Leveraging YOLOv9 for Advanced Colorectal Polyp Detection

Zeeshan Haider
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
zhaider@psu.edu.sa (corresponding author)

Ahmad Taher Azar
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia | Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt
aazar@psu.edu.sa

Samah ALmutlaq
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
salmutlaq@psu.edu.sa

Received: 24 February 2025 | Revised: 20 April 2025 | Accepted: 8 May 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.10701*

ABSTRACT

The increasing prevalence of colorectal cancer has necessitated improved diagnostic tools, which has spurred significant research efforts into Artificial Intelligence (AI)-assisted polyp detection and localization methods. Missed diagnoses due to human factors, such as fatigue or inexperience, are recognized to have severe consequences. This study investigates the efficacy of state-of-the-art object detection models for enhanced polyp identification, focusing on the performance of four variants of the YOLOv9 model (gelan-e, gelan-c, yolov9-c, and yolov9-e) for colorectal polyp detection and localization. These models were trained and tested using two distinct datasets: a combined dataset comprised of CVC-CLinicDB and Kvasir-SEG, and the LDPolypVideo dataset. The impact of different YOLOv9 architectures on detection accuracy and localization precision is analyzed. The YOLOv9 variants achieved mAP@50 scores up to 99.1% on CVC-ClinicDB (a 16% improvement over YOLOv8), outperforming YOLOv8 and other models, and 55.56% mAP@50 on LDPolypVideo, demonstrating enhanced accuracy and efficiency in colorectal polyp detection. This study highlights the potential of YOLOv9 to enhance the accuracy and efficiency of colorectal polyp detection.

Keywords-polyp detection; localization; medical imaging; colorectal cancer; deep learning; YOLOv9; artificial intelligence

I. INTRODUCTION

Artificial Intelligence (AI) is revolutionizing medical imaging and is affecting every stage of the imaging process. From image acquisition and reconstruction to analysis and interpretation, AI-driven tools are enhancing diagnostic accuracy, enabling personalized treatment planning, and improving the prediction of patient outcomes [1-3]. AI and Machine Learning (ML) are revolutionizing industries from finance to agriculture by streamlining procedures and decision-making. In finance, AI systems examine large datasets to detect fraud and forecast market trends, thus improving security and investment strategies. ML models help farmers with precision farming by monitoring crop health and optimizing resource utilization, resulting in higher yields and greater sustainability [4-7]. This integration of AI and ML is critical for accelerating innovation and efficiency across different disciplines.

Colorectal polyps are abnormal growths or outgrowths that form in the lining of the colon. These growths occur due to excessive cell proliferation within the bowel. Although most colorectal polyps are benign at their inception, they can evolve into Colorectal Cancer (CRC), a leading cause of cancer-related mortality, if left unnoticed and untreated over time [8-10]. According to the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO), CRC was diagnosed in approximately 2 million individuals around the world in 2020, resulting in nearly 1 million deaths [11]. That is why automatic detection and localization of

polyps is a very important problem in the scope of medical imaging. Colorectal polyps can form anywhere along the length of the large intestine, including the rectum. However, a higher prevalence of polyps has been observed in the distal colon or on the left side of the large intestine. In contrast, polyps in the proximal colon, or the right side, are often more difficult to detect during colonoscopy due to the anatomical structure of this region [11-13].

Medical image detection is quite valuable in planning and pre/post-treatment of polyps. In ML, the detection of colorectal polyps involves training a model to identify and localize polyp features within images or live video streams. These models may be refined through fine-tuning to differentiate between various polyp types [14]. Colorectal polyp detection models are designed to accomplish one or more of the following objectives: polyp classification, polyp detection, or polyp segmentation. Although some models are trained for a single task, others are capable of performing multiple tasks sequentially, such as detection followed by classification. For polyp detection models, training involves providing the network with images and the corresponding labels. These labels typically consist of bounding box coordinates and an indicator of polyp presence. Sometimes, polyp class labels are also provided if available in the dataset.

YOLO models have consistently advanced real-time object detection. From YOLOv1's [15] single-stage approach to YOLOv9's PGI approach [16], the series has evolved; YOLOv2 and YOLOv3 introduced FPN and multi-scale training [17, 18], YOLOv4 with CSPDarknet53 and PANet-SPP modifications to the architecture [19], YOLOv5 retained YOLOv4 techniques with PyTorch implementation [20], multi-task learning in YOLO-R [21], YOLO-X with anchor-free detection and decoupled heads [22], YOLOv6 [23], and YOLOv7 with re-parameterization and model scaling [24]. YOLOv8 [25] has shown a potential for high performance with minimal architectural changes. YOLO-NAS [26] explored quantization for faster and more accurate detection. YOLOv9 introduces PGI to address information bottleneck challenges and ensure complete input information for the target task.

In the context of medical imaging, and particularly polyp detection, YOLOv9 presents distinct advantages. The task of accurately identifying polyps in colonoscopy images or CT scans is often hindered by factors such as low contrast, variability in polyp size and shape, and the presence of noise or artifacts. Previous YOLO models, including YOLOv8, have faced challenges due to the information bottleneck principle, where crucial spatial and textural details can be lost in deeper layers, leading to suboptimal detection of these subtle yet clinically significant structures. YOLOv9's PGI module directly addresses this by enabling the network to retain more of the original input information throughout its processing. This is critical in polyp detection because preserving fine-grained details enhances the model's ability to differentiate polyps from surrounding tissue. In addition, YOLOv9 incorporates the Generalized Efficient Layer Aggregation Network (GELAN), which optimizes the network architecture for improved feature extraction and integration. GELAN's efficient design allows the model to better capture complex features relevant to polyp identification while maintaining computational efficiency, which is important for real-time or high-throughput clinical applications. The enhanced feature representation and gradient flow facilitated by PGI and GELAN contribute to YOLOv9's superior accuracy and robustness in polyp detection, potentially reducing false negatives and improving diagnostic outcomes.

This study employs YOLOv9 model variants and examines their effectiveness in the detection of CRC polyps. Two datasets were utilized to train Deep Learning (DL) models: one containing images from CVC-ClinicDB [27] and Kvasir-SEG [28], and the LDPolypVideo [29] dataset. A comprehensive analysis was conducted on various variants of YOLOv9 (gelan-e, gelan-c, yolov9-c, yolov9-e) to evaluate their performance in polyp detection.

II. RELATED WORK

This section provides an overview of contemporary approaches to the detection of colorectal polyps. In [30], a real-time polyp detection algorithm used modified YOLOv3 and YOLOv4 architectures. In [31], the original DarkNet53 backbone in YOLOv3 was replaced by a CSPNet, and a modified CSPDarkNet53 architecture was introduced as the backbone for the YOLOv4 model. In [32], a saliency detection network was proposed for the identification of polyps in static images. This study employed Neutrosophic theory to mitigate the impact of white light reflections. Using a single-value Neutrosophic set (SVNS), this study presented an image-suppression technique to reconstruct colonoscopy images without the interference of white light reflections. In [33], a novel DL algorithm was proposed for polyp characterization, which was integrated into the GI Genius V2 endoscopy system [34]. The classification module of this system employed two pre-trained ResNet18 networks, with the first discriminating between adenoma and non-adenoma polyps in each frame, and the second generating a descriptor for every identified polyp. The model was trained on a proprietary dataset of unfiltered colonoscopy videos. In [35], a YOLOv3 network was initially pre-trained on the PASCAL VOC dataset for general object detection, and then fine-tuned on a dataset of 28,576 colonoscopy images to adapt it for real-time polyp detection.

In [36], a novel polyp detection method used VGG16 and MobileNet. To enhance data quality, input colonoscopy images were preprocessed by removing non-informative black regions. Random Forest (RF), a classic ML algorithm, was utilized in [37] for the detection of premalignant colorectal polyps. Due to the scarcity of open-source polyp datasets, data augmentation techniques for polyp segmentation were explored in [38]. To achieve better accuracy, various optimizers and a unique augmentation strategy were adopted, in which the augmentation percentage was varied for each epoch. In [39], the YOLOv5m object detection model was integrated with the SWIN transformer for real-time polyp detection in colonoscopy videos. The YOLOv5m backbone was used to extract features from individual frames, which were then combined with features extracted by the SWIN transformer. This study concluded that this combined approach of incorporating a local feature extraction network with the SWIN transformer led to enhanced overall performance.

In [40], a real-time polyp detection method utilized the YOLOv4 architecture, which was initially pre-trained on the ImageNet dataset and subsequently optimized for GPU performance using the TensorRT tool. The samples were resized and enlarged before being fed into the YOLOv4 network for further processing. In [41], a plug-in module was introduced that can be integrated with any object detection algorithm. This module, named the Instance Tracking Head (ITH), incorporated a feature-based mechanism to enable multi-task training objectives within existing detection frameworks.

Previous YOLO models, including YOLOv8, struggled with the information bottleneck principle, where deeper layers could lose critical data, leading to suboptimal gradient updates. YOLOv9's PGI directly counters this by preserving essential data across network layers, ensuring more reliable gradient generation and better model convergence. The Single Shot multibox Detector (SSD) is a one-stage detector known for balancing speed and accuracy, but it struggles with small objects and complex scenes due to its reliance on predefined anchor boxes and less effective feature extraction. The DC-SSDNet used for polyp detection in [42] was found to have a performance inferior to that of YOLOv9. Unlike YOLOv9's PGI, EfficientDet [43] does not introduce a specific mechanism to ensure reliable gradient generation or prevent error accumulation under deep supervision.

III. METHODOLOGY

A. Dataset Used

This study utilized two datasets for the task of polyp detection and localization. The first dataset is a combination of CVC-ClinicDB [27] and Kvasir-SEG [28]. The Kvasir-SEG dataset, consisting of 1,000 images associated with ground truth masks, was partitioned into a training set of 900 images and a testing set of 100 images. Similarly, the CVC-ClinicDB dataset, comprising 612 images, was divided into a training set of 548 images and a testing set of 64 images. To evaluate the generalizability of the model, the supplementary experiments used three external datasets: EndoScene [44], ColonDB [45], and ETIS-LaribDB [46]. These datasets, comprising 60, 380, and 196 images, respectively, were not utilized during the training phase. All datasets had only binary masks for polyps in the image. To obtain proper bounding boxes from these masks for the problem of detection and localization, traditional computer vision-based techniques were applied, such as eroding, dilation, finding mask contours, and obtaining bounding boxes from these contours.

LDPolypVideo [29] is the second dataset employed for training YOLOv9 variants, which is quite large compared to the first dataset. This dataset includes a comprehensive collection of colonoscopy videos and features a diverse range of polyps and intricate bowel environments. Comprising 160 colonoscopy videos and 40,266 frames with polyp annotations, LDPolypVideo is four times larger than the previously largest colonoscopy video database, CVC-ClinicVideoDB.

To enhance the robustness of the model, each image fed into the YOLOv9 model was resized to a standard input size for the first dataset and to 480×480 for the LDPolypVideo dataset. Additionally, a variety of data augmentation techniques were applied during training, including HSV-hue, HSV-value, saturation, translation, scale, image mixup, left-right flipping, and mosaic. These augmentations introduce variations into the training data, helping the model to generalize better to unseen images. However, these augmentations were not applied during testing to ensure consistent results with real-world input.

B. Polyp Detection Model Architecture

Deep neural networks often suffer from information bottlenecks in which input data gradually lose information as it propagates through the network [47]. Several strategies have been proposed to mitigate it. Reversible architectures explicitly preserve input information by repeatedly using it throughout the network [33, 48]. Masked modeling employs reconstruction loss to implicitly maximize feature extraction [49]. Deep supervision introduces auxiliary losses at intermediate layers to ensure that crucial information is preserved and transferred to the deeper layers [50].

YOLOv9 [16] introduces PGI to address the limitations of existing detection methods, particularly the increased inference cost associated with reversible architectures. PGI employs an auxiliary reversible branch to generate reliable gradients, ensuring that deep features capture essential task-specific characteristics. Unlike traditional deep supervision methods, the auxiliary branch of PGI avoids semantic loss by focusing on gradient propagation. This innovative approach enables efficient training by distributing gradient information across various semantic levels without incurring additional computational overhead. Furthermore, PGI overcomes the limitations of mask modeling by freely selecting task-specific loss functions. YOLOv9's PGI mechanism applies to various deep neural network sizes, surpassing the limitations of deep supervision, which is restricted to extremely deep networks. YOLOv9 introduces GELAN, based on ELAN [40], which prioritizes a balance between model complexity, accuracy, and real-time inference speed. This modular design allows users to select computational blocks suitable for different inference devices. Experiments on the MS COCO dataset demonstrate YOLOv9's state-of-the-art performance across all metrics.

In summary, YOLOv9 contributes to the field in several key ways. A theoretical analysis of existing DNN architectures, conducted from a reversible functions perspective, successfully explains previously challenging phenomena and guides the design of PGI and an auxiliary reversible branch. This approach overcomes the limitations of deep supervision, typically restricted to exceptionally deep networks, paving the way for the practical application of novel lightweight architectures in real-world scenarios. The GELAN architecture, characterized by a higher parameter utilization rate than state-of-the-art depth-wise convolution designs, achieves a more efficient framework with a lightweight model size, faster processing speed, and enhanced detection accuracy.

1) Programmable Gradient Information (PGI)

YOLOv9 introduces a novel auxiliary supervision approach, motivated by the limitations of existing methods. This approach comprises three key components: a main branch, an auxiliary reversible branch, and multi-level auxiliary

information. During inference, only the main branch is utilized, incurring no additional computational cost. The auxiliary reversible branch addresses the information bottleneck problem inherent in deep neural networks. This bottleneck can hinder the loss function's ability to generate reliable gradients. Additionally, multi-level auxiliary information is incorporated to mitigate error accumulation, particularly in multi-prediction branch architectures and lightweight models. The next subsections dive into these two critical components.

2) Auxiliary Reversible Branch

The auxiliary branch provides essential information to guide the learning process and prevent misleading correlations caused by incomplete features. A fully reversible architecture can ensure complete information flow as it significantly increases inference costs. To mitigate this, YOLOv9 proposes an auxiliary reversible branch that acts as an extension of the deep supervision branch. This branch provides reliable gradient information to the main branch, mitigating information loss due to bottlenecks. This approach ensures that the network learns to extract relevant information for the specific task without compromising the efficiency of inference. To further enhance the learning process, YOLOv9 incorporates multi-level auxiliary information inspired by the concept of feature pyramids. This approach combines gradients from multiple prediction heads, ensuring that the main branch receives complete information for objects of various sizes.

3) Generalized ELAN (GELAN)

The GELAN architecture is a novel approach inspired by the gradient path planning principles of CSPNet and ELAN [51]. GELAN prioritizes a balance between lightweight design, rapid inference, and high accuracy. By generalizing ELAN's structure, GELAN allows for the integration of various computational blocks beyond stacked convolutional layers.

C. YOLOv9 Variants

This study trained four variants of the YOLOv9 model for polyp detection: Gelan-c, Gelan-e, Yolov9-c, and Yolov9-e. These variants differ in their architectures. Gelan-c and Yolov9-c belong to the compact YOLOv9 architecture, while Gelan-e and Yolov9-e belong to the elaborate YOLOv9 architecture. Table I outlines the number of layers, parameters, and computational requirements for these variants along with YOLOv8 variants.

TABLE I. SPECIFICATIONS OF YOLO V8 AND V9 VARIANTS

Variant	Layers	Parameters	GFLOPs
Gelan-c	621	25441698	103.2
Gelan-e	1228	58057378	192.2
Yolov9-c	962	51011140	238.9
Yolov9-e	1475	69415556	244.9
Yolov8l	365	43630611	165.4
Yolov8x	365	68153571	258.1

IV. EXPERIMENTAL SETUP

The experiments aimed to study performance by training and testing different variants of YOLOv9 on two CRC polyp datasets. The first dataset was a combination of CVC-ClinicDB and Kvasir-SEG, while the second dataset was the LDPolypVideo. Additionally, a variety of data augmentation techniques were applied during training, including HSV-hue, HSV-value, saturation, translation, scale, image mixup, left-right flipping, and mosaic. The models were trained for 50 epochs for each optimization experiment. The learning rate used for all experiments was 0.0001 with the SGD optimizer, as it gave the best results. All experiments were conducted using an NVIDIA RTX 3060TI.

V. RESULTS AND ANALYSIS

This study evaluated model performance using Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP). Precision quantifies the accuracy of positive predictions, while recall measures the model's ability to identify all actual positive cases. Average precision captures the area under the Precision-Recall curve, providing an overall assessment of detection performance. The F1-score balances precision and recall, offering a comprehensive measure of detection accuracy [52]. mAP calculates the average AP across all classes.

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{1}{N}\sum_{n=1}^{N} \mathrm{AP_n} \quad (3)$$

where $AP = \int_0^1 P(R)\, dR$, TP denotes True Positives, FP denotes False Positives, and FN denotes False Negatives.

This study used mAP@50 as the primary performance metric, which is widely used in object detection to evaluate the precision-recall trade-off at an Intersection over Union (IoU) threshold of 0.5. The mean Average Precision at the threshold of 0.5 (mAP@0.5), unlike simple classification, evaluates the model's ability to pinpoint the precise location and size of defects within images. By setting the IoU threshold to 0.5, this metric demands a substantial overlap (at least 50%) between the predicted and actual defect areas for a true positive. This stringent criterion is essential in the context of polyp detection, where even small undetected defects can significantly impact diagnosis. Consequently, mAP@0.5 provides a more comprehensive assessment of the model's performance, encompassing not only the presence of polyps but also their severity and spatial extent, making it a crucial metric for accurate diagnosis of polyps.

Figure 1 illustrates the comparison between actual labels and predicted labels for the combined Kvasir and CVC-ClinicDB dataset. The actual labels are shown on the left side of the image, while the predicted labels are shown on the right side. Similarly, Figure 2 shows the comparison between actual labels and predicted labels for the LDPolypVideo dataset.

Figures 3 and 4 show the progression of the loss during training for both datasets separately. Figures 5 and 6 show the progression of Precision, Recall, map@50, and map@50:95, employed for validation of the models during training on separate datasets.

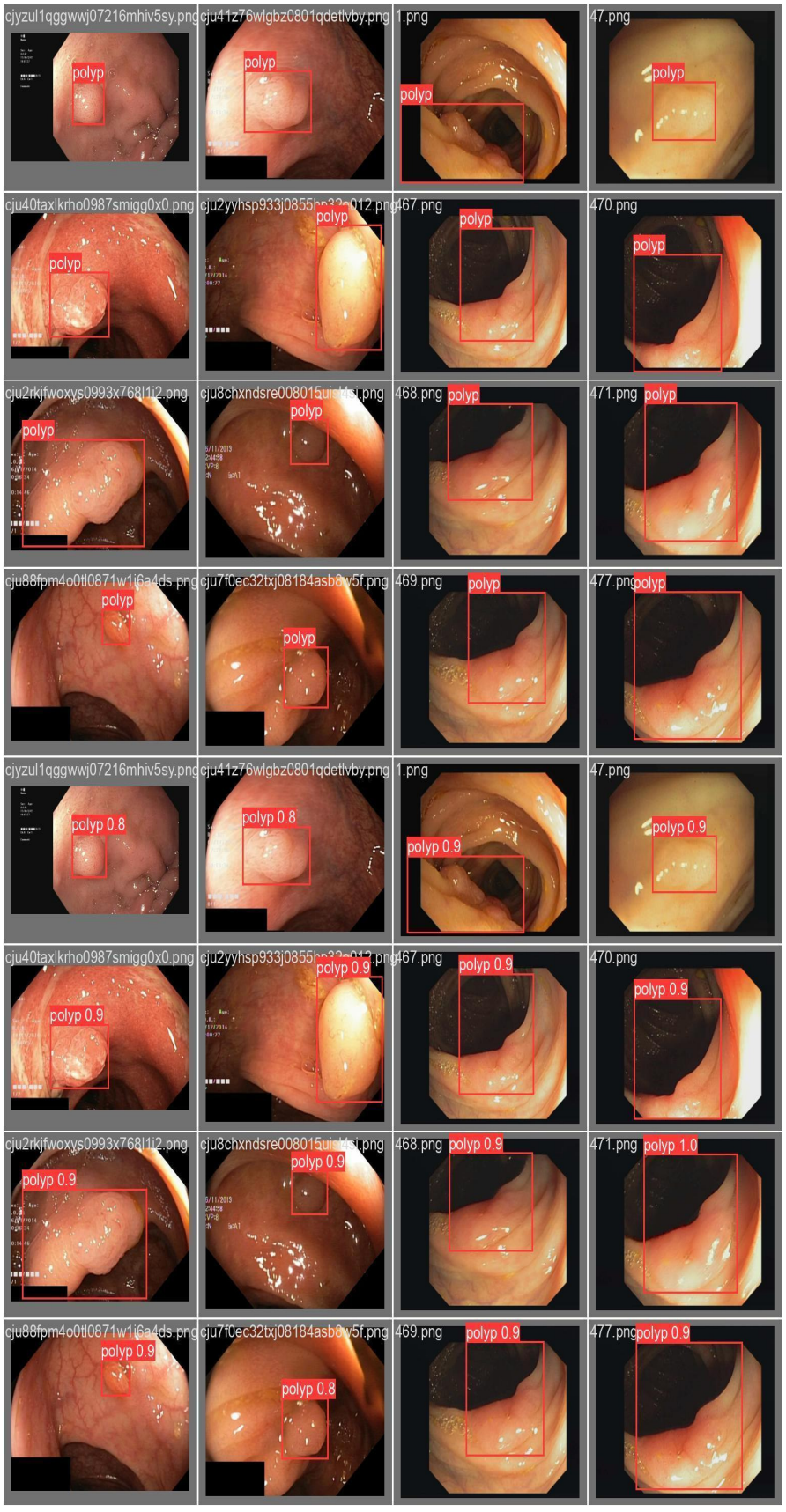

Fig. 1. Actual polyp labels (a) and predicted polyp detections (b) for training on the Kvasir-SEG and CVC-ClinicDB dataset.

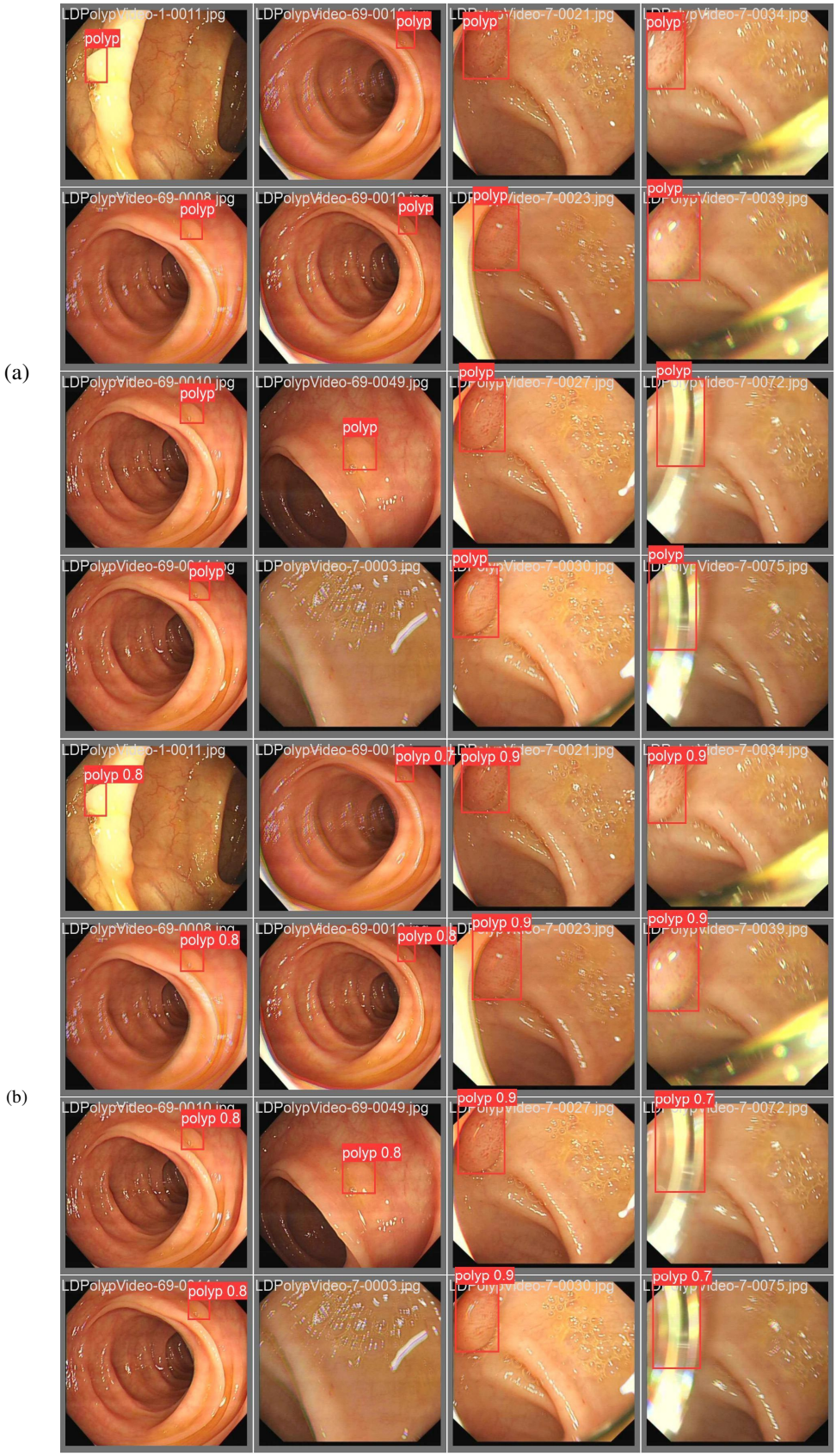

Fig. 2. Actual polyp labels (a) and predicted polyp detections (b) for training on the LDPolypVideo dataset.

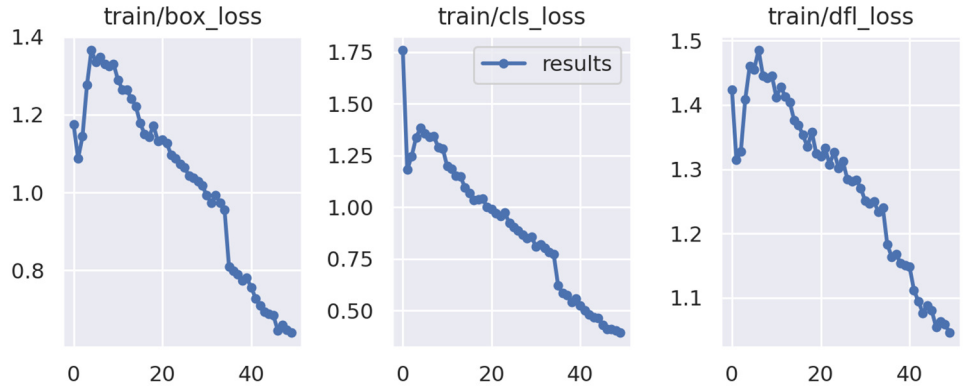

Fig. 3. Loss progression for training on the Kvasir-SEG and CVC-ClinicDB combined dataset

Fig. 4. Loss progression for training on the LDPolypVideo dataset.

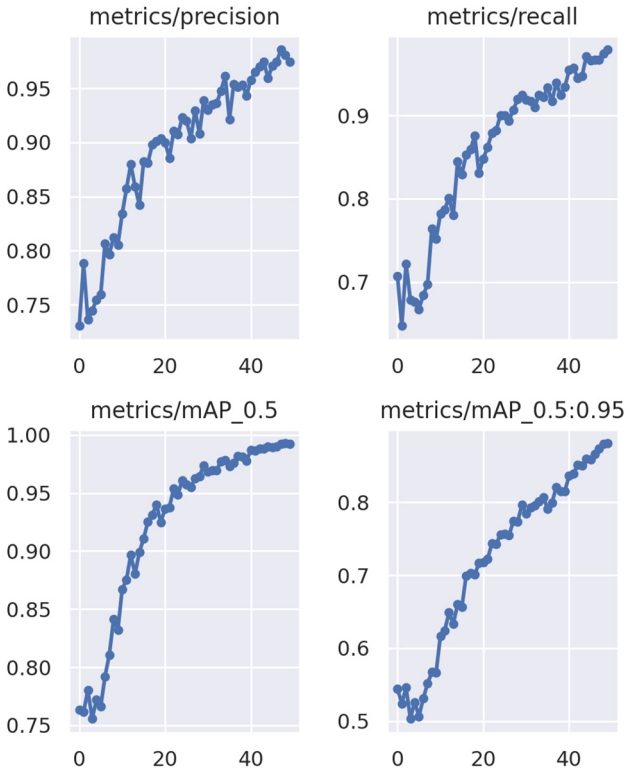

Fig. 5. Precision, Recall, and mAP at various thresholds for Kvasir-SEG and CV-ClinicDB validation dataset.

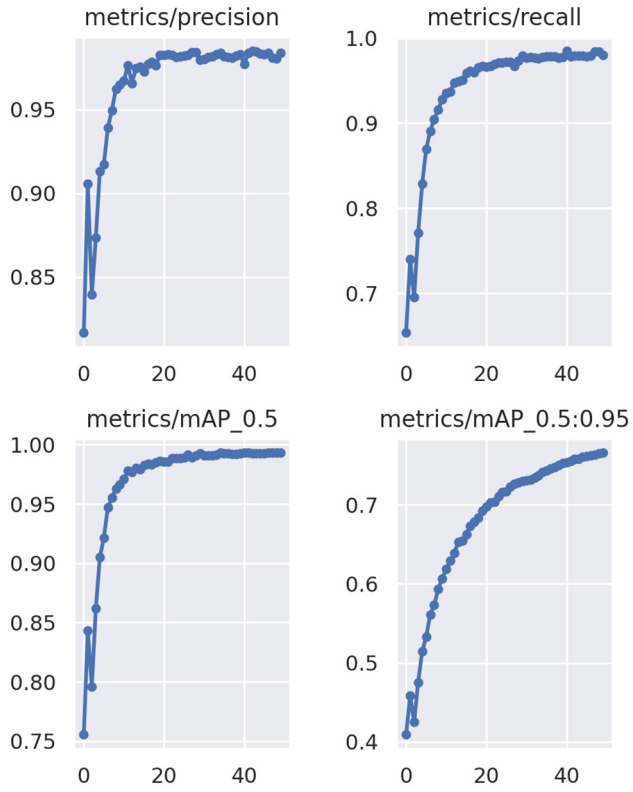

Fig. 6. Precision, Recall, and mAP at various thresholds on LDPolypVideo validation dataset.

Table II shows the precision, recall, and mAP@50 scores for all YOLO variants on the five test datasets. Figure 7 shows mAP@50 for all YOLOv9 variants trained on the combined Kvasir-SEG and CVC-ClinicDB dataset. In general, the best or close to the best results were obtained by the Gelan-e variant.

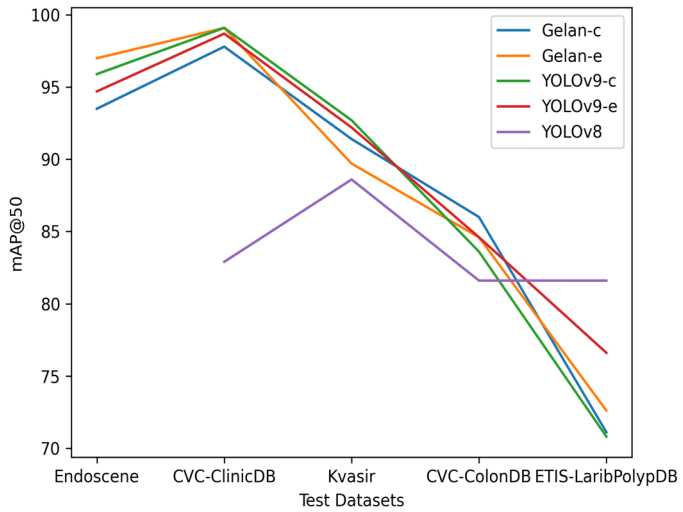

Fig. 7. mAP@50 of all evaluation datasets for different YOLOv9 variants.

For Endoscene and CVC-ClinicDB, Gelan-e outperformed all other models with mAP scores of 97 and 99.1%, respectively. For the Kvasir-SEG dataset, the YOLOv9-c model was on top, acquiring an mAP@50 of 92.7%. Gelan-c outperformed other models with an mAP@50 of 86% when evaluated across the ColonDB dataset.

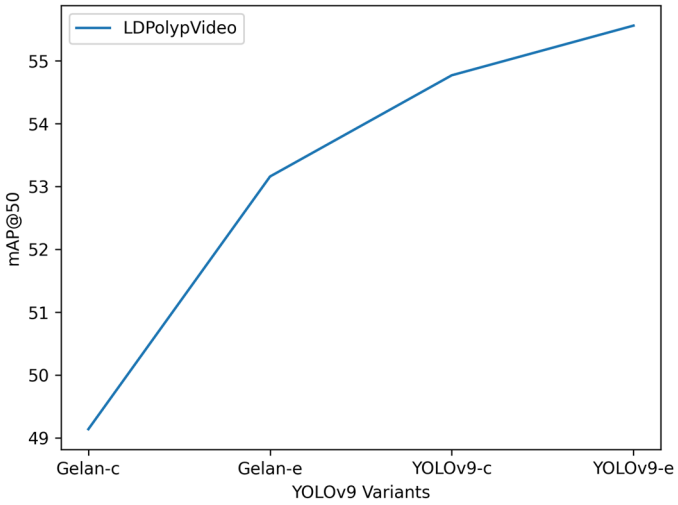

Fig. 8. mAP@50 on the LDPolypVideo dataset for different YOLOv9 variants.

The results show that the DC-SSDNet model showed a higher performance than the YOLOv9 variants only in the case of the ETIS-LaribDB dataset, with an mAP@50 of 81.6%. The superiority of the YOLOv9 variants is quite evident in the results, since the YOLOv9 variants beat YOLOv8 and other models in two out of three common datasets on which these models were compared. Also, the difference in the values of

mAP@50 is quite high in some cases, as for CVC-ClinicDB, this difference is almost 16%. Table III delineates precision, recall, and mAP@50 scores for the LDPolypVideo dataset. This is one of the largest polyp datasets, with around 24000 training/validation and 15000 testing images. Figure 8 visually depicts the performance of mAP@50 across variants of YOLOv9. Results show that YOLOv9-e tops the list in producing the best mAP@50 at 55.56%.

The YOLOv9 Gelan-e variant was the top performer, offering the best balance of precision, recall, and mAP@50 across all datasets, particularly excelling on CVC-ClinicDB and Kvasir-SEF. Models such as Gelan-c and YOLOv9-c prioritize Precision (fewer false positives) but sometimes sacrifice Recall (missed detections). Gelan-e and YOLOv9-e offer a better balance, which is critical for medical applications where missing a polyp (low Recall) can be more harmful than false positives. The YOLOv9 architecture includes a deeper network with Gelan and an advanced PGI mechanism, making it robust for medical image analysis. However, all variants struggle with ETIS-LaribDB, highlighting the need for further improvements in handling complex, low-quality images. The YOLOv9-e model, characterized by increased architectural complexity, is likely not to generalize due to the limited training set for the first dataset. It is well-established that more complex models necessitate larger datasets to effectively learn generalizable features, which can be clearly seen in the LDPolypVideo dataset results in Table III.

TABLE II. PRECISION, RECALL, AND MAP@50 SCORE FOR FIVE DIFFERENT POLYP DETECTION TEST DATASETS FOR VARIOUS YOLO MODELS

Models	EndoScene			CVC-ClinicDB			Kvasir-SEG			ColonDB			ETIS-LaribDB		
	P	**R**	**mAP@50**	**P**	**R**	**mAP@50**	**P**	**R**	**mAP@50**	**P**	**R**	**mAP@50**	**P**	**R**	**mAP@50**
YOLOv9 Gelan-c (this study)	**99.7**	85.0	93.5	96.6	95.5	97.8	90.9	82.9	91.4	85.1	72.1	**86.0**	79.8	64.9	71.1
YOLOv9 Gelan-e (this study)	96.1	**93.3**	**97.0**	**97.4**	**97.0**	**99.1**	86.6	80.8	89.7	**89.5**	71.3	84.6	86.6	60.6	72.6
YOLOv9-c (this study)	99.5	90.0	95.9	98.5	96.9	99.1	89.9	**89.0**	**92.7**	85.8	73.4	83.6	74.9	64.6	70.8
YOLOv9-e (this study)	97.7	88.3	94.7	96.9	97.0	98.7	89.1	85.0	92.2	83.8	**73.6**	84.6	85.5	63.9	76.6
DC-SSDNet [43]	-	-	-	-	-	92.24	**-**	-	-	-	-	-	-	**-**	**90.86**
YOLOv8n [53]	-	-	-	86.7	75.1	80.9	83.2	78.9	82.7	-	-	-	82.5	73.7	79.2
YOLOv8s [53]	-	-	-	89.1	72.3	81.4	91.4	74.2	82.9	-	-	-	84.1	76.9	80.7
YOLOv8m [53]	-	-	-	81.6	75.2	79.6	**94.6**	77.1	88.6	-	-	-	83.3	80.6	81.0
YOLOv8l [53]	-	-	-	89.7	76.9	82.9	91.1	80.8	87.2	-	-	-	80.0	75.9	79.9
YOLOv8x [53]	-	-	-	82.8	80.0	81.1	92.8	75.4	88.0	-	-	-	76.1	78.8	79.5
EfficientDet [53]	-	-	-	-	-	-	**-**	-	88.3	-	-	-	-	**-**	**-**
YOLOv7 [54]	-	-	-	80.6	83.3	75.0	**-**	-	-	-	-	-	-	**-**	**-**
YOLO-SRPD [55]	-	-	-	90.2	82.7	89.0	86.7	88.6	89.4	-	-	-	**89.6**	**87.4**	88.8

TABLE III. PRECISION, RECALL, AND MAP@50 SCORE FOR LDPOLYPVIDEO DATASET FOR YOLOV9 VARIANTS

Models	LDPolypVideo		
	P	**R**	**mAP@50**
Gelan-c	72.35	39.19	49.14
Gelan-e	**76.85**	42.77	53.16
YOLOv9-c	75.56	45.09	54.77
YOLOv9-e	76.38	**46.93**	**55.56**

VI. CONCLUSION AND FUTURE WORK

This study presented a comprehensive analysis of various YOLOv9 models on polyp detection. The best mAP@50 results were obtained in most cases for the Gelan-e YOLOv9 model variant. The Gelan-e model showed the best results for Endoscene and CVC-ClinicDB with mAP@50 values of 97 and 99.1%, respectively. YOLOv9-c acquired a mAP@50 value of 92.7% on Kvasir-SEG, whereas Gelan-c acquired the first place with a mAP@50 value of 86% when evaluated across the ColonDB dataset. In addition, the performance of different YOLOv9 variants was investigated for polyp detection on the LDPolypVideo dataset. Experiments showed that YOLOv9-e was the best model with a mAP@50 value of 55.56%. Future work will investigate the effect of other object detection models (YOLOv11, YOLOv12) on polyp detection. The authors also intend to increase the number of datasets and curate a large dataset for generalized model training.

DATA AVAILABILITY

Data will be available on request.

ACKNOWLEDGMENTS

This paper is based on a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 13382-psu-2023-PSNU-R-3-1-EI-. The authors would like to acknowledge the support of Prince Sultan University, Riyadh, Saudi Arabia, in

paying the article processing charges for this publication. This research was also supported by the Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia.

REFERENCES

[1] U. Hameed, M. Ur Rehman, A. Rehman, R. Damaševičius, A. Sattar, and T. Saba, "A deep learning approach for liver cancer detection in CT scans," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 7, Jan. 2024, Art. no. 2280558, https://doi.org/10.1080/21681163.2023.2280558.

[2] H. Inbarani H., A. T. Azar, and J. G, "Leukemia Image Segmentation Using a Hybrid Histogram-Based Soft Covering Rough K-Means Clustering Algorithm," *Electronics*, vol. 9, no. 1, Jan. 2020, Art. no. 188, https://doi.org/10.3390/electronics9010188.

[3] S. Al-Otaibi, M. Mujahid, A. R. Khan, H. Nobanee, J. Alyami, and T. Saba, "Dual Attention Convolutional AutoEncoder for Diagnosis of Alzheimer's Disorder in Patients Using Neuroimaging and MRI Features," *IEEE Access*, vol. 12, pp. 58722–58739, 2024, https://doi.org/10.1109/ACCESS.2024.3390186.

[4] P. K. N. Banu, A. T. Azar, and H. H. Inbarani, "Fuzzy firefly clustering for tumour and cancer analysis," *International Journal of Modelling, Identification and Control*, vol. 27, no. 2, pp. 92–103, Jan. 2017, https://doi.org/10.1504/IJMIC.2017.082941.

[5] S. R. Waheed, N. M. Suaib, M. S. M. Rahim, A. R. Khan, S. A. Bahaj, and T. Saba, "Synergistic Integration of Transfer Learning and Deep Learning for Enhanced Object Detection in Digital Images," *IEEE Access*, vol. 12, pp. 13525–13536, 2024, https://doi.org/10.1109/ACCESS.2024.3354706.

[6] A. Koubaa, A. Ammar, M. Alahdab, A. Kanhouch, and A. T. Azar, "DeepBrain: Experimental Evaluation of Cloud-Based Computation Offloading and Edge Computing in the Internet-of-Drones for Deep Learning Applications," *Sensors*, vol. 20, no. 18, Sep. 2020, Art. no. 5240, https://doi.org/10.3390/s20185240.

[7] H. I. Elshazly, A. M. Elkorany, A. E. Hassanien, and A. T. Azar, "Ensemble classifiers for biomedical data: Performance evaluation," in *2013 8th International Conference on Computer Engineering & Systems (ICCES)*, Cairo, Egypt, Nov. 2013, pp. 184–189, https://doi.org/10.1109/ICCES.2013.6707198.

[8] J. H. Bond and P. P. C. of the A. C. of Gastroenterology, "Polyp Guideline: Diagnosis, Treatment, and Surveillance for Patients With Colorectal Polyps," *Official journal of the American College of Gastroenterology | ACG*, vol. 95, no. 11, pp. 3053-3063, Nov. 2000.

[9] Y. Hao, Y. Wang, M. Qi, X. He, Y. Zhu, and J. Hong, "Risk Factors for Recurrent Colorectal Polyps," *Gut and Liver*, vol. 14, no. 4, pp. 399–411, Jul. 2020, https://doi.org/10.5009/gnl19097.

[10] N. Shussman and S. D. Wexner, "Colorectal polyps and polyposis syndromes," *Gastroenterology Report*, vol. 2, no. 1, pp. 1–15, Feb. 2014, https://doi.org/10.1093/gastro/got041.

[11] "Global burden of cancer attributable to HIV: a worldwide incidence analysis," *IARC*. https://www.iarc.who.int/cancer-type/colorectal-cancer.

[12] L. L. Marchand, L. R. Wilkens, L. N. Kolonel, J. H. Hankin, and L. C. Lyu, "Associations of Sedentary Lifestyle, Obesity, Smoking, Alcohol Use, and Diabetes with the Risk of Colorectal Cancer1," *Cancer Research*, vol. 57, no. 21, pp. 4787–4794, Nov. 1997.

[13] K. Hicham, S. Laghmati, B. Cherradi, S. Hamida, and A. Tmiri, "Enhancing Colorectal Polyps Detection using Transfer Learning on DICOM Metadata," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19417–19423, Feb. 2025, https://doi.org/10.48084/etasr.9024.

[14] K. ELKarazle, V. Raman, P. Then, and C. Chua, "Detection of Colorectal Polyps from Colonoscopy Using Machine Learning: A Survey on Modern Techniques," *Sensors*, vol. 23, no. 3, Jan. 2023, Art. no. 1225, https://doi.org/10.3390/s23031225.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91.

[16] C. Y. Wang, I. H. Yeh, and H. Y. Mark Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," in *Computer Vision – ECCV 2024*, 2025, pp. 1–21, https://doi.org/10.1007/978-3-031-72751-1_1.

[17] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525, https://doi.org/10.1109/CVPR.2017.690.

[18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018, https://doi.org/10.48550/arXiv.1804.02767.

[19] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv, Apr. 23, 2020, https://doi.org/10.48550/arXiv.2004.10934.

[20] G. Jocher *et al.*, "ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements." Zenodo, Oct. 29, 2020, https://doi.org/10.5281/ZENODO.4154370.

[21] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "You Only Learn One Representation: Unified Network for Multiple Tasks." arXiv, May 10, 2021, https://doi.org/10.48550/arXiv.2105.04206.

[22] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021." arXiv, Aug. 06, 2021, https://doi.org/10.48550/arXiv.2107.08430.

[23] C. Li *et al.*, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications." arXiv, 2022, https://doi.org/10.48550/ARXIV.2209.02976.

[24] 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 *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, Jun. 2023, pp. 7464–7475, https://doi.org/10.1109/CVPR52729.2023.00721.

[25] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, https://doi.org/10.3390/make5040083.

[26] J. B. Nadar, "YOLO-NAS: A Game-Changer in Object Detection with Deci AI's Neural Architecture Search Technology," *Medium.com*, May 30, 2023. https://medium.com/aimonks/yolo-nas-a-game-changer-in-object-detection-with-deci-ais-neural-architecture-search-technology-66e41ee9b3a0.

[27] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, Jul. 2015, https://doi.org/10.1016/j.compmedimag.2015.02.007.

[28] D. Jha *et al.*, "Kvasir-SEG: A Segmented Polyp Dataset," in *MultiMedia Modeling*, 2020, pp. 451–462, https://doi.org/10.1007/978-3-030-37734-2_37.

[29] Y. Ma, X. Chen, K. Cheng, Y. Li, and B. Sun, "LDPolypVideo Benchmark: A Large-Scale Colonoscopy Video Dataset of Diverse Polyps," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021, pp. 387–396, https://doi.org/10.1007/978-3-030-87240-3_37.

[30] I. Pacal *et al.*, "An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets," *Computers in Biology and Medicine*, vol. 141, Feb. 2022, Art. no. 105031, https://doi.org/10.1016/j.compbiomed.2021.105031.

[31] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 1571–1580, https://doi.org/10.1109/CVPRW50498.2020.00203.

[32] K. Hu *et al.*, "Colorectal polyp region extraction using saliency detection network with neutrosophic enhancement," *Computers in Biology and*

Medicine, vol. 147, Aug. 2022, Art. no. 105760, https://doi.org/10.1016/j.compbiomed.2022.105760.

[33] C. Biffi, P. Salvagnini, N. N. Dinh, C. Hassan, P. Sharma, and A. Cherubini, "A novel AI device for real-time optical characterization of colorectal polyps," *npj Digital Medicine*, vol. 5, no. 1, Jun. 2022, Art. no. 84, https://doi.org/10.1038/s41746-022-00633-6.

[34] "GI Genius™ Intelligent Endoscopy Module." https://www.medtronic.com/en-us/healthcare-professionals/products/digestive-gastrointestinal/gastrointestinal-artificial-intelligence/gi-genius-intelligent-endoscopy-module.html.

[35] A. Nogueira-Rodríguez *et al.*, "Real-time polyp detection model using convolutional neural networks," *Neural Computing and Applications*, vol. 34, no. 13, pp. 10375–10396, Jul. 2022, https://doi.org/10.1007/s00521-021-06496-4.

[36] A. Ellahyani, I. E. Jaafari, S. Charfi, and M. E. Ansari, "Fine-tuned deep neural networks for polyp detection in colonoscopy images," *Personal and Ubiquitous Computing*, vol. 27, no. 2, pp. 235–247, Apr. 2023, https://doi.org/10.1007/s00779-021-01660-y.

[37] S. Grosu *et al.*, "Machine Learning–based Differentiation of Benign and Premalignant Colorectal Polyps Detected with CT Colonography in an Asymptomatic Screening Population: A Proof-of-Concept Study," *Radiology*, vol. 299, no. 2, pp. 326–335, May 2021, https://doi.org/10.1148/radiol.2021202363.

[38] Z. Haider, A. T. Azar, and T. Saba, "Data Augmentation and Optimizer Tuning for Polyp Segmentation," in *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, Paris, France, May 2024, pp. 1–6, https://doi.org/10.1109/ICCAD60883.2024.10554050.

[39] C. Ma, H. Jiang, L. Ma, and Y. Chang, "A Real-Time Polyp Detection Framework for Colonoscopy Video," in *Pattern Recognition and Computer Vision*, 2022, pp. 267–278, https://doi.org/10.1007/978-3-031-18907-4_21.

[40] P. Carrinho and G. Falcao, "Highly accurate and fast YOLOv4-based polyp detection," *Expert Systems with Applications*, vol. 232, Dec. 2023, Art. no. 120834, https://doi.org/10.1016/j.eswa.2023.120834.

[41] T. Yu *et al.*, "An end-to-end tracking method for polyp detectors in colonoscopy videos," *Artificial Intelligence in Medicine*, vol. 131, Sep. 2022, Art. no. 102363, https://doi.org/10.1016/j.artmed.2022.102363.

[42] G. Polat, E. Işık Polat, K. Kayabay, and A. Temizel, "Polyp Detection in Colonoscopy Images using Deep Learning and Bootstrap Aggregation," in *IEEE International Symposium on Biomedical Imaging, Endoscopy Detection and Segmentation Workshop (EndoCV2021)*, Apr. 2021.

[43] M. Souaidi, S. Lafraxo, Z. Kerkaou, M. El Ansari, and L. Koutti, "A Multiscale Polyp Detection Approach for GI Tract Images Based on Improved DenseNet and Single-Shot Multibox Detector," *Diagnostics*, vol. 13, no. 4, Jan. 2023, Art. no. 733, https://doi.org/10.3390/diagnostics13040733.

[44] D. Vázquez *et al.*, "A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images," *Journal of Healthcare Engineering*, vol. 2017, no. 1, 2017, Art. no. 4037190, https://doi.org/10.1155/2017/4037190.

[45] 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, Oct. 2016, https://doi.org/10.1109/TMI.2015.2487997.

[46] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, no. 2, pp. 283–293, Mar. 2014, https://doi.org/10.1007/s11548-013-0926-3.

[47] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," in *2015 IEEE Information Theory Workshop (ITW)*, Jerusalem, Israel, Apr. 2015, pp. 1–5, https://doi.org/10.1109/ITW.2015.7133169.

[48] Y. Cai *et al.*, "Reversible Column Networks." arXiv, Feb. 01, 2023, https://doi.org/10.48550/arXiv.2212.11696.

[49] Y. Chen *et al.*, "SdAE: Self-distillated Masked Autoencoder," in *Computer Vision – ECCV 2022*, vol. 13690, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Springer Nature Switzerland, 2022, pp. 108–124.

[50] Z. Shen, Z. Liu, J. Li, Y. G. Jiang, Y. Chen, and X. Xue, "Object Detection from Scratch with Deep Supervision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 398–412, Feb. 2020, https://doi.org/10.1109/TPAMI.2019.2922181.

[51] C. Y. Wang, H. Y. M. Liao, and I. H. Yeh, "Designing Network Design Strategies Through Gradient Path Analysis." arXiv, Nov. 09, 2022, https://doi.org/10.48550/arXiv.2211.04800.

[52] S. Fan *et al.*, "On line detection of defective apples using computer vision system combined with deep learning methods," *Journal of Food Engineering*, vol. 286, Dec. 2020, Art. no. 110102, https://doi.org/10.1016/j.jfoodeng.2020.110102.

[53] Md. F. Ahamed, Md. R. Islam, Md. Nahiduzzaman, Md. J. Karim, M. A. Ayari, and A. Khandakar, "Automated Detection of Colorectal Polyp Utilizing Deep Learning Methods With Explainable AI," *IEEE Access*, vol. 12, pp. 78074–78100, 2024, https://doi.org/10.1109/ACCESS.2024.3402818.

[54] L. T. T. Hong *et al.*, "Real-time detection of colon polyps during colonoscopy using YOLOv7," *Journal of Military Science and Technology*, no. CSCE7, pp. 122–134, Dec. 2023, https://doi.org/10.54939/1859-1043.j.mst.CSCE7.2023.122-134.

[55] S. Wang, J. Xie, Y. Cui, and Z. Chen, "Colorectal Polyp Detection Model by Using Super-Resolution Reconstruction and YOLO," *Electronics*, vol. 13, no. 12, Jan. 2024, Art. no. 2298, https://doi.org/10.3390/electronics13122298.