

# AN AUTOMATIC COLORECTAL POLYPS DETECTION APPROACH FOR CT COLONOGRAPHY

Mohamed Yousuf<sup>\*‡</sup>    Islam Alkabbany<sup>\*</sup>    Asem Ali<sup>\*</sup>    Salwa Elshazley<sup>†</sup>    Albert Seow<sup>§</sup>  
Gerald Dryden<sup>§</sup>    Aly Farag<sup>\*</sup>

<sup>\*</sup> Computer Vision and Image Processing Laboratory (CVIP), University of Louisville, Louisville, KY

<sup>†</sup> Kentucky Imaging Technologies Louisville, KY

<sup>‡</sup> Faculty of Engineering, Ain Shams University, Cairo, Egypt

<sup>§</sup> School of Medicine, University of Louisville, Louisville, KY

## ABSTRACT

In this work, we propose an automatic colorectal polyps detection approach that consists of two cascade stages. In the first stage, a CNN model is trained to detect polyps in axial CT slices. The CNN model has been fed by the segmented colon wall CT slices instead of the original CT slices. Using the segmented images as an input to the CNN model has drastically improved the detection and localization results, e.g., the mAP is increased by 36%. To reduce the false positives generated by the detector, the second stage classifier is deployed to exploit the different views of the CT scans instead of the axial view only. So, the classifier is trained using the 2D images of axial views, i.e., the candidate polyps generated by the detector, as well as their corresponding 2D images of sagittal and coronal views. The experimental results of this approach were validated by 3 radiologists and the approach successfully identified polyps after the classification stage with an AUC  $\sim 98.6\%$ .

**Index Terms**— Colorectal polyp, CT, CNN, MAP-MRF.

## 1. INTRODUCTION

Colorectal cancer (CRC) originates as small growths (polyps) attached to the luminal wall of the colon and rectum. If polyps are not diagnosed and treated, they may grow in size and become cancerous. The American Cancer Society (ACS) recommends that people at average risk of colorectal cancer should start regular screening at age 45 [1]. Optical colonoscopy (OC) is the standard screening approach. However, it is an invasive and expensive procedure. On the other hand, Computed Tomographic Colonography (CTC) provides clinically acceptable performance [2, 3, 4], and it always came out ahead from a cost standpoint with a cost advantage up to 58% vs. OC [5]. Also, among other non-invasive CRC screening approaches, CTC is much more

accurate in terms of sensitivity and specificity for polyp detection [6]. Therefore, accurate CTC will benefit the entire population recommended for CRC screening.

Computed Tomographic Colonography (CTC) is a screening modality that can be used for the detection of colorectal polyps. Radiologists can detect polyps from CTC images at a sensitivity comparable to that of optical colonoscopy. The American College of Radiology Imaging Network (ACRIN) [2] performed standard CTC and OC on 2531 patients at 15 study centers in the US. The study showed that per-patient sensitivity and specificity for CTC were:  $0.9 \pm 0.03$  and  $0.86 \pm 0.02$ , respectively. These figures are very close to OC.

During the past two decades, our research team developed a front-end CTC system (e.g., [7, 8, 9, 10, 11]) and we still work on its enhancement. The main goal of the proposed work is to automatically identify polyp candidates in CT slices. This helps radiologists in reading CT scans and it may reduce the probability of missing polyps during CT scan reading.

Related work on polyp detection methods includes: M. Summers et al., [12] trained a model to detect polyps based on their shape with sensitivity = 71%. While Van Wijk et al., [13] proposed a shape-invariant classical algorithm to detect the colon's polyp. Hong et al., 2006 [14] introduced a pipeline to detect polyps using integrating texture and shape analysis in addition to volume rendering and conformal colon flattening.

Learned features using convolutional neural networks (CNN) boosted image processing algorithms in terms of accuracy. For instance, Liew et al., [15] used convolutional neural networks (CNN) to detect polyps using endoscopic images. On the other hand, Godkhindi and Gowda, [16] trained a CNN classifier to identify CT slices, which have polyps. However, their approach does not localize the polyp in the CT slices. Uemura and Näppi's work [17] shows a way to classify polyps using 3D networks, but it needs high computational power in order to feed the CNN network with 3D data alongside with high-time computation needed due to

large data size. Detecting and localizing polyps in CT scans is still a challenging problem. In this work, we tackle this challenge. The main contributions of this work include:

1. Developing a two stages cascade detector for colorectal polyps. The first stage is a CNN-based detector from axial CT scans.
2. The CNN-based detector's effective receptive field is controlled by using a segmented colon to focus on colon regions.
3. Exploiting the three different views (axial, coronal, and sagittal) of DICOM data to reduce the false positives of the detector.

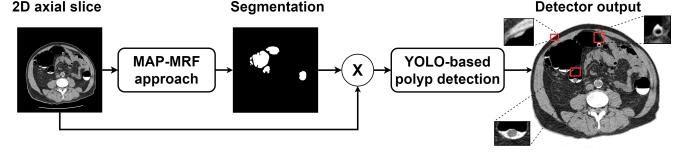
## 2. PROPOSED APPROACH

To develop an efficient approach that overcomes polyps detection and localization challenges, the algorithm should focus on colon regions. To investigate this issue, first, a state-of-the-art object detection algorithm [18] had been applied in order to detect polyps. However, detecting polyps from raw CT scans is very challenging as confirmed by experiments in the results section below. So, the CNN detector should be guided to the regions of interest only, which are the colon wall. So, we change the network effective receptive field to focus on colon regions. Our proposed approach for colon polyp detection and localization from CT scans consists of two stages. The first stage has two main steps. The first step is to segment the colon in order to reduce any dispensable information other than colon regions. Then the segmented colon regions are fed into CNN to localize any potential polyp. These two steps are considered as the first stage of our approach. Figure 1 shows the first stage of the proposed pipeline.

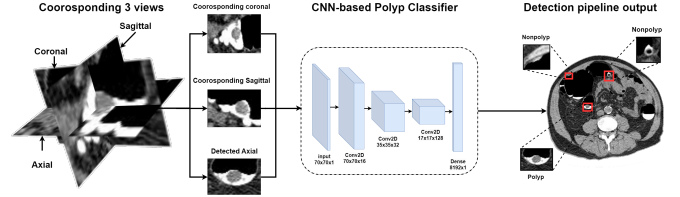
However, training a detector using axial CT slices generates many false positives. So, another step exploits the unique feature of DICOM files (i.e., DICOM can be accessed through volume) not only the axial view. Therefore, we train a classifier using the three views to reduce the false positive rate. To avoid the drawbacks of using a volume in the training process (e.g., [17]), our proposed classifier uses three 2D images (sagittal, coronal, and axial views) for each candidate generated by the detector. This approach in the second stage would consume less time and memory compared to the 3D network. Figure 2 shows the second stage of the pipeline of the proposed algorithm, and the following three subsections describe the details of the approach.

### 2.1. Colon segmentation:

Colon segmentation is a challenging problem because the colon has asymmetric topology. Also, uncertainties appear due to the presence of Hounsfield intensity regions consisting of air, soft tissue, and high-attenuation structures similar to



**Fig. 1.** The first stage of the proposed approach. An input axial CT slice is segmented. Then a CNN-based detector is fed by a segmented colon to localize polyp candidates. The output shows the location and the size of any potential polyp that needs to be verified.



**Fig. 2.** The second stage of the proposed approach. The detected axial candidate and its corresponding sagittal and coronal views are fed into a CNN-based classifier to classify a candidate using a weighted vote from each view. The output shows the location and the size of the final detected polyp after removing false positives.

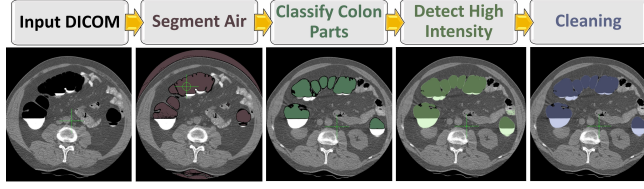
bone. In addition, complications result due to the presence of residual stool, lesions, and disconnected colon segments.

The proposed segmentation approach involves multiple steps, as shown in Figure. 3. The first step is to detect the probable air regions in the DICOM image based on thresholding. The second step is to collect the segmented potential air regions as connected components. The third step uses the Markov Random Field (MRF)-based algorithm to determine which segment could belong to the colon. Then, the high-intensity (fluid) regions below the air regions inside the colon have been detected by region growing to the gravity direction in the DICOM's high-intensity threshold.

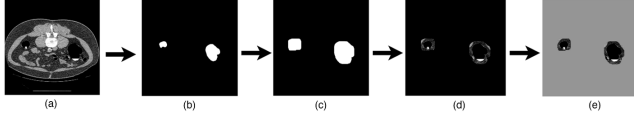
The problem is formulated as the Maximum-A-Posterior (MAP) estimation of a Markov Random Field (MRF). To segment a volume, we initially labeled the volume based on its grey-level probabilistic model. Then we create a weighted undirected graph with vertices corresponding to the set of volume voxels  $\mathcal{P}$ , and a set of edges connecting these vertices. Consider a neighborhood system  $\mathcal{N}$  of all unordered pairs  $\{i, j\}$  of neighboring voxels in  $\mathcal{P}$ . Let  $\mathcal{L}$  the set of labels corresponding to the colon and its background and denote the set of labeling by  $f$ . The goal is to find the optimal segmentation; best labeling  $f$ , by minimizing the function:

$$E(f) = \sum_{\{i,j\} \in \mathcal{N}} V(f_i, f_j) + \sum_{i \in \mathcal{P}} D(f_i), \quad (1)$$

where  $D(f_i)$  measures how much assigning a label  $f_i$  to voxel



**Fig. 3.** Colon segmentation approach.



**Fig. 4.** The preprocessing procedure. Segmented colon regions (b), from an input slice, is dilated (c) and is used to mask the input slice to show colon ROIs (d). Then, the background color is changed to differentiate between the air in the colon and the background (e).

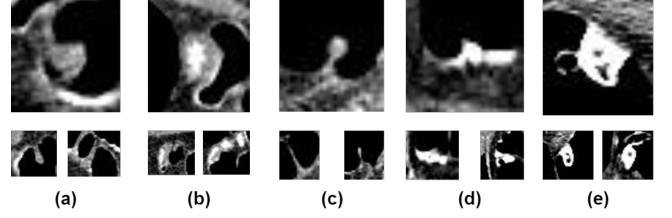
$i$  disagrees with the voxel intensity  $I_i$  and this can be estimated from the log-likelihood of each class e.g.,  $D(f_i) = \log(p(I_i|f_i))$ . The first term is the pairwise interaction model which represents the penalty for the discontinuity between voxels  $i$  and  $j$  (see [19] for more details). Optimization is carried out using the Graph Cut approach [20].

## 2.2. Polyp detection

After getting the segmented colon, a preprocessing step is done. Figure 4 shows how each 2D axial slice is preprocessed for the training. The segmented regions are dilated to include the tissues of the colon wall, not just the wall boundaries. Then the CT slice is masked and the background is changed to not interfere with the regions of interest features. The output of the preprocessing step is the masked axial view, which is fed into a CNN-based detector. The end-to-end network, YOLO model [21], is trained to predict bounding boxes with class labels "polyp" or "nonpolyp". Since the goal of the first stage is to detect all the true polyps even with a high number of false positives, the threshold of the confidence score, which identifies if the detected box contains any polyp, is chosen to detect all possible true positive samples in addition to false positive samples, which will be rejected in the next stage.

## 2.3. Polyp classification

The detector, in the first stage, is trained to identify all the true positive samples with a number of false positives. To reject these false positive samples, the CNN-based classifier, which its architecture is shown in Figure. 2, is developed to check if the detected candidate is a polyp or not. Since a DICOM scan is a volume, this feature is used to get additional information about the candidate detected from the axial CT slice. Since most polyps look like a small protruding mound,



**Fig. 5.** Different false positive samples. The top is the axial view showing the geometric feature of polyps. Bottom corresponding sagittal and coronal views reject these candidates (a-d). In hard cases (e.g., e) the three views confirm the presence of the geometric feature of a polyp, so the proposed approach may fail.

this geometric feature should appear in the three views unlike the nonpolyp tissues e.g. colon folds. Figure 5 shows the axial views of different false positives, which have a geometric feature similar to polyps and they can deceive the detector, but the other two views show the absence of these geometric features, so the candidate is not a polyp. This is the hypothesis of the proposed classifier. However, in hard samples (e.g. Figure. 5-e) the three views confirm the presence of the geometric feature of a polyp, so the proposed approach may fail. The sagittal and coronal views, which correspond to the detected bounding box in the original axial view, Figure. 2, are extracted and used to train the proposed CNN-based classifier to categorize the candidate region as a colorectal polyp or not. In the training phase, the three views are used as independent samples to train the classifier. But in the inference phase, the outputs of the three corresponding views are used by their weights to vote on the final decision.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Dataset

The dataset that is used for training and testing the proposed approach consists of CT scans for 49 patients, that were originally obtained from the University of Wisconsin hospitals, provided by Dr. Pickhardt, in the supine and prone positions. The scans have 59 annotated polyps bigger than 6 mm. Lesions annotation was done by one of three experienced radiologists. The scans are in DICOM format and 2D images can be read in the primary window setting (1500, -250 HU).

### 3.2. Training and testing procedure

First, the colon segmentation approach is applied to the axial view of CT slices to generate binary masks. Since some polyps could be extended inside the colon wall, binary masks are dilated and used to extract regions of interest from the colon wall. Next, data augmentation is used to generate different variations from the current dataset. In the data augmentation, we use different exposure, saturation, and bright-

**Table 1.** The results of the detection stage.

| Model          | Parameters | Sensitivity | mAP    |
|----------------|------------|-------------|--------|
| YOLO-V5 only   | 7,050,500  | 44.4 %      | 43.5 % |
| YOLO-V5 + Seg. | 7,050,500  | 86.67 %     | 79.7 % |
| YOLO-V7 + Seg. | 36,481,772 | 88.05%      | 85.7 % |

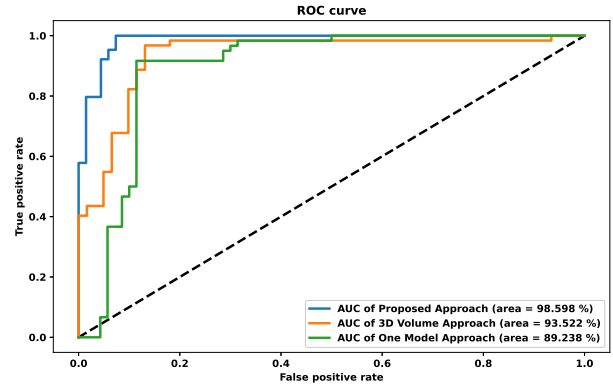
ness in addition to flipping horizontally. After that, the output of preprocessing step is used to train a YOLO network. Two types of YOLO architectures are investigated YOLO-V5 and YOLO-V7[18]. To highlight the importance of the preprocessing step, we train a YOLO-V5 model using the original 2D CT slices without preprocessing. Then, the proposed preprocessing step is applied on the data and it is used to train both YOLO-V5 and YOLO-V7 models. For better evaluation, a 5-fold cross-validation method is used to evaluate the performance of the detector. Table 1 shows the results of the three models. Note that YOLO-V7 needs more epochs than YOLO-V5 to be trained due to  $\sim 36.5M$  model parameters versus  $\sim 7M$  parameters for YOLO-V5. As expected, the proposed preprocessing step significantly enhances the performance of the detector, e.g., the YOLO-V5 mAP increased from 43.5% to 79.7%, which was boosted using YOLO-V7 to 85.7%. The YOLO-V7 detector is chosen to be the first stage of the proposed pipeline since it has good sensitivity.

To evaluate the proposed classifier, the three corresponding views are separately fed into the classifier model, then the final decision is calculated by using weighted votes of the three scores. This is done using a 5-fold cross-validation method, then the area under the ROC curve is calculated as shown in Table 2 and Figure. 6. In order to investigate more classification architectures, two additional experiments are done. In the first experiment, instead of using the detected 2D axial view ( $70 \times 70$ ) to find other corresponding 2D views (coronal  $70 \times 70$  and sagittal  $70 \times 70$ ), we use the detected 2D axial view to find the corresponding volume ( $70 \times 70 \times 15$ ). Then a 3D-CNN model is trained using the extracted volumes. In the second experiment, instead of using the three corresponding views separately, the three views are used as a single input with three channels to train a Depthwise convolution-based model. Since the spatial relations are different in the three views, the standard CNN architecture cannot be used with an input with the three views as three channels, so depthwise convolution is used because it uses different weights for each channel. Table 2 shows the performance of each model.

The proposed 2D CNN model has an area under curve 98.59%, which means that the classifier rejects most of the false positive results from the detector step, and most of the true positive results are correctly classified. Figure 6 shows the ROC curve (receiver operating characteristic curve) for each experiment. As expected, the 3D-CNN model needs high computational power to train and test procedures, e.g. it is almost 7 times the size of the 2D-CNN model as shown

**Table 2.** The results of the classification stage.

| Model           | Parameters | Sensitivity | AUC    |
|-----------------|------------|-------------|--------|
| 2D CNN          | 173, 202   | 94.80%      | 98.59% |
| 3D CNN          | 1,296,577  | 87.80%      | 93.52% |
| Depthwise CONV. | 399,314    | 84.73%      | 90.14% |

**Fig. 6.** The ROC curve and area under the curve after the classification step.

in Table 2.

To illustrate the effect of each stage, an experiment is conducted using 19 2D axial slices with polyps. The best-trained detector misses 11 polyps, i.e.,  $FN = 11$ , in addition to three false positive samples. To detect all the polyps, i.e., make  $FN = 0$ , we decrease the threshold of the confidence score of the detector. So, the detector identifies all the polyps in addition to 13 false positive samples. After the classification stage, these false positive samples are decreased to one sample. Finally, we investigated the false positive samples result from the two stages pipeline. These samples are hard to be classified since they have polyp features in the three views. As shown in Figure. 5, cases (a) to (d) are correctly classified because at least one of their views does not have polyp features, but in case (e), the classifier identifies it as polyp because its three views have polyp features.

## 4. CONCLUSION

In this work, we developed an approach for colorectal polyp detection to identify CT slices that have polyps and then highlight the polyp's locations. The proposed approach consisted of two cascade stages: one to detect polyps and then refine this detection with a classifier. The CNN-based detector was guided by changing its effective receptive field using a segmented colon wall. This drastically enhanced the detection performance. The proposed classifier exploited the different views of the CT scans to successfully refine the detection by rejecting false positive samples. The high performance, i.e.,  $AUC \sim 98.6\%$ , encourages radiologists to use the proposed approach for reading CT scans in a short time.

## 5. REFERENCES

- [1] "The american cancer society," <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/acs-recommendations.html>, Accessed: 2021-12-28.
- [2] C D Johnson, M Chen, A Y Toledano, J P Heiken, A Dachman, M D Kuo, C O Menias, B Siewert, J I Cheema, R G Obregon, et al., "Accuracy of ct colonography for detection of large adenomas and cancers," *NEJM*, vol. 359, no. 12, 2008.
- [3] Cesare Hassan and Perry J Pickhardt, "Cost-effectiveness of ct colonography," *Radiologic Clinics*, vol. 51, no. 1, pp. 89–97, 2013.
- [4] Perry J Pickhardt, "Ct colonography for population screening: ready for prime time?," *Digestive diseases and sciences*, vol. 60, no. 3, pp. 647–659, 2015.
- [5] Bruce Pyenson, Perry J Pickhardt, Tia Goss Sawhney, and Michele Berrios, "Medicare cost of colorectal cancer screening: Ct colonography vs. optical colonoscopy," *Abdominal imaging*, vol. 40, no. 8, pp. 2966–2976, 2015.
- [6] P J Pickhardt, P M Graffy, B Weigman, N Deiss-Yehiely, C Hassan, and J M Weiss, "Diagnostic performance of multitarget stool dna and ct colonography for noninvasive colorectal cancer screening," *Radiology*, vol. 297, no. 1, pp. 120–129, 2020.
- [7] M S Hassouna, A A Farag, and R Falk, "Virtual fly-over: A new visualization technique for virtual colonoscopy," in *MICCAI*. Springer, 2006, pp. 381–388.
- [8] Mostafa Mohamed, Amal Farag, Asem M Ali, Salwa Elshazly, Aly A Farag, and Mohamad Ghanoum, "Fly-in visualization for virtual colonoscopy," in *ICIP*. IEEE, 2018, pp. 2062–2066.
- [9] Mostafa Mohamad, Amal Farag, Asem M Ali, Salwa Elshazly, Aly A Farag, and Mohamad Ghanoum, "Enhancing virtual colonoscopy with a new visualization measure," in *ISBI 2018*. IEEE, 2018, pp. 294–297.
- [10] Mostafa Mohamed, Asem Ali, Salwa Elshazly, and Aly Farag, "Stabilising visualisation by reducing camera movements in virtual colonoscopy methods," *Comput Methods Biomech Biomed Eng Imaging Vis*, vol. 9, no. 3, pp. 253–260, 2021.
- [11] S. Elshazly A. Dachman P. Pickhardt J. Yee R. Falk A. Farag, A. Aly and A. Seow, "'fly-in: A robust visualization approach for ct colonography," in *RSNA*, 2021.
- [12] Ronald M Summers, Christopher F Beaulieu, Lynne M Pusanik, James D Malley, R Brooke Jeffrey Jr, Daniel I Glazer, and Sandy Napel, "Automated polyp detector for ct colonography: feasibility study," *Radiology*, vol. 216, no. 1, pp. 284–290, 2000.
- [13] Cees van Wijk, Vincent F van Ravesteijn, Frank M Vos, Roel Truyen, Ayso H de Vries, Jaap Stoker, and Lucas J van Vliet, "Detection of protrusions in curved folded surfaces applied to automated polyp detection in ct colonography," in *MICCAI*. Springer, 2006, pp. 471–478.
- [14] Wei Hong, Feng Qiu, et al., "A pipeline for computer aided polyp detection," *IEEE Trans Vis Comput Graph*, vol. 12, no. 5, pp. 861–868, 2006.
- [15] Win Sheng Liew, Tong Boon Tang, Cheng-Hung Lin, and Cheng-Kai Lu, "Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches," *COMPUT METH PROG BIO*, vol. 206, pp. 106114, 2021.
- [16] Akshay M Godkhindi and Rajaram M Gowda, "Automated detection of polyps in ct colonography images using deep learning algorithms in colon cancer diagnosis," in *ICECDS*. IEEE, 2017, pp. 1722–1728.
- [17] Tomoki Uemura, Janne J Näppi, Toru Hironaka, Hyounseop Kim, and Hiroyuki Yoshida, "Comparative performance of 3d-densenet, 3d-resnet, and 3d-vgg models in polyp detection for ct colonography," in *Medical Imaging 2020: Computer-Aided Diagnosis*. SPIE, 2020, vol. 11314, pp. 736–741.
- [18] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, 2022.
- [19] Asem M Ali and Aly A Farag, "A novel framework for nd multimodal image segmentation using graph cuts," in *ICIP*. IEEE, 2008, pp. 729–732.
- [20] Yuri Boykov and Marie-Pierre Jolly, "Interactive organ segmentation using graph cuts," in *MICCAI*. Springer, 2000, pp. 276–286.
- [21] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.