Real-time detection of colon polyps during colonoscopy using YOLOv7

Le Thi Thu Hong[1*], Le Huu Nhuong[2], Ngo Toan Thang[2],
Doan Quang Tu[1], Nguyen Sinh Huy[1], Nguyen Duc Hanh[1],
Trinh Tien Luong[1], Ngo Duy Do[1], Le Anh Dung[2]

[1]Institute of Information Technology, Academy of Military Science and Technology;
[2]Military Medical Hospital 354, General Department of Logistics.
*Corresponding author: lethithuhong1302@gmail.com
Received 12 Sep. 2023; Revised 11 Dec. 2023; Accepted 12 Dec. 2023; Published 30 Dec. 2023
DOI: https://doi.org/10.54939/1859-1043.j.mst.CSCE7.2023.122-134

ABSTRACT

Deep learning has made brilliant achievements in detecting colonic polyps in colonoscopy videos in recent years. However, the detection of colonic polyps in colonoscopy videos is problematic because of the complex environment of the colon and the various shapes of polyps. Therefore, researchers need to spend a lot of time searching for real-time detection systems with good performance and that are suitable for the current equipment and working environment. This paper aimed to investigate the polyp detection potential of the state-of-the-art deep learning model You Only Look Once version 7. We implemented, trained, and tested the polyp detection model using open public datasets: Kvasir-Seg, CVC-ClinicDB, CVC_ColonDB, and ETIS-LaribPolypDB. Validation of the test set utilizing Recall, Precision, F1 Score, and Average Precision (AP) showed that the model achieved the highest performance on CVC-ClinicDB with 83.3% Recall, 80.6% Precision, 81.9% F1 Score, 75% AP@0.5, 51.8% AP and the mean processing time per frame was 20 ms. The automatic polyp detection model exhibited good performance, as evidenced by the high detection sensitivity and rapid processing. This model can help endoscopists improve polyp detection performance during the colonoscopy procedure.

Keywords: Colorectal cancer; Deep learning; Object detection; Polyp detection.

1. INTRODUCTION

Colorectal cancer (CRC) is one of the most common causes of cancer-related death worldwide for both men and women. Colonoscopy is considered the gold-standard investigation for colorectal cancer screening. The studies indicated that colonoscopy was associated with a 60% reduction in CRC mortality [1] and a 70% reduction in late-stage CRCs [2]. Colorectal polyp is one of the significant causes of gastroenterology, which leads to colorectal cancer. Colorectal polyps are tissue growth from a mucous membrane, projecting itself into the intestine lumen. Polyp detection is vital for cancer prevention since early-stage detection significantly increases the chance of an effective treatment. Accurate detection of polyps is the most critical issue during a colonoscopy. The polyp detection rate is an essential quality indicator during colonoscopy. Therefore, increasing the polyp detection rate for adequate CRC screening via colonoscopy is vital. However, accurate polyp detection is difficult because of 1) the various sizes and shapes of polyps, bowel preparation quality, the time dedicated to mucosal inspection during withdrawal, and 2) the experience of the colonoscopist. It is reported that a polyps detection miss-rate of up to 20% - 24% puts patients at a high risk of death from CRC [3]. This presents an opportunity to leverage computer-aided (CAD) systems to detect polyps automatically, supporting clinicians and reducing the number of polyps missed.

Polyp detection in colonoscopy images refers to the process of identifying and

locating polyps using computer vision and image analysis techniques. During a colonoscopy, the camera captures real-time images of the colon. Polyps can vary in size, shape, and color, making their detection challenging for physicians. Therefore, CAD systems have been developed to assist in identifying polyps and improve the accuracy of diagnosis. Polyp detection in colonoscopy images aims to assist medical professionals by highlighting potential polyp regions for further examination. This can help reduce the risk of missed polyps and improve the overall effectiveness of colon cancer screening. Figure 1 is an example of polyp detection in colonoscopy images present in the Kvasir-Seg [4] dataset.

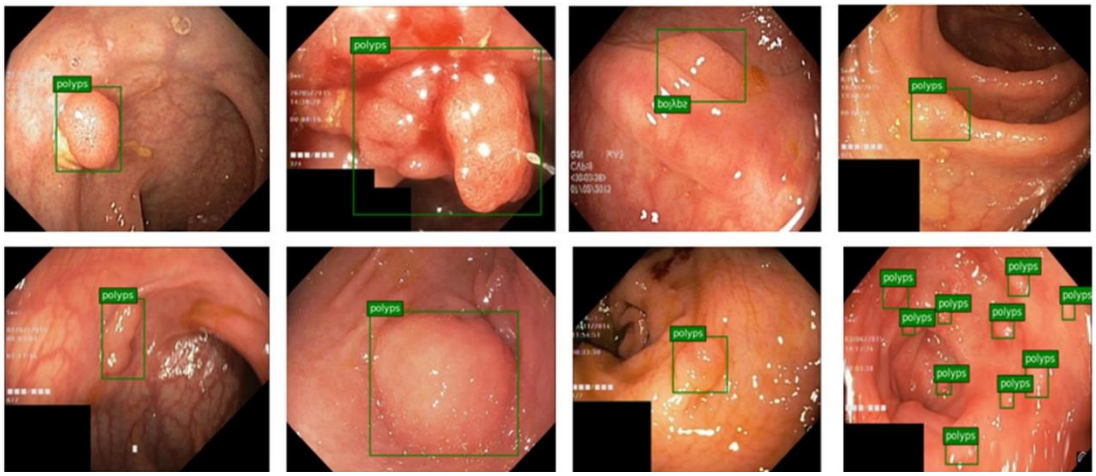


Figure 1. *Examples of polyp detection on colonoscopy image.*

Automatic polyp detection systems using deep-learning methods have been proposed for detecting colorectal polyps in real-time colonoscopy videos [5-7]. Despite the optimistic results of previous studies, further investigations are necessary to show the generalizability of deep-learning algorithms. Single-stage detectors and two-stage detectors are two commonly used approaches in the field of object detection in computer vision. Single-stage detectors are object detection models that directly predict bounding boxes and class labels in a single pass through the input image. Unlike two-stage detectors, which consist of a region proposal stage followed by a classification and refinement stage, single-stage detectors perform object detection in a simpler and more efficient manner. They eliminate the need for generating region proposals, resulting in faster inference times. One famous example of a single-stage detector is the You Only Look Once (YOLO) [8] family of models. YOLO models are known for their real-time performance and have been widely used in real like autonomous driving, surveillance, and object tracking. In this paper, we investigated the potential of Only Look Once (YOLO) [8] in detecting polyps in colonoscopy videos. You Only Look Once (YOLO) is one famous example of a single-stage detector that divides the input image into a grid and directly predicts bounding boxes and class probabilities for each grid cell. It uses anchor boxes of different scales and aspect ratios to handle objects of various sizes and shapes. YOLO models are known for their real-time performance and have been widely used in applications like autonomous driving, surveillance, and object tracking. We implemented the YOLOv7 [16] model for polyp detection. We trained and evaluated the model with open, public datasets: Kvasir-Seg, CVC-ClinicDB, CVC_ColonDB, and

ETIS-LaribPolypDB. Our main contribution is that we present a polyp detection system that utilizes a state-of-the-art single-stage detector network YOLOv7 and evaluates the performance on multiple public image datasets using standard metrics for object detection such as Recall, Precision, and Average Precision (AP). This will help researchers gain insights into the capability of detecting polyps in colonoscopy images using YOLOv7. This model demonstrates excellent performance with high accuracy and fast processing time. This indicates that the model can be used to develop real-time automated polyp detection systems to assist physicians during endoscopic procedures.

The rest of the paper is organized as section 2 reviews related research. In section 3, we describe our proposed methodology of polyp detection using YOLOv7 in detail. Section 4 outlines our experiment settings, experimental results, and discussion. Finally, in section 5, we summarize and conclude this work.

2. RELATED WORK

Over the past years, researchers have made several efforts to develop CADx prototypes for automated polyp detection. Most of the prior polyp detection approaches were based on analyzing the color, texture, shape, or edge information of polyps to detect polyp regions. More recent approaches have used deep convolution neural networks for polyp detection.

Conventional polyp detection methods were based primarily on handcrafted features followed by a separate classifier. The features used for polyp detection are mainly geometric, including shape, size, and texture information. For example, geometric and texture features have been utilized in the literature for polyp detection. Tajbakhsh et al. [9] presented a method based Canny edge detector in each of the three RGB channels. This is done to produce edge maps, and then oriented patches for each pixel are extracted to classify them as polyp or non-polyp. Tajbakhsh et al. [10] also proposed a feature extraction method to extract sub-patch with a 50% overlap and calculate their average vertically, resulting in a one-dimensional signal. After that, they use DCT coefficients as a feature for each extracted patch. Finally, they use a two-stage random forest classifier to label each patch. The author of [11] proposed an automated polyp detection by using edge cross-section profiles, and a Supported Vector Machine (SVM) was used for the classification.

Recently, CNN has brought large attention to the computer vision community and made great progress in various tasks. Similar to other medical imaging applications, CNN-based approaches have gained much attention in recent years for automating the feature extraction process to detect polyp regions with unprecedented Precision. In [12], the authors propose a deep learning-based approach for polyp detection in colonoscopy images. They use convolutional neural networks to detect and classify polyps automatically. The model achieves high accuracy and demonstrates the potential of CNNs for assisting endoscopists in real-time polyp detection. The publication [13] introduces an ensemble approach for polyp detection in colonoscopy videos. The authors employ multiple convolutional neural networks in parallel to improve detection performance. The proposed ensemble model demonstrates enhanced sensitivity and specificity, highlighting its potential as an assistance tool for endoscopists. Zhang et al. [14] presented a novel regression-based CNN pipeline for polyp detection during

colonoscopy. The proposed pipeline was constructed in two parts. A fast object detection model, ResYOLO, was used to learn the spatial features of colorectal polyps, and then temporal information was incorporated via a tracker to refine the detection results. Liu et al. [15] investigated the potential of the Single Shot Detector (SSD) framework for detecting polyps in colonoscopy videos. Three feature extractors were assessed, including ResNet50, VGG16, and InceptionV3. They validated this method on the 2015 MICCAI polyp detection challenge datasets and compared it with teams that attended the challenge, YOLOV3, and the two-stage method, Faster-RCNN. Their results demonstrated that the proposed method surpassed all the teams in the MICCAI challenge and YOLOV3 and was comparable with the two-stage method.

These publications reflect the advancements in using computer vision techniques for polyp detection in different endoscopic imaging modalities. They highlight the potential of these approaches in improving detection accuracy and assisting endoscopists during colonoscopy procedures.

3. MATERIALS AND METHODS

3.1. Datasets

To train the models and evaluate their performance, we use publicly available datasets of polyp images and videos:

- ETIS-LaribPolypDB [19]: This is a dataset of 196 still images extracted from 34 colonoscopy videos. In total, there are 44 examples of different polyps presented in various sizes and viewpoints. The images have an HD (high definition) resolution of 1225 x 966 pixels. Some images contain two or three polyps, making the total number of polyp appearances 208.

- CVC-ColonDB [17]: This dataset comprises 300 still images presenting 15 unique polyps coming from 15 different studies. The images have an SD (standard definition) resolution of 574x500. In every image, there exists only one polyp.

- CVC-ClinicDB [18]: This contains 31 unique polyps extracted from 29 colonoscopy videos and presented 646 times in 612 still images with a pixel resolution of 384x288 in SD (standard definition).

- Kvasir-Seg [4], publicized by Simula Research Laboratory, includes 1000 polyp images with varying sizes from 332 × 482 to 1920× 1072 and their corresponding ground truth masks manually annotated by expert endoscopists from Oslo University Hospital (Norway)

Regarding the type of ground truth provided, three of them provide polyps annotated with binary masks: CVC-ClinicDB, CVC-ColonDB, and ETIS-Larib. In these cases, we converted the binary masks into bounding boxes to analyze them with our model (the procedure is described below in this section). Finally, Kvasir-SEG provides both segmentation and localization information.

3.2. YOLOv7 for polyp detection

The pipeline presented in figure 2 shows the different phases executed during this work. The public datasets were used to train and evaluate the polyp detection network. The images have suffered some preprocessing, where alphanumeric characters were

removed. Data augmentation was performed on the fly, including vertical flipping, horizontal flipping, random rotation, random scaling, random shearing, random Gaussian blurring, random brightness, and random cropping and padding for training the polyp detector. All the images were labeled with polyp bounding boxes, and last, the data was split into the train, validation, and test sets. YOLOv7 detectors were trained on the train and validation sets and evaluated in the test set.

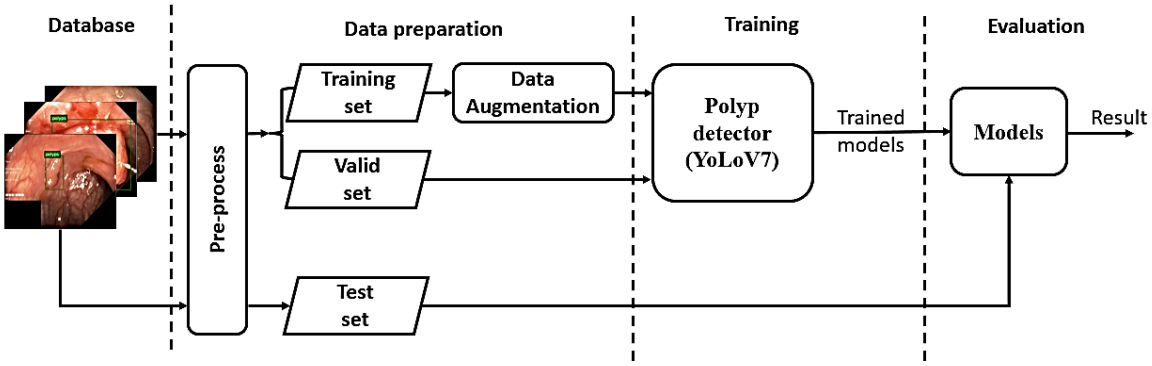


Figure 2. *Polyp detection pipeline.*

YOLO[8] is a popular real-time object detection algorithm that revolutionized the field of computer vision by providing an efficient and accurate approach to detecting objects in images and videos. The main idea behind YOLO is to treat object detection as a regression problem, where a single neural network is trained to directly predict bounding boxes and class probabilities from input images. This is in contrast to traditional object detection algorithms that use complex multi-stage pipelines. YOLO achieves real-time performance by making predictions on the entire image in a single pass. YOLO's architecture is presented in figure 3, and it uses a deep convolutional neural network (CNN) as its backbone. The network takes an input image and processes it through multiple convolutional layers to extract features at different spatial scales.

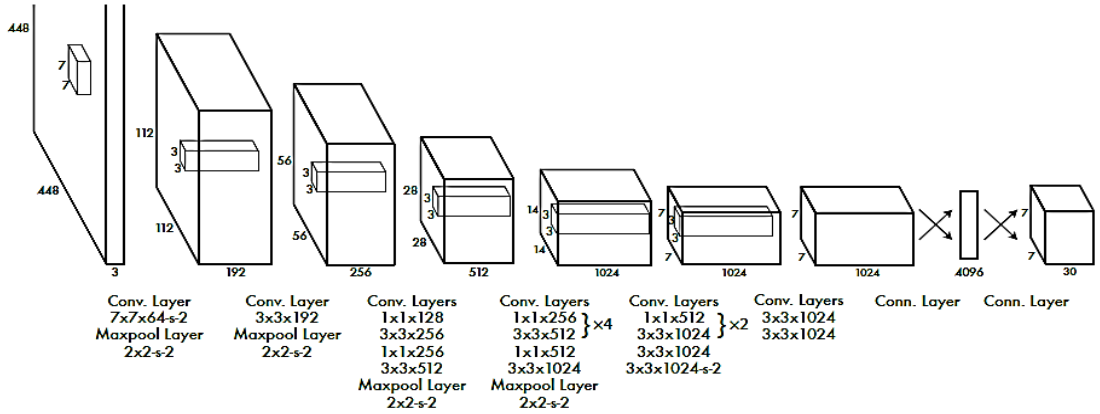


Figure 3. *YOLO Architecture from the original paper [8].*

YOLO divides the input image into a grid of cells. Each cell is responsible for predicting a fixed number of bounding boxes. The spatial dimensions of the grid determine the size of the predicted bounding boxes. Each grid cell predicts bounding boxes by regressing their coordinates relative to the cell location. YOLO predicts

multiple bounding boxes per grid cell, along with their associated class probabilities. Along with the bounding box coordinates, YOLO predicts the probability of each class being present in the bounding box. It uses softmax activation to compute the class probabilities. Since YOLO predicts multiple bounding boxes for each object, it employs a post-processing step called non-max suppression to remove duplicate detections. This step keeps only the most confident bounding box for each object based on a threshold value. YOLO is trained on labeled datasets with annotated bounding boxes. The training process involves optimizing the network to minimize the error between predicted bounding boxes and ground truth bounding boxes. It uses a combination of localization loss (related to bounding box coordinates) and classification loss (related to class probabilities). Since the first release of YOLO in 2015, it has evolved a lot with different versions, including YOLO, YOLOv2, YOLO9000, YOLOv3, YOLOv4, YOLOR, YOLOX, YOLOv5, YOLOv6, YOLOv7.

In this work, we choose YOLOv7 [16] for polyp detection. YOLOv7 reformed its architecture by integrating the Extended Efficient Layer Aggregation Network (E-ELAN), which allows the model to learn more diverse features for better learning. In addition, YOLOv7 scales its architecture by concatenating the architecture of the models it is derived from, such as YOLOv4, Scaled YOLOv4, and YOLO-R. This allows the model to meet the needs of different inference speeds.

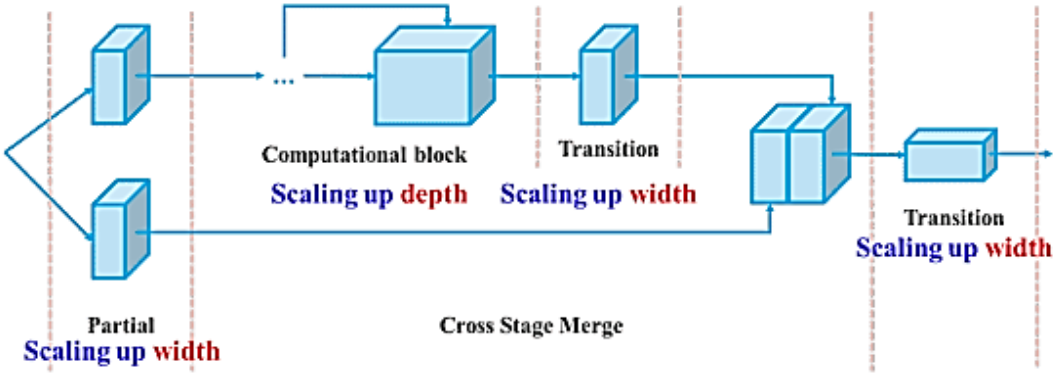


Figure 4. *Compound scaling up depth and width for concatenation-based model [17].*

3.3. Evaluation method

The polyp detection accuracy was evaluated as true positive(TP), false positive(FP), false negative(FN), and true negative(TN). TP, TN, FP, and FN are determined using the IOU (Intersection Over Union) metric, which is simply the ratio of the intersection area over the union area of the two bounding boxes. Figure 5 depicts the method for calculating IOU (Intersection over Union). To be specific, if the IOU between the predicted polyp bounding box and the ground truth polyp bounding box is greater than or equal to the threshold, only one TP would be considered per polyp. Any detection that the IOU between the predicted polyp bounding box and the ground truth polyp bounding box is smaller than the threshold is counted as FP. The absence of alarm in images with a polyp was considered a false negative (FN), counting one per polyp in the image that had not been detected. Regarding images without polyps, we defined them as true negatives (TN) whenever any output was not provided for this particular image.

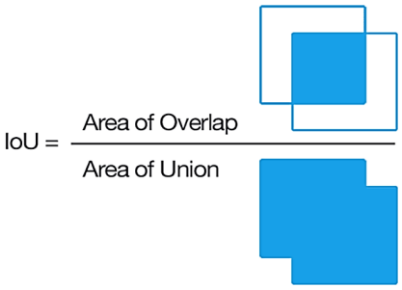


Figure 5. *IOU (Intersection Over Union) metric.*

We use TP, TN, FP, and FN terminologies to evaluate the performance of the models in terms of:

Recall (Sensitivity): It measures the ratio of true detection outputs to the total number of polyps in the test dataset. This metric shows the detection ability of a specific model and is computed as Eq. (1):

$$Recall = 100 \times \frac{TP}{TP + FN} \tag{1}$$

Precision: It measures the ratio of true detection outputs to the total number of predicted outputs, including false alarms. This metric shows the ability of a model to make correct predictions model and is computed as Eq. (2):

$$Precision = 100 \times \frac{TP}{TP + FP} \tag{2}$$

F1 score: The F1 score is the harmonic mean of Precision and Recall, given by Eq.(3).

$$F_1 = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

Average Precision (AP): The average Precision (AP) is a way to summarize the precision-recall curve (PR curve) into a single value representing the average of all precisions. From Precision and Recall, we can draw the PR curve for each class. The PR curve indicates the trade-off between Precision and Recall for different IOU thresholds. The Average Precision (AP) is the area under the PR curve. A larger area under the curve implies higher Precision and recall, which signifies a good-quality model. This work uses 101 Point Interpolation AP, which was introduced by the COCO Detection Challenge [22]. MS COCO defines AP as mAP@[0.5:0.95], evaluated with IoU threshold ranges from 0.5 to 0.95 at a step frequency of 0.05. In addition, we also use $AP@0.5$, where TP and FP are determined using an IoU threshold value of 0.5.

Mean Processing Time per Frame (MPT): It is the amount of time a detection model needs to process a single frame.

3.4. Implementation

We used the Kvasir-Seg dataset to train the polyp detection model and tested the model using three separate datasets: CVC-ClinicDB, CVC-ColonDB, and ETIS_Labris. The proposed models are implemented using PyTorch and TensorFlow. All algorithms have been programmed/trained on a PC with a GeForce GTX 1080 Ti GPU. The detection network is updated via the Adam optimizer, and the learning rate is set to 0.0001. All the training data is divided into mini-batches for network training, and the

mini-batch size is set as four during the training stage. The model has trained 100 epochs, and the model generated at the epoch with a max AP value on the validation set is the final polyp detection model.

4. RESULTS AND ANALYSIS

4.1. Accuracy of YOLOv7 polyp detector

Detection of colonic polyps is a challenging task since they appear in various sizes and shapes. Meanwhile, in clinical applications, high demand for real-time speed is also necessary. In this paper, we investigated the YOLOv7 model, which has relatively high accuracy and fast detection speed**.** We installed YOLOv7, trained the model using the KVar-Seg dataset, and then tested the model with various independent datasets. The values of the loss function and the accuracy of the model during the training process are shown in figure 6. This figure demonstrates that the model converges very well during the training process with 100 epochs.

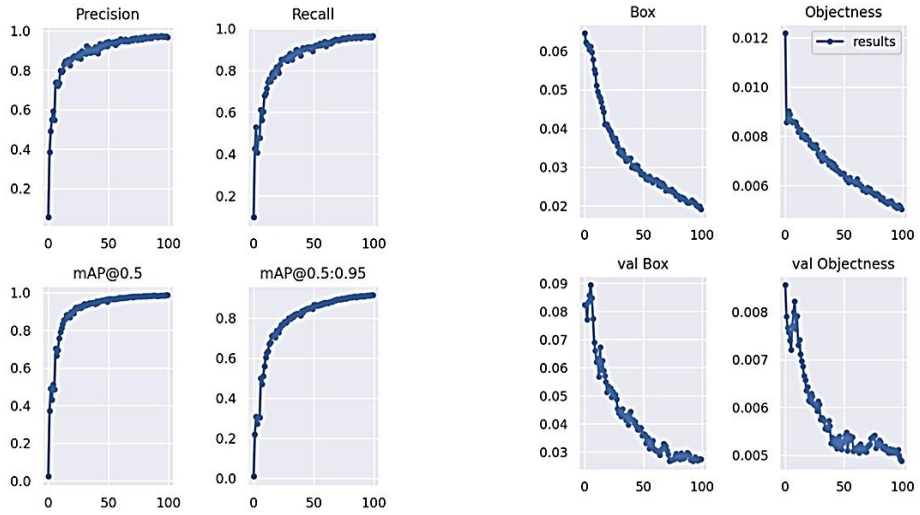


Figure 6. *The model's accuracies and the values of the loss function during the training process.*

Detailed results on different datasets are listed in tables 1, 2, and 3. The threshold in the tables demonstrated the confidence level of the detected region with polyps.

Table 1. *Accuracy of networks with different threshold values on the ETIS-LaribPolypDB.*

Threshold	**Precision (%)**	**Recall (%)**	**F1 score (%)**	AP@0.5 **(%)**	**AP (%)**
0.1	94	67.6	78.6	67	40.8
0.2	72.2	59.7	65.4	54.9	40.5
0.3	73.5	58.2	65.0	53.8	40
0.4	74.8	56.1	64.1	52.3	38.9
0.5	*77.4*	*54.1*	*63.7*	*50.8*	*38.1*
0.6	83.2	48.5	61.3	47.6	36.2
0.7	83.2	48.5	61.3	46.8	35.6
0.8	85.9	43.4	57.7	42.1	32.9
0.9	88.9	32.7	47.8	32.7	26.9

Table 2. *Accuracy of networks with different threshold values on the CVC-ClinicDB.*

Threshold	Precision (%)	Recall (%)	F1 score (%)	AP@0.5 (%)	AP (%)
0.1	**83.3**	**80.6**	**81.9**	**75.2**	**51.8**
0.2	85.1	78.3	81.6	73.5	50.9
0.3	85.7	76.6	80.9	71.7	50
0.4	86.6	74.8	80.3	70.3	49.3
0.5	*86.7*	*73.5*	*79.6*	*70*	*48.7*
0.6	87.5	70.8	78.3	66.7	47.3
0.7	88	66.8	75.9	63.2	45.4
0.8	88.3	61.4	72.4	58.4	42.6
0.9	91	47.7	62.6	46.1	34.9

Table 3. *Accuracy of networks with different threshold values on the CVC-ColonDB.*

Threshold	Precision (%)	Recall (%)	F1 score (%)	AP@0.5 (%)	AP (%)
0.1	87.3	80.7	83.9	80.6	46.7
0.2	87.3	80.7	83.9	79.3	46.3
0.3	87.2	79.7	83.3	77.8	45.8
0.4	87.2	77.3	82.0	75.8	45.1
0.5	*90.7*	*74.7*	*81.9*	*74.1*	*44.3*
0.6	92.4	73.3	81.7	72.9	43.7
0.7	92.7	72	81.0	71	42.7
0.8	94	67.6	78.6	67	40.8
0.9	98.2	54.7	70.3	54.5	34.2

From these tables, the following conclusions can be drawn:

- The lower the threshold value, the higher AP and AP@0.5. AP and AP and AP@0.5 decreased along the increased threshold.

- Under the same confidence level of 0.5, the dataset in which the model gets the worst was ETIS-LaribPolypDB. The low performance in the ETIS-LaribPolypDB dataset is not surprising, as the authors state in their publication [21] that the dataset contains images selected to include a high degree of diversity in polyp morphology, multiple polyps, motion blur, and specular reflections to create a challenging dataset. Figure 7 presents the accuracy of the YOLOV7 Polyp Detector on different test sets of the polyp detection model with a confidence level of 0.5, and some detection samples by the Polyp Detector are shown in figure 8.

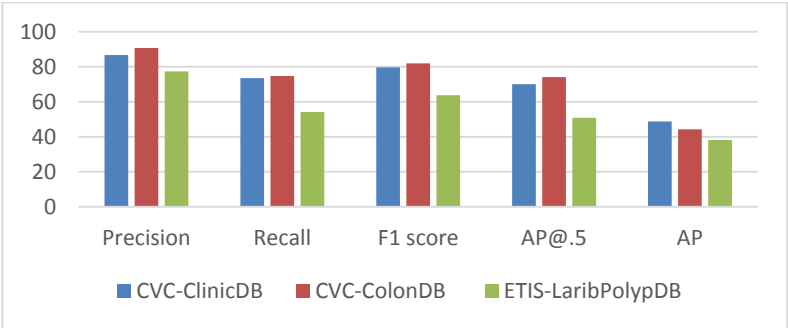
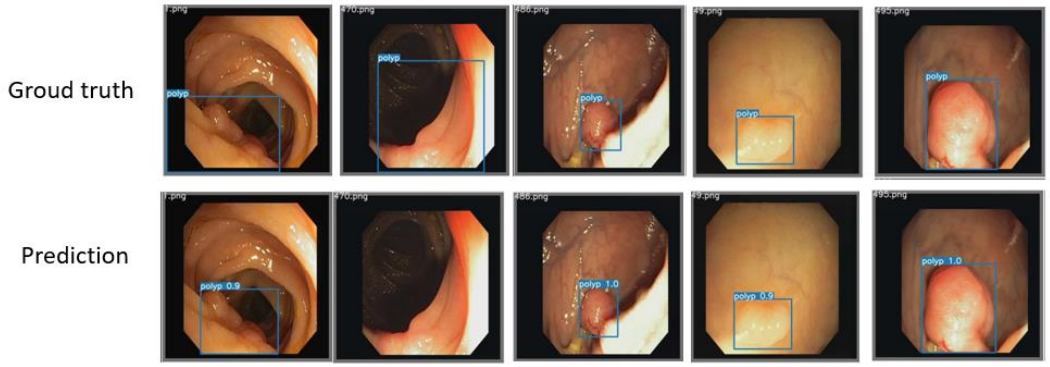
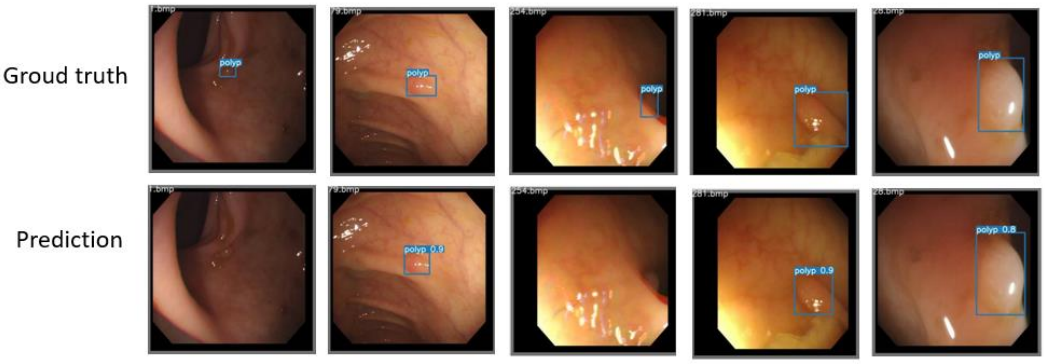


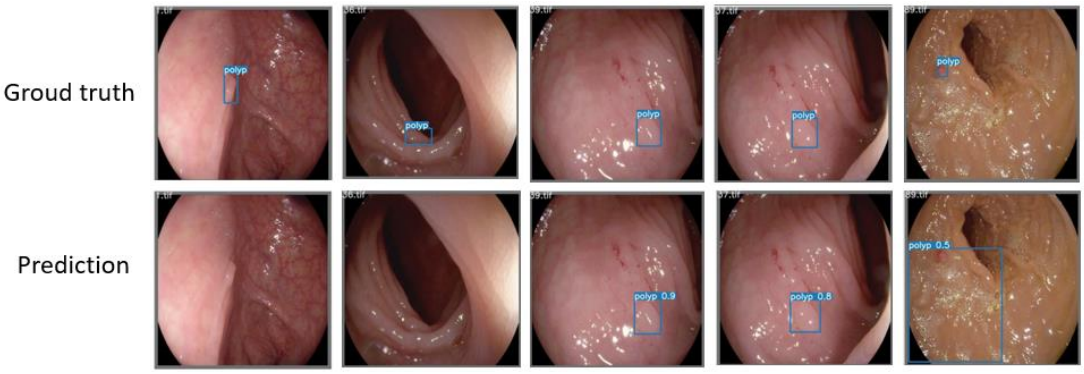
Figure 7. *Accuracy of YOLOv7 Polyp Detector with threshold = 0.5.*



(a) Results of Polyp detection on the CVC-ClinicDB dataset.



(b) Results of Polyp detection on the CVC-ColonDB dataset.



(c) Results of Polyp detection on the ETIS-LaribPolypDB dataset.

Figure 8. *Polyp detection results of YOLV7 Polyp Detector with threshold = 0.5.*

- The YOLOV7 Polyp Detector achieved the highest accuracy on the CVC-ClinicDB dataset with 83.3% Recall, 80.6% Precision, 81.9% F1 Score, 75% AP@0.5, 51.8% AP. The processing time per frame (MPT) is approximately 20 ms. The model achieved the highest accuracy on the CVC-ClinicDB dataset with 83.3% Recall, 80.6% Precision, 81.9% F1 Score, 75% AP@0.5, and 51.8% AP. In addition, the processing time per frame (MPT) is about 20 ms. The polyp detector achieved good performance and can be used for real-time endoscopy applications.

4.2. Comparison of the model's performance with other methods

The polyp detection performance of YOLOv7 was compared with other methods in

table 4, which includes all published studies reporting bounding-box-based performance metrics (i.e., comparing predicted bounding boxes against the true bounding boxes of the ground truth) in at least one of the selected public colonoscopy image datasets. These data were used then to analyze the performance of various models on the public datasets included in this study and compare our detection model with them. The table includes one row for each study experiment with the following information: training set, testing set, recall, Precision, and F1-score. In general, we can observe that the proposed YOLOv7 methods significantly improved the detection of true positives and produced fewer false positives, thus achieving better Precision, recall, and F1 scores.

Table 4. *Comparing performance results of deep learning models for polyp detection.*

Paper	**Training**	**Test**	**Results**		
			Precision (%)	**Recall (%)**	**F1 score (%)**
Park et al., 2016 [12]	Private	Private	85	86	85.5
Brandao et al., 2018 [22]	CVC-ClinicDB +ASU-Mayo	**ETIS-Larib**	73	90	**81**
		CVC-ColonDB	80	90	85
Zheng Y. et al., 2018[23]	CVC-ClinicDB+ CVC-ColonDB	ETIS-Larib	77	74	76
Tian Y. et al., 2019 [24]	Private	ETIS-Larib	64	74	69
Jia, Xiao, et al. 2020[25]	CVC-ColonDB	**CVC-ClinicDB**	**85**	92	**88**
	CVC-ClinicDB	ETIS-Larib	64	82	72
Qadir et al., 2021 [26]	CVC-ClinicDB	ETIS-Larib	86	87	86
		CVC-ColonDB	88	91	**90**
Pacal et al., 2022	SUN	SUN	96	86	91
	PICCOLO	PICCOLO	93	80	86
Proposed YOLOV7	Kvasir-Seg	*CVC-ClinicDB*	*83.3*	80.6	81.9
		CVC-ColonDB	*87.3*	80.7	*83.9*
		ETIS-Larib	**94**	67.6	78.6

5. CONCLUSIONS

This paper investigated a YOLOv7 framework-based method for polyp detection in colonoscopy videos/images. A novel polyp detection framework using YOLOv7 architecture is proposed, which includes three fundamental steps: Data preprocess, Training model, and Evaluation. Extensive experiments conducted on four public datasets showed that YOLOv7 models achieved the best detection results of true positives and comparable F1 scores. Its detection speed especially outperformed all the

existing methods dramatically. Our results indicated the feasibility of the YOLOv7-based method to provide complementary information for endoscopists while performing a colonoscopy procedure. The development of an automatic polyp detection platform would have a positive impact on future CRC management.

REFERENCES

[1]. Baxter, Nancy N., et al. *"Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty."* Journal of Clinical Oncology 30.21: 2664, (2012).

[2]. Doubeni, Chyke A., et al. *"Screening colonoscopy and risk for incident late-stage colorectal cancer diagnosis in average-risk adults: a nested case–control study."* Annals of internal medicine 158.5_Part_1: 312-320, (2013).

[3]. Leufkens, A. M., et al. *"Factors influencing the miss rate of polyps in a back-to-back colonoscopy study."* Endoscopy 44.05: 470-475, (2012).

[4]. D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, *"Kvasir-SEG: A segmented polyp dataset,"* in Proc. Int. Conf. Multimedia Modeling. Springer, pp. 451-462, (2020).

[5]. Misawa, Masashi, et al. *"Artificial intelligence-assisted polyp detection for colonoscopy: initial experience."* Gastroenterology 154.8: 2027-2029, (2018).

[6]. Wang, Pu, et al. *"Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy."* Nature biomedical engineering 2.10: 741-748, (2018).

[7]. Urban, Gregor, et al. *"Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy."* Gastroenterology 155.4: 1069-1078, (2018).

[8]. Redmon, Joseph, et al. *"You only look once: Unified, real-time object detection."* Proceedings of the IEEE conference on computer vision and pattern recognition, (2016).

[9]. Tajbakhsh, Nima, Suryakanth R. Gurudu, and Jianming Liang. *"Automated polyp detection in colonoscopy videos using shape and context information."* IEEE transactions on medical imaging 35.2: 630-644, (2015).

[10]. Tajbakhsh, Nima, Suryakanth R. Gurudu, and Jianming Liang. *"A classification-enhanced vote accumulation scheme for detecting colonic polyps."* Abdominal Imaging. Computation and Clinical Applications: 5th International Workshop. Proceedings 5. Springer Berlin Heidelberg, (2013).

[11]. Wang, Yi, et al. *"Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy."* IEEE Journal of Biomedical and Health Informatics 18.4: 1379-1389, (2013).

[12]. Park, Sun Young, and Dusty Sargent. *"Colonoscopic polyp detection using convolutional neural networks."* Medical Imaging: Computer-Aided Diagnosis. Vol. 9785. SPIE, (2016).

[13]. Tajbakhsh, Nima, Suryakanth R. Gurudu, and Jianming Liang. *"Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks."* IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE, (2015).

[14]. R. Zhang,Y. Zheng, C. C.Y. Poon, D. Shen, and J.Y.W. Lau, *"Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker"*, Pattern Recognit., vol. 83, pp. 209-219, (2018).

[15]. Liu, Ming, Jue Jiang, and Zenan Wang. *"Colonic polyp detection in endoscopic videos with single shot detection based deep convolutional neural network."* IEEE Access 7: 75058-75066, (2019).

[16]. Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. *"YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors."* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2023).

[17]. Bernal, Jorge, Javier Sánchez, and Fernando Vilarino. *"Towards automatic polyp detection with a polyp appearance model."* Pattern Recognition 45.9: 3166-3182, (2012).

[18].Bernal, Jorge, et al. *"WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians."* Computerized medical imaging and graphics 43: 99-111, (2015).
[19].Silva, Juan, et al. *"Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer."* International journal of computer assisted radiology and surgery 9: 283-293, (2014).
[20].Ma, Yiting, et al. *"LDPolypVideo benchmark: a large-scale colonoscopy video dataset of diverse polyps."* Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, Proceedings, Part V 24. Springer International Publishing, (2021).
[21].COCO Detection Challenge (Bounding Box). Available online: https://competitions.codalab.org/competitions/20794.
[22].Brandao, Patrick, et al. *"Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks."* Journal of Medical Robotics Research 3.02: 1840002, (2018).
[23].Zheng, Yali, et al. *"Localisation of colorectal polyps by convolutional neural network features learnt from white light and narrow band endoscopic images of multiple databases."* 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, (2018).
[24].Tian, Yu, et al. *"One-stage five-class polyp detection and classification."* IEEE 16th international symposium on biomedical imaging (ISBI 2019). IEEE, (2019).
[25].Jia, Xiao, et al. *"Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction."* IEEE Transactions on Automation Science and Engineering 17.3: 1570-1584, (2020).
[26].Qadir, Hemin Ali, et al. *"Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction."* Medical Image Analysis 68:101897, (2021).

TÓM TẮT

Phát hiện polyp đại tràng thời gian thực sử dụng mạng YOLOv7

Trong những năm gần đây, học sâu đã đạt được những thành tựu xuất sắc trong việc phát hiện polyp trực tràng trong video nội soi đại tràng. Tuy nhiên, việc phát hiện polyp trực tràng trong video nội soi đối mặt với nhiều vấn đề do môi trường phức tạp của ruột già và hình dạng đa dạng của các polyp. Do đó, các nhà nghiên cứu cần dành nhiều thời gian tìm kiếm những hệ thống phát hiện thời gian thực với hiệu suất tốt và phù hợp với thiết bị, môi trường làm việc tại các cơ sở y tế. Bài báo này nhằm mục đích nghiên cứu khả năng phát hiện polyp của mô hình học sâu tiên tiến YOLOv7. Chúng tôi đã đề xuất mô hình học sâu phát hiện polyp dựa trên YOLOv7, cài đặt, huấn luyện và kiểm thử mô hình bằng các bộ dữ liệu công khai được sử dụng rộng rãi trong cộng đồng nghiên cứu: Kvasir-Seg, CVC-ClinicDB, CVC_ColonDB và ETIS-LaribPolypDB. Kiểm thử mô hình sử dụng các chỉ số Recall, Precision, F1 Score và Average Precision (AP) cho thấy mô hình đạt được hiệu suất cao nhất trên CVC-ClinicDB với tỷ lệ Recall là 83.3%, Precision là 80.6%, F1 Score là 81.9%, AP@0.5 là 75%, AP là 51.8% và thời gian xử lý trung bình mỗi khung hình là 20 ms. Mô hình phát hiện polyp đại tràng đề xuất cho hiệu năng tốt, có độ nhạy cao trong việc phát hiện polyp và thời gian xử lý nhanh chóng. Mô hình này có thể sử dụng để xây dựng các ứng dụng thời gian thực trợ giúp các bác sĩ nội soi nâng cao tỉ lệ phát hiện polyp trong quá trình nội soi đại tràng.

Từ khóa: Ung thư đại trực tràng; Học sâu; Phát hiện đối tượng; Phát hiện polyp.