Automatic Detection of Polyps Using Deep Learning

Francisco Oliveira[1], Dalila Barbosa[1,2(✉)], Ishak Paçal[3], Danilo Leite[1], and António Cunha[1,2]

[1] UTAD—University of Trás-os-Montes and Alto Douro, 5001-801 Vila Real, Portugal
al67530@alunos.utad.pt, dalila.i.barbosa@inesctec.pt, {danilol, acunha}@utad.pt
[2] INESC TEC—Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
[3] Igdir University, Iğdir, Turkey
ishak.pacal@igdir.edu.tr

Abstract. Colorectal cancer is a leading health concern worldwide, with late detection being a primary challenge due to its often-asymptomatic nature. Routine examinations like colonoscopies play a pivotal role in early detection. This study harnesses the potential of Deep Learning, specifically convolutional neural networks, in enhancing the accuracy of polyp detection from medical images. Three distinct models, YOLOv5, YOLOv7, and YOLOv8, were trained on the PICCOLO dataset, a comprehensive collection of polyp images. The comparative analysis revealed YOLOv5's submodel S as the most efficient, achieving an accuracy of 92.2%, a sensitivity of 69%, an F1 score of 74% and a mAP of 76.8%, emphasizing the effectiveness of these networks in polyp detection.

Keywords: Machine learning · polyp detection · colonoscopy · YOLO

1 Introduction

According to GLOBOCAN2020, colorectal cancer is the third most common cancer worldwide, accounting for approximately 10% of all cases. This malignancy often originates from cells within the intestinal lining, which, over time, may increase into a benign tumor or polyp. As these polyps grow, they possess an increased risk of malignancy, a process often influenced by hereditary or spontaneous gene mutations governing cellular regulation [1].

The presence of such polyps in the colon is a significant public health concern due to their potential as colorectal cancer precursors. Consequently, the emphasis on early detection methods, such as colonoscopy, becomes paramount in identifying and addressing these polyps before their malignant transformation. Effective detection and prevention strategies are essential for mitigating the incidence of colorectal cancer and enhancing public health outcomes [2].

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 254–263, 2024.
https://doi.org/10.1007/978-3-031-60665-6_19

Polyp detection is primarily through colonoscopy, an invasive imaging procedure for the large intestine [3]. However, given its reliance on human interpretation, its effectiveness can vary. Deep learning has emerged as a tool that promises enhanced accuracy in polyp identification.

Deep learning's application in medical imaging is intriguing yet demanding. Legal considerations, especially concerning patient data rights and healthcare providers responsibilities, pose challenges [4]. Additionally, the success of deep learning often hinges on vast data sets, making data scarcity a hurdle to optimize its potential. These challenges, while formidable, invigorate researchers. The allure lies in optimizing medical practices ensuring timely and effective patient care.

This work aims to delve into deep learning models for polyp detection in colonoscopy images and juxtapose their efficacies. It entails exploring deep learning techniques for object detection, sourcing and curating public polyp databases, and training and assessing detection models.

Our research group has been involved in the development of projects using Deep Learning and Convolutional Neural Networks. In [5] is developed an unsupervised method for homography estimation in video capsule endoscopy frames, to be later applied in capsule localization systems. The pipeline is built on an unsupervised convolutional neural network, utilizing a VGG Net architecture, that estimates the homography between two images.

In [6] a variety of CNN models, such as, AlexNet, VGG16, and ResNet, were evaluated, using a transfer learning approach to maximize their efficacy, achieving a precision of 94,0% in lesion detection.

In [7] was explored abnormality classification within an unbalanced dataset of images from capsule endoscopy, using vector features extracted from the pre-trained CNNs to assess the impact of transfer learning with limited samples.

Within this paper, we have structured our content into three primary sections. Section 2 delves into an extensive literature review. Section 3 delineates the research methodology, encapsulating database preparation, training processes, polyp detection techniques, and culminating in model evaluation. Section 4 articulates the findings derived from this research.

2 Related Works

Recent advancements in deep learning techniques have substantially improved pattern recognition capabilities within medical imaging, particularly concerning polyp detection in colonoscopy videos. It follows a brief review of the principal methodologies published for automatic polyp detection in colonoscopy examinations.

A study in [8] revamped the YOLOv4 algorithm for real-time polyp detection using CSPNet and swapped activation and loss functions. Their method achieved 91.6% accuracy on the ETIS-LARIB dataset and 96.0% on the CVC-ColonDB dataset.

In [9], the YOLOv3 and YOLOv4 algorithms use data augmentation and transfer learning. By adjusting activation and loss functions, they achieved an accuracy of 97,0% on the SUN polyp dataset and 92.6% on PICCOLO.

Zhang et al. introduced a two-part CNN pipeline for polyp detection in colonoscopies [10]. Using the ResYOLO detection algorithm refined with colonoscopic images and the Effective Convolution Operators tracker, their method achieved an accuracy of 88.6% and a speed of 6.5 FPS.

Ma et al. [11] highlighted the variance in colonoscopy accuracy due to doctors' expertise and fatigue. They proposed a model based on the SSD_Inception_v2 network, achieving 94.9% accuracy and 93.7% sensitivity in polyp detection.

Nogueira-Rodríguez et al. [12] developed a YOLOv3-based deep learning model for real-time polyp detection, enhanced with object tracking. Trained on 28,576 images, it attained an F1 score of 88,0% for single images and 72.6% sensitivity for videos. The tracking addition improved the model's specificity for CAD system integration. They utilized six diverse datasets, including ImageNet and colonoscopy images. Despite its performance, the model's F1 score dropped by 13.7% on the PICCOLO and SUN datasets.

Qadir et al. [13] devised a polyp detection system using F-RNC, incorporating 2D Gaussian masks for improved polyp segmentation. This method demonstrated 86.6% sensitivity on the ETIS-LARIB dataset and 91,0% sensitivity on the CVC-ColonDB dataset. Another unnamed study using YOLOv3 achieved a 72.6% sensitivity rate.

Wan et al. [14] employed YOLOv5 for polyp detection using the Kvasir-SEG dataset and a local collection named WCY. Their results showed over 90% accuracy for both datasets. Meanwhile, Karaman et al. [15] optimized YOLOv4 with the artificial bee colony algorithm, achieving around 80,0% accuracy on the SUN and PICCOLO polyp datasets.

Gao et al. [16] presented the YOLOv5x-CG model for colorectal lesion detection using an expanded 4949-image dataset from Shanghai Sixth People's Hospital. The model employed image enhancement, k-means clustering for lesion localization, and integrated the Coordinated Attention (CA) and "Ghost" modules. It achieved accuracies of 92,3%, 95,5%, and 87,0% for polyps, adenomas, and cancer, respectively.

In a related study by Gan et al. [17], the YOLOv5x6 model was employed on the EndoCVC dataset of 2910 images, incorporating strategies like data mosaicking and Test Time Augmentation. The model achieved detection scores of 79,5% and 88,2% in two test rounds, indicating its robust performance.

3 Methodology

The methodology used in this work is outlined in the pipeline, shown in Fig. 1.

3.1 Database

The database selected was PICCOLO. It was decided on this type of dataset because it can be considered as an intermediate-level set. It does not suffer from a lack of data, but it will not give us the results that a more extensive database would provide. This dataset provided 3433 images.

The PICCOLO database presents injuries recorded between October 2017 and December 2019 at Basurto University Hospital (Bilbao, Spain). In total, the PICCOLO

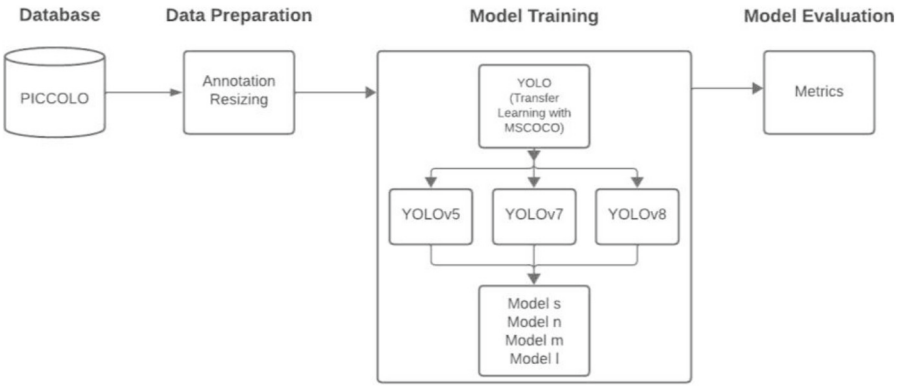

Fig. 1. Methodology pipeline

dataset included 76 lesions from 48 patients. 62 of these 76 lesions have image frames from white light (WI) and narrow-band imaging (NBI). The remaining 14 lesions were recorded using LB only (Fig. 2).

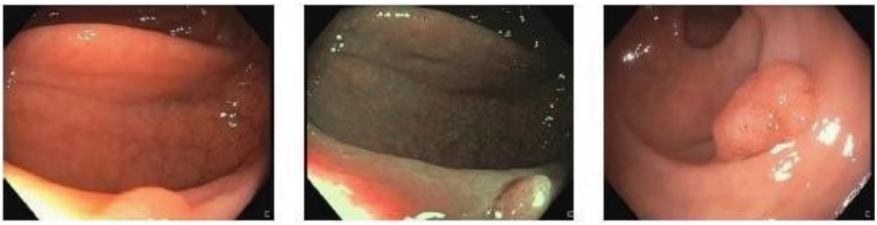

Fig. 2. Sample images from the PICCOLO database.

3.2 Data Preparation

In this study, we divided the images from the PICCOLO dataset in three sets, where 333 images were used in the test set, 2203 in the training set, and 897 in the validation set. Multiple YOLO models were utilized, necessitating the resizing of annotations to the prescribed format. This adaptation was required for the effective functioning of YOLO v5, v7, and v8. Roboflow, a dedicated tool, was used to facilitate these modifications and ensure the annotations were in the correct format. An example of these annotations is illustrated in Fig. 3.

3.3 Model Training

Transfer learning was utilized on YOLO detection models with the MSCOCO dataset, a comprehensive resource with 80 object categories and 330,000 images. This technique

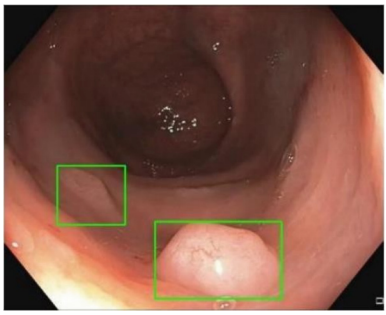

Fig. 3. Annotations on dataset images.

leveraged the MSCOCO-trained network parameters as an initial set for further training on our target datasets.

CNN models were selected for polyp detection, with a focus on the YOLO algorithm because of its speed, accuracy in crowded images, and real-time multi-class detection capabilities. This has made it a favorite in many computer vision tasks. The newer YOLO versions were specifically evaluated, including YOLOv7, v8, and v5, due to its proven success in detecting anomalies like polyps. In this study, models 'n', 's', 'm', and 'l' were tested to compare their performances, from the least to the most computationally intensive.

Models n, s, m, and l have parameters of 1.9M, 7.2M, 21.2M, and 7.2M, respectively. Their GPU processing times range from 45 ms to 224 ms, and they achieve mAP@50 values between 45.7% and 67.3%. Additionally, their mAP50-95 values vary from 28,0% to 49,0%.

YOLOv7, which shares a dataset with YOLOv5, boasts an accuracy of 69.7%, operating at a GPU speed of 3.2 ms.

Meanwhile, the latest YOLOv8 offers five pre-trained models like YOLOv5. These models, n through l, possess parameters ranging from 3.2M to 25.9M. They achieve mAP50-95 values up to 52.9%, and their GPU speeds vary between 0.99 ms and 2.39 ms.

3.4 Model Evaluation

The accuracy of a machine learning model, often defined as the ratio of correctly classified samples, is a commonly used metric for evaluation.

Standard metrics derived from the confusion matrix include sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, accuracy, and precision.

Considering the True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) results obtained by each model in the test set, the following metrics were calculated to evaluate their performance (Table 1).

Mean average precision (mAP) is a key metric for evaluating object detection, like polyp detection. It calculates the average precision per class considering prediction confidence.

Table 1. Metrics and their algebraic expressions.

Metrics	Expression
Precision	$P = \frac{TP}{TP+FP}$
Sensitivity	$P = \frac{TP}{TP+FN}$
F1 Score	$\frac{2\times TP}{2\times TP+FN+FP}$

The precision-recall curve shows how well a model classifies in binary tasks, with its shape influenced by classification thresholds. The Area Under the Curve (AUC) is a widely used metric that captures the overall performance of the model across these thresholds.

4 Results and Discussion

The result from this experiment is shown in Table 2, where it is showcased, the results obtained from state-of-the-art methodologies alongside the results from this study. This arrangement offers a comprehensive perspective, allowing for a direct comparison between the current benchmarks in the field and our contributions. The metrics used for the evaluations were accuracy, precision, sensitivity, F1-score, and mAP.

In Fig. 4 we can see the detection made by the models, in YOLOv5, where there are bounding boxes, red and green, that represent the detection made by each model, and the ground truth, respectively.

To perform a comprehensive assessment and identify the YOLO model with the best results a comparative approach was taken. Figure 5 shows the recall-precision curve of each model and reveals YOLOv5 as our dataset's top-performing model for polyp detection despite its training on smaller image sizes.

Table 2. Comparison with the state-of-the-art.

		Results			
Reference	RNC (Recurrent Neural Network)	Precision	Sensibility	F1-Score	mAP
Wan et al. [14]	YOLOv	91,3%	92,1%	91,7%	–
Pacal & Karaboga [8]	YOLOv5	96,0%	96,7%	96.4%	–
Pacal et al. [9]	YOLOv3	92,6%	79,9%	85,8%	97,0%
Karaman et al.[18]	YOLOv5x+ABC	93,3%	77,4%	–	84,5%
YOLOv5	Modelo n	98,0%	70,0%	70,0%	71,1%

(*continued*)

Table 2. (*continued*)

		Results			
Reference	RNC (Recurrent Neural Network)	Precision	Sensibility	F1-Score	mAP
YOLOv5	Modelo s	92,2%	69,0%	74,0%	**76,8%**
YOLOv5	Modelo m	98,9%	67,0%	72,0%	74,6%
YOLOv5	Modelo l	92,8%	58,0%	65,0%	70,4%
YOLOv7	-	97,5%	**75,0%**	**77,0%**	71,3%
YOLOv8	Modelo n	**99,9%**	6,0%	73,0%	71,9%
YOLOv8	Modelo s	**99,9%**	68,0%	73,0%	75,8%
YOLOv8	Modelo m	98,7%	67,0%	69,0%	66,7%
YOLOv8	Modelo l	99,6%	63,0%	70,0%	70,0%

It suggests that reducing image resolution can simplify the network's learning process for specific tasks, enhancing model speed and results. YOLO versions 7 and 8 were trained at a 640 × 640 resolution and showed a slight performance decline of 5.5% and 1%, respectively.

In general, the results were promising when it came to detecting polyps, but the best models were the YOLOv5s, with 76.8% AP, and the YOLOv8s with 75.8% mAP. This can be explained by the size of the data set used, i.e., for model n, if our set were smaller, it would possibly have better results compared to the others. For models larger than models, however, the lack of sufficient data can cause a drop in performance.

One possible reason to the version 5 of YOLO has outperformed more advanced architectures is because the dataset used in the study is not very large, and simpler models, such as YOLOv5, may do better because they have a lower risk of overfitting. More recent versions of YOLO, such as v7 and v8, can be more complex and therefore require larger datasets for effective training.

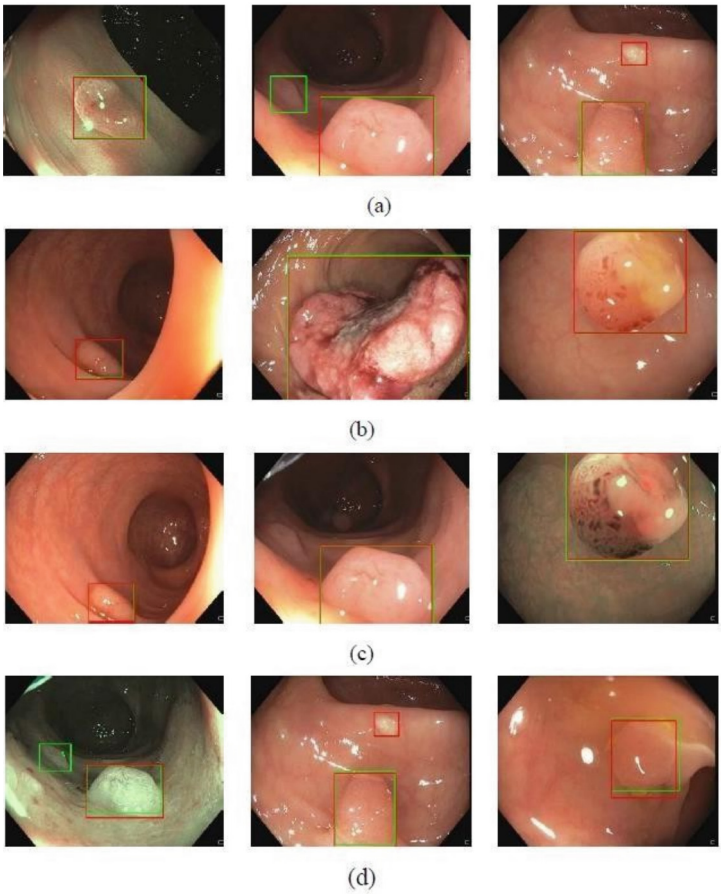

Fig. 4. Polyp Detection by YOLOv5. a) model l; b) model m; c) model s; d) model n; Red box – Prediction by the model; Green box – Ground Truth (Color figure online)

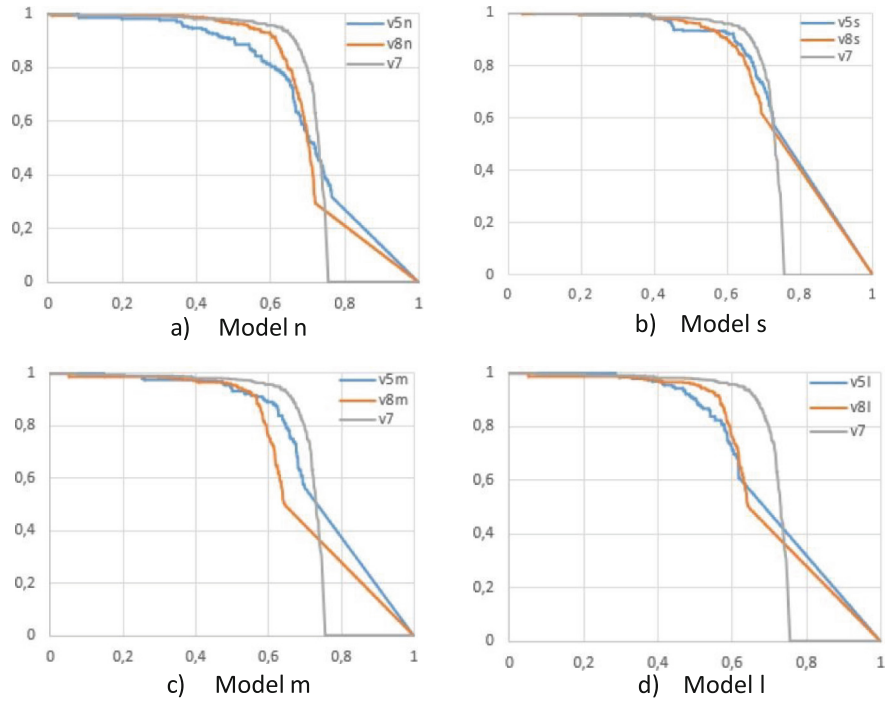

Fig. 5. Comparison of detection models. grey line – YOLOv7; orange line – YOLOv8; blue line – YOLOv5. (Color figure online)

5 Conclusions

Early detection of colon polyps is vital to prevent colorectal cancer, a major global cause of cancer deaths. Recent advances in artificial intelligence, especially convolutional neural networks, have enhanced automated polyp detection in medical imaging. This study is aimed to apply and compare deep learning methods for the identification of colon polyps. Following a brief literature survey, we selected YOLO architectures (YOLOv5, YOLOv7, and YOLOv8) for evaluation, utilizing the Piccolo dataset. Evaluations across the three YOLO iterations identified YOLOv5s as the most proficient, registering a precision of 92.2%, sensitivity of 69,0%, F1 score of 74,0%, and an AP of 76.8%. The scarcity of literature assessing YOLOv7 and YOLOv8 in polyp detection is worth noting, but our findings underscore their prospective utility.

Future research should emphasize training on dedicated colonoscopy datasets, refining pre-trained models for this niche, and leveraging strategies such as K-Fold for optimal parameter tuning.

Acknowledgements. National Funds finance this work through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020.

References

1. Henriksen, F.L.: Polyp detection using neural networks-data enhancement and training optimization. MS thesis (2017)
2. Ladabaum, U., et al.: Strategies for colorectal cancer screening. Gastroenterology **158**(2), 418–432 (2020)
3. Sakdyyah, A., Bestari, M.B., Suryanti, S.: Description of colonoscopy and histopathology of chronic diarrhea causes in non-neoplasm: literature review. Indones. J. Gastroenterol. Hepatol. Dig. Endosc. **22**(1), 52–59 (2021)
4. Castiglioni, I., et al.: AI applications to medical images: from machine learning to deep learning. Physica Med. **83**, 9–24 (2021)
5. Gomes, S., et al.: Unsupervised neural network for homography estimation in capsule endoscopy frames. Proc. Comput. Sci. **164**, 602–609 (2019)
6. Ribeiro, J., Nóbrega, S., Cunha, A.: Polyps detection in colonoscopies. Proc. Comput. Sci. **196**, 477–484 (2022)
7. Fonseca, F., et al.: Abnormality classification in small datasets of capsule endoscopy images. Proc. Comput. Sci. **196**, 469–476 (2022)
8. Pacal, I., Karaboga, D.: A robust real-time deep learning based automatic polyp detection system. Comput. Biol. Med. **134**, 104519 (2021)
9. Pacal, I., et al.: An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. Comput. Biol. Med. **141**, 105031 (2022)
10. Zhang, R., Poon, C.C.Y.: Regression-based convolutional neural network with a tracker. Comput.-Aided Anal. Gastrointest. Videos 133–139 (2021). https://doi.org/10.1007/978-3-030-64340-9_16
11. Ma, Y., et al.: Polyp location in colonoscopy based on deep learning. In: 2019 8th International Symposium on Next-Generation Electronics (ISNE). IEEE (2019)
12. Nogueira-Rodríguez, A., et al.: Real-time polyp detection model using convolutional neural networks. Neural Comput. Appl. **34**(13), 10375–10396 (2022)
13. Qadir, H.A., et al.: Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction. Med. Image Anal. **68**, 101897 (2021)
14. Wan, J., Chen, B., Yu, Y.: Polyp detection from colorectum images by using attentive YOLOv5. Diagnostics **11**(12), 2264 (2021)
15. Karaman, A., et al.: Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection. Appl. Intell. **53**(12), 15603–15620 (2023)
16. Gao, J., et al.: White-light endoscopic colorectal lesion detection based on improved YOLOv5. Comput. Math. Methods Med. **2022** (2022)
17. Gan, T., et al.: Detection of polyps during colonoscopy procedure using YOLOv5 network. In: EndoCV@ ISBI (2021)
18. Karaman, A., et al.: Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC). Expert Syst. Appl. **221**, 119741 (2023)