

DETECTION AND SEGMENTATION OF COLORECTAL POLYPS IN GASTROINTESTINAL COLONOSCOPY IMAGES USING MASK R-CNN FRAMEWORK

Ashok Bekkanti

Research Scholar, Department of CSE, Annamalai University, Tamil Nadu, India

Sumathi Ganesan

Assistant Professor, Department of CSE, Annamalai University, Tamil Nadu, India

Narayana Satyala

Professor, Department of CSE, Gudlavalleru Engineering College, Andhra Pradesh, India

Abstract

Computer-aided polyp segmentation is an important activity that assists gastroenterologists in evaluating and removing abnormal tissue from gastrointestinal system. Disease polyps arise mostly in colorectal area of gastrointestinal system and in mucous membrane, that includes micro-abnormal tissue protrusions that raise risk of incurable illnesses such as cancer. Hence, early evaluation of polyps can reduce likelihood of polyps developing into cancer, such as adenomas, that can develop into cancer. Deep learning-based diagnostic tools are critical for early illness detection. To segregate polyps using colonoscopy frames, a deep learning approach known as Mask R-CNN is presented. Mask R-CNN having RPN-CNN backbone is a modified version of standard F-CNN having 3 stages: feature extraction, ROI pooling, and segmentation for severity analysis.

Mask R-CNN technique outperforms previous deep learning techniques in segmentation. trials utilized a single dataset that is obtained from big intestine of gastrointestinal system via a colonoscopy technique. predicted technique exceeds having a mean Dice of 98 percentage, accuracy of 96.6 percentage, precision of 97 percentage, recall of 96.6 percentage, and an F1-score of 92 percentage.

Keywords: Segmentation, Colorectal Polyps, Feature Extraction, Neural Network, Pooling Process

1. Introduction

Ulcerative colitis as well as Crohn disease are 2 chronic systemic inflammatory illnesses known collectively as IBD [1]. They are cauterized by an incorrect immune response to commensally bacteria in genetically vulnerable people. Ulcerative colitis is incurable, necessitates lifetime medical treatment, and develops from frequent to full digestion disorder [2]. Endoscopy is critical in making an initial diagnosis, monitoring disease sizes or activity, assessing disease consequences, providing cancer surveillance, and creating a firm endpoint in clinical trials testing novel therapies [3].

Therapeutic strategy has shifted to pursuing integrated hard objectives (like clinical and endoscopic remission). Mucosal repair has been linked to better long-term results. Endoscopic grading methods for ulcerative colitis, on or hand, are varied and subjective, having significant inter- and intra-observer variability, and are not commonly employed in clinical practise [4]. Even in randomised controlled studies, is substantial diversity in use and interpretation. Scoring standardisation by impartial remote central reading is an ideal option, but it is not possible in daily clinical practise [5]. Developments in artificial intelligence approaches are increasingly being applied to automate picture processing in medical domains. Deep learning algorithms for identifying melanoma, diabetic retinopathy, polyps, and or disorders from still photos are recent examples [6].

Application of comparable technologies to pictures collected from movies, such as those produced during colonoscopy, is still in its early stages of development. Still pictures, on or hand, have potential to be beneficial in illnesses such as UC since y might provide an accurate, widely accessible, and low-cost resource for study and applications [7]. We studied viability of using deep learning algorithms to rate severity of UC endoscopic and applied m to full-motion video footage of colonoscopies [8]. use of mask R-CNN is significant in numerous deep learning methods like neural networks having convolutional, recurrent neural networks having gated recurrent modules, alexnet, and inception technique because discriminator is challenging due to correct identification of all objects in an picture even while precisely segregating every instance. It thus incorporates elements from traditional computer vision duties of object detection, in that goal is to categorise individual objects and localise every to use a bounding box, and semantic segmentation, in that primary objective is to classify every pixel into a fixed category having out distinguishing object instances¹. Considering this, one could assume a sophisticated procedure to be necessary to provide satisfactory results. [9] Current strategies that penalise technique coefficients to assure that they do not over fit to training data were utilized to prevent overfitting and to strive for simpler techniques having interpretable coefficients. Efficient Net tries to differentiate patients in training set who had unfavourable health result from those who did not by as wide a margin as feasible. Every tree is trained and assessed on random subsets of data generated having replacement to capture subtleties in data. Furthermore, mask branch adds just a minor computational burden, allowing for a quick system and rapid experimentation. In light of this, following work contributions are made:

- Using manually labelled pictures from films of real endoscopic colonoscopies done at a single centre, we trained Mask R-CNN on a dataset and fine-tuned system.
- Mask R-CNN having RPN-CNN backbone may extract colonoscopy picture features automatically than manually having substantial pre-processing.
- Suggested system can detect and segment adenomatous polyps from colonoscopy pictures, and it can do so on a configurable scale.

Following are remaining portions of this chapter: Part 2 discusses a few past research studies, Section 3 demonstrates proposed technique and approach, Section 4 contains experiment findings and discussion, and Section 5 finishes having a conclusion and future study proposals.

2. Related Works

Based on picture textures, several writers have quantified and graded severity of UC in optical colonoscopy video frames. Others presented a feature extraction approach based on accumulation of pixel differences in values, that gave greater accuracy and speed for 'severe' and 'moderate' classes than current methods. Following are some similar works:

In [10] provides an intelligent technique to diagnose alimentary canal disorders by employing EWT and CNN. EWT aids in decomposition of pictures into several modes and extraction of specific patterns in pictures. se deconstructed pictures are n sent into proposed deep CNN for 2-level illness categorisation. result reveals 96.65 percentage accuracy, 0.9298 MCC in 1st level, and 94.25 percentage accuracy, 0.8108 MCC in 2nd level of classification. Transfer-learning techniques using several pre-trained techniques are proposed in [11] to identify and categorise digestive illnesses.

EfficientNetB0 obtained highest accuracy, precision, and recall scores of 98.01 percentage, 98 percentage, and 98 percentage, respectively. GastroEffNetV1 is a deep learning network introduced in [12] for automated categorisation of anomalies in capsule endoscopy pictures. classifier has a validation accuracy of 99.15 percentage, a validation loss of 0.0918, a specificity of 99.25 percentage, a sensitivity of 99.25 percentage, and an AUC of 0.991. In [13], a deep learning-based hybrid trying to stack ensemble method for identifying and characterising gastrointestinal system abnormalities is presented, with the goal of early diagnosis with high accuracy and sensitive measurements while reducing workload to support specialists and objectivity in endoscopic diagnosis. Stacking ensemble approaches worked well, with 98.42 percent ACC and 98.19 percent MCC in the KvasirV2 dataset and 98.53 percent ACC and 98.39 percent MCC in the HyperKvasir dataset.

In evaluation, [14] Current strategies that penalise technique coefficients to assure that they do not over fit to training data were utilized to prevent overfitting and to strive for simpler techniques having interpretable coefficients. Efficient Net tries to differentiate patients in training set who had unfavourable health result from those who did not by as wide a margin as feasible. Every tree is trained and assessed on random subsets of data generated having replacement to capture subtleties in data, total precision is 88.91 percentage. [16] introduce UC-NfNet, an automated UC classification approach, together having a syntic data production pipeline for categorising colonoscopy UC pictures.

EfficientNet architecture is utilized to train a deep learning network in [17], and test results were provided in article. Research have demonstrated that gastrological pictures may be correctly categorised having an accuracy of 0.935. [18] developed and validated a 101-layer CNN technique using 90 percentage of data, leaving 10 percentage for a holdout test set. overall accuracy rate is 77.2 percentage. Among MES one, two, and three, average specificity is 85.7 percentage, sensitivity is 72.4 percentage, PPV is 77.7 percentage, and NPV is 87.0 percentage. [19] offer a method for classifying UC severity in colonoscopy movies by recognising vascular (vein) patterns, that are particularly specified in this study as no., of blood vessels in video frames.

We apply CNN and picture pre-processing approaches to recognise se vascular patterns. Experiments reveal that suggested strategy for identifying UC severity by recognising se vascular patterns considerably improves classification performance. CNN techniques were considered for training and testing at 2 distinct classification levels in [20] to recognise its

projected value. experimental result reveals 97.25 percentage and 93.75 percentage of accuracy for initial level and 2nd of classification.

Current strategies that penalise technique coefficients to assure that y do not over fit to training data were utilized to prevent overfitting and to strive for simpler techniques having interpretable coefficients. Efficient Net tries to differentiate patients in training set who had unfavourable health result from those who did not by as wide a margin as feasible. Every tree is trained and assessed on random subsets of data generated having replacement to capture subtleties in data. To prevent having too many coupled networks that chose same best predictors, just a subset of predictors is considered at every split in layer. Mask R-CNN is meant to circumvent se problems, as mentioned below.

3. System Technique

Dataset is 1st loaded having endoscopy as well as colonoscopy movies, as seen in figure-1. To eliminate noise, se datasets are pre-processed having Wienmed filter. Pre-processed data is passed into feature extractor, that employs ResNet having Region proposal network (RPN). Lastly, using Mask-RCNN and severity analysis, collected features are segmented.

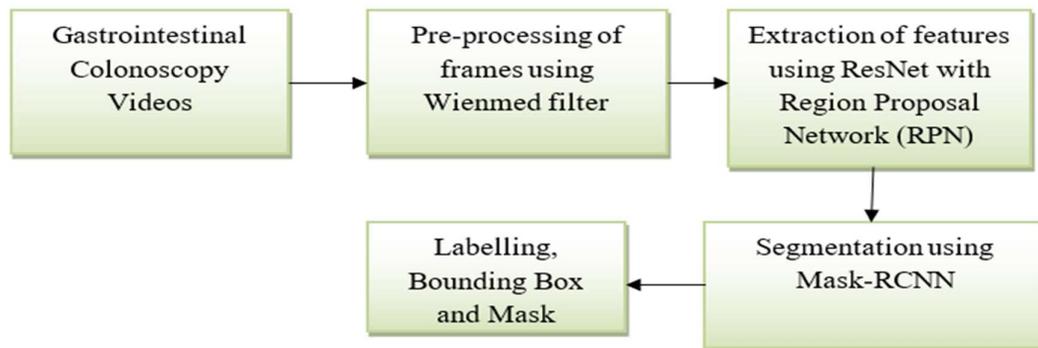


Figure-1 Overall block diagram for proposed segmentation process

- **Video to frames and pre-processing of frames**

Colonoscopy and endoscopy footage to be inspected for suspected polyps has been imported onto computer. footage is n sent into planned automated system. In reality, every video is a running sequence of static pictures known as frames. Nevertheless, not all of original video's sections are significant; there are some useless regions carrying descriptions and or information. Studying such areas is a complete waste of time.

As a result, unwanted areas of original video frame are eliminated, resulting in frames. Wienmed filter is combined version of 2 filters, Wiener [21] and median filters [22]. These 2 filters work together to effectively decrease noise distribution and frame faults. primary goal of such a filter is to substitute noisy and nearby picture pixels, that were previously organised based on intensity of picture. This pre-processing phase enhances video frame part that contains unwanted distortions or improves a variety of picture properties that are important for further processing. Wienmed filter is applied to dataset, and pre-processed frame is obtained. Equation (1) is utilized to determine mean m in every pixels using PQ.

$$m = \frac{1}{PQ} \sum_{p,q \in A} a(p, q)$$

Here, p, q is indicates pixel dimensions of every picture, a symbolises every picture and A is an picture. Additionally, Gaussian noise variation is provided in Equation (2), σ^2 is variance.

$$\sigma^2 = \frac{1}{PQ} \sum_{p,q \in A} n(p, q) - m^2$$

$$w(p, q) = \sigma^2 [n - a(p, q)]$$

Thus, Wienmed filter remove noise in frames and attained pictures are moved to feature extraction process.

4. Extraction of features using RPN having CNN

Figure-2 shows how CNN having Region proposal network (RPN) collects features from distinct hierarchical levels having varied sizes, such that every level includes knowledge about higher-level and lower-level characteristics.

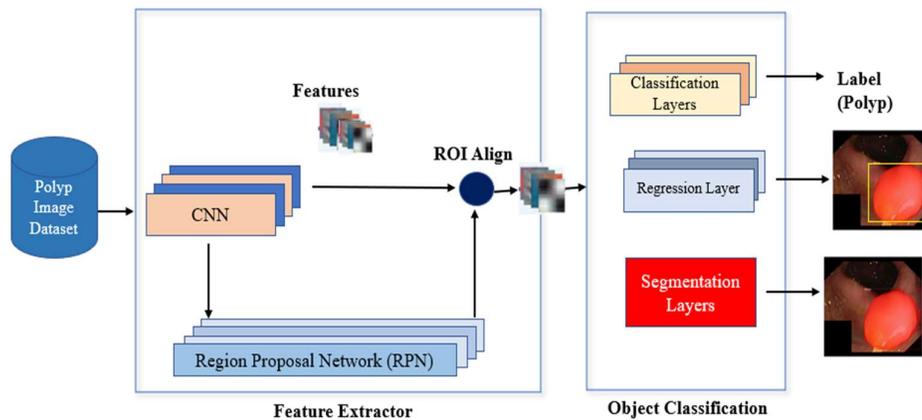


Figure-2 architecture of Region proposal network (RPN) having CNN backbone

First, CNN back-bone is utilized for extraction of feature-map in incoming pictures in sizes $C_0 H_0 W_0$ ($C_0 = 3$ for RGB pictures), as feature derived by CNN shown to be extremely resilient to many types of visually related tasks. Convolutional layer (Conv1) converts a repressor and a classifier head, forecasts of a boundary box array and confident arrays are n obtained for every feature vector ($H_2 W_2$ in total). A repressor head encodes offsets between usual anchor boxes and corresponding expected bounding boxes (Conv2). A classifier head (Conv4) with a Softmax function delivers a confidence score reflecting whether projected bounding box belongs to background or not. To facilitate similarity learning and increase classification performance as four-tuple that comprises a coordinates pairs for it's own top left corner, as well as height and breadth. $C_3 = 4$ anchor, here anchor is amount of anchors to forecast for every position. $C_4 =$ anchor dim embedding, $C_5 =$ anchors, trials is set at 20.

5. Region Proposals-

Several region suggestions (sets of boxes) were created at every sliding window point, that are referred to as anchors; y have varying sizes as well as aspect ratio. Anchor serve to connect

elements to its original place in picture. For generating box likelihood ratings (article or backdrop) and bounding box deltas, RPN contain 2 different fully linked layers (box refinement). RPN's goals are ground-truth categories and ground-truth cluster centres. Every anchors is assign to matched targets after being evaluated based on IoU values. Anchors' IoU values are determined even against picture's regression coefficients objects.

Loss function- Given a set of embedding $\epsilon_1 = \{(\epsilon_i, p_i^*), i \in Z^+\}$, here ϵ_i gives embedding for i^{th} anchor and $p_i^* \in \{0,1\}$ indicates it's anchor label, for transform ϵ_1 to 1 as sets of embeddings $\epsilon_2 = \{(\epsilon_i, \epsilon'_i, s_i), i \in Z^+\}$. Here $s_i \in \{0,1\}$ gives similarities among embeddings ϵ_i , and ϵ'_i a set of embedding triplets $\epsilon_3 = \{(\epsilon_i^a, \epsilon_i^p, \epsilon_i^n), i \in Z^+\}$:

$$L_{pair}(\epsilon, \epsilon', s) = \frac{1}{2} s \|\epsilon - \epsilon'\|^2 + \frac{1}{2} (1 - s) [\max(m - \|\epsilon - \epsilon'\|, 0)]^2$$

Here m is margin. During loss function minimisation procedure, distance between 2 samples from m .

ROI pooling: Anchors boxes are encode as 4-tuple $[x_a, x_b, h_a, w_a]$, here (x_a, y_a) gives coordinate of left top, where (h_a, w_a) give height with width. Offset among projected bound boxes to refine anchor boxes $t = [t_x, t_y, t_h, t_w]$ here:

$$\begin{aligned} t_x &= (x - x_a)/w_a \\ t_y &= (y - y_a)/h_a \\ t_h &= \log(h/h_a) \\ t_w &= \log(w/w_a) \end{aligned}$$

here $[x, y, h, w]$ is 4-tuple for final predicted bounding box similar to $[x_a, x_b, h_a, w_a]$, as in a supervised learning scenario. offsets here represented as $t^* = [t_x^*, t_y^*, t_h^*, t_w^*]$ here:

$$\begin{aligned} t_x^* &= (x^* - x_a)/w_a \\ t_y^* &= (y^* - y_a)/h_a \\ t_h^* &= \log(h^*/h_a) \\ t_w^* &= \log(w^*/w_a) \end{aligned}$$

here $[x^*, y^*, h^*, w^*]$ is having all definition is given by:

$$L_{smooth L_1}(t, t^*) = \sum_{j \in \{x, y, h, w\}}^k f(t_j - t_j^*) \cdot L1 \text{ loss}$$

here, $f(\cdot)$ is smooth L_1 function:

$$f(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

Smooth $L1$ loss Equation (5) is tested as localization.

6. Segmentation using Mask-RCNN

Mask R-CNN [20] is a generic framework for segmenting object instances. It is a natural development of Faster R-CNN [26], a cutting-edge object detector. Mask RCNN uses same initial stage as Faster R-CNN, that is a region suggestion network (RPN). CNN having Region proposal network (RPN) collects features from distinct hierarchical levels having varied sizes, such that every level includes knowledge about higher-level and lower-level characteristic In addition to Mask R-CNN fixes misalignment problem using RoIAlign, a

quantization-free layer. We employ architecture described in Fig 3 to compare Mask R-CNN having various extractor for polyp identification and segmentation.

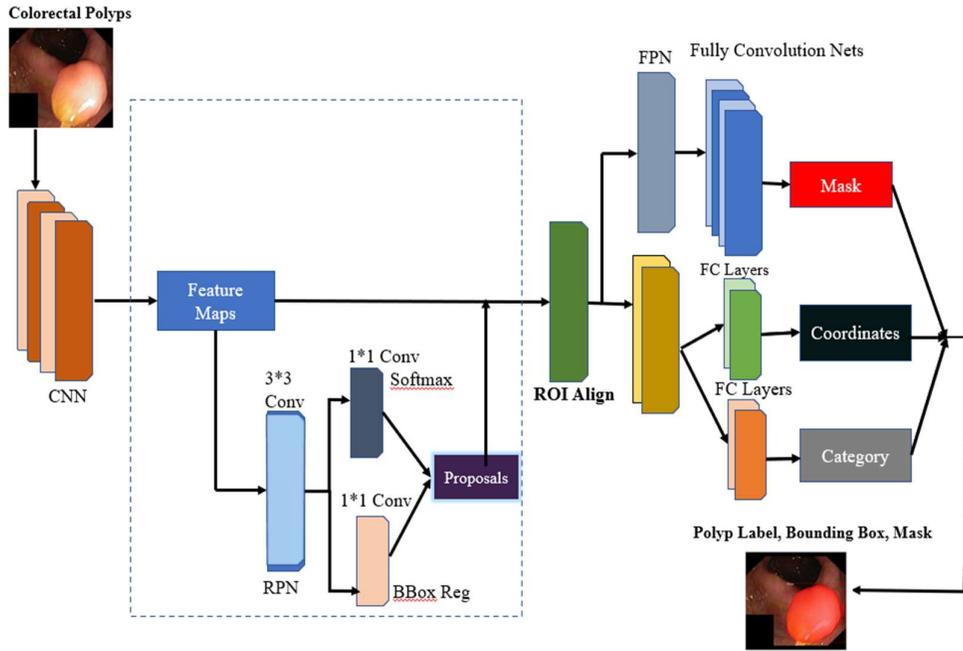


Figure-3 Functional diagram of Mask RCNN for segmentation and classification

For training our techniques, we employ a multi-task loss known anchor developed by RPN on every region of interest. We pick best matched ground-truth box b for every anchor a . If there is match, anchor provide $y_a=1$, as well as vector $(\varphi(b_a; a))$ encoding box b in relation to anchor a . If no match is found $y_a=0$. Every anchor has a 1414 dimensional output from mask branch. Every anchor a 's loss n comprises of 3 losses: location based loss l_{loc} for predicted box $f_{loc}(I, a, \theta)$, classification loss l_{cls} for predicted box $f_{cls}(I, a, \theta)$, l_{mask} predicted box $f_{mask}(I, a, \theta)$, here I is picture and θ is technique parameter.

$$L(a, I; \theta) = \frac{1}{m} \sum_{i=1}^m \frac{1}{N} \sum_{j=1}^N \mathbf{1}[a \text{ is positive}], l_{loc}(\varphi(b_a; a) - f_{loc}(I, a, \theta)) + l_{cls}(y_a f_{cls}(I, a, \theta)) + l_{mask}(mask_a, f_{mask}(I, a, \theta)),$$

Here m gives size, N gives no., of anchors every frames. L1 is utilized to technique localisation loss, softmax is utilized to technique classification loss, and binary cross-entropy is utilized to technique mask loss.

7. Performance analysis:

Accuracy, precision, recall, F1-score, computing speed, and energy consumption are all utilised. se parameters are compared to those of convolutional neural networks (CNNs) [10], ResNet50 [15], and EfficientNet [17], as well as proposed Mask RCNN RPN. Table-1 contains information on dataset.

DETECTION AND SEGMENTATION OF COLORECTAL POLYPS IN GASTROINTESTINAL COLONOSCOPY
IMAGES USING MASK R-CNN FRAMEWORK

S.No	Description of Samples	Platform	# Images
1	Endoscopy Images of Colorectal Polyps collected	The Hyper Kvasir Dataset www.kaggle.com	1000
2	Endoscopy Images of Colorectal Polyps used for Model Selection	The Hyper Kvasir Dataset www.kaggle.com	560
3	Endoscopy Images of Colorectal Polyps used for Model Evaluation	The Hyper Kvasir Dataset www.kaggle.com	56

- **Accuracy** reflects proposed deep learning technique's total prediction ability. ability of classifier techniques to predict absence and presence of polyp is given by TP and TN. no., of FP and FN predictions generated by techniques is identified by FP and FN.

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

- **Precision** gives success of attack categorisation technique, and so on. Precision reflects a classifier's likelihood of correctly predicting a positive outcome when illness is present. TP may be calculated as shown in equation below.:

$$Precision (P) = \frac{TP}{TP + FP}$$

- **Recall** When re is no traffic, a classifier's chance of predicting a negative outcome is 1. TN may be calculated as follows:

$$Recall(R) = \frac{TP}{TP + FN}$$

- **F1- Score** is employed in order to compute prediction performance. It is determined by calculating the weighted average (or harmonic average) of accuracy and recall. A score of one is considered the best, hereas a score of zero is considered the worst. TNs are not included in F-measures. F1-score can be computed as:

$$F1 - Score = \frac{2*P*R}{P+R}$$

- We utilise IoU evaluate polyp segmentation, as shown below.:

$$J(A, B) = \frac{A \cap B}{A \cup B} = \frac{A \cap B}{|A| + |B| - |A \cap B|}$$

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Here, A gives output picture of method and B is actual ground-truth.

Table-1: Comparison of Accuracy

No of epochs	CNN	ResNet50	EfficientNet	Mask_RCNN_RPN
50	89	84	74	96
100	87.9	81	75.3	96.6
150	89.7	81	73	96.6
200	89.5	82	72	96
250	89	81	74	96.4

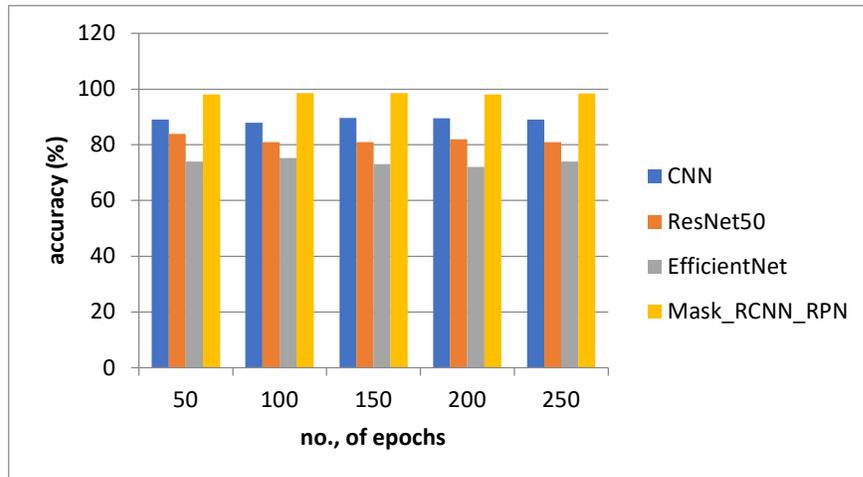


Figure-4 analysis of accuracy

Figure 4 compares accuracy of current CNN, ResNet50, and EfficientNet techniques to proposed Mask RCNN RPN approach, where X axis represents number of epochs utilised for analysis and Y axis represents accuracy values achieved in percentage. When compared to current CNN, ResNet50, and EfficientNet techniques, suggested Mask RCNN RPN approach achieves 96.6 percentage accuracy, which is 9.5 percentage better than CNN, 14.6 percentage better than ResNet50, and 23 percentage better than EfficientNet method.

Table-2: Comparison of Precision

No. of epochs	CNN	ResNet50	EfficientNet	Mask_RCNN_RPN
50	86	78	61	95
100	88	77	65	93
150	87	73	64	96

DETECTION AND SEGMENTATION OF COLORECTAL POLYPS IN GASTROINTESTINAL COLONOSCOPY
IMAGES USING MASK R-CNN FRAMEWORK

200	87	75	63	97
250	86	76	65	97.3

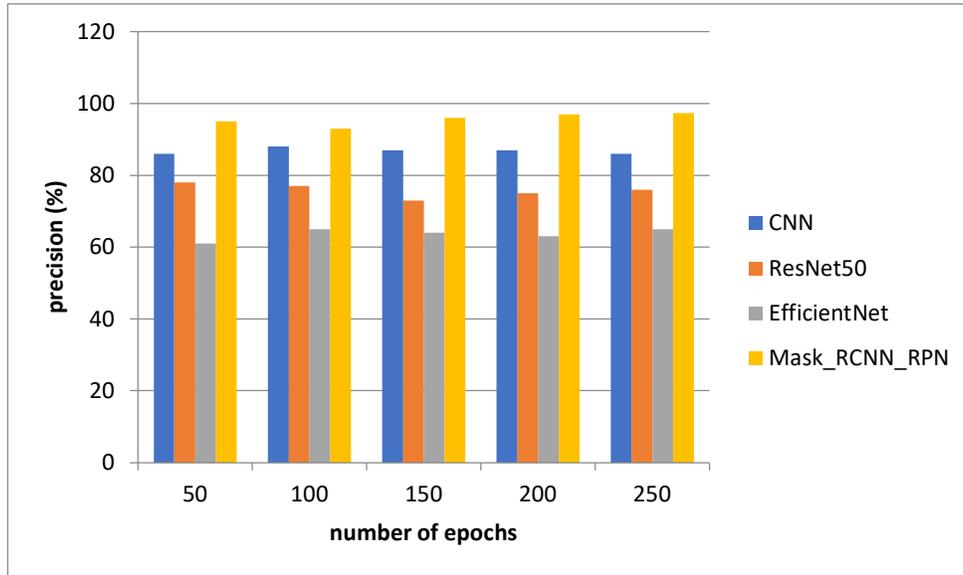


Figure-5: Analysis of Precision

Figure 5 depicts a precision comparison between current CNN, ResNet50, and EfficientNet techniques and proposed Mask RCNN RPN approach, where X axis represents no., of epochs utilised for analysis and Y axis represents accuracy values achieved in percentage. When compared to current CNN, ResNet50, and EfficientNet techniques, suggested Mask RCNN RPN approach achieves 97 percentage accuracy, which is 11 percentage better than CNN, 20 percentage better than ResNet50, and 32 percentage better than EfficientNet method.

Table-3: Comparison of Recall

No., of epochs	CNN	ResNet50	EfficientNet	Mask_RCNN_RPN
50	86	78	61	90
100	88	77	65	91
150	87	73	64	90.1
200	87	75	63	92
250	86	76	65	91.1

DETECTION AND SEGMENTATION OF COLORECTAL POLYPS IN GASTROINTESTINAL COLONOSCOPY IMAGES USING MASK R-CNN FRAMEWORK

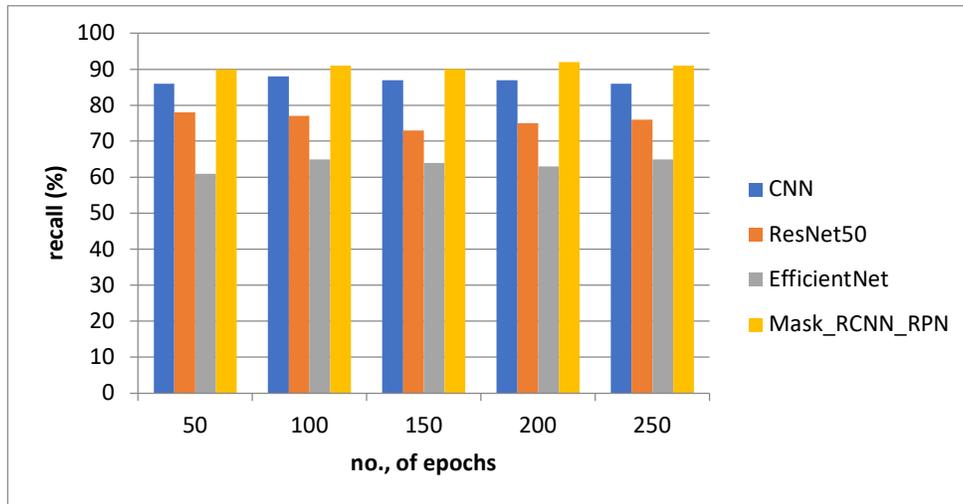


Figure-6: Analysis of Recall

Figure 6 depicts a recall comparison between current CNN, ResNet50, and EfficientNet techniques and proposed Mask RCNN RPN approach, where X axis represents no., of epochs utilised for analysis and Y axis represents recall values acquired in percentage. When compared to current CNN, ResNet50, and EfficientNet techniques, suggested Mask RCNN RPN approach obtains 96 percentage recall, which is 11 percentage better than CNN, 20 percentage better than ResNet50, and 31 percentage better than EfficientNet method.

Table-4: Comparison of F1-score

No., of epochs	CNN	ResNet50	EfficientNet	Mask_RCNN_RPN
50	82	77	72	92
100	80	71	74	92.1
150	83	74	75	92
200	85	75	76	92
250	86	76	75	91.1

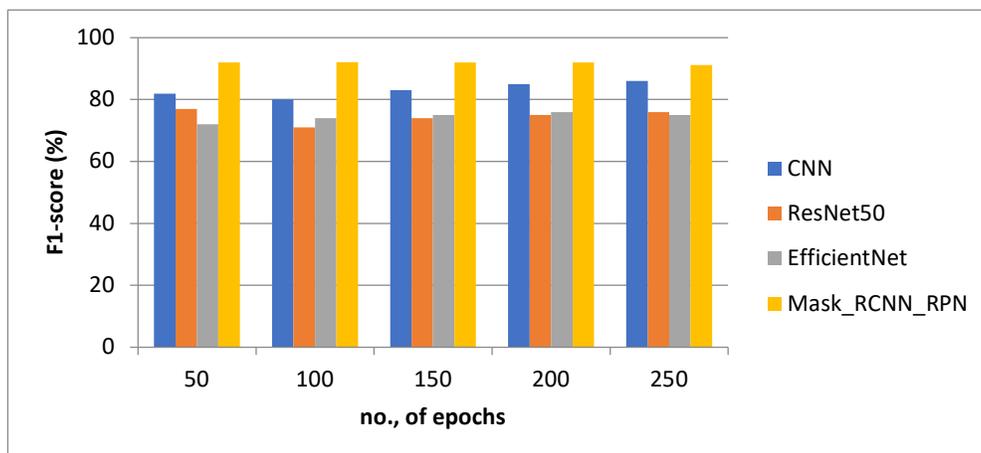


Figure-7: Analysis of F1-score

Figure 7 depicts a comparison of F1-score values obtained in percentage between existing CNN, ResNet50, and EfficientNet methods and proposed Mask RCNN RPN method, where X axis represents no., of epochs used for analysis and Y axis represents F1-score values obtained in percentage. When compared to current CNN, ResNet50, and EfficientNet techniques, suggested Mask RCNN RPN approach obtains 92 percentage of F1-score, which is 17 percentage better than CNN, 27 percentage better than ResNet50, and 16 percentage better than EfficientNet.

Table-5: Comparison of Dice score

No., of epochs	CNN	ResNet50	EfficientNet	Mask_RCNN_RPN
50	82	77	72	98
100	80	71	74	98.7
150	83	74	75	98
200	85	75	76	98
250	86	76	75	98.5

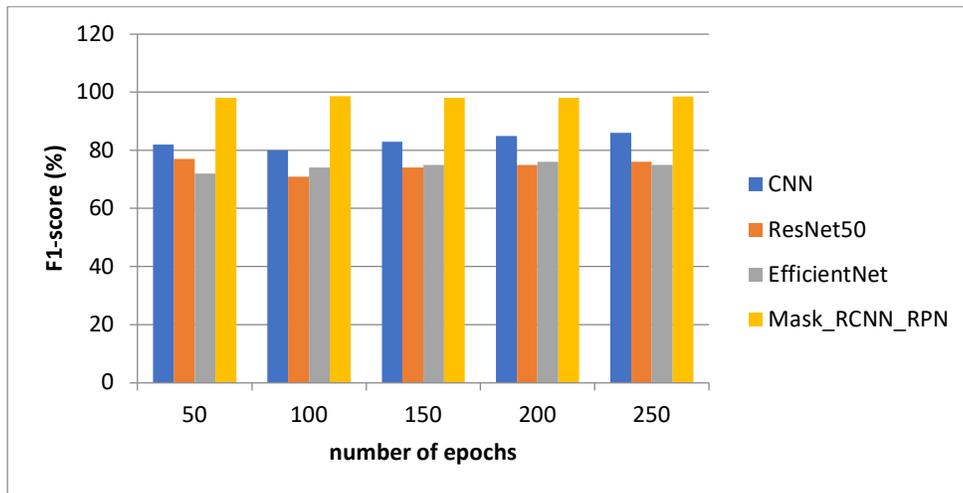


Figure-8: Analysis of Dice score

Figure 8 depicts a dice score comparison between current CNN, ResNet50, and EfficientNet techniques and proposed Mask RCNN RPN method, where X axis represents no., of epochs utilised for analysis and Y axis represents dice score values produced in percentage. When compared to current CNN, ResNet50, and EfficientNet techniques, suggested Mask RCNN RPN approach obtains 98 percentage of F1-score, which is 17 percentage better than CNN, 27 percentage better than ResNet50, and 16 percentage better than EfficientNet.

Table- 6 Overall comparative analysis

Methods	Accuracy (percentage)	Precision (percentage)	Recall (percentage)	F1-score (percentage)	Dice score (percentage)
CNN	89.9	86	87	82	82
ResNet50	84	77	76	77	77
EfficientNet	75	65	65	75	75
Mask_RCNN_RPN	96.6	97	96	92	98

8. Conclusion

Polyp segmentation for colorectal cancer is a difficult process because the variety of polyp forms, colours, and intensities. Another difficulty would be a lack of access to a big public polyp dataset having annotations. We discussed our methods to polyp identification difficulty in this work. Using RPN, we modified and assessed R-CNN Masking has modern CNN feature extractors. With an improved training dataset, the results demonstrate that Mask R-CNN could serve as a feasible technique for polyp segmentation and detection, and that required a deeper or more intricate CNN extracted features may become unneeded. In the future, we will investigate multisource data and applying our technique to colonoscopy video processing. Additionally, we will examine dual-branch hybrid technology's non-substitutability with implementations of higher-order concentration mechanisms.

References:

1. Baumgart, D. C. et al. Crohn's disease. *Lancet* 380, 1590–1605. [https://doi.org/10.1016/S0140-6736\(12\)60026-9](https://doi.org/10.1016/S0140-6736(12)60026-9) (2012).
2. Danese, S. et al. Ulcerative colitis. *N. Engl. J. Med.* 365, 1713–1725. <https://doi.org/10.1056/NEJMra1102942> (2011).
3. Graham, D. B. et al. Pathway paradigms revealed from genetics of inflammatory bowel disease. *Nature* 578, 527–539. <https://doi.org/10.1038/s41586-020-2025-2> (2020).
4. Baumgart, D. C. et al. Newer Biologic and Small-Molecule Therapies for Inflammatory Bowel Disease. *N Engl J Med* 385, 1302–1315. <https://doi.org/10.1056/NEJMra1907607> (2021).
5. Pariente, B. et al. Development of Crohn's disease digestive damage score, Lemann score. *Infamm. Bowel Dis.* 17, 1415–1422. <https://doi.org/10.1002/ibd.21506> (2011).
6. Takenaka, K. et al. Development and validation of a deep neural network for accurate evaluation of endoscopic pictures from patients having ulcerative colitis. *Gastroenterology* 158, 2150–2157. <https://doi.org/10.1053/j.gastro.2020.02.012> (2020).
7. LeCun, Y. et al. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539> (2015).
8. Ratner, A. et al. in *Weak Supervision: A New Programming Paradigm for Machine Learning* (The Stanford AI Lab Blog, 2019).
9. S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection having region proposal networks. In *NIPS*, 2015

10. Mohapatra, S., Pati, G. K., Mishra, M., & Swarnkar, T. (2023). Gastrointestinal abnormality detection and classification using empirical wavelet transform and deep convolutional neural network from endoscopic pictures. *Ain Shams Engineering Journal*, 14(4), 101942
11. Thomas Abraham, J. V., Muralidhar, A., Sathyarajasekaran, K., & Ilakiyaselvan, N. (2023). A Deep-Learning Approach for Identifying and Classifying Digestive Diseases. *Symmetry*, 15(2), 379.
12. Rajkumar, S., Sairam, V. A., Krithika, G. K., Harini, C. S., Dhanusha, P., Chandrasekar, G. E., & Saphthagirivasan, V. (2023). GastroEffNetV1-CNN based Automated detection of Gastrointestinal abnormalities from capsule endoscopy pictures.
13. Sivari, Esra, Erkan Bostanci, Mehmet Serdar Guzel, Koray Acici, Tunc Asuroglu, and Tulin Ercelebi Ayyildiz. "A New Approach for Gastrointestinal Tract Findings Detection and Classification: Deep Learning-Based Hybrid Stacking Ensemble Techniques." *Diagnostics* 13, no. 4 (2023): 720.
14. Ruan, G., Qi, J., Cheng, Y., Liu, R., Zhang, B., Zhi, M., ... & Wei, Y. (2022). Development and validation of a deep neural network for accurate identification of endoscopic pictures from patients having ulcerative colitis and Crohn's disease. *Frontiers in Medicine*, 9, 854677.
15. Qi, J., Ruan, G., Liu, J., Yang, Y., Cao, Q., Wei, Y., & Nian, Y. (2022). PHF3 Technique: A Pyramid Hybrid Feature Fusion Framework for Severity Classification of Ulcerative Colitis Using Endoscopic Pictures. *Bioengineering*, 9(11), 632.
16. Turan, M., & Durmus, F. (2022). UC-NfNet: deep learning-enabled assessment of ulcerative colitis from colonoscopy pictures. *Medical Picture Analysis*, 82, 102587.
17. Uçan, M., Kaya, B., & Kaya, M. (2022, October). Multi-Class Gastrointestinal Pictures Classification Using EfficientNet-B0 CNN Technique. In *2022 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 1-5). IEEE.
18. Bhambhani, H. P., & Zamora, A. (2021). Deep learning enabled classification of Mayo endoscopic subscore in patients having ulcerative colitis. *European Journal of Gastroenterology & Hepatology*, 33(5), 645-649.
19. Mokter, M. F., Oh, J., Tavanapong, W., Wong, J., & de Groen, P. C. (2020). Classification of ulcerative colitis severity in colonoscopy videos using vascular pattern detection. In *Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction having MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 11* (pp. 552-562). Springer International Publishing.
20. Mohapatra, S., Nayak, J., Mishra, M., Pati, G. K., Naik, B., & Swarnkar, T. (2021). Wavelet transform and deep convolutional neural network-based smart healthcare system for gastrointestinal disease detection. *Interdisciplinary Sciences: Computational Life Sciences*, 13, 212-228.
21. dos Santos JCM, Carrijo GA et al (2020) Fundus picture quality enhancement for blood vessel detection via a neural network using CLAHE and wiener filter. *Res Biomed Eng*:1–13. <https://doi.org/10.1007/s42600-020-00046-y>
22. Sheela CJJ, Suganthi G (2020) An efficient denoising of impulse noise from MRI using adaptive switching modified decision based unsymmetric trimmed median filter. *Biomed Signal Process Control* 55:101657. <https://doi.org/10.1016/j.bspc.2019.101657>