BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Deep learning model applied to real-time delineation of colorectal polyps

Moana Gelu-Simeon[1,2*], Adel Mamou[3], Georgette Saint-Georges[1], Marceline Alexis[1], Marie Sautereau[1], Yassine Mamou[4] and Jimmy Simeon[5]

Abstract

Background Deep learning models have shown considerable potential to improve diagnostic accuracy across medical fields. Although YOLACT has demonstrated real-time detection and segmentation in non-medical datasets, its application in medical settings remains underexplored. This study evaluated the performance of a YOLACT-derived Real-time Polyp Delineation Model (RTPoDeMo) for real-time use on prospectively recorded colonoscopy videos. Methods: Twelve combinations of architectures, including Mask-RCNN, YOLACT, and YOLACT++, paired with backbones such as ResNet50, ResNet101, and DarkNet53, were tested on 2,188 colonoscopy images with three image resolution sizes. Dataset preparation involved pre-processing and segmentation annotation, with optimized image augmentation. Results: RTPoDeMo, using YOLACT-ResNet50, achieved 72.3 mAP and 32.8 FPS for real-time instance segmentation based on COCO annotations. The model performed with a per-image accuracy of 99.59% (95% CI: [99.45 – 99.71%]), sensitivity of 90.63% (95% CI: [78.95 – 93.64%]), specificity of 99.95% (95% CI: [99.93 – 99.97%]) and a F1-score of 0.94 (95% CI: [0.87–0.98]). In validation, out of 36 polyps detected by experts, RTPoDeMo missed only one polyp, compared to six missed by senior endoscopists. The model demonstrated good agreement with experts, reflected by a Cohen's Kappa coefficient of 0.72 (95% CI: [0.54–1.00], $p < 0.0001$). Conclusions: Our model provides new perspectives in the adaptation of YOLACT to the real-time delineation of colorectal polyps. In the future, it could improve the characterization of polyps to be resected during colonoscopy.

Keywords Artificial intelligence, Convolutional neural network, Deep learning model, Instance segmentation

Background

Colorectal cancer (CRC) is currently the third most common cancer in men and the second in women, as well as the second leading cause of all deaths from cancer [1]. Incidence and mortality rates differ across regions, which can be partially attributed to economic inequalities and differences in healthcare access [2, 3]. While effective screening measures has fallen the incidence of CRC in France, it has markedly increased in the French Caribbean islands during the past twenty years and continues to increase in women [4].

Colonoscopy is effective in preventing CRC and lowering its mortality rate (up to 60%), by detecting all

*Correspondence:
Moana Gelu-Simeon
moana.simeon@chu-guadeloupe.fr
[1]Service d'Hépato-Gastroentérologie, CHU de la Guadeloupe, Pointe- à-Pitre F-97100, France
[2]Univ Antilles, Univ. Rennes, INSERM, EHESP, IRSET (Institut de Recherche en Santé, Environnement et Travail) - UMR_S 1085, Pointe-à-Pitre F-97100, France
[3]Biostatistic department, Univ. Montpellier, Montpellier F-34000, France
[4]Service de Médecine Nucléaire, CHU de la Guadeloupe, Pointe-à- Pitre F-97100, France
[5]Informatic Department, CHU de la Guadeloupe, Pointe-à-Pitre F-97100, France

types and sizes of colorectal polyps, and enabling their resection during the procedure [5]. Many studies have confirmed an association between the quality of a colonoscopy and CRC incidence or mortality [6, 8]. The three main quality criteria for a colonoscopy are good preparation of the colon, sufficient time to inspect the colonic mucosa and a high adenoma detection rate (ADR) [7]. The ADR may range from 17 to 47% depending on the endoscopist, whereas a 1% increase in the ADR can enable a 3% reduction in the risk of interval CRC [9]. Consequently, missed lesions may be the cause of 58% of post-colonoscopy CRC [7], thus ranking CRC as the fourth most commonly reported medical diagnostic error [10].

Since 2012, the introduction of artificial intelligence (AI) has contributed to improve diagnostic accuracies and reduce error rates in all areas of medicine [11, 13]. Since 2016, early studies in experimental or retrospective settings, based on deep learning models such as convolutional neural network (CNN) models, have produced promising results regarding automated colorectal polyp detection with instance segmentation [14, 19]. Indeed, in endoscopy, we need to characterize each polyp independently in order to validate their therapeutic management and resecability, while semantic segmentation simply detects whether the observed abnormality is a polyp or not. More recently, academic researchers, professional societies and industry have demonstrated that AI improved the ADR during randomized control trials, thus making this technology available to more users [20, 23]. In the future, computer-aided diagnosis should contribute to enhance the performance of endoscopists, by lowering the missed adenoma rate, determining the reliability of the resection and ensuring its precision.

Most CNN models adopt a two-stage architecture, which has demonstrated high performance precision, as reported for Fast R-CNN and Mask R-CNN [24, 25]. However, this makes them too slow for real-time use. As the choice between speed with one-stage architecture or precision with two-stage architecture is determined by the needs of the application and the material resources involved (graphics processing unit (GPU) and computer processing unit (CPU)). The YOLACT model is a full CNN that was originally designed with one-stage architecture to enable the real-time detection of non-medical image datasets [26]. YOLACT + + was secondarily developed from YOLACT architecture to allow real-time instance segmentation with improved accuracy [27, 28]. Our choice of YOLACT was based on its recognized performance in real-time instance segmentation, as demonstrated by Bolya et al. [26, 27]. Since initial work on YOLACT comparing it to YOLOv3, which showed that YOLACT was faster than YOLO, no comparison has been made with other YOLO series [29, 33]. Moreover, YOLACT allows flexible mask generation thanks to a learned set of prototype masks, which has proven beneficial for polyp delineation. In 2023, SOLOv2 was compared to YOLACT, confirming YOLACT's superior accuracy at equivalent speed, thus significantly reducing the number of parameters and floating-point operations per second (FLOPs) [34].

YOLACT + + was furthermore used with different backbones and successfully applied for detection of colorectal polyps during the endoscopy computer vision challenge in 2020 [35].

The aim of the present study was to select, design and evaluate the performance precision of a YOLACT derived model, the Real-time Polyp Delineation Model (RTPoDeMo), with respect to speed detection, instance segmentation and delineation of colorectal polyps. RTPoDeMo was trained on the validation dataset of colon images for the two-class "polyp" and "background" problem with a supervised learning process, and then challenged on 18 prospectively recorded colonoscopy videos.

Methods

Selection and training process of the model

First step: Architecture application design

We identified a global architecture application enable to real-time instance segmentation with three frameworks method architectures: YOLACT and YOLACT + + based on "one-stage architecture", and Mask-RCNN on "two-stage architecture". The YOLACT and YOLACT + + models were initialized with weights pretrained on the ImageNet dataset [36], consistent with the approach reported in the original YOLACT + + implementation by Bolya et al. [27], Mask-RCNN was initialized with COCO database [37].

These method architectures with instance functionality were associated with three different backbones: ResNet-50, ResNet-101 and DarkNet-53 to extract features of the data images. These backbones are most commonly adopted in the literature in association with YOLACT or YOLO [15, 26, 27]. Beyond algorithmic modifications, diverse architectural structures like ResNet or VGG for YOLACT and DarkNet for YOLOv3 and YOLOv4 enhance their performance [38]. The training model was conducted using a parallel architecture with an NVIDIA Tesla P100 GPU and an Intel Xeon 2.20 GHz processor, enabling real-time processing. We implemented the model using Python 3.7 with PyTorch v1.4 and OpenCV. Regarding hyperparameters, we selected the Stochastic Gradient Descent optimizer as it provides more stability during the training process. We empirically adjusted a learning rate between 1e-4 and 1e-3 and chose the batch size based on the GPU memory

constraints. The model was trained for 50 epochs to ensure stable convergence.

We used the Gradient-weighted Class Activation Mapping (GRAD-CAM) heat map technique to visualize and understand the areas of an image considered by the model for its predictions 39. By highlighting these regions, it helps in identifying and correcting errors and biases in the model. This visualization technique ensures that the model's focus is on relevant features and that the model learns on the polyp area, improving both the model's reliability and interpretability.

Second step: Preparation of the medical images

Preprocessing dataset. A dataset of 2,188 endoscopic images from four public datasets (Kvasir-SEG 1,000 images, resolution varies from 332 × 487 to 1920 × 1072, CVC-ClinicDB 612 images 384 × 288, CVC-ColonDB 380 images 574 × 500 and ETIS-LaribPolypDB 196 images 1225 × 966) was used [40, 43]. They were used for grounding mask in order to train the model. These datasets essentially contain polyp images. Each image was supplied with its corresponding ground truth segmentation mask because these datasets are designed for semantic segmentation, not for instance segmentation (Fig. 1). When a frame contains multiple polyps, the CNN needs to differentiate between individual polyps (named instances) to correctly process the loss function. Since we used previously annotated datasets with only one ground truth mask per image, a preprocessing step, using the COCO annotation format, was required, to differentiate between individual polyps within the same binary mask [44]. For this, and extract the corresponding list of polygon vertices, we used a border following algorithm based on topological structural analysis as described by Satoshi et al. [45]. The extracted polygon vertices are then placed in a JSON file according to the COCO annotation format [44].

Dataset augmentation consisted in rotated and randomly flip inputs horizontally and vertically. The dataset was randomly split into training dataset (80% of the data), validation dataset (10% of the data), and test sets (10% of the data) using an 80:10:10 ratio. All images extracted from the datasets were included only in one of the three sets (training, validation, or test), thus ensuring complete independence between the datasets. This achieved to configure the weight of the hyper-parameters of the models and increase the accuracy of the model to 1.4%.

The third step: Training process and benchmark of each model

The models were evaluated using the mean average precision (mAP) metric (defined lower) to identify the most accurate configuration. The best-performing model was then used to calibrate the image resolution with the selected backbone, and its performance was compared to that of Mask R-CNN and YOLACT + + methods. We benchmarked Mask R-CNN and YOLACT + + with nine YOLACT models by combining three backbone configurations (ResNet50, ResNet101, DarkNet53) with three image resolutions (400, 550, and 700 pixels).

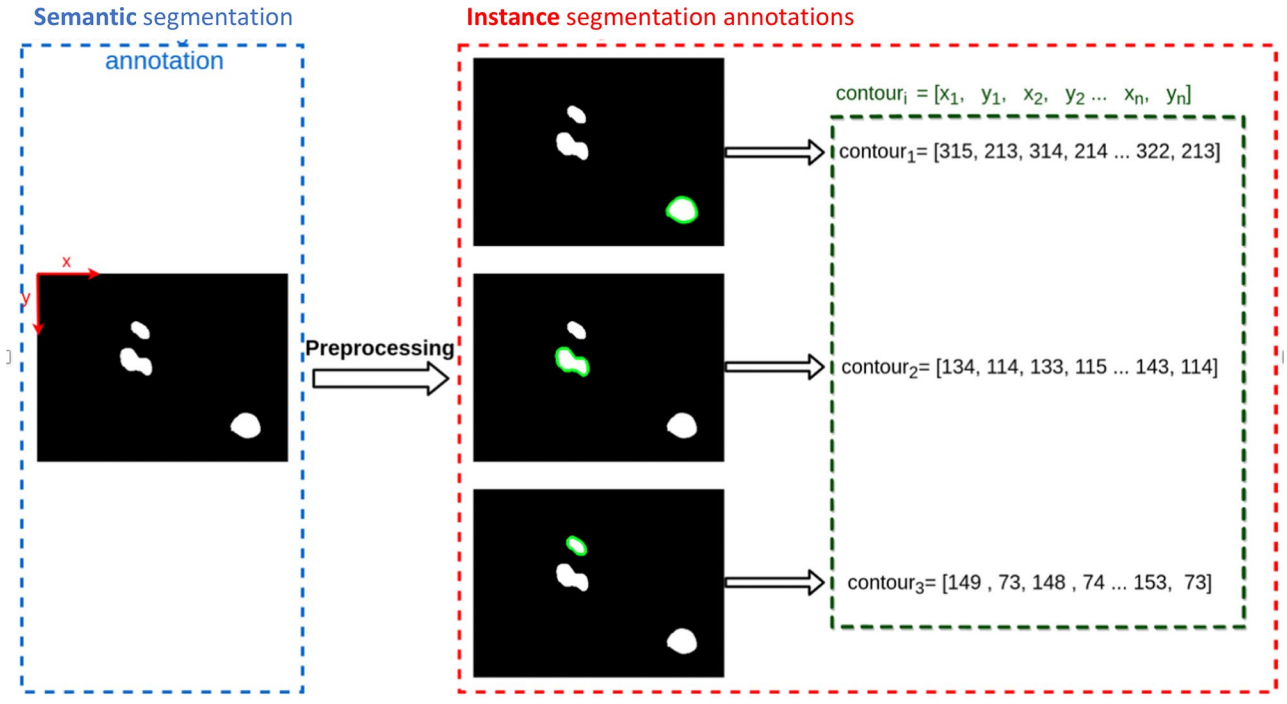

Fig. 1 Conversion from semantic segmentation annotation to instance segmentation annotation using the border following algorithm

Mask-RCNN was to slow compared with YOLACT et and YOLACT ++ (Table 1). To ensure a fair comparison, YOLACT, YOLACT++, and Mask R-CNN models were trained and evaluated under identical experimental conditions: same dataset splits, image resolutions (400, 550, 700 pixels), backbone pairings (ResNet-50, ResNet-101, DarkNet-53), training epochs, optimizer settings, and loss functions. All models were executed using the same hardware (NVIDIA Tesla P100 GPU) and software environment (Python 3.7, PyTorch 1.4). This standardized setup allowed us to isolate the architectural differences in performance (accuracy and real-time processing speed).

The Jaccard index or intersection over the union (IoU) quantifies the extent of the overlap between the predicted bounding box over the ground truth bounding box, providing an essential measure of segmentation accuracy. The Average Precision (AP), calculated at various IoU thresholds, offers insights into the precision of segmentation at different levels of accuracy. The mAP, computed across multiple IoU thresholds, provides a comprehensive evaluation of the model's overall performance. We evaluate the model efficiency by calculating the frames per second (FPS). The "Backbone" uses convolutional, pooling layers, and activation functions to capture the essential features from the image.

Based on two performance criteria: delineation accuracy (mAP) and real-time processing ability (FPS > 30) (Table 1), the YOLACT-ResNet50 model with a 550 pixels resolution was the most effective, achieving a mAP of 72.3 and an FPS of 32.8 on COCO annotation. However, when compared to YOLACT + + with the ResNet50 and ResNet101 backbones at the same 550 pixels resolution, although these models provide better accuracy with mAP gains of + 2.15 and + 2.83, respectively, they exhibit significant FPS losses of -9.19 and – 13.79 and do not meet the FPS > 30 criterion. Additionally, the Mask RCNN_550-ResNet50 model, despite having the highest mAP of the twelve benchmarked models, displays the lowest FPS at 9.4, with a gap of -23.42 compared to the reference model and – 20.6 compared to the required threshold. Finally, the YOLACT-Res-Net 50_550 model, named "RTPodeMo", resulted with the highest performance at the end of the evaluation phase of the 12 models (Fig. 2). One-stage architecture framework was confirmed with YOLACT method for instance segmentation, ResNet50 backbone with the ReLU activation function [46] and 550 pixels for the polyp images resolutions.

Description of the model's architecture with YOLACT framework method

The internal architecture of YOLACT is composed of functions that provide various processes (Fig. 3). The Feature Pyramid Network (FPN) combines feature maps from different layers of the backbone to create a rich, multi-scale feature representation. The "Detected Small Object" process ensures that smaller objects are accurately represented and detected by leveraging the enhanced multi-scale features from the FPN. ProtoNet uses convolutional layers to produce fixed prototype masks, which serve as a basis for generating instance-specific masks.

The "Prediction Head" provides:

- Bounding Boxes that predicts the coordinates of the object.
- Confidence Scores that predicts confidence scores for object classification and applies the softmax function to convert these scores into class probabilities.
- Mask Coefficients that predicts coefficients used to combine with prototype masks and to generate instance masks.

The "Assembly" uses the predicted mask coefficients to create a weighted combination of the prototype masks, resulting in initial instance masks for each detected

Table 1 Training phase for the different backbone/input size combinations

Method	Backbone	input size	fps	mAP	AP50	AP75	AP95
YOLACT	ResNet50	550	32.82	72.32	93.28	82.16	6.64
YOLACT	ResNet50	700	25.83	72.09	93.16	82.75	7.62
YOLACT	ResNet50	400	41.85	71.35	93.56	81.38	4.84
YOLACT	ResNet101	550	23.16	73.53	94.52	79.94	5.79
YOLACT	ResNet101	700	18.72	73.58	93.20	80.59	7.13
YOLACT	ResNet101	400	30.08	72.44	93.69	79.58	6.61
YOLACT	DarkNet53	550	29.41	73.31	94.67	82.80	10.54
YOLACT	DarkNet53	700	23.55	73.04	93.30	83.00	6.60
YOLACT	DarkNet53	400	38.86	72.12	92.75	82.56	8.15
YOLACT++	ResNet50	550	23.63	74.47	94.18	84.81	12.35
YOLACT++	ResNet101	550	19.03	75.15	95.25	86.11	13.70
Mask-RCNN	ResNet50	550	9.4	78.29	99.6	91.56	10.81

Abbreviations: fps, frame per second; mAP, mean average precision; AP, average precision

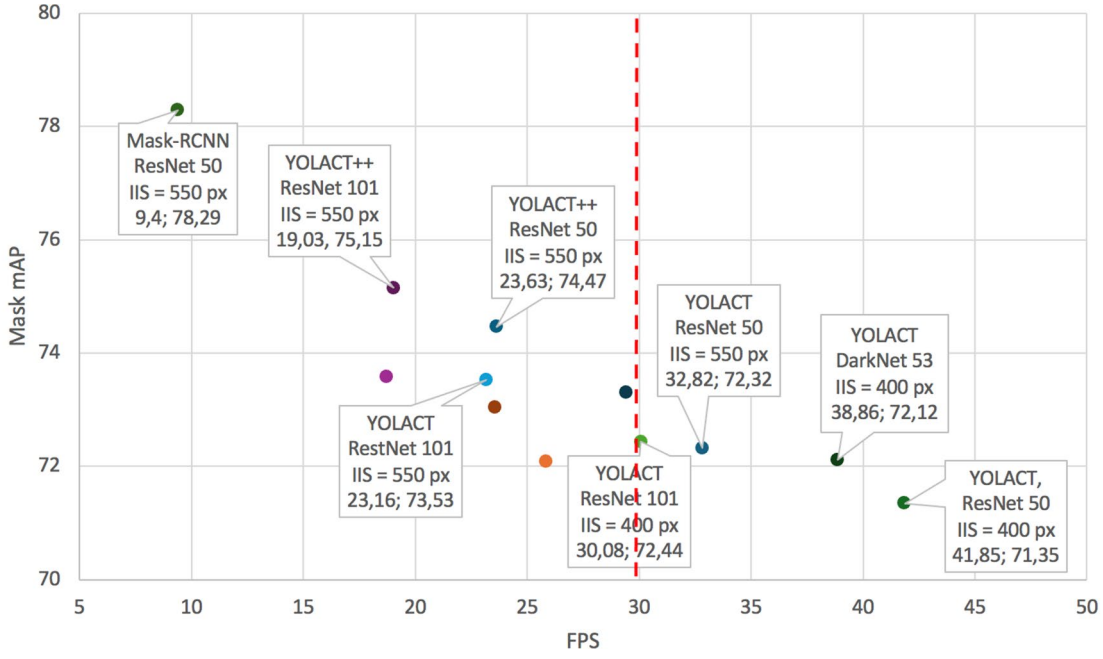

Fig. 2 Trade-off analysis between efficiency and performance for various instance segmentation methods and backbones. In each bubble is indicated the corresponding model (YOLACT or Mask-RCNN), the backbone (ResNet or DarkNet), the image input size (IIS), the mAP score and the associated FPS (frame per second)

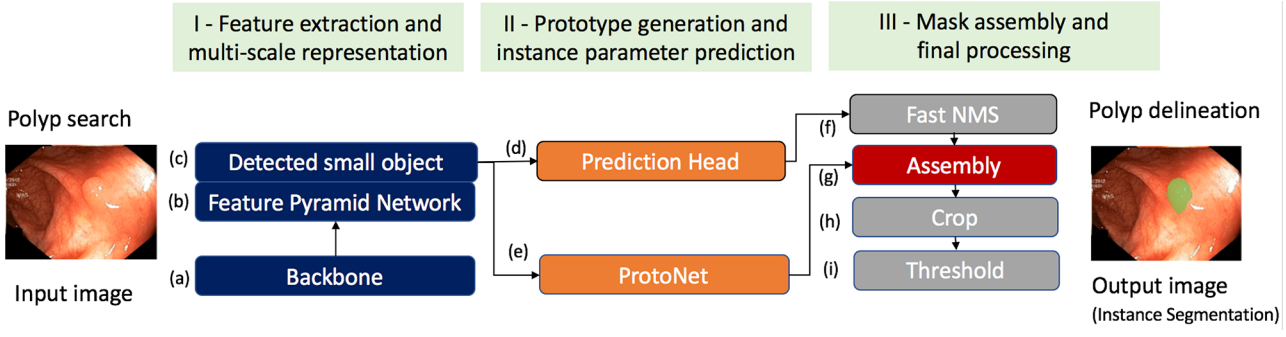

Fig. 3 RTPoDeMo architecture: Image segmentation process. (**a**) Backbone extracts low-level and high-level features from the input image; (**b**) Feature pyramid network enhances multiscale feature representation to handle objects of different sizes; (**c**) Detected small object handles the detection of smaller objects within the feature maps; (**d**) Predicted head predicts bounding boxes, confidence scores and mask coefficient for each detected instance; (**e**) ProtoNet generates a set of prototype masks that are independent of instances; (**f**) Fast NMS (Non-Maximum Suppression) filters out redundant bounding box predictions and retains only the most relevant ones on their confidence scores; (**g**) Assembly combines the prototype masks generated by ProtoNet with the predicted mask coefficients for each detected instance; (**h**) Crop refines the boundaries of the instance masks; (**i**) Treshold applies a binary treshold to the instance masks to convert them into clear, segmented regions

object. Fast Non-Maximum Suppression (NMS) applies NMS to the bounding boxes by calculating IoU and suppressing overlapping boxes with lower confidence scores, ensuring that only the most confident, non-overlapping detections are kept.

The "Crop" process refines the boundaries of the instance masks to fit precisely within the detected bounding boxes, ensuring the mask accurately represents the object. The "Threshold" function applies a threshold to the mask values, setting pixels above the threshold to 1 (foreground) and those below to 0 (background), resulting in binary masks that clearly delineate the objects. Finally, the "Final Instance Masks" are obtained after applying the threshold, clearly segmenting the detected objects, and ready for visualization or further processing.

Deployment and validation of the selected model in real-time conditions

Study design

A non-interventional pilot study was conducted in the Gastroenterology Department at Guadeloupe University Hospital, France. Colonoscopy videos were collected during standard colonoscopy procedures to screen for CRC between 1st June 2023 and 9 July 2023, in patients

aged 18 years or older. Any colonoscopies performed in a context of lower gastrointestinal bleeding were excluded from the study.

Colonoscopy procedures and video record

Three senior gastroenterologists (over 10 years of experience) performed the colonoscopies and recorded the videos during procedures. The senior gastroenterologists identified any polyps during these standard procedures and images were captured using high-definition colonoscopes (CF-HQ190 or PCF-H190, Olympus®, Tokyo, Japan) and monitors (EVIS Exera III CLV-190, Olympus®, Tokyo, Japan). The criteria for video quality were good bowel preparation (BOSTON > 7) and an optimal colon exploration time (more than 6 min). Only colonoscopies involving a complete exploration of the colon from caecum to anus and compliant with these quality criteria were recorded. The video stream speed was 30 frames per second.

Endoscopists also provided the following clinical metadata associated to each polyp:

- Polyp size are clustered in three categories according to: (a) small (6 size ≤ 10 mm); (b) Large (> 10 mm).
- Polyp classification according to Paris criteria [47]: six groups are defined: Ip (protruded, pedunculated), Is (protruded, sessile), IIa (superficial, elevated), IIb (flat), IIc (superficial shallow, depressed) and III (excavated).

Application

As the training datasets essentially contain polyp images, we test the ability of the model to detect and delineate the polyps in the context of images that does not contain any lesion directly on the recorded videos coming from the same center. We applied RTPoDeMo to the videos recorded during colonoscopies (Fig. 4*), through a computer interface created between the colonoscopy video processor and the model, which ran the AI model during video processing (Fig. 4**). We used the same hardware for training and inference. In order to have sufficient execution power for real-time detection in a clinical environment, we chose a hardware that consumes little energy and that would be easily transportable in an operating room. This AI model has the ability to delineate the shape of the polyp with superimposed green fluorescence rather than standard classical bounding. An expert, who had been trained to read RTPoDeMo, retrospectively reviewed all the anonymized recorded colonoscopy videos (Fig. 4***). He assigned each polyp appropriately, determined the number of detected polyps, annotated them in each frame and numbered the frames where the polyps were detected.

Metrics and statistical analysis

Intersection over Union (IoU) measures the overlap between the predicted bounding box or mask Mp over the ground truth bounding box or mask Mgt based on Jaccard Index:

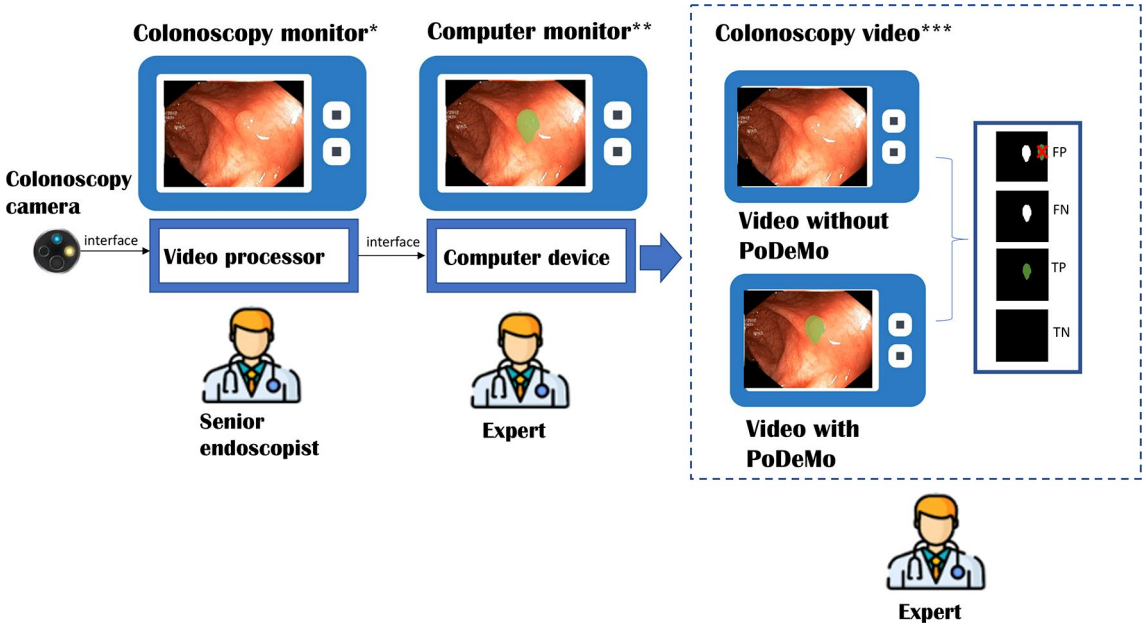

Fig. 4 *Colonoscopy procedure. **Recording of the colonoscopy video by the expert during the procedure with a computer monitor connected to the video processor of the colonoscopy monitor and running of RTPoDeMo under real time conditions. ***Review process for the anonymized colonoscopy videos by the expert, the annotation of polyps detected in each frame and the evaluation of frame quality obtained with RTPoDeMo

$$\text{IoU value} = \frac{\text{area } (\text{Mgt} \cap \text{Mp})}{\text{area } (\text{Mgt} \cup \text{Mp})} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \in [0,1]$$

The IoU Threshold (noted as α) is set to determine the confusion matrix by comparing the result with the predicted IoU score. If α > IoU value, the prediction is considered as a false positive (FP). Otherwise, it is considered as a true positive (TP).

We evaluated model performance using multiple IoU thresholds ranging from 0.50 to 0.95, in steps of 0.05, following the standard COCO evaluation protocol [44]. This approach allows for a comprehensive and balanced assessment of segmentation performance, capturing both lenient and strict overlap scenarios.

The AP was calculated at different IoU thresholds (0.50 to 0.95) and mAP (mean Average Precision) was computed as the average of APs across these IoU thresholds

$$\text{AP} = \sum{}_{\text{i}} \left(\text{Ri} - \text{Ri} - 1\right) .\text{Pinterp } (\text{Ri})$$

Where R_i and R_{i-1} are recall levels at successive thresholds, and P_{interp} (R_i) is the interpolated precision at R_i

$$\text{mAP} = \frac{1}{\text{N}} \sum {}^{\text{N}}_{\text{i}=1} \text{APi}$$

Where N is the number of classes and APi is the AP for class i.

The polyp detection rates reported following the expert's evaluation with RTPoDeMo (Polyp RTPoDeMo) were compared with the expert's evaluation without AI (Polyp Expert), as the reference. Contingency analysis compared the number of agreements and disagreements on individual video frames.

In the context of video analysis, we defined: (a) a True Positive (TP) as a predicted polyp in a video frame that overlaps with the expert annotation with IoU ≥ 0.5; (b) False Positive (FP) as either: a predicted mask that has IoU < 0.5 with any ground truth annotation, or a prediction in a frame where no polyp was annotated by the expert; (c) False Negative (FN) as a polyp annotated by the expert but missed by the model in the corresponding frame.

The FPS was calculated using the mean inference time:

$$\text{FPS} = \frac{1}{\text{mean inference time}}$$

Per-image accuracy, sensitivity, specificity and precision were calculated for each polyp in each colonoscopy video as follows: per-image accuracy was defined as the number of frames with the correct presence or correct absence of AI detection labels divided by the total number of frames; per-image sensitivity was defined as the number of frames in which a polyp was correctly detected by the model divided by the number of overall image frames with a polyp; and per-image specificity was defined as the number of frames without AI detection labels and without polyps divided by the total number of frames without polyps. The bootstrapped 95% confidence interval and associated p-value were calculated.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{P}} = \text{Recall}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{\text{TN}}{\text{N}}$$

To summarize the balance between sensitivity and precision, F1 Score was calculated:

$$\text{F1 score} = 2^{*}\frac{\text{Precision}^{*}\text{Recall}}{\text{Precision} + \text{Recall}}$$

Furthermore, Cohen's Kappa coefficient assessed the level of agreement between segmentations performed by the model and the expert. It is more robust than simple percent agreement calculation since it takes into account the agreement occurring by chance.

Cohen's Kappa is defined as:

$$\text{k} = \frac{\text{Po} - \text{Pe}}{1 - \text{Pe}}$$

Where Po is the observed agreement among raters and Pe is the expected agreement by chance.

Cohen's Kappa can be interpreted as follows: values between 0 and 0.20 indicate no agreement, 0.21–0.40 slight agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement, and 0.81-1 almost perfect agreement.

Quantitative variables were presented as median, minimum and maximum values, while qualitative variables were presented by size and percentage. All analyses were performed using RStudio Version 1.2.5001.

Ethical and regulatory aspects

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki and was approved by the Guadeloupean University Hospital Research Ethics Committee (May 2021). All participants gave their written informed consent to permit the recording and review of their colonoscopy videos and consent to participate to the study. Clinical trial number: not applicable.

Results

The validated model was achieved with the best mAP score for bounding IoUs ranging from 0.5 to 0.95 and FPS > 30 (Fig. 2). It consisted of the YOLACT method plus the ResNet-50 backbone, with FPS = 32.82, image size = 550, mAP = 72.32 and APs associated with IoUs of 50, 75, and 95% (Table 2). The validated model (YOLACT, ResNet-50), named "RTPoDeMo", on a parallel architecture computer (NVIDIA Tesla P100 GPU and Intel Xeon 2.20 GHz CPU) was then connected to the endoscopy processor to enable the visualization of polyp delineation during the colonoscopy procedure (Fig. 4).

Nineteen colonoscopies were performed between 1st June and 9 July 2021, one colonoscopy had been excluded because carried out in the context of lower gastrointestinal bleeding and 18 unaltered colonoscopy videos were analyzed. During the 18 colonoscopy procedures, the senior gastroenterologists identified 29 polyps. The median video duration was 37 min, with a minimum duration of 7 min and a maximum of 66 min.

From the 18 colonoscopy videos, 36 polyps were identified by the expert with or without RTPoDeMo (Table 3). When the model detected a polyp, a precise outlining appeared on the screen around the polyp (Fig. 5). On four colonoscopy videos, lasting less than 12 min, the expert found no polyp; on the other 14 videos, he detected one to nine polyps. As shown in Fig. 5, polyps of different sizes and at different segments of the colon were precisely delineated using RTPoDeMo. They included 29 polyps, which had been identified by the senior endoscopists during the colonoscopy procedures and six additional polyps missed by them. They contained 18 polyps less than 1 cm and 11 more than 1 cm. Among them: 7 pedunculated and 6 sessile Paris I, 9 flat Paris II and 7 ulcerated Paris III. As reported in Table 3, one additional polyp was falsely detected by the expert review with RTPoDeMo in video number 13, in which no polyp was found by the expert. This falsely detected polyp was identified in 18 frames of the RTPoDeMo overlaid colonoscopy video. One other polyp was missed by the expert review with RTPoDeMo in colonoscopy video number 10, which found nine polyps. This missed polyp was identified in 177 frames. We present different situations in which the model was able to delineate the polyp despite a partial view of the digestive wall, as the presence of bubbles in front of the polyp, a polyp in a fold or of very small size (Fig. 6). The accuracy of RTPoDeMo was calculated from FP, FN, TP and TN frames and is reported in Table 3. The per-image accuracy was 99.59% (95% CI [99.45 – 99.71%]), per-image sensitivity reached 90.63% (95% CI [78.95 – 93.64%]), and per-image specificity 99.95% (95% CI [99.93 – 99.97%]). The average number of frames with FP per video was 19 frames, with a minimum of 0 frames and a maximum of 33 frames. The average number of frames with FN per video was 227 frames, with a minimum of 0 frames and a maximum of 1036 frames. The Cohen's Kappa coefficient between RTPoDeMo and the expert was 0.72 (95% CI [0.54–1.00], $p < 0.0001$), which indicated good agreement regarding 95% of the observations. The model achieved a F1-score of 0.94 (95% CI: [0.87–0.98]).

RTPoDeMo and the expert only disagreed on two polyps: one missed, and one falsely detected by the model (Table 3).

Discussion

This study evaluated the performance of YOLACT derived RTPoDeMo designed for real-time application on prospectively recorded colonoscopy videos. We have reported here all the steps that led to the selection and validation of the model, including its design, the preparation of medical images, the training process and the benchmarking of each model and evaluation of their performances. We have demonstrated that RTPoDeMo deep-learning model was able to perform, real-time polyp detection and instance segmentation, but also the delineation of colorectal polyps by means of green fluorescence. This preliminary study, conducted on 18 prospectively recorded colonoscopy videos, achieved high levels of per-image specificity, sensitivity and accuracy, and good agreement with the expert's review.

Concerning the delineation of polyps, various segmentation techniques are available. These encompass approaches such as considering the entire scene (semantic segmentation), identifying individual objects (instance segmentation), amalgamating both strategies (panoptic segmentation), relying on object borders (boundary-based segmentation), or segmenting the image into sections (hierarchical segmentation). In our study, we used instance segmentation, which is the more precise approach. RTPoDeMo has the ability to delineate the shape of the polyp with superimposed green fluorescence rather than standard classical bounding. This could be very useful for gastroenterologists in order to predict the limits of the lesions and assess the conditions for a complete resection. Some authors presented a 2D delineation of colorectal polyps, but the precision of delineation did not achieve that seen with our model [48].

Table 2 Mean average precision and average precision of several IoU thresholds in the validated model

Model	mAP	AP50	AP55	AP60	AP65	AP70	AP75	AP80	AP85	AP90	AP95
ResNet50_550	71.23	90.04	89.94	89.94	88.31	86.83	85.50	75.69	62.56	37.01	6.51

Abbreviations: mAP, mean average precision; AP, average precision

Table 3 Performances of RTPoDeMo on 18 unaltered colonoscopy videos

Colonoscopy number	Video Duration (minutes)	Polyp* Expert *N* = 36	Polyp** RTPoDeMo *N* = 36	TP (*n*)	FP (*n*)	FN (*n*)	TN (*n*)	Accuracy (%)	Sensitivity per image (%)	Specificity per image (%)
1	24	1	1	2210	24	245	40,721	99.38%	90.02%	99.94%
1	24	2	2	3588	24	189	39,399	99.51%	95%	99.94%
2	9	1	1	409	33	36	15,722	99.57%	91.91%	99.79%
3	15	1	1	2593	16	80	24,311	99.64%	97.01%	99.93%
4	33	1	1	1748	31	112	57,509	99.76%	93.98%	99.95%
4	33	2	2	3013	28	30	56,329	99.9%	99.01%	99.95%
4	33	3	3	2345	24	204	56,827	99.62%	92%	99.96%
5	29	1	1	4935	22	548	46,695	98.91%	90.01%	99.95%
5	29	2	2	2974	25	124	49,077	99.71%	96%	99.95%
5	29	3	3	439	19	60	51,682	99.85%	87.98%	99.96%
6	12	0	0	0	0	0	21,600	100%	-	100%
7	17	1	1	2886	18	59	27,637	99.75%	98%	99.93%
8	41	1	1	3001	28	297	70,474	99.56%	90.99%	99.96%
8	41	2	2	239	22	26	73,513	99.93%	90.19%	99.97%
8	41	3	3	4087	30	308	69,375	99.54%	92.99%	99.96%
9	7	1	1	324	32	36	12,208	99.46%	90%	99.74%
10	58	1	1	7017	20	292	97,071	99.7%	96%	99.98%
10	58	2	2	3047	10	229	101,114	99.77%	93.01%	99.99%
10	58	3	3	3800	11	78	100,511	99.91%	97.99%	99.99%
10	58	4	0	0	0	177	104,223	99.83%	0%	100%
10	58	5	4	1915	19	39	102,427	99.94%	98%	99.98%
10	58	6	5	2466	10	274	101,650	99.73%	90%	99.99%
10	58	7	6	4618	16	192	99,574	99.8%	96.01%	99.98%
10	58	8	7	376	12	56	103,956	99.93%	87.04%	99.99%
10	58	9	8	9325	14	1036	94,025	98.99%	90%	99.99%
11	8	0	0	0	0	0	14,400	100%	-	100%
12	22	1	1	5132	20	446	34,002	98.82%	92%	99.94%
13	9	0	1	0	18	0	16,182	99.89%	-	99.89%
14	51	1	1	5573	31	420	85,776	99.51%	92.99%	99.96%
14	51	2	2	3751	24	239	87,786	99.71%	94.01%	99.97%
14	51	3	3	6261	33	128	85,378	99.82%	98%	99.96%
14	51	4	4	8224	30	715	82,831	99.19%	92%	99.96%
15	19	1	1	8032	11	993	25,164	97.06%	89%	99.96%
16	27	1	1	4434	24	386	43,756	99.16%	91.99%	99.95%
16	27	2	2	1708	22	35	46,835	99.88%	97.99%	99.95%
17	66	1	1	4039	16	351	114,394	99.69%	92%	99.99%
17	66	2	2	6797	19	512	111,472	99.55%	92.99%	99.98%
17	66	3	3	2251	13	94	116,442	99.91%	95.99%	99.99%
17	66	4	4	208	16	21	118,555	99.97%	90.83%	99.99%
18	10	0	0	0	0	0	18,000	100%	-	100%

Abbreviations: AI, artificial intelligence; N: number of polyps; n, number of frames; TP: true positive frames detected with RTPoDeMo; FP: false positive frames detected; TN: true negative frames detected; FN: false negative frames detected; *polyp numbered by the expert, **polyp numbered with RTPoDeMo

The choice of the YOLACT architecture with ResNet-50 for the detection and delineation of polyps using instance segmentation techniques can be justified by several key factors. Bolya et al. [27] have shown that YOLACT and YOLACT + + had real-time performances in order to perform object detection but with a less efficient mAP delimitation than Mask-RCNN which has not real-time abilities. Although YOLACT is less accurate than YOLACT + + in delineating areas of interest, with a mAP lower by 2.15, it compensates by offering significantly higher speed, with a 9.19 FPS advantage, reaching 32.83 FPS against 23.63 FPS for YOLACT++. Our study supports Bolya's findings that YOLACT outperforms YOLACT + + in terms of frame rate (evaluated by FPS) but lags behind in terms of accuracy (IoU and mAP).

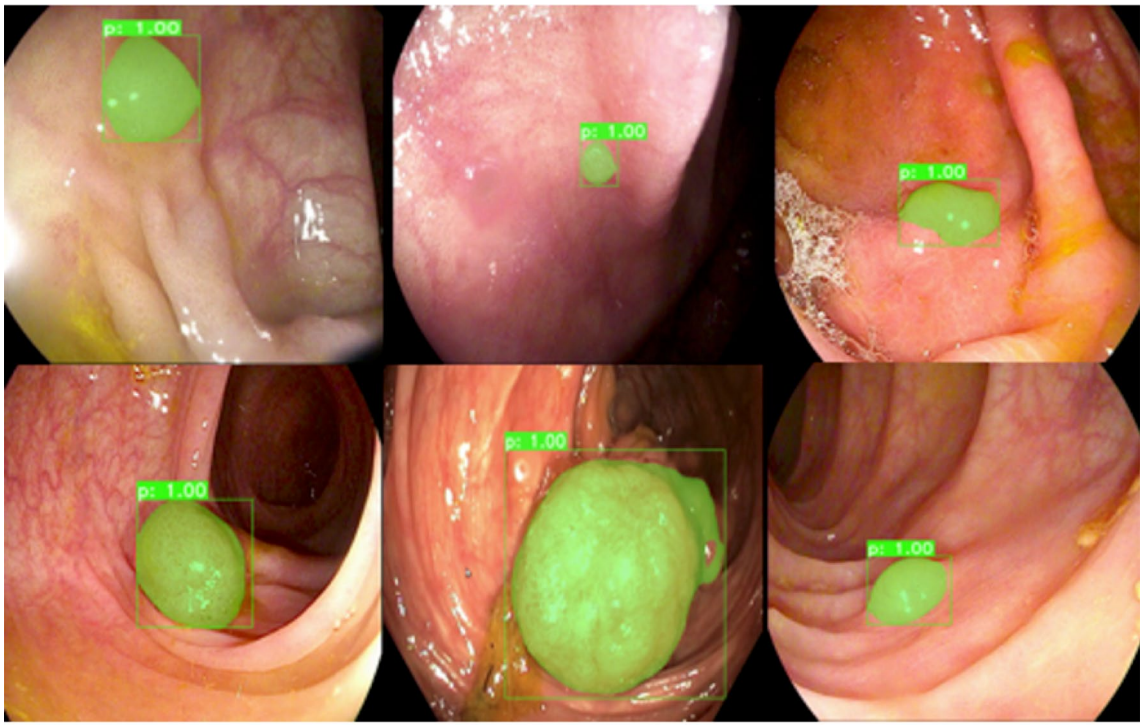

Fig. 5 Delineation of colorectal polyps of different sizes with RTPoDeMo

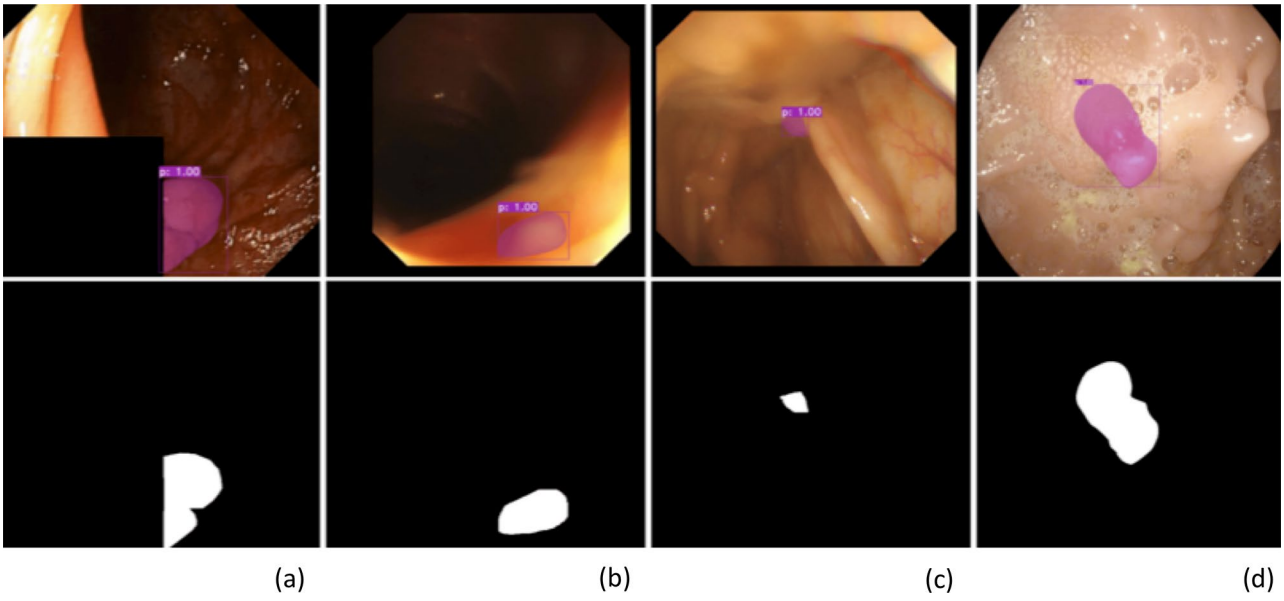

Fig. 6 RTPoDeMo was able to delineate the polyp despite (**a**) partial view of the digestive wall, (**b**) polyp in a fold, (**c**) polyp of very small size, (**d**) presence of bubbles in front of the polyp

The backbone ResNet-50 was chosen for its optimal balance between high-quality feature extraction and real-time processing abilities. ResNet-50 provides superior delineation accuracy and faster frame rates than ResNet-101 and DarkNet-53, making it ideal for clinical applications requiring precise and efficient real-time segmentation. Overall, the combination of YOLACT with ResNet-50 achieves high delineation accuracy while maintaining a high frame rate, ensuring efficient and accurate real-time analysis. It offers a practical and efficient solution for the real-time detection and delineation of polyps, providing a valuable tool for enhancing the accuracy and speed of colorectal cancer screening procedures. However, unlike Bolya's results

using non-medical images (550 pixels) with a ResNet50 backbone [27], our analysis of medical polyp videos shows that YOLACT + + performs slightly below the real-time FPS threshold. In contrast, YOLACT consistently maintains a frame rate above 30 FPS, enabling real-time processing for polyp instance segmentation in these videos.

The training polyp dataset had previously been resized (400, 550 and 700 pixels) with appropriate image dimensions to reduce processing power, implement the process in real time conditions and prevent any time variations. Because we had not solved the overfitting problem on a relatively small data set, we chose not to increase the number of images but using data augmentation techniques that involved digital image processing, such as rotation and random horizontal and vertical flips. We applied these techniques to ensure that the shape of the polyps remained unchanged. Introducing artificial variations of the images through data augmentation improved the model's ability to generalize, thereby mitigating the risk of overfitting and data leakage. In order to move the dataset from semantic segmentation to instance segmentation COCO annotations, we have chosen to implement these data in the model, a decision that was influenced by the fact that a single image could encompass more than one polyp.

In order to assess the performance of the model, we used specialized computer vision metrics, which offer valuable insights for medical professionals. The confusion matrix is a visual representation that effectively categorizes predictions and actual outcomes, facilitating the calculation of sensitivity, specificity and accuracy scores. The IoU, AP, and mAP are inherently linked with the confusion matrix, as they draw upon the information it encapsulates. Per-image sensitivity and specificity are particularly relevant in a medical context, assessing the effectiveness of a model to detect TP and TN frames, and per-image accuracy scores quantify a model's precision in segmenting individual polyps. To evaluate the quality of predicted masks in polyp localization, specificity, sensitivity, precision and F1-score are commonly used metrics. Additionally, the Area Under the Curve - Receiver Operating Characteristic curve is computed for an overall score. However, these metrics are not well suited for evaluating localization performance. By setting an IoU threshold, predictions are classified as true or false. Sensitivity can be defined for localization but defining TN consistently remains problematic. Some studies define specificity on an image basis, but this approach varies and can be inconsistent. Therefore, mAP is preferred, as it provides a more consistent evaluation of segmentation performance by averaging over multiple IoU thresholds. We chose IoU as the primary metric because it directly measures both the localization and the contour accuracy of segmentation, unlike simpler metrics such as pixel wise accuracy, which can be biased by class imbalance. The multi-threshold evaluation is more informative than a single cut-off, as it reflects model robustness under various levels of localization precision. IoU is the standard in instance segmentation tasks and offers more clinically meaningful insights for our specific objective, precise polyp delineation to support therapeutic resection. Our model achieved high F1-Score, confirming its ability to detect true polyps while minimizing false positives. This high F1-score reflects the model's strong clinical relevance as a real-time assistive tool, maintaining both reliability and precision in polyp detection.

Cohen's Kappa coefficient provides insight into the level of agreement between algorithmic segmentations and those performed by the expert and highlights the reliability of a model's results. The weighted Cohen's Kappa coefficient was computed at the polyp level rather than per frame. This choice was guided by the clinical objective of our study, which is not to evaluate pixel-level segmentation or frame-by-frame consistency, but to assess whether the AI system can reliably detect the presence of a polyp during a colonoscopy. In clinical practice, what matters is whether the system is able to trigger an alert to assist the endoscopist in real time. Therefore, evaluating the agreement between the AI and the expert on a per-polyp basis provides a more meaningful measure of diagnostic performance. This approach also avoids over-representing agreement on non-informative frames and allows for a direct interpretation of the model's relevance in supporting polyp detection during endoscopic procedures.

During our study, the expert review with RTPoDeMo outperformed the senior endoscopists by identifying six additional polyps that had been missed during the colonoscopies. The performances of RTPoDeMo achieved in our study were close to those currently being reported using other commercially available AI algorithms for polyp detection [20, 23]. In addition, the average number of FP frames in the colonoscopy videos (19 frames) was similar to that reported during previous studies [17]. RTPoDeMo falsely detected one polyp in a video without polyp, but this detection only involved 18 frames out of 16,200, which is negligible. False positive frames were essentially related to endoscopic features such as bubbles, poor bowel preparation or rapid movements during the exploration. Diminishing the size of the detection window would enable a reduction in the number of FP frames in the future.

The medical benefits of detecting additional polyps using RTPoDeMo should be tempered by the fact that most of them were detected in the left colon and were quite small (median size = 3 mm). However, other authors have also reported more very small adenomas

being detected with AI without a significant statistical difference regarding larger adenomas [20, 22, 49, 50], or higher resection rate of adenomas. Indeed, Areia et al. [51 recently reported that a CRC screening strategy involving AI could reduce its incidence by 6% and its mortality by 5%, thus rendering it cost-effective. In our study, we could not assess the ADR, a prospective control study would be appropriate for this. Moreover, new challenges consist in incorporating the real-time optical characterization of polyps during colonoscopy into a deep-learning model that enables an accurate differentiation between adenomatous and non-adenomatous polyps [52, 54]. This is clinically relevant to select optimal treatment, which could be associated with the volume of the lesion in future studies.

Moreover, like the French Caribbean region of Guadeloupe, many countries worldwide suffer from a lack of gastroenterologists and poor compliance with CRC screening using fecal immunological tests [2]. Consequently, enhancing the performance of gastroenterologists in preventing CRC through the use of innovative endoscopic tools would be of considerable value and clinically relevant in a CRC screening program [55]. The high accuracy of this model and the very good agreement between the expert and the model offer additional reasons to imagine how such a model could also assist inexperienced endoscopists in achieving higher ADR.

Conclusions

We found high performance of RTPoDeMo with a novel approach of multi-instance segmentation, real-time polyp detection and delineation. Its feasibility in clinical practice is supported by its short processing time, so that it could be easily employed during colonoscopy procedures with negligible latency and without slowing exploration of the colon. This model displayed clinical applicability and simplicity of use. The innovative value of our system relies on a precise delineation of a polyp, which could provide more precision about the characteristics of the polyp. This will be more accurate to evaluate the colorectal polyps before resection with a lower risk of missing a part of the lesion. These potential values in recognizing the shape and the pathology of colorectal lesions should be improved in further studies. We hope to confirm the good results obtained with this model in a future experiment on our regional image databases collected as part of a health data warehouse, in order to evaluate the model's performance in specific populations.

Abbreviations

RTPoDeMo	Real-time Polyp Delineation Model: CRC, colorectal cancer
ADR	Adenoma detection rate
AI	Artificial intelligence
CNN	Convolutional neural network
GPU	Graphics processing unit
CPU	Computer processing unit
FLOPS	Floating-point operations per second
GRAD-CAM	Gradient weighted Class Activation Mapping
mAP	Mean average precision
IoU	Intersection over the union
AP	Average precision
FPS	Frame per second
FPN	Feature Pyramid Network
NMS	Non-Maximum Suppression
FP	False positive
TP	True positive
TN	True negative
FN	False negative

Acknowledgements

We address a particular thank to Walid Bousselham (National University of Singapore) for his technical help on the deep-learning model used for this study and Younes El Hazibi for his involvement in the project through his medical thesis.

Author contributions

MG-S designed the study, obtained technical and material support, analyzed and interpreted the data and drafted the manuscript. JS analyzed the colonoscopy videos, interpreted the data, supervised the study and drafted the manuscript. MA, MS and GS-G performed the colonoscopies and critically revised the manuscript for important intellectual content. AM and YM analyzed and interpreted the data. All authors read and agreed the final version of the manuscript.

Funding

Not applicable.

Data availability

Data are provided in the manuscript.

Declarations

Ethics approval and consent to participate

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki and was approved by the Guadeloupean University Hospital Research Ethics Committee (May 2021). All participants gave their written informed consent to participate to the study.

Consent for publication

All participants gave their written informed consent to permit the recording and review of their colonoscopy videos.

Relevant guidelines and regulations

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 December 2024 / Accepted: 22 May 2025
Published online: 04 June 2025

References

1. Baidoun F, et al. Colorectal Cancer epidemiology: recent trends and impact on outcomes. CurrDrugTargets. 2021;22:998–1009.
2. Menvielle G, Dugas J, Richard J-B, Luce D. Socioeconomic and healthcare use-related determinants of cervical, breast and colorectal cancer screening practice in the French West Indies. Eur J Cancer Prev. 2018;27:269–73.
3. Kelly DM, Estaquio C, Léon C, Arwidson P, Hermann N. Temporal trend in socioeconomic inequalities in the uptake of cancer screening programmes in France between 2005 and 2010: results from the Cancer barometer surveys. BMJ Open. 2017;7:e016941.
4. Joachim C. al. Trends in colorectal cancer in the Caribbean: A population-based study in Martinique, 1982–2011. Rev Epidemiol Sante Publique. 2017:65181–188.

5. Kahi CJ, Hewett DG, Norton DL, Georges JE, Douglas KR. Prevalence and variable detection of proximal colon serrated polyps during screening colonoscopy. Clin Gastroenterol Hepatol. 2011:942–46.
6. Doubeni CA. al. Screening colonoscopy and risk for incident late-stage colorectal cancer diagnosis in average-risk adults: a nested case-control study. Ann Intern Med. 2013:158; 312–320.
7. Baxter NN, Warren JL, Barrett MJ, Stukel TA, Doria-Rose VP. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. J Clin Oncol. 2012:302664–2669.
8. Kaminski MF et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med. 2010: 362;1795–1803.
9. Corley DA, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med. 2014;370:1298–306.
10. Schiff GD, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009;169:1881–7.
11. Langlotz CP, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The academy workshop. Radiology. 2019;291:781–91.
12. Deng J, Dong W, Socher R, Li L-J, Kai L, Fei-Fei L, ImageNet. A large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition. 2009:248–255 https://doi.org/10.1109/CVPR.2009.5206848
13. Russakovsky O, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis. 2015;115:211–52.
14. Misawa M. al. Artificial Intelligence-Assisted polyp detection for colonoscopy: initial experience. Gastroenterology 2018:154;2027–2029.e3.
15. Urban G, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology. 2018;155:1069–78.
16. Wittenberg T, Zobel P, Rathke M, Mühldorfer S. Computer aided detection of polyps in Whitelight- colonoscopy images using deep neural networks. Curr Dir Biomed Eng. 2019;5:231–4.
17. Lee JY et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. Sci Rep. 2020:108379.
18. Park SY, Sargent D. Colonoscopic polyp detection using convolutional neural networks. Proc. SPIE 9785, Medical Imaging. Computer-Aided Diagnosis.2016; https://doi.org/10.1117/12.2217148
19. Wang P, Xiao X, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng. 2018;2:741–8.
20. Repici A et al. Efficacy of Real-Time Computer-Aided detection of colorectal Neopplasia in a randomized trial. Gastroenterology 2020:159;512–520.e7.
21. Repici A, et al. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. Gut. 2022;71:757–65.
22. Wang P, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut. 2019;68:1813–9.
23. Xu H et al. Artificial Intelligence-Assisted colonoscopy for colorectal Cancer screening: A multicenter randomized controlled trial. Clin Gastroenterol Hepatol. 2023:1337–e3463.
24. Ren S, He K, Girshick R, Sun Faster J. R-CNN: Towards Real-Time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2017: 39;1137–1149.
25. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. IEEE Trans Pattern Anal Mach Intell. 2020;42:386–397.
26. Bolya D, Zhou C, Xiao F, Lee YJ. Yolact: Real-time instance segmentation. IEEE Int Conf Comput Vis (ICCV). 2019. https://doi.org/10.1109/ICCV.2019.00925.
27. Bolya D, Zhou C, Xiao F, Lee YJ. YOLACT++Better Real-Time instance segmentation. IEEE Trans Pattern Anal Mach Intell. 2022;44:1108–21.
28. Angeles Ceron JC, Chang L, Ruiz GO, Ali S. Assessing YOLACT++for real time and robust instance segmentation of medical instruments in endoscopic procedures. Annu Int Conf IEEE Eng Med Biol Soc. 2021;1824–7. https://doi.org/10.1109/EMBC46164.2021.9629914.
29. Nogueira-Rodríguez A, Domínguez-Carbajales R, Campos-Tato F, Herrero J, Puga M, Remedios D, Rivas L, Sánchez E, Iglesias A, Cubiella J, et al. Real-time polyp detection model using convolutional neural networks. Neural Comput Appl. 2022;34:10375–96.
30. Pacal I, Karaboga D. A robust real-time deep learning based automatic polyp detection system. Comput Biol Med. 2021;134:104519. https://doi.org/10.1016/j.compbiomed.2021.104519.
31. Ghose P, Ghose A, Sadhukhan D et al. Improved polyp detection from colonoscopy images using finetuned YOLO-v5. Multimed Tools. 2024:8342929–42954.
32. Sahafi A, Koulaouzidis A, Lalinia M. Polypoid lesion segmentation using YOLO-V8 network in wireless video capsule endoscopy images. Diagnostics (Basel). 2024;22:474.
33. Gong D, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomized controlled study. Lancet Gastroenterol Hepatol. 2020;5:352–61.
34. Luo Y, Wang Y, Zhao Y, Guan W, Shi H, Fu C, Jiang H. A lightweight network based on dual-stream feature fusion and dual-domain attention for white blood cell segmentation. Front Oncol. 2023;13:1223353.
35. Ali S, et al. Deep learning for detection and segmentation of artefact and disease instances in Gastrointestinal endoscopy. Med Image Anal. 2021;70. https://doi.org/10.1016/j.media.2021.102002.
36. Deng J, Dong W, Socher R, J Li L, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. Inst Electr Electron Eng (IEEE). 2010;248–55. https://doi.org/10.1109/cvpr.2009.5206848.
37. He K, Gkioxari G, Dollár P et al. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017:2961–2969.
38. Sahafi A, Koulaouzidis A, Lalinia M. Polypoid Lesion Segmentation Using YOLO-V8 Network in Wireless Video Capsule Endoscopy Images. Diagnostics (Basel). 2024 22;14(5):474.
39. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. IEEE International Conference on Computer Vision (ICCV). 2017:618–626. https://doi.org/10.1109/ICCV.2017.74
40. Pogorelov C, Randel KR, Griwodz C, De Lange T. Kvasir: A multi-class image dataset for computer aided Gastrointestinal disease detection. Proc MMSys'17. 2017;164–9. https://doi.org/10.1145/3083187.3083212.
41. Bernal J, Sánchez J, Fernández-Esparrach G, Gil D, Rodriguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput Med Imaging Graph. 2015;43:99–111.
42. Bernal J, Sánchez J, Vilariño F. Towards automatic polyp detection with a polyp appearance model. Pattern Recogn. 2012;45:3166–82.
43. Silva J, Histace A, Romain O, Dray X, Granado B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. Int J Comput Assist Radiol Surg. 2014;9:283–93.
44. Lin TY et al. Microsoft COCO: common objects in context. Lect Notes Comput Sci. 2014: 8693740–755.
45. Satoshi Suzuki, Keiichi A, Be. Topological structural analysis of digitized binary images by border following. Comput Vis Graphics Image Process. 1985:3032–46.
46. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10). 2010:807–814.
47. The Paris endoscopic. Classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc. 2003;58(6 Suppl):S3–43.
48. Qadir HA, Shin Y, Solhusvik J, Bergsland J, Aabakken L, Balasingham I. Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction. Med Image Anal 2021:68101897.
49. Wallace MB et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. Gastroenterology. 2022:163;295–304.e5.
50. Quan SY et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. Sci Rep. 2022:216598.
51. Areia M et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. Lancet Digit Health. 2022: e4436–444.
52. Biffi C et al. A novel AI device for real-time optical characterization of colorectal polyps. NPJ Digit Med. 2022: 5;84.
53. Neumann H, Kreft A, Sivanathan V, Rahman F, Galle PR. Evaluation of novel LCI CAD EYE system for realtime detection of colon polyps. PLoS ONE. 2021:16e0255955.
54. Kliegis L, Obst W, Bruns J, Weigt J. Can a polyp detection and characterization system predict complete resection?? Dig dis. 2022: 40;115–118.

55. Shaukat A et al. Computer-Aided detection improves adenomas per colonoscopy for screening and surveillance colonoscopy: A randomized trial. Gastroenterology. 2022: 163732–741.

Publisher's note