Effective Early Polyp Detection from Medical Images with YOLO-V7

Filippos Iakovidis
School of Science and Technology
International Hellenic University
Thermi, Thessaloniki, Greece
Email: fiakovidis@ihu.edu.gr

Leonidas Akritidis
School of Science and Technology
International Hellenic University
Thermi, Thessaloniki, Greece
Email: lakritidis@ihu.gr

Panayiotis Bozanis
School of Science and Technology
International Hellenic University
Thermi, Thessaloniki, Greece
Email: pbozanis@ihu.gr

Abstract—**Polyps constitute a potentially harmful cluster of cells appearing in the colon lining. Without early diagnosis and treatment, they can lead to Colorectal cancer (CRC), a quite common and highly lethal type of cancer. The early detection of a polyp, via a colonoscopy screening, can prevent CRC. However, the traditional colonoscopies exhibit high polyp-miss detection rates, leading to late CRC diagnoses. Therefore, studies for computer-aided detection systems for polyps are highly important. In our study, we employed the YOLOv7 model for the detection of polyps using transfer learning, utilizing a well-established endoscopic dataset. We take advantage of non-polyp images while training, to render our model more robust, achieve generalization, mitigate class imbalance and reduce noise. By training the model to recognize not only objects but also background elements, we reduce the risk of false positives. Compared to other researches, our model demonstrates a notable performance in both accuracy and real-time detection speed, achieving Precision and Recall values equal to 0.877 and mean average precision equal to 0.929.**

Index Terms—**object recognition, computer vision, YOLO, polyp detection, CRC, colorectal cancer**

I. INTRODUCTION

Colorectal cancer (CRC) is the second most lethal type of cancer in the United States. According to cancer statistics, new colon and rectum cases are expected to be 152.810 for both sexes in the year 2024, with 53.010 estimated deaths [1]. In the same article, the probability of developing invasive colon and rectum cancer in the USA is calculated to be 4.3% (1 in 23) for males and 3.9% (1 in 25) for females during their lifespan. Moreover, CRC is the main cause of cancer death for males aged 20–49 and the third most lethal for all ages. Meanwhile, it is the primary cause of cancer death for females aged 40–49 and the third for all ages. These numbers are indicative of the high frequency and mortality of colorectal cancer.

A colon polyp appears on the lining of the colon or rectum when the population of nearby cells grows abnormally. Most polyps are benign; however, the rapid cell division rates may render them malignant, leading to CRC. The standard procedure of CRC diagnosis is colonoscopy screening. An early diagnosis followed by the removal of the polyp, is an essential clinical procedure to prevent CRC [2]. As important as a colonoscopy screening can be, the polyp miss-detection rate is up to 25% [3], so the requirement for additional and more effective polyp detection techniques is imperative.

The recent revolution in the area of computer vision offered an attractive alternative for medical diagnosis tasks. In this context, deep learning architectures are trained to identify and localize specific objects within images or video frames. Traditional approaches on this task, relied on manually crafted features and heuristic algorithms, limiting their efficacy in handling complex and varied data.

The object detection models are divided in two categories, based on how many times the same input image is passed through the underlying network: single-shot and two-stage detectors. The members of the former category process the input images in a single pass, whereas the latter perform two passes. This difference renders single-shot detectors significantly faster and computationally efficient, at the cost of slightly degraded accuracy.

The YOLO (You Only Look Once) models lie in the single-shot detector category. They comprise an entire family of models renowned for their efficiency and accuracy in real-time applications [4]. They implement deep learning structures based on Convolutional Neural Networks (CNNs) in order to identify multiple objects within an image. But in contrast to the traditional classifiers, YOLO models not only predict class probabilities, but they also output bounding boxes around the recognized objects. This unique property allows them to combine remarkable accuracy with high inference speed, making them superior to other approaches, and especially in real-time applications. Over the years, YOLO has evolved through successive updates, each introducing its own performance and architectural enhancements.

Inspired by its performance, multiple authors employed YOLO for detecting polyps in medical images. Liu et al. proposed an approach for processing colonoscopic images based on YOLOv3 [5]. Their method integrates spatio-temporal information into a 2-D CNN-based object detection model with the aim of improving the polyp misprediction rates. The experiments demonstrated a remarkable F1 performance of 77% on the Etis-Larib dataset. Similarly, Misawa et al. developed an advanced polyp detection system also based on YOLOv3, which achieves real-time detection with over 90% sensitivity and specificity [6]. More recently, Karaman et al. presented an approach based on YOLOv5 that utilized the Artificial Bee Colony (ABC) optimization algorithm in

order to specify the optimal activation functions and hyper-parameters for the YOLOv5 detector [7].

In this paper we used a more recent model of the YOLO family, YOLOv7 [8]. According to its authors, YOLOv7 is the fastest and most accurate object detector in the range of 5-160 frames per second. To the best of our knowledge, this is among the first works to employ this version in polyp detection tasks. The combination of YOLOv7 with transfer learning approaches makes our research achieve more precise detection and much higher inference speed. By using non-polyp images and treating them as background in the training process, we obtain a more robust model that reduces the risk of False Positives. In addition, we decided to merge the adenomatous and hyperplastic polyps into a single class. By simplifying the detection like that, we achieved both higher accuracy and real-time detection speed.

The rest of this paper is organized as follows: In Section II we refer to the most significant advances in the area of polyp detection. Next, Section III describes our methodology for polyp detection with the YOLOv7 model. The results of the experimental evaluation are presented in Section IV. The conclusions of this paper are discussed in Section V along with some interesting elements for future research.

II. RELATED WORK

Numerous state-of-the-art models have emerged in the research field of object detection, each offering unique strengths and innovations. In earlier studies, the work of [9] used computer-aided tumor detection using colour wavelet features. On the other hand, the method of [10] detected polyps based on texture, whereas [11] introduced elliptical shape features. Despite their significant contribution on the field, these studies were based on feature patterns that are occasionally common between polyp and polyp-like structures. This similarity often leads to degraded performance.

The authors in [12] proposed a region segmentation, description and classification approach to detect polyps in colonoscopy videos, taking advantage of the novel Sector Accumulation Depth of Valleys Accumulation (SA-DOVA) as their region descriptor. Moreover, [13] introduced a binary classification algorithm which focuses on the geometrical features of the polyps and on the texture content of the frame. Their basic geometrical assumption is that polyps are protrusions, having a round-like shape.

During the last years, the adoption of CNN architectures has grown exponentially across diverse imaging tasks and domains. The effectiveness of these models rendered deep learning increasingly appealing and facilitated its integration into critical sectors like medicine. Regarding the object detection field, Region-based CNN (R-CNN) [14], Fast R-CNN [15] and Faster R-CNN [16] have shown profound progress in both accuracy and speed. These region-based CNN models learn rich feature representations automatically, in contrast to the detectors that employ feature hand-crafting methods. Another CNN model, called SSD-GPNet, was presented in [17]. It attempts to retrieve the lost information from the Max-Pooling

layers and exploit it as extra feature maps, leading to a notable improvement of mean average precision.

The Regional CNNs though, are two-stage detectors, which makes them comparatively slower at training and inference as well as more computationally demanding. On the other hand, the YOLO models have the ability to perform real-time object detection in a single pass through the neural network, making them significantly faster and more efficient. Among a large collection of YOLO variants, here we focus on YOLOv7, a model that combines high detection performance and inference speed [8]. In contrast, its ancestors are slower and less accurate, whereas the more recent variants (e.g. YOLOv8 [18], YOLOv9 [19], etc.) offer increased accuracy but they are slower to train and computationally more intensive.

YOLOv7 achieves its robustness through the introduction of three novel elements: i) an architecture called E-ELAN that allows the network to constantly improve its learning ability without destroying the original gradient path, ii) a model re-parameterization strategy [20], and ii) a new dynamic label assignment method [21].

In the domain of polyp detection, several competitive models demonstrated promising results. In this context, [22] introduced a multi-scale CNN architecture for polyp detection in colonoscopy videos, achieving high sensitivity and specificity. Furthermore, [23] developed a novel deep learning-based method for real-time polyp detection and segmentation, showcasing promising results on challenging datasets. The work of [24] modified YOLOv4-Tiny to improve its performance, by using the Inception-ResNet-A block as a replacement of the CSPDarknet-53-tiny backbone structure of the original YOLOv4-tiny. The authors of [25] used data augmentation techniques and fine-tuned the YOLOv5 model on the Kvasir-SEG and Shutter stock polyps datasets. Their method outperformed several other architectures like R-CNN, Faster R-CNN, and YOLOv4, in terms of detection accuracy. These models represent significant advancements in the field of polyp detection, offering potential solutions to enhance diagnostic accuracy and streamline clinical workflows.

Finally, several works have tested the polyp detection performance of more recent YOLO versions by using multiple publicly available benchmark datasets. In this context, [26] trained YOLOv7 on CVC-ClinicDB, CVC_ColonDB, ETIS-LaribPolypDB and others, achieving the highest performance on the former. On the other hand, [27] employed YOLOv8 across five distinct datasets.

III. METHODOLOGY

In this work we employed YOLOv7 for fast and accurate polyp detection. The model comprises several stacks of two-dimensional convolutional layers of different input/output channels, kernel sizes and stride values. Each 2d convolutional layer is typically followed by a Batch Normalization layer and employs the Sigmoid Linear Unit (SiLU) function for activation. YOLOv7 also includes several upsampling, concatenation and max-pooling modules.

For the requirements of this study, we employed the Colonoscopy Polyp Detection and Classification dataset (CPDC) [28]. Compared to other polyp detection datasets (e.g. CVC colon DB), CPDC is larger and more diverse. Each sample image is accompanied by an annotation file with information about the polyp class and the bounding box that indicates its presence in the image.

Initially, we developed a simple routine in order to translate the dataset annotation files to YOLO-compatible data files. In particular, this routine extracted: i) the coordinates of the boxes, and ii) their dimensions (width w_b and height h_b). With this data we were able to compute the horizontal and vertical centres of the boxes, C_x and C_y. Subsequently, the box dimensions and the center coordinates were normalized by dividing them by the image width w and height h, respectively:

$$\hat{w_b} = w_b/w \qquad (1)$$

$$\hat{h_b} = h_b/h \qquad (2)$$

$$\hat{C_x} = C_x/w \qquad (3)$$

$$\hat{C_y} = C_y/h \qquad (4)$$

Finally, the dataset annotation files were transformed to YOLO-compatible annotation files by recording the tuple:

$$[\; y \quad \hat{C_x} \quad \hat{C_y} \quad \hat{w_b} \quad \hat{h_b} \;] \qquad (5)$$

where y represents the polyp class. Note that the CPDC dataset in its annotation files, also includes information about the type of the polyp. The two types of the polyps are adenomatous and hyperplastic. In our study, we merged these two types into a single category (polyps), which allowed for a more generalized approach to polyp detection, since we solely focused on identifying and locating the presence of polyps. This simplified approach yielded improved performance in terms of both accuracy and real-time detection speed.

The handling of non-polyp images is of a high importance as well. Following the same pattern during the transformation of the CPDC annotation files, we created blank YOLO annotation files for those types of images. By adopting this strategy, we feed the model not only with polyp images, but also with non-polyp ones (which the model itself handles as background). Since these background images are treated as negative samples, we achieve a reduction in detection of False Positive samples. The absence of polyps in training is almost as important as their presence, as this offers a more realistic training scenario.

IV. EXPERIMENTS

In this section we report the results of our experimental evaluation. All the experiments have been conducted on the Google Colab platform. This specific virtual environment offered L4 GPUs with 62.8 GB of available System RAM, 22.5 GB GPU RAM and 201.2 GB of disk space.

A. Dataset

The effectiveness of our proposed methodology was attested by using the Colonoscopy Polyp Detection and Classification dataset (CPDC)[1] [28]. It consists of a collection of training, validation and test images along with their corresponding annotations files. Specifically, the training set comprises 28773 images including 1725 images that do not contain polyps. In the validation set, there are 4254 images with 40 non-polyp images. Finally, the test set consists of 4872 images, with 153 of those being non-polyp. The original dimensions of the images vary in a range between 384×288 and 768×576 pixels.

As mentioned earlier, this variance in the image dimensions is not significant, since YOLOv7 implements an image resizing process that reliably preserves the aspect ratio.

B. Evaluation Measures

To evaluate detection performance, we employed four well-established measures: Precision, Recall, F-score, and Mean Average Precision.

Precision P is a measure that estimates the percentage of correct predictions. In the problem of polyp identification, it is representative of the prediction confidence when a positive detection occurs. On the other hand, Recall R computes the fraction of the identified objects. Therefore, high values of Recall will alarm more patients to proceed to extended checks and eventually receive early treatment. For this reason, the importance of Recall is very high; this is valid in most predictive models in the area healthcare.

F-score takes both Precision and Recall into consideration. It measures the balanced performance of a model between these two measures in the following fashion:

$$F1 = \frac{2PR}{P + R} \qquad (6)$$

Another way of jointly examining Precision and Recall is via the Precision-Recall curve; a diagram that illustrates how these two measures correlate in various thresholds. Average Precision (AP) introduces a single value to effectively represent the entire Precision-Recall curve; it does so by computing the area below the curve. Moreover, Mean Average Precision (mAP) evaluates an object detector by calculating the average AP values across multiple object classes.

Finally, Intersection over Union (IoU) constitutes one of the most popular ways of quantifying detection performance; it computes the overlap between the predicted bounding boxes and the actual ones.

C. Model Training and Fine-tuning

We conducted a limited grid search (in terms of training epochs) in order to determine the optimal hyper-parameters for our YOLOv7 model. The selected values are depicted in Table I.

[1]https://doi.org/10.7910/DVN/FCBUOR

TABLE I: Model Hyperparameters.

Parameter	Value	Parameter	Value
lr0	0.01	anchor_t	4.00
lrf	0.1	fl_gamma	0.00
momentum	0.937	hsv_h	0.015
weight_decay	0.0005	hsv_s	0.70
warmup_epochs	3.00	hsv_v	0.40
warmup_momentum	0.80	degrees	0.00
warmup_bias_lr	0.10	translate	0.20
box	0.05	scale	0.90
cls	0.30	fliplr	0.50
cls_pw	1.00	mosaic	1.00
obj	0.70	mixup	0.15
obj_pw	1.00	paste_in	0.15
iou_t	0.20	loss_ota	1.00

Regarding the training of the model, we took advantage of the `img` flag, that automatically handles the rescaling of the input images before they are fed into the model during training. The aspect ratio is maintained as well. In our study, we set the image dimensions to be 416×416 and we used a batch size of 16 examples. These dimensions were adequate to achieve both high accuracy, and fast training time. We also experimented with larger image sizes (i.e., 640×640), without significant gains in detection performance, but with significant deceleration in training.

Under these circumstances, the YOLOv7 model comprised of 415 layers (as it was described in Section III) and about 37.2 million parameters. The training process was completed in roughly 7.4 hours. We trained for 50 epochs, with each epoch being completed in the span of 8 to 9 minutes, performing approximately 3.5 iterations per second.

D. Numeric Results

Table II reports the values of the aforementioned measures for the validation and test sets of the CPDC dataset. Regarding inference in the test set, our model achieved a remarkable Precision value of 0.877, exhibiting a satisfactory capability of correctly detecting polyps within the test images. In addition, the value of Recall was also 0.877; this signifies an equal effectiveness in locating a significant fraction of the existing polyps. In addition, Mean Average Precision, mAP@.50 was found equal to 0.929. YOLOv7 also reports mAP over different IoU thresholds, from 0.5 to 0.95; this value was measured equal to $mAP@0.5 - 0.95 = 0.583$.

By comparing these values with the ones achieved in the validation images, we observed: i) a small decrease of about

TABLE II: Polyp detection performance of the employed model in the validation and test sets.

Measure	Validation	Test
Precision	0.914	0.877
Recall	0.813	0.877
F-Score	0.861	0.877
mAP@0.5	0.876	0.929
mAP@0.5-0.95	0.632	0.583
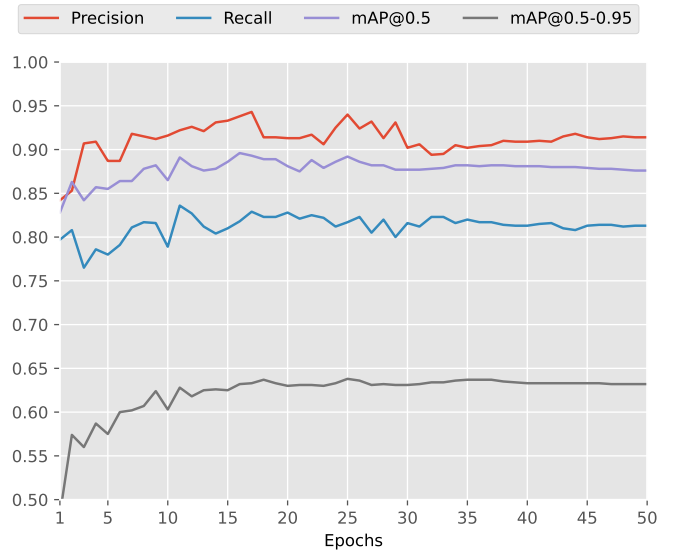
Fig. 1: Fluctuation of values of Precision, Recall, mAP@0.5, mAP@0.5-0.95 during the 50 epochs of YOLOv7 training process.

4% in Precision, and ii) a more significant increase of roughly 8% in Recall. These differences indicate that our model has been trained effectively and its generalization ability is satisfactory. Similar small changes have also been observed in the two aforementioned mAP values.

Figure 1 illustrates the fluctuation of the metric values during the 50 training epochs. The diagram depicts that all four values begin to stabilize between epochs 31-33 and they remain almost stable afterwards. This verifies the adequacy of training, since no essential benefits derive after epoch 33.

More specifically, the value of Precision fluctuates between 0.842 and 0.933 before getting stable at roughly 0.914. Recall oscillates in the range 0.765–0.829 before it becomes almost constant at the value of 0.813. In addition, mAP@0.5 begins from 0.828, peaks at 0.892 and stabilizes at 0.876 Finally, the value of mAP@0.5-0.95 moves in the range 0.483–0.638 before it converges a little bit lower at 0.632. As mentioned earlier, convergence is achieved at epochs 30-33; after this threshold is reached, the fluctuation in the values of all evaluation measures becomes infinitesimal.

E. Visual Results

After the inference on the test images, we collected visual information about the performance of the YOLOv7 model.

We organize the presentation in four figures. Figure 2 depicts 12 representative cases of successful polyp detection, along with the predicted bounding boxes and the confidence levels returned by the model. In contrast, Figure 3 illustrates three endoscopic images where our model failed to correctly detect a polyp. In particular, although only one polyp exists in each image, YOLOv7 incorrectly identifies a second one (accompanied by a lower confidence score).
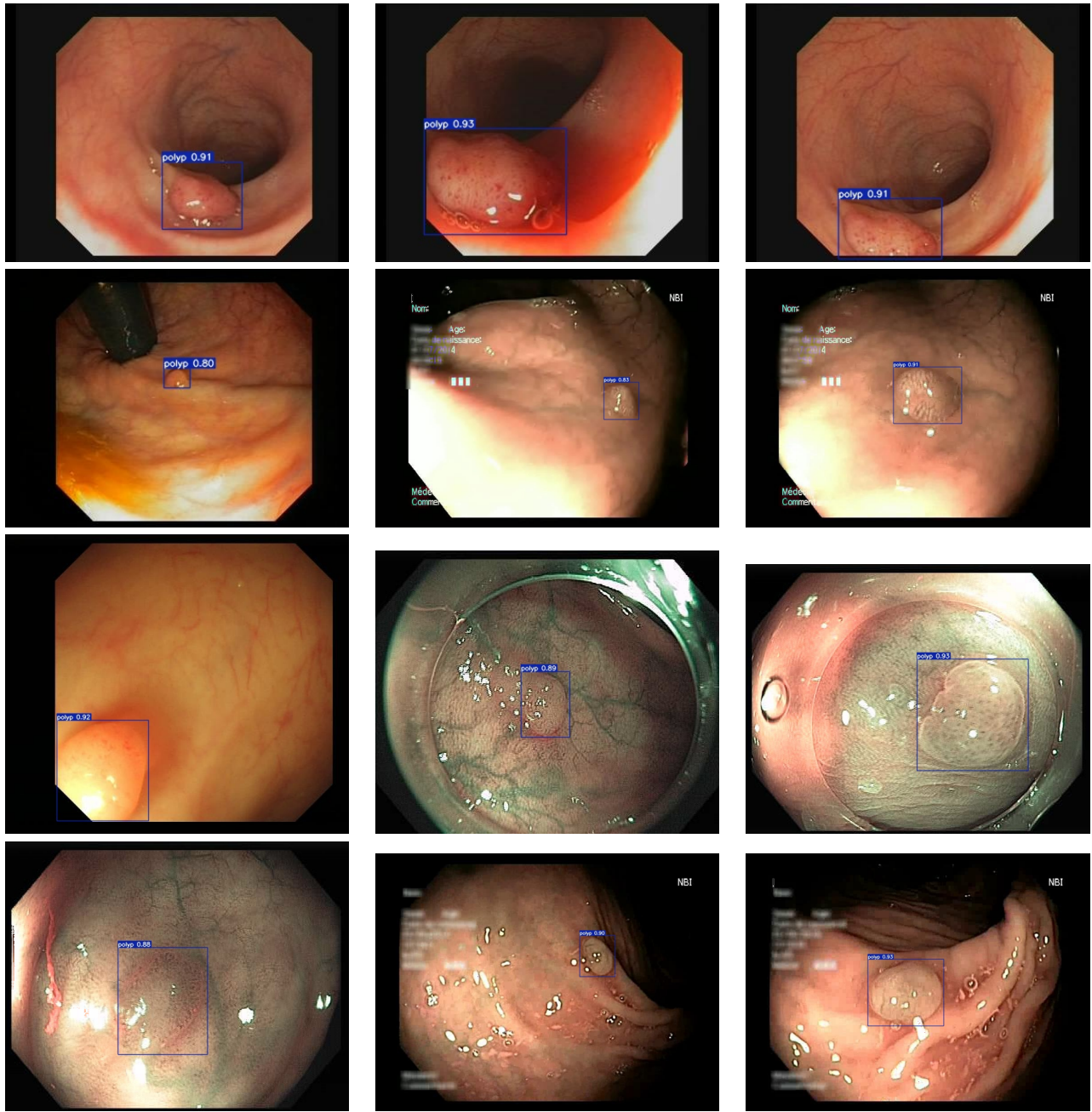
Fig. 2: Examples of successful polyp detection in 12 images of the Colonoscopy Polyp Detection and Classification dataset.
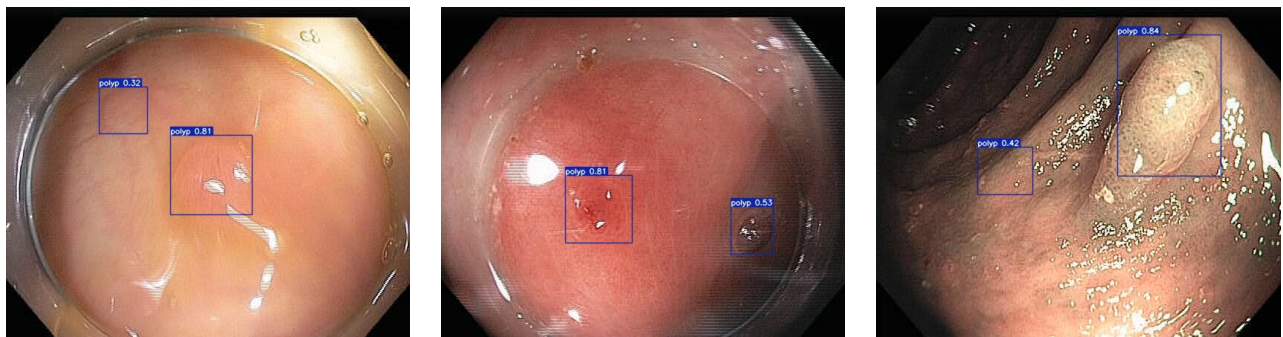

Fig. 3: Examples of failed polyp detection in 3 images of the Colonoscopy Polyp Detection and Classification dataset.
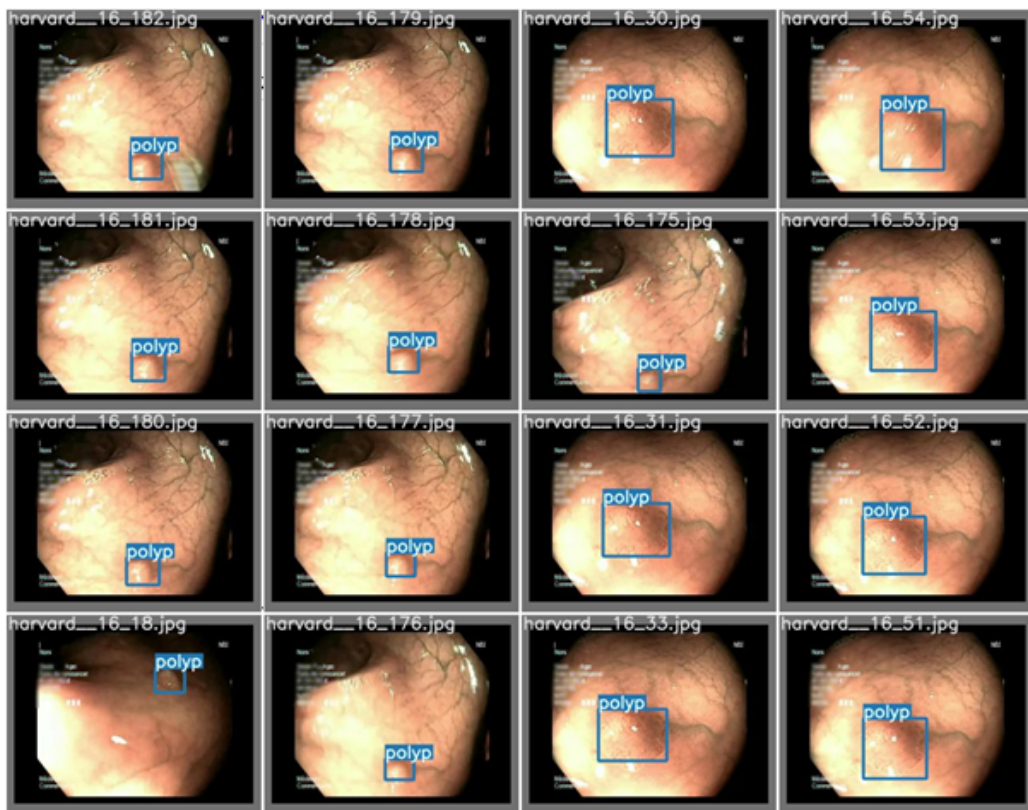
Fig. 4: A batch of 16 polyp images; the bounding boxes indicate the ground truth position and size of the polyps.
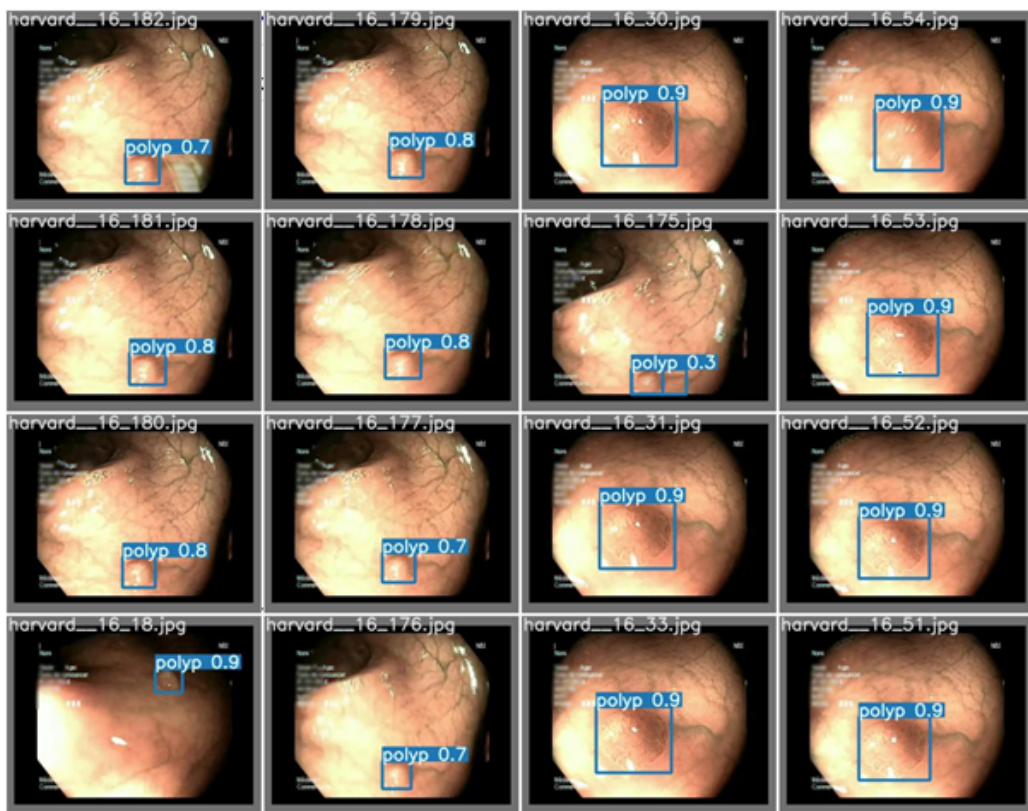

Fig. 5: The images of Fig. 4, along with the predicted bounding boxes and their respective confidence scores.

The next two diagrams in Figures 4 and 5 allow a direct comparison between the groud truth and the predicted size and position of 16 polyps. Both figures display the same images, but Fig. 4 depicts the ground truth bounding boxes, whereas Fig. 5 illustrates the predicted ones along with the confidence score returned by the model.

In general, the detection performance of our model is very satisfactory. This is verified from both the high confidence scores and the visual inspection. Regarding the former, their values lie into the range 0.7–0.9. The only exception to that was the image with filename harvard_16_175.jpg, where the model incorrectly identified two polyps (instead of one), and with a low confidence score (roughly 0.3).

V. Conclusions and Future Work

In this paper we studied effective data preprocessing techniques for early polyp detection with the YOLOv7 model. More specifically, we transformed the annotations off the original dataset into a specific format, so that the model could receive its inputs properly. With over 28000 training images and the inclusion of background non-polyp images, we achieved very satisfactory results in very high detection speeds. The final result showed high Precision and Recall, with values of 0.877 and mAP@.50 of 0.929.

Although this performance is promising, there are certainly several avenues for future work. Experiments with different architectures and deep learning techniques could be conducted on the same dataset, in order for us to actively compare the outcomes. Approaches like R-CNN, or even different versions of the YOLO family would provide us with better insight. Additionally, expanding or merging the dataset with several other public ones and also using several data augmentation techniques for both polyp and non-polyp images, could possibly lead to even better accuracy metrics. Finally, the use of image segmentation could be proven to be a crucial improvement in the task of polyp detection, as it seems to be a way-to-go approach in recent researches in the generic medical field.

References

[1] D. S. Dizon and A. H. Kamal, "Cancer statistics 2024: All hands on deck," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 1, 2024.

[2] M. Gschwantler, S. Kriwanek, E. Langner, B. Göritzer, C. Schrutka-Kölbl, E. Brownstone, H. Feichtinger, and W. Weiss, "High-grade dysplasia and invasive carcinoma in colorectal adenomas: a multivariate analysis of the impact of adenoma and patient characteristics," *European Journal of Gastroenterology & Hepatology*, vol. 14, no. 2, pp. 183–188, 2002.

[3] A. Leufkens, M. Van Oijen, F. Vleggaar, and P. Siersema, "Factors influencing the miss rate of polyps in a back-to-back colonoscopy study," *Endoscopy*, pp. 470–475, 2012.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[5] X. Liu, X. Guo, Y. Liu, and Y. Yuan, "Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images," *Medical Image Analysis*, vol. 71, p. 102052, 2021.

[6] M. Misawa, S.-e. Kudo, Y. Mori, T. Cho, S. Kataoka, A. Yamauchi, Y. Ogawa, Y. Maeda, K. Takeda, K. Ichimasa *et al.*, "Artificial intelligence-assisted polyp detection for colonoscopy: Initial experience," *Gastroenterology*, vol. 154, no. 8, pp. 2027–2029, 2018.

[7] A. Karaman, I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, S. Coskun, O. Sahin, and D. Karaboga, "Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC)," *Expert Systems with Applications*, vol. 221, p. 119741, 2023.

[8] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.

[9] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 3, pp. 141–152, 2003.

[10] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung für die Medizin 2009*, 2009, pp. 346–350.

[11] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. De Groen, "Polyp detection in colonoscopy video using elliptical shape feature," in *Proceedings of the 2007 IEEE International Conference on Image Processing*, vol. 2, 2007, pp. II – 465–II – 468.

[12] J. Bernal, J. Sánchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, vol. 45, no. 9, pp. 3166–3182, 2012.

[13] A. V. Mamonov, I. N. Figueiredo, P. N. Figueiredo, and Y.-H. R. Tsai, "Automated polyp detection in colon capsule endoscopy," *IEEE Transactions on Medical Imaging*, vol. 33, no. 7, pp. 1488–1502, 2014.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[15] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[17] X. Zhang, F. Chen, T. Yu, J. An, Z. Huang, J. Liu, W. Hu, L. Wang, H. Duan, and J. Si, "Real-time gastric polyp detection using Convolutional Neural Networks," *PloS one*, vol. 14, no. 3, p. e0214133, 2019.

[18] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with YOLOv8," *arXiv preprint arXiv:2305.09972*, 2023.

[19] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv preprint arXiv:2402.13616*, 2024.

[20] M. Hu, J. Feng, J. Hua, B. Lai, J. Huang, X. Gong, and X.-S. Hua, "Online convolutional re-parameterization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 568–577.

[21] S. Li, C. He, R. Li, and L. Zhang, "A dual weighting label assignment scheme for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9387–9396.

[22] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, 2015.

[23] P. Wang, T. M. Berzin, J. R. G. Brown, S. Bharadwaj, A. Becq, X. Xiao *et al.*, "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study," *Gut*, vol. 68, no. 10, pp. 1813–1819, 2019.

[24] M. Doniyorjon, R. Madinakhon, M. Shakhnoza, and Y.-I. Cho, "An improved method of polyp detection using custom YOLOv4-tiny," *Applied Sciences*, vol. 12, no. 21, p. 10856, 2022.

[25] P. Ghose, A. Ghose, D. Sadhukhan, S. Pal, and M. Mitra, "Improved polyp detection from colonoscopy images using finetuned YOLO-v5," *Multimedia Tools and Applications*, pp. 1–26, 2023.

[26] N. T. Thang, D. Q. Tu, N. S. Huy, N. D. Hanh, T. T. Luong, N. D. Do *et al.*, "Real-time detection of colon polyps during colonoscopy using yolov7," *Journal of Military Science and Technology*, no. CSCE7, pp. 122–134, 2023.

[27] M. Lalinia and A. Sahafi, "Colorectal polyp detection in colonoscopy images using yolo-v8 network," *Signal, Image and Video Processing*, vol. 18, no. 3, pp. 2047–2058, 2024.

[28] K. Li, M. I. Fathan, K. Patel, T. Zhang, C. Zhong, A. Bansal, A. Rastogi, J. S. Wang, and G. Wang, "Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations," *Plos one*, vol. 16, no. 8, p. e0255809, 2021.