps.
- Strategies based on either Supervised classifier such as fully convolutional networks (FCN), random forest (RF), SVMs or unsupervised clustering strategies.

In one of the initial automatic polyp segmentation method, an appearance modeling in the segmentation step is proposed as in [5]. The main idea is a region segmentation descriptor named as Sector Accumulation-Depth of Valleys Accumulation (SA-DOVA). The main effect and contribution of proposed descriptor is more on polyp region segmentation than instant segmentation of individual polyps. The performance of their descriptor in comparison with other competitive descriptors like MPEG-7, shape, texture and color descriptors has been demonstrated to be higher. The highest polyp segmentation accuracy achieved by this method is 75.79% which was higher than NCuts and TurboP segmentation methods. The authors applied their own proposed segmentation method to their own provided dataset named CVC-ColonDB database [11]. The authors of [5] also proposed another polyp localization method which is named as Window Median Depth of Valleys Accumulation (WM-DOVA) energy maps [6]. The polyp localization results for this method is at most about 72.33% for 265 out of 300 CVC-colonDB database images [10] and about 70.26% for 430 out of 612 CVC-clinicBD database images [11]. At all, the methods in [5] and [6] proposed approaches for more determining the presence and location of polyps but not detecting them in an accurate pixel-wise manner.

Li et al. [7] proposed a new deep learning network with FCN based architecture in an end-to-end hierarchy. The proposed network comprises of 31 layers and its difference with conventional FCNs is that it does not include cropping 2D layers. The proposed method was applied to CVC-ClinicDB database images and achieved an accuracy of 96.98%. The main deficiency of the proposed method is that it is designed only for the specific size of endoscopy images in CVC-ClinicDB due to the specific deconvolutional size of intermediate layers.

In [8], several CNNs with different pretrained CNN backbones such as Alexnet, Googlenet and VGG are employed for FCN architecture and implementation. The proposed strategy was applied to ETIS-Larib dataset and the highest efficiency was achieved for FCN-VGG architechture which was about 60.2% F_1-Score as the harmonic mean of precision and recall. The proposed method did not include a very novel contribution and suffers from the same shortcomings in [7].

The proposed method in [9] has also the same basic deep learning CNN as proposed in [7]. The proposed network has an FCN-8S structure with almost the same deep neural network graph. However, Akbari et al. in [8] applied a patch selection and data augmentation solution to modify the shortcoming of methods in [7] and [8]. The implementation is done on a popular convolutional architecture known as caffe [12]. The evaluation results for implementing over CVC-colonDB database images reaches to an accuracy of 97.7%.

In this paper, the proposed method improves the shortcomings of the competitive methods for automatic detection of colorectal polyps such as compatibility with all standard database images with various sizes. Moreover, it modifies the common architecture of U-Net for being employed with any image or video formats associated with optical colonoscopy modalities.

The rest of this paper is organized as follows. In section II, the details of the proposed method and its stages are described with full information about proposed network and its various stages. In section III, the employed databases for applying proposed method and consequent implementation results are introduced. Section IV concludes remarkable achievements of our proposed method and discusses possible future works.

2 Proposed Method

AS the main purpose of our proposed method is to apply a flexible, robust and efficient semantic segmentation approach to various optical colonoscopy databases so that it can be employed as a reliable computer-aided diagnosis tool for colorectal studies. According to this fact, we proposed our method to improve previously proposed semantic segmentation strategies like image size flexibility and performing with high accuracy along with low computational costs. Therefore, it is designed to consist of three stages. These stages are described in the following steps.

2.1 Pre-processing stage: Color Space Augmentation

In this stage, the input colonoscopy slice images or frames are transformed to three different color spaces known as CMYK, La*b* and gray-level. This image color transformation removes RGB color space redundancies due to being more uncorrelated and ads more useful color features to assist discrimination. The proposed image trajectory is as follows:

1) 4 channels related to CMYK color space transformation,
2) 3 channels related to La*b* color space transformation,
3) 1 gray-level image.

2.1.1 Color Space Augmentation Advantages

All the above color transformed channels are concatenated to each other from their 3rd dimension and form an augmented image with M×N×8 dimensions. By this approach, the images are prepared for being imported as inputs to the designed and built U-Net convolutional neural network in the next stage. To make the input layer of the proposed network independent of input image size, a size patching solution is employed. The employed patching strategy even helps adding more rotational patches to the number of dataset images. This approach increases the number of training images and will assist the network to overcome the shortage of input data for a convergent training procedure.

2.2 Proposed U-Net Architecture

In this stage, an innovative deep learning convolutional neural network with adaptive architecture as a robust and well-known object-wise segmentation strategy is employed. The proposed CNN has a U-Net layer graph. Most significant benefits of U-Net over conventional FCNs or even Mask RCNN are its faster running ability and its compatibility with various input dimensionality [13]. In fact, U-Net belongs to FCN networks since it does not include fully connected layers. However, there is a characteristic which makes a distinction between U-Net and other FCNs and it is related to its specific architecture which is symmetric and can be divided into three distinct parts known as:

1) The Downsampling/contracting path.
2) Bottleneck.
3) The Upsampling/Expanding path.

U-Nets unlike FCNs have skip connections between downsampling and upsampling paths [14]. This structure reinforces U-Nets against overfitting due to low number of datastore images. Figure 2 illustrates the depicts sample FCN and U-Net and provides a graphical comparison between them.

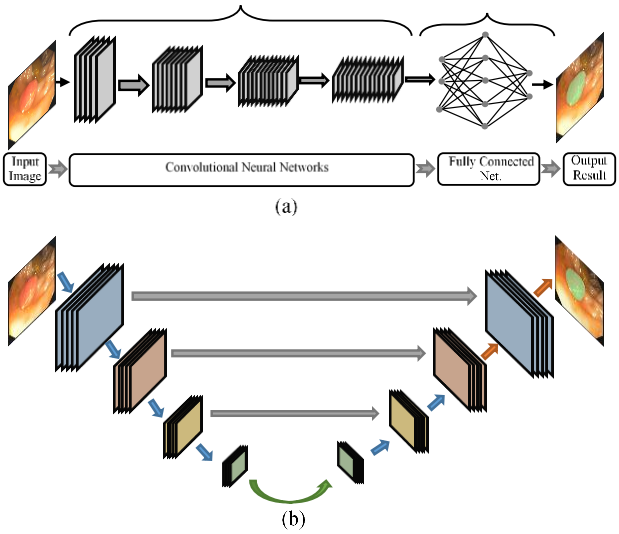


Fig. 2 Comparative schemes between for : (a) conventional deep learning neural networks and (b) U-NETs.

The U-Net privileges over other deep learning CNN semantic segmentation approaches are as follows:

(a) U-Net provides general information about context of pixels belonging to the target objects by combining localization and context features and therefore leads to better segmentation map predictions.

(b) There is no dense layer in its architecture; so, it is adaptive to different input image sizes.

(c) It is suitable for applying various types of augmentation due to limited number of annotated samples in biological applications.

For describing the whole construction of proposed U-Net, it could be start from a general overview of whole architecture and then going through layers, respectively. According to the first stage of the proposed method, some patches with various rotations are extracted from augmented training database. The dimension of patches is M×M×N in which M is usually equal to a power of 2 and N is a multiplication of 8. Such patch sizes lead to no need for padding in proposed U-Net CNN and in fact make it adaptive to input image datastores with various size. The input datastore images will be fed into the convolutional layers in the contracting path of the proposed U-Net which can be named as encoding layers. Each encoding path includes three layers which are Convolution, batch normalization and ReLU layers, respectively. In the bottleneck of proposed network, there are some deconvolution and ReLU layers. The deconvolution layers construct the first decoding approach and includes rational strides. After this path, the decoding layers with expanding procedures are employed. These layers have reflective roles against the mirror layers in contracting path and provide upsampling abilities. Finally, the last two layers have the responsibility of concluding the semantic segmentation procedure. The output layer is also employed Dice loss approach.

As it can be inferred from Table I, the proposed U-Net comprises of layers in two stages. The first stage is related to feature extraction (layers 2 to 19) known as encoding layers and the second stage associates with reconstructive predictions (layers 20 to 42) and are named as decoding layers. The first and the last two layers are input layer and output conclusive layers, respectively which can be considered as common layers in the same conventional CNNs.

One of the important attributes of the proposed network is that there are connections between the same layers of encoding and decoding stages.

Table 1. Fully Detailed Hirearchitecture Of Proposed U-Net

Layer No.-Name	Kernel size	No. Kernels/ Channels	Stride	Activations
1-Input	-	-	-	M×M×N×1
2-En1_Conv1	3×3×3	P	[1 1 1]	M×M×N×P
3-En1_BN1	-	P	-	M×M×N×P
4-En1_ReLU1	-	-	-	M×M×N×P
5-En1_Conv2	3×3×3	Q=2×P	[1 1 1]	M×M×N×Q
6-En1_ReLU2	-	-	-	M×M×N×Q
7-En1_maxpool	2×2×2	-	[2 2 2]	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times Q$
8-En2_Conv1	3×3×3	Q	[1 1 1]	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times Q$
9-En2_BN1	-	Q	-	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times Q$
10-En2_ReLU1	-	-	-	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times Q$
11-En2_Conv2	3×3×3	R=2×Q	[1 1 1]	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times R$
12-En2_ReLU2	-	-	-	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times R$
13-En2_maxpool	2×2×2	-	[2 2 2]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times R$
14-En3_Conv1	3×3×3	R	[1 1 1]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times R$
15-En3_BN1	-	R	-	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times R$
16-En3_ReLU1	-	-	-	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times R$
17-En3_Conv2	3×3×3	S=2×R	[1 1 1]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
18-En3_ReLU2	-	-	-	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
19-En3_maxpool	2×2×2	-	[2 2 2]	$\frac{M}{8}\times\frac{M}{8}\times\frac{N}{8}\times S$
20-De4_Conv1	3×3×3	S	[1 1 1]	$\frac{M}{8}\times\frac{M}{8}\times\frac{N}{8}\times S$
21-De4_ReLU1	-	-	-	$\frac{M}{8}\times\frac{M}{8}\times\frac{N}{8}\times S$
22-De4_Conv2	3×3×3	T=2×S	[1 1 1]	$\frac{M}{8}\times\frac{M}{8}\times\frac{N}{8}\times T$
23-De4_ReLU2	-	-	-	$\frac{M}{8}\times\frac{M}{8}\times\frac{N}{8}\times T$
24-De4_UpConv	2×2×2	T	[2 2 2]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times T$
25-Concat3	-	-	-	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times(T+S)$
26-De3_Conv1	3×3×3	S	[1 1 1]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
27-De3_ReLU1	-	-	-	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
28-De3_Conv2	3×3×3	S	[1 1 1]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
29-De3_ReLU2	-	-	-	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
30-De3_UpConv	2×2×2	S	[2 2 2]	$\frac{M}{4}\times\frac{M}{4}\times\frac{N}{4}\times S$
31-Concat2	-	-	-	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times(S+R)$
32-De2_Conv1	3×3×3	R	[1 1 1]	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times R$
33-De2_ReLU1	-	-	-	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times R$
34-De2_Conv2	3×3×3	R	[1 1 1]	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times R$
35-De2_ReLU2	-	-	-	$\frac{M}{2}\times\frac{M}{2}\times\frac{N}{2}\times R$
36-De2_UpConv	2×2×2	R	[2 2 2]	M×M×N×R
37-Concat1	-	-	-	M×M×N×(R+Q)
38-De1_Conv1	3×3×3	Q	[1 1 1]	M×M×N×Q
39-De1_ReLU1	-	-	-	M×M×N×Q
40-De1_Conv2	3×3×3	Q	[1 1 1]	M×M×N×Q
41-De1_ReLU2	-	-	-	M×M×N×Q
42-Final_Conv	1×1×1	2	[1 1 1]	M×M×N×2
43-Softmax	-	-	-	M×M×N×2
44-Ouput_Loss	-	-	-	-

The mentioned U-Net in table 1 comprises of 44 layers; however, the number of layers can vary due to the decomposition coefficient. For instance, the number of decompositions or downsampling stages for the U-Net depicted in table 3 is equal with 3, i.e. there would be 3 max-pooling with the rate of 2 (kernel size 2), which leads to a geometric progression with progression coefficient 2. This parameter can be any positive integer and therefore, the number of U-Net layers will be proportional with this number.

2.3 Post-Processing Stage

To modify output segmented images and remove the redundant pixels scattered away, a morphological modifying approach is applied as a post processing stage. A sample result of applying post-processing to the proposed U-Net semantic segmentation is shown in figure 3.

(a) (b) (c)

Fig. 3. Porposed polyp segmentation results: (a) sample polyp image, (b) instant Segmentation before post-process and (c) segmentation results after post-process

The whole training process is conducted on MATLAB 2019a with GPU of NVIDIA GTX 2080.

In general, the full proposed method can be shown in the following pseudo code format:

Table 2: pseudo code for the proposed method

Inputs: images as Train, Test and Validation data-stores (Trds, Teds, Valds); input patch dimensions M and N; initial sampling rate P (Q=2×P, R=2×Q, S=2×R and T=2×S; No. of Training epochs as Termination Criteria = NE, No. of decompositions = NN.
Result: semantic segmented images

Pre-processing stage:
Read Trds, Teds and Valds for color transformation

def Color_transfer_preprocess():
[Trdsn,Valdsn,Tedsn] = Concat(rgb2cmyk(Trds,Valds,Teds), rgb2lab(Trds,Valds,Teds),rgb2gray(Trds,Valds,Teds))
return Trdsn, Valdsn, Tedsn

Training Procedure:
def U_NET_semantic_Seg([M, N], Trds_n, Valds_n, sampling_rates =P, NN, NE)
For i 1 to NE
In_params = Train_input_layers(M,N)
Down_params = Train_Downsapling_Layers(NN)
Up_params = Train_Upsampling_Layers(NN)
UNET_learnable_params = concat(In_params, Down_params, Up_params)
return UNET_learnable_params
}

Test Procedure:
def Test_procedure(Tedsn,UNET_Learnable_params)
calculate No. of TP, FP, FN and TN pixels
Accuracy = (TP+TN)/(FP+FN+TP+TN)
F1_Score = (2×TP) / (2×TP + FP + FN)
return Accuracy, F_1-Score

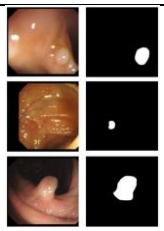
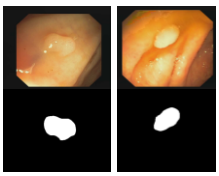
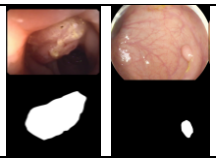
3 Experimental Results

To demonstrate the efficient application of proposed method for semantic segmentation of colorectal polyps, it is necessary to apply the network to standard optical colonoscopy databases. Such standard databases are introduced in a separate subsection. In addition, it is more important to choose appropriate evaluation criteria for assessment purposes. These criteria and their associated parameters are defined under second subsection. The last subsection of this part is dedicated to evaluation results representation.

3.1 Optical Colonoscopy Datasets

The standard optical colonoscopy databases which can be employed for proposed method evaluation are introduced in Table II. The related information to each of these databases and their sample images alongside their related ground truths are described briefly in separate column of this table. The most important note about these databases is that they are all provided by well-known clinical institutes and biomedical research centers.

Table 2. Employed Datasets

Dataset Name	Sample images and related ground truths	Brief Description
CVC-ColonDB [10]		300 region of interest (ROI) frames selected from 15 short colonoscopy sequences, coming from 15 different studies.
CVC-ClinicDB [11]		612 still images from 29 different sequences. Each image has its associated manually annotated ground truth covering the polyp.
ETIS-Larib [15]		196 frames extracted from colonoscopy videos

3.2 Evaluation Criteria

There are general criteria to evaluate and assess the test results of proposed method's implementation to standard colorectal polyp database images. These criteria are pixel-wise, and their related parameters are defined as follows:

True positive (*TP*): Real polyp pixels denoted correctly as polyp pixels.

True negative (*TN*): Real non-polyp pixels predicted correctly as non-polyp pixels.

False positive (*FP*): Non-polyp pixels predicted wrongly as polyp pixels.

False negative (*FN*): Real polyp pixels ignored in a false manner as non-polyp pixels.

The above defined parameters denote the following evaluation criteria:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F}_1\text{-Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

where F_1-Score, also known as F-measure, is the harmonic mean of Precision and Recall criteria.

However, there are two other popular criteria for considering as similarity evaluation metrics which are defined as follows:

$$\text{Dice} = \frac{2\times TP}{(2\times TP+FP+FN)} \quad (5)$$

$$\text{Jaccard} = \frac{TP}{(TP+FP+FN)} \quad (6)$$

Jaccard is also known as intersection over union (IoU) criterion and it is considered as a metric for managing classification stage in semantic segmentation CNNs.

The Relationship between Jaccard and Dice similarity coefficients is as follows:

$$\text{Dice} = \frac{2\times\text{Jaccard}}{(1+\text{Jaccard})} \quad (7)$$

$$\text{Jaccard} = \frac{\text{Dice}}{(2-\text{Dice})} \quad (8)$$

As it is inferred from the definition of equations (7) and (8), the two mentioned assessment criteria are hidden inside the F_1-Score. In fact, Dice metric is like F1-Score and Jaccard observes the following inequality

$$F_1\text{-Score}/2 \leq \text{Jaccard} \leq F_1\text{-Score} \quad (9)$$

So, Jaccard or IoU and F_1-Score metrics are always within a factor 2 of each other.

3.3 Comparative Results

As a concluding step, the proposed method is applied distinctly to each of the introduced databases and the results depicted in Tables 3 to 5 are achieved.

Table 3 shows that our proposed method can overperform the previously proposed method in [7] in the case of producing more TN and reducing the number of FN so that it reaches to higher accuracy and recall rates than the comparative method.

Table 3. Comparative Results For Cvc-Clinicdb Dataset

Criteria / Method	Evaluation Criteria (%)			
	Accuracy	**Precision**	**Recall**	**F_1-Score**
FCN+U-Net Method [7]	97.0	**90.0**	77.0	**83.0**
Proposed Method without Post-Process	**98.32**	86.1	76.5	81.0
Proposed Method with Post-Process	**99.02**	70.2	**82.7**	76.0

In figure 4, the bar chart showing the comparative values of the accuracy and F_1-Score criteria for the method proposed in [7] and our proposed method (without post-process and with Post-process) are depicted.

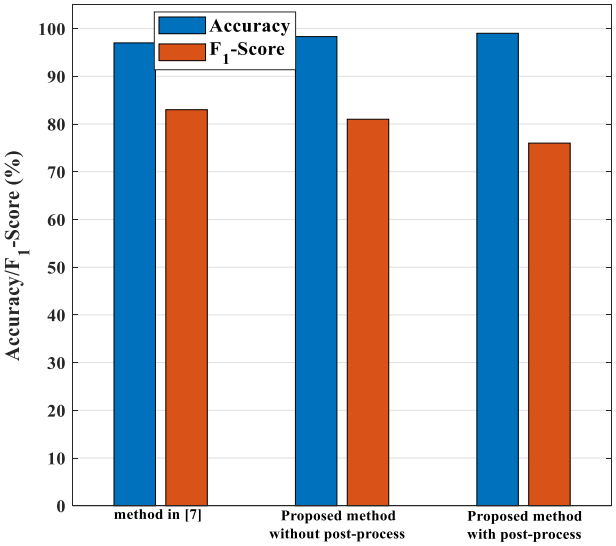


Fig. 4. Accuracy and F1-Score comparison between the proposed method in [7] and the proposed one in this paper (without and with post-processing stage)

Table 4 shows the evaluation results for both our proposed method and the one proposed in [8].

According to the results, our proposed method with no post-process produces less FPs than the competitive method in [8] but it includes more FNs. By applying post-processing stage, the rate of FNs decreases, and the rate of FP and TN rise so that the accuracy and recall increase, but the rate of precision falls down. However, the visual results in the second approach (with post-process) seems to be better.

Table 4. Comparative Results For Etis-Larib Dataset

Criteria / Method	Evaluation Criteria (%)			
	Accuracy	**Precision**	**Recall**	**F_1-Score**
Proposed Method [8]	-	73.6	86.3	79.4
Proposed Method without Post-Process	**99.1**	**92.82**	77.67	**84.57**
Proposed Method with Post-Process	**99.6**	70.2	**90.9**	79.23

As the results shown in table 4 and also in figure 5, the main reason for higher value of F_1-Score criteria for the proposed method in comparison with those of the proposed one in [8] is that U-Net is symmetric and so it provides local with global feature combination.

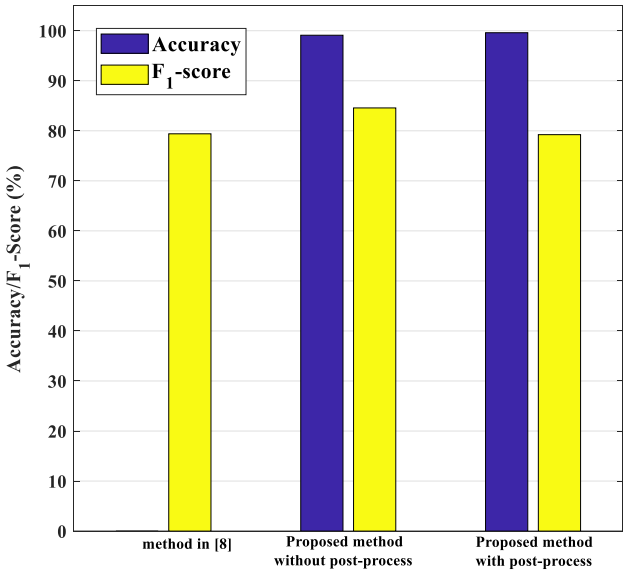


Fig. 5. Accuracy and F1-Score comparison between the proposed method in [8] and the proposed one in this paper (without and with post-processing stage)

Table V shows the comparative implementation results for our proposed method and the one in [9] in the terms of the same criteria applied for evaluating the other two database images. The results demonstrate that the proposed method resemble the competitive method proposed in [9] with less computational complexity. The reason for lower computational complexity is related to the lower number of layers in the proposed CNNs. The proposed method in [9] comprises of at least 10 convolutional layers which have more computational and connection complexities.

Table 5. Comparative Results For Cvc-Colondb Dataset

Criteria / Method	Evaluation Criteria (%)			
	Accuracy	**Precision**	**Recall**	**F_1-Score**
FCN-8s [9]	97.7	**88.3**	74.8	**81.0**
Proposed Method without Post-Process	97.3	80.5	71.2	75.6
Proposed Method with Post-Process	**98.2**	62.0	**82.4**	70.7

The results shown in figure 6 demonstrates that however our proposed method in this paper is not able to outperform the proposed method in [9] from F_1-Score aspect, but its accuracy exceeds the accuracy in [9].

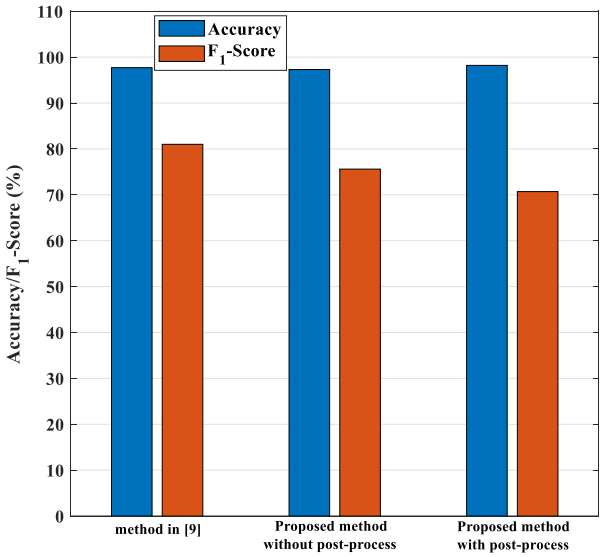


Fig. 6. Accuracy and F1-Score comparison between the proposed method in [9] and the proposed one in this paper (without and with post-processing stage)

4 Conclusion

This paper proposed a novel approach for fully automatic polyp detection. The proposed approach includes three main steps. First, a preprocess step is applied to the dataset images. The preprocess comprises of three distinct color transformations known as La*b*, CMYK and gray-level. The extracted color channels are concatenated to each

other to form multicolor channel datastore. In the second step, an innovative convolutional neural network with U-Net architecture is proposed. The main attribution of the proposed network is that it is adaptive to colorectal images with various heights and width. Moreover, the network is organized to extract features from either stationary images or video streams at once. The third step is a post processing for improving the pixel-wise classification outcomes.

The proposed method in this paper can detect and localize polyps from various optical colonoscopy databases, in a fully automatic manner. So, it is well-acquainted with several colorectal databases which include optical colonoscopy images with different dimensionalities. The implementation results of proposed method are promising in terms of both illustrative contributions and experimental evaluations. It gives higher performance in terms of Accuracy (at least 98.2%) than the other previously proposed methods.

References

[1] Special issue: NORDCAN Cancer data from the Nordic countries.

[2] Cancer statistics – specific cancers: Source: Statistics Explained (http://ec.europa.eu/eurostat/statisticsexplained) - 17/10/2018.

[3] A. Tresca, "The Stages of Colon and Rectal Cancer," New York Times , About.com, 2010.

[4] J. P. Hassinger, S. D. Holubar and et al. "Effectiveness of a Multimedia-Based Educational Intervention for Improving Colon Cancer Literacy in Screening Colonoscopy Patients," Diseases of the Colon & Rectum, vol. 53, no. 1301, 2010.

[5] J. Bernal, J. Sánchez, F. Vilariño, "Towards Automatic Polyp Detection with a Polyp Appearance Model," Elsevier Pattern Recognition, vol. 45, no. 9, pp. 3166-82, September, 2012.

[6] J. Bernal, J. Sánchez, G. Fernández-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.

[7] Q. Li, G. Yang, Z. Chen, B. Huang, L. Chen, D. Xu, X. Zhou, S. Zhong, H. Zhang and T. Wang, "Colorectal Polyp Segmentation Using A Fully Convolutional Neural Network," 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI2017), Shanghai, China, 14-16 October, 2017.

[8] P. Brandao, E. Mazomenos, G. Ciuti, R. Caliò, F. Bianchi, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo, and D. Stoyanov1, "Fully Convolutional Neural Networks for Polyp Segmentation in Colonoscopy," SPIE Medical Imaging - Computer-Aided Diagnosis Proceedings, vol. 10134, Orlando, Florida, United States, 2017.

[9] M. Akbari, M. Mohrekesh, E. Nasr-Esfahani, S.M. Reza Soroushmehr, N. Karimi, S. Samavi, K. Najarian, "Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18-21 July, 2018.

[10] J. Bernal, J. Sánchez, F. Vilariño, Cvc-colondb: A database for assessment of polyp detection, Database, 2012.

[11] http://mv.cvc.uab.es.

[12] Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," 22nd ACM international conference on Multimedia, pp. 675–8, Orlando, FL., 3-7 November, 2014.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI2015), pp. 234-41, 2015.

[14] R. Lguensat, M. Sun, R. Fablet, E. Mason, P. Tandeo and Ge Chen, "EddyNet: A Deep Neural Network For Pixel-Wise Classification of Oceanic Eddie," 10 November, 2017.

[15] 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.