

## Review Article

# Artificial intelligence algorithms for real-time detection of colorectal polyps during colonoscopy: a review

Meng-Yuan Nie<sup>1,2</sup>, Xin-Wei An<sup>1,2</sup>, Yun-Can Xing<sup>3</sup>, Zheng Wang<sup>3</sup>, Yan-Qiu Wang<sup>4</sup>, Jia-Qi Lü<sup>1,2</sup>

<sup>1</sup>Center for Advanced Laser Technology, Hebei University of Technology, Tianjin, China; <sup>2</sup>Hebei Key Laboratory of Advanced Laser Technology and Equipment, Tianjin, China; <sup>3</sup>Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; <sup>4</sup>Langfang Traditional Chinese Medicine Hospital, Langfang, Hebei, China

Received July 17, 2024; Accepted November 14, 2024; Epub November 15, 2024; Published November 30, 2024

**Abstract:** Colorectal cancer (CRC) is one of the most common cancers worldwide. Early detection and removal of colorectal polyps during colonoscopy are crucial for preventing such cancers. With the development of artificial intelligence (AI) technology, it has become possible to detect and localize colorectal polyps in real time during colonoscopy using computer-aided diagnosis (CAD). This provides a reliable endoscopist reference and leads to more accurate diagnosis and treatment. This paper reviews AI-based algorithms for real-time detection of colorectal polyps, with a particular focus on the development of deep learning algorithms aimed at optimizing both efficiency and correctness. Furthermore, the challenges and prospects of AI-based colorectal polyp detection are discussed.

**Keywords:** Colorectal polyp, colorectal cancer, deep learning, convolutional neural networks

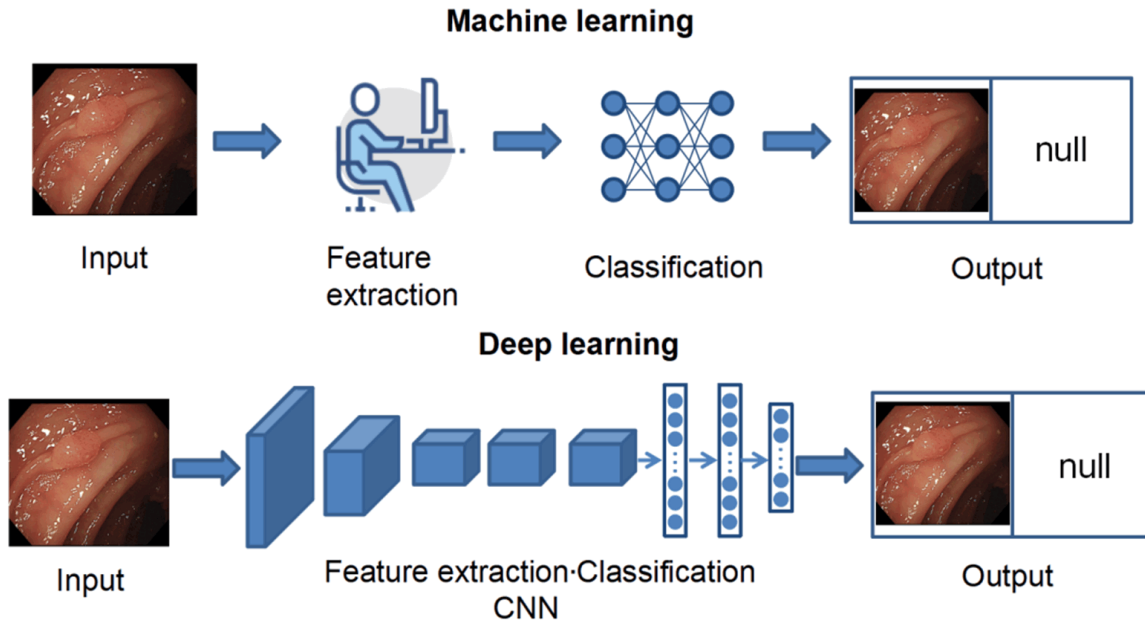
## Introduction

Colorectal cancer (CRC) accounts for 10% of all cancer types [1]. It has the third-highest incidence and second-highest mortality rate among cancers globally. The five-year survival rate for advanced-stage CRC is merely 14%. Colorectal polyps are precancerous lesions that have the potential to develop into colorectal cancer over a span of 5-10 years [2]. Therefore, early diagnosis of colorectal polyps can effectively prevent colorectal cancer. The adenoma detection rate (ADR) is a widely recognized quality indicator for colonoscopy. Medical evidence suggests that each percentage point increase in ADR correlates with a 3% to 6% decrease in interstitial colorectal cancer incidence [3]. Therefore, an effective increase in ADR can potentially reduce the incidence of colorectal cancer [4].

With the advancement of medical imaging technology, colonoscopy has emerged as the predominant method for detecting and diagnosing polyps. This procedure uses a camera and a

flexible tube to examine the bowel. It provides high-definition video of the bowel and is suitable for visually detecting pathological inflammation and colonic diseases [5]. However, traditional endoscopy relies on the physician's experience, making it time-consuming and prone to missing small and atypical polyps [6]. As artificial intelligence technology rapidly advances, computer-aided detection systems (CADE) have emerged as formidable tools to improve the detection rate and accuracy of polyps [7]. Repici et al. conducted a multicenter, randomized controlled study involving 685 subjects to compare high-definition colonoscopy with and without the CADE system. The results showed that the group using the CADE system exhibited a significantly higher ADR than the control group. This suggests that the CADE system can analyze endoscopic video and accurately identify and localize polyps, thus aiding physicians in decision-making [8].

In clinical applications, it is crucial for doctors to promptly detect colorectal polyps. However, gastroenterologists currently exhibit only 76%



**Figure 1.** Machine Learning and Deep Learning. Machine learning involves manual feature extraction, which is subsequently fed into a classifier, whereas deep learning has both feature extraction and classification done automatically by a convolutional neural network (CNN).

accuracy in detecting small polyps (less than 1 cm) in real-time optics [9], posing significant challenges in real-time polyp identification. Real-time processing of colorectal polyp detection algorithms can assist gastroenterologists in improving polyp detection rates [10]. Therefore, there is an increasing demand for real-time algorithms, and frame rate has become an important metric for measuring the merit of detection algorithms. This paper adopts frames per second (FPS) as an assessment metric for detection rates.

Traditional machine learning algorithms rely heavily on hand-crafted descriptors for feature learning. The hand-crafted features such as color, shape, texture, and edges are extracted and fed into a machine-learning classifier, which separates the lesion from the background [11]. In recent years, most colonoscopy studies on polyp detection have utilized convolutional neural networks (CNNs). Unlike traditional machine-learning algorithms, deep learning employs CNNs for both feature extraction and classification, enabling the automatic and efficient identification of complex patterns and features (Figure 1). This process eliminates the need for the tedious task of manually designing features. Moreover, the deep learning model consists of multiple layers of neural networks,

allowing them to learn complex non-linear relationships. This architecture facilitates the progressive extraction of high-level features from low-level ones, thereby enhancing extraction capabilities. As a result, deep learning has far surpassed traditional machine learning methods in accuracy.

This review explores recent advances in the real-time detection of colorectal polyps. We present the application of traditional machine learning and deep learning [12] algorithms in real-time colorectal polyp detection, emphasizing the potential of deep learning methods to improve both detection frame rates and detection accuracy. The integration of previous studies is used to analyze the future challenges and perspectives of real-time colorectal polyp detection algorithms.

#### Detecting and locating colorectal polyps

Colorectal polyp detection [13] is the process of identifying and localizing polyps in the colorectum using medical imaging techniques. The primary objective of polyp detection algorithms is to determine the presence of polyps in medical images accurately. These algorithms are helpful during colonoscopy when the detection system promptly notifies the physician

examining an acoustic signal or a logo as soon as the detection system suspects that the image or video may contain a polyp. Polyp detection algorithms can be broadly classified into two categories: (a) machine learning methods and (b) deep learning methods.

### *Machine learning methods*

Before the popularity of deep learning methods, traditional machine learning algorithms predominantly relied on hand-crafted descriptors for feature learning. Previous methods for detecting colorectal polyps primarily used shape features [14, 15], texture features [16-19], color features [20], edge features, or combinations of these features [21]. These features were then fed into a machine learning classifier designed to distinguish lesions from the background.

Hwang proposed an ellipse shape-based polyp detection method to identify the polyp region by least squares fitting of ellipses. This method achieved a detection rate of 15 FPS [15]. Ameling et al. analyzed more than four hours of high-resolution colonoscopy videos using four texture feature extraction methods based on grey-level co-occurrence matrices (GLCMs) and local binary patterns (LBPs). The results of the study indicated that the authors achieved classification results of up to 0.96 for the area under the receiver operating characteristic (ROC) curve [17]. Ševo proposed a model that uses texture analysis for the automatic detection of inflammation, suitable for real-time application and parallel processing. Experimental results demonstrate that the method can detect inflammatory regions in real time with more than 84% accuracy. In some video frame segments, the detection accuracy can reach more than 90% [18]. Iakovidis et al. conducted a comparative study of texture features for gastric polyp detection in endoscopic videos. Among the four texture feature extraction methods, texture spectrum (TS) histogram, texture spectrum and color histogram statistics (TSCHS), LBP histogram, and color wavelet covariance (CWC), the CWC gave the best results with an area under the ROC curve of 88.6% [19].

Giritharan et al. utilized various color and texture features to detect bleeding lesions in video frames. The experimental results showed their

method exhibited high sensitivity and subjectivity [21]. Wang et al. introduced a software system called "Polyp-Alert", which aims to assist endoscopists in detecting polyps by providing visual feedback during colonoscopy. The system employs previous edge cross-section visual features and a rule-based classifier to detect polyp edges. The software correctly detected 97.7% (42/43) of polyp shots on 53 randomly selected video files of complete colonoscopies, achieving polyp detection at rates up to 10 FPS [23]. Kominami et al. researched and developed a real-time image recognition system for predicting the histological diagnosis of colorectal lesions in narrow-band imaging. This system analyzes the region of interest (ROI) in narrow-band imaging (NBI) video endoscopic images and displays the output values of the Support Vector Machine (SVM) in real time. It achieves an accuracy of 94.9% at a frame rate of 20 FPS [22]. However, the extraction of these features is typically performed manually, resulting in a lack of robustness and being time-consuming. Furthermore, these methods do not effectively detect polyps in real time and are associated with a high false-positive rate. Therefore, reliably and accurately detecting colorectal polyps in real time remains a significant challenge.

### *Deep learning methods*

With the wide application of deep learning in medical image processing [24-28], deep learning-based polyp detection algorithms have been proposed in recent years. CNN represents a critical architecture within deep learning. CNN-based detectors can automatically extract abstract and discriminative features compared to manual feature extraction. Bernal et al. compared the effectiveness of manually crafted features with CNN-extracted features in detecting polyps. They claimed that the CNN-based approach exhibits superior performance [29]. Tajbakhsh et al. proposed a unique three-way image representation that captures the color, texture, shape, and temporal information of colon polyps. The method employs CNN to learn polyp features at multiple scales to improve the accuracy of polyp localization. This method significantly reduced the false alarm rate and the delay in polyp detection compared to earlier techniques [30]. In 2018, Zhang et al. proposed a new regression-based CNN pipeline for polyp detection in colonoscopy.

However, it achieved a processing speed of only 6.5 FPS, which was inadequate for real-time applications [31]. In 2019, Jiang et al. designed a deep learning model called the “Artificial Intelligence Endoscopist (AI-doscopist)” for localization during colonoscopy. The system was constructed based on ResNet50 and YOLO (you only look once). YOLOv2 has certain advantages in real-time detection. Combined with the powerful feature extraction capability of ResNet50, the system achieves a good balance between detection speed and accuracy. The results indicated that the AI-doscopist successfully localized 124 out of 128 polyps (polyp base sensitivity of 96.9%) in 144 complete colonoscopies with 93.3% specificity [32].

Although the above studies are more robust than traditional manual extraction methods, their diagnostic performance and processing speed remain inadequate for real-time clinical applications. With the increasing development of deep learning, more algorithms have been proposed to meet the clinical real-time requirements. Among these real-time algorithms, the optimization can be divided into two directions: improving the detection speed and improving the detection correctness.

### *Algorithms optimization for improving targeting frame rate*

In colonoscopy, real-time algorithms typically require a detection rate of 25 FPS or higher. This detection rate is essential for the algorithm to process each image or video frame promptly, thereby providing timely feedback to the physician. In this section, we summarize the methods used to improve the speed of algorithmic detection in the reviewed papers. The following methods can be used alone or in combination to increase the frame rate of the algorithm detection.

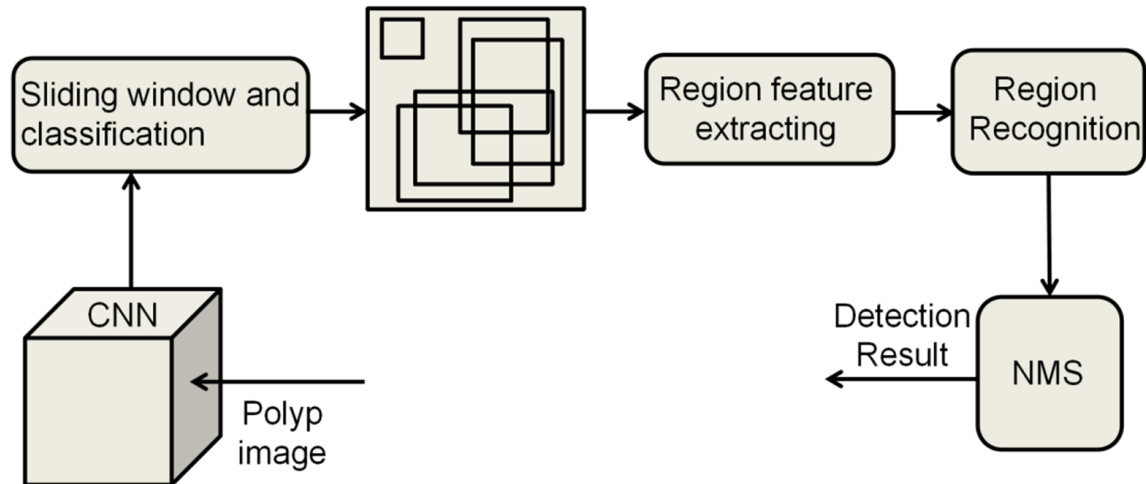
*Use of anchor-free detection algorithms:* An anchor-based detection algorithm uses predefined anchor frames to regress and classify targets. These anchor frames are generated by clustering through methods such as K-means before training and represent the main distribution of targets in the dataset. While anchor-based detection algorithms are the current mainstream direction, anchor-free detection algorithms eliminate the need for predefined

anchor frames. These algorithms detect targets by directly predicting the critical point or center of the object, thereby reducing computational load and enhancing detection speed. Consequently, they have gained significant attention in recent years [33-35]. Some researchers have successfully applied the anchor-free method to real-time colonoscopy detection and achieved satisfactory results.

Yang et al. proposed the you only look once-objectbox (YOLO-OB) model, which employs the ObjectBox detection head and utilizes a center-based anchor-free frame regression strategy. The algorithm achieved a detection rate of 39 FPS on the RTX3090 graphics card [36]. Wang et al. proposed a new anchor-free polyp net (AFP-Net) that formulates objects as centroids. They removed the center pool in CenterNet [37] and replaced it with a context-enhanced module and feature pyramid design for real-time response. Additionally, a particular cosine truth projection strategy was designed to compensate for the decrease in recall due to the removal of the anchor mechanism. The method can achieve a detection speed of 52.6 FPS [38].

*Use of lightweight network architecture:* Lightweight networks seek to decrease the number of model parameters and complexity while preserving model accuracy, thereby emerging as a prominent research focus in computer vision [39-41]. These models are capable of operating on resource-constrained devices and enabling real-time detection, which is crucial for the swift identification of polyps during colonoscopy. Consequently, several researchers have started to implement lightweight models for the real-time detection of colorectal polyps.

Since YOLOv5s exhibits superior feature extraction capability, Ou et al. chose to build on this algorithm by reducing the number of convolutional kernels and removing the large target detection head to design a lightweight model that enables real-time detection of polyps called Polyp-YOLOv5-Tiny. Although this model has a slight loss of accuracy, it has a significant advantage in model size and detection rate. The algorithm can achieve a detection rate of 113.6 FPS [42]. Yoo et al. proposed a new lightweight model that replaces the neck of a CNN by incorporating a Token-Sharing Transformer (TST) into YOLOv5, called YOLOv5-TST. The TST module utilizes the attention mechanism of the



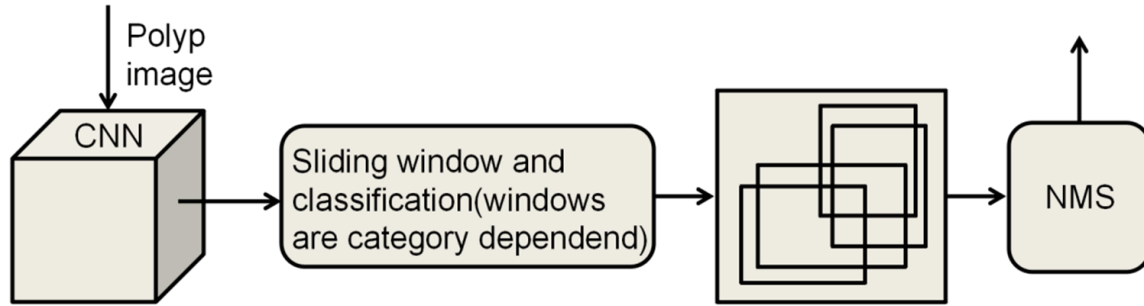
**Figure 2.** Principle of two-stage object detection algorithm. Two-stage Object Detection algorithms usually consist of two main stages: Region Proposal and Classification and Regression. CNN, convolutional neural network; NMS, non-maximum suppression.

Transformer to fuse local and global features. This efficient feature fusion reduces redundant computations and improves the overall efficiency of the model. The Transformer architecture naturally supports parallel computation, which allows for faster processing of large-scale data compared to traditional CNNs. This model reduces the number of parameters in the neck without significant performance degradation based on YOLOv5, achieving a detection speed of 138.3 FPS [43]. The Polyp-YOLOv5-Tiny model was reproduced in that literature, and the two models were compared using a unified dataset. The results indicate that on the Kvasir dataset, YOLOv5m-TST achieves a precision of 0.9369, while Polyp-YOLOv5-Tiny achieves a precision of 0.9072.

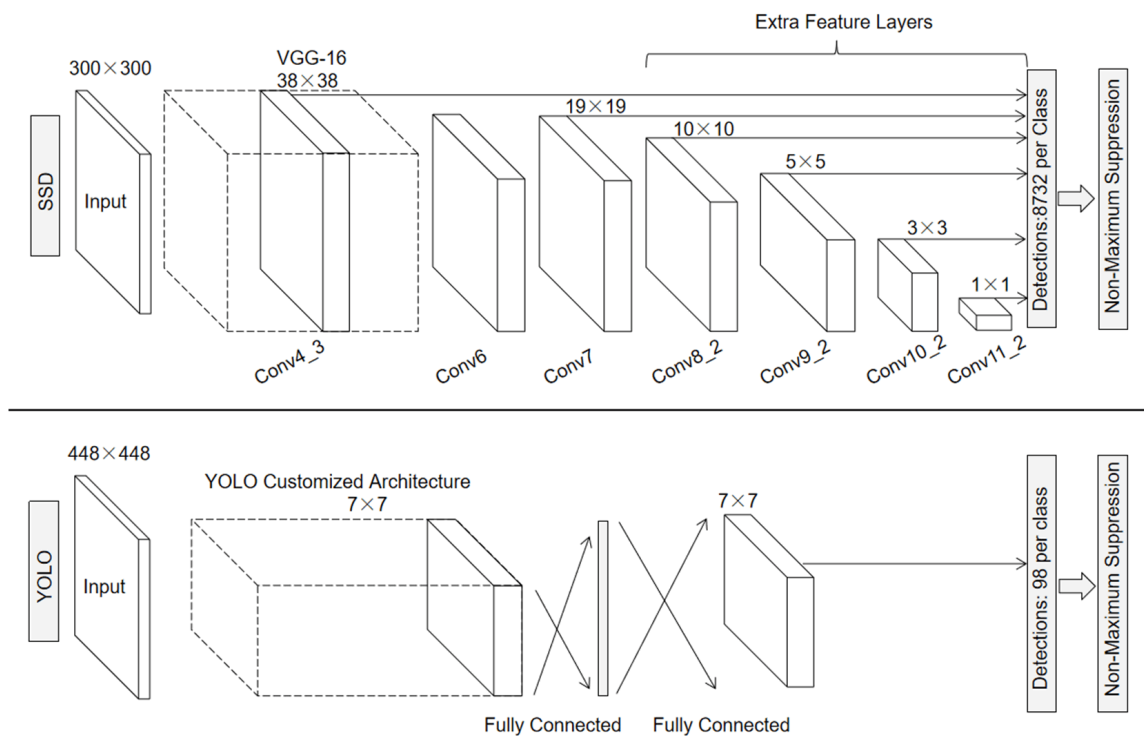
*Use of algorithms based on one-stage detection methods:* The current object detection algorithms are divided into two main types: one-stage [44] and two-stage algorithms [45]. The two-stage object detection algorithm operates in two distinct phases to accomplish detection. Initially, it generates a set of candidate regions that are likely to encompass the target object, a process known as region proposal. Subsequently, each of these regions undergoes precise classification and localization, allowing the algorithm to attain a high level of detection accuracy. Finally, non-maximum suppression (NMS) is used to remove redundant bounding boxes (**Figure 2**). Classical

two-stage object detection algorithms include region-based convolutional neural network (R-CNN) and Faster R-CNN. In contrast, one-stage object detection algorithms perform target classification and localization in a single forward pass. The primary characteristic of these algorithms is the omission of the region proposal stage, allowing for the direct generation of class probabilities and positional coordinates of the object, thereby achieving faster detection speeds (**Figure 3**).

The main classical one-stage detectors include YOLO, SSD (Single Shot MultiBox Detector), and RetinaNet [48]. A study in 2021 comparing the performance of different algorithms for polyp detection showed that, in this benchmarking, the YOLO network achieves state-of-the-art performance while operating in real-time [49]. YOLO's network structure consists of 24 convolutional layers and two fully connected layers, with the convolutional layers dedicated to feature extraction. The  $1 \times 1$  and  $3 \times 3$  convolutions are used alternately in the convolutional layers, where  $1 \times 1$  convolution is used to reduce the dimensionality of the feature space and  $3 \times 3$  convolution is used to extract the features. The convolutional layer is followed by two fully connected layers to further process the features and generate the final detection results. The network accepts input images sized at  $448 \times 448$  pixels, and the final output is a tensor with dimensions of  $7 \times 7 \times 30$ . The primary



**Figure 3.** Principle of the one-stage object detection algorithm. One-stage Object Detection algorithm is faster as it generates the class and position coordinates of the target directly from the input image, eliminating the region proposal stage. CNN, convolutional neural network; NMS, non-maximum suppression.



**Figure 4.** Network structure of YOLO (you only look once) algorithm and SSD (single shot multibox detector) algorithm (Adapted with permission from [47]). The YOLO algorithm extracts feature by means of 24 convolutional layers and 2 fully-connected layers, the input image is usually of size 448×448. The detection layer divides the image into S×S grids (usually 7×7), each of which predicts B bounding boxes (usually 2) along with their confidence and category probabilities. Finally, the most appropriate bounding boxes are filtered to remove overlapping boxes. SSD uses VGG-16 to extract the feature maps, the input image is typically 300×300 in size, and it adds extra feature layers on top of the base network to generate multi-scale feature maps.

YOLO mode processes images in real time at a detection rate of 45 FPS [46]. In comparison to the YOLO network, SSD adds multiple feature layers at the end of the base network, which are used to predict the default frames at different scales and aspect ratios (Figure 4). This enables SSD to effectively handle objects of different sizes and detect them at various

scales, achieving a detection rate of 59 FPS [47].

Lee et al. developed and validated a deep learning algorithm for polyp detection utilizing YOLOv2 on a dataset comprising 8,075 images, which included 503 polyps. The algorithm was validated using three datasets, which detected

**Table 1.** Comparison of deep learning algorithms for improving frame rate

Categories	Method	Dataset	Result	Reference
Machine learning methods	Manual extraction of features	-	FPS: <20	[14-23]
Early deep learning algorithms	CNN	-	FPS: <28	[30-32]
Use of anchor-free detection algorithms	YOLO-OB	both	FPS: 39	[36]
	AFP-Net	Public	FPS: 52.6	[38]
Use of lightweight network architecture	Polyp-YOLOv5-Tiny	Public	FPS: 113.6	[42]
	YOLOv5m-TST	Public	FPS: 138.3	[43]
Use of algorithms based on one-stage detection methods	SSD based CNN	Public	FPS: 32	[52]
	SSDGPNet	Private	FPS: 50	[53]
	MP-FSSD	Public	FPS: 62.5	[54]
	YOLOv2	Both	FPS: 67.16	[50]
	YOLOv4	Public	FPS: 122	[51]

CNN, convolutional neural network; YOLO, you only look once; YOLO-OB, you only look once-objectbox; AFP-Net, anchor-free polyp net; TST, token-sharing transformer; SSD, single shot multibox detector; SSDGPNet, single shot multibox detector for gastric polyps; MP-FSSD, multiscale pyramidal fusion single-shot multibox detector network.

all 38 polyps and seven additional polyps detected by the endoscopist. The operation speed was 67.16 FPS [50]. Pacal et al. proposed a method for real-time automated polyp detection on a small dataset using a model based on YOLOv4. The method significantly improves the performance of polyp detection using various data augmentations, NVIDIA TensorRT, and integrated learning models. Unlike YOLOv2 and YOLOv3, the architecture of this network is modified on Darknet-53, and the Cross-Stage-Partial-connections (CSP) added to Darknet-53 helps to reduce the computational effort of the model while maintaining the same accuracy. Their proposed method achieves a deferral time of 8.2 ms and a detection rate of 122 FPS on a single RTX 2080 TI graphics card. When utilizing the Nvidia TensorRT framework, this delay decreases to below 4 milliseconds, achieving speeds exceeding 250 FPS [51].

In addition to the YOLO algorithm, SSD has shown excellent performance in real time. Liu et al. implemented a polyp detection method based on SSD, where they evaluated three different feature extractors, including ResNet50, VGG-16, and Inception-V3, with an additional feature layer on top of the base network. Among these feature extractors, Inception-V3 obtained the highest precision and recall, achieving a detection speed of 32 FPS, exceeding the minimum requirements for clinical applications [52]. Zhang et al. reported a CNN for polyp detection built on an SSD architecture called single shot multibox for gastric polyps

(SSDGPNet). The results indicated that the enhanced SSD can achieve real-time gastric polyp detection at 50 FPS [53]. Souaidi et al. proposed a deep polyp detection method based on multiscale pyramidal fusion single-shot multibox detector network (MP-FSSD). This model introduces an edge pooling layer, a splicing module, and a down sampling block on top of the SSD to generate a new pyramid layer and improve the detection accuracy and speed. The results show that the algorithm can achieve a test speed of 62.5 FPS [54]. In this section, we summarise several methods to improve the detection rate of the algorithm and review related papers. For better comparison, we present the results in **Table 1**.

*Algorithms optimization for improving detection correctness*

In real-time colorectal polyp detection, a high correct rate is essential for minimizing resource waste associated with repeated examinations and enhancing the efficiency of the healthcare system. So many methods dedicated to improving the correct rate of real-time colorectal polyp detection algorithms have been proposed. We present these algorithms in five categories. Before that, we provide a brief overview of the evaluation metrics utilized in the reviewed literature.

The accuracy of a model refers to the percentage of correct predictions made across all instances. Precision indicates the proportion of samples predicted by the model to be posi-

tive cases that are actually positive cases. Sensitivity, or recall, indicates the proportion of positive cases that are correctly predicted by the model in all samples that are actually positive cases. Specificity is the proportion of negative samples correctly identified by the model. Intersection over Union (IoU) is used to measure the degree of overlap between the predicted frame and the real frame. The higher the IoU, the closer the predicted frame is to the real frame. By changing the threshold value, the relationship curve between precision and recall is plotted, and the area under the precision-recall curve is the average precision (AP). AP is calculated as follows. Taking the average of the APs for all categories is the mean Average Precision (mAP). mAP@0.5 indicates that the precision-recall curve is plotted, and the area under the curve is computed under an IoU threshold of 0.5.

*Centralizing multiple models in a single algorithm:* Integrated learning is an approach aimed at enhancing efficiency by integrating the results of multiple classification models into a single high-quality classifier. This method strikes a balance between training speed and accuracy by combining the prediction results of multiple weak learners, thereby significantly improving the overall performance of the model and compensating for the shortcomings of a single model [55].

Zhao et al. developed the Adaptive Small Object Detection Ensemble (ASODE) model. They combined SSD with excellent feature extraction capability and detection rate with adaptive lightweight YOLOv4 to improve the accuracy of target polyp detection without significantly increasing the memory footprint of the model. The model achieves an adenoma detection accuracy of 92.70% in video analysis [56]. Ma et al. propose a method for real-time polyp inspection from colonoscopy videos. They enhanced the overall performance of their algorithm by integrating Swin transformer blocks into a CNN-based YOLOv5m network. Their method demonstrated a 5.3% improvement in accuracy compared to the baseline network, achieving an accuracy of 83.6% on the CVC-ClinicalVideoDB dataset [57]. Sharma et al. integrated ResNet, GoogLeNet, and Xception into a powerful model for the prediction of video frames extracted from colonoscopy, achieving

a precision of 98.6% and an accuracy of 98.3%. Furthermore, this model can distinguish between cancerous and non-cancerous polyps [58].

*Use of three-dimensional method:* Although deep learning has made significant progress in medical image processing tasks, the majority of existing research has predominantly focused on utilizing two-dimensional (2D) deep learning to solve 2D image analysis challenges. Recently, some researchers have advocated for the adoption of three-dimensional (3D) deep learning methods for detection and segmentation tasks involving medical data [59-63]. These works have shown that 3D methods can achieve better performance than 2D methods when dealing with 3D medical data [64], as they can capture more spatial information and fully utilize the 3D spatial information to generate more discriminative features, thereby enhancing detection accuracy. These works have motivated researchers to explore the feasibility of 3D deep learning methods in colonoscopy video processing.

Yu et al. proposed a three-dimensional fully convolutional network (3D-FCN), which converts the fully-connected layer in three-dimensional convolutional neural network (3D-CNN) to a convolutional layer. Their approach reduces redundant computations and speeds up detection compared to traditional sliding window methods. By leveraging an integrated approach that merges online and offline strategies, this method enhances the encoding of spatiotemporal information within videos. The results indicated that the algorithm achieved a precision of 88.1% [65]. Misawa et al. developed an AI-assisted system using a 3D-CNN, which demonstrated superior performance on video datasets compared to other deep learning methods and could be applied in a clinical setting [66]. The following year, the team reconstructed the CADe system and evaluated its performance. The results showed that the developed AI system exhibited high sensitivity regardless of polyp size and morphology, suggesting its potential for automated polyp detection [67].

*Use of CNN-based models:* Although contemporary algorithms such as SSD and YOLO demonstrate superior speed, they exhibit lower detec-



tion accuracy. Additionally, one-stage frameworks typically perform worse than two-stage architectures when detecting small objects. Therefore, some CNN-based models have been proposed that are capable of automatically extracting multi-level features from images, thereby enhancing the precision of detection results [68].

Urban et al. utilized CNNs for computer-aided image analysis to improve polyp detection. The algorithm was tested on manually labeled images with an accuracy of 96.4% for polyp identification [69]. Most existing CNN methods detect polyps on each frame independently, which may lead to a jittery effect and reduce the detection accuracy. Some studies [31, 70, 71] exploited the temporal dependency between consecutive frames to improve detection performance, thereby increasing accuracy and reducing false alarms. In addition, Zhang et al. employed a tracker to refine the detection results of each frame generated by the CNN, which can significantly reduce jitter. However, their proposed method requires accurate polyp detection in the initial frame, which is difficult to achieve [31]. Therefore, Zheng et al. proposed an optical flow model combined with an on-the-fly trained convolutional neural network (OptCNN), which combines a real-time trained CNN model with a spatial voting algorithm to improve the detection results of a single-frame polygon detector [72]. Livovsky et al. proposed a novel RetinaNet-based polyp detection system DEEP<sup>2</sup>, which uses a CNN for target detection, outputs a list of bounding boxes containing polyps, and filters and aggregates the bounding boxes through a temporal logic layer. This approach utilizes knowledge from the previous frame to assist in current detection [73].

*Detection of small-sized polyps:* Small-sized polyps are more likely to be overlooked in real-time colonoscopy detection. As a result, more methods have been proposed for the detection of small-sized polyps. SSD mainly utilizes shallow features for predictions that lack semantic information. Consequently, traditional SSD architectures are not well-equipped to capture both local detailed features and global semantic features, resulting in poor performance on small objects. Souaidi et al.'s polyp detection method, based on MP-FSSD, utilizes a multi-scale feature fusion approach that enhances

the network's ability to detect small targets. The mean average precision can reach 91.56% [54]. Fu et al. proposed a network for real-time colorectal polyp detection and diagnosis called D2polyp-Net. This network utilizes a double pyramid structure that combines shallow spatial information and deep semantic information to improve polyp localization accuracy. The detection precision can be up to 80.1% and is particularly effective for small-sized polyp detection [74]. Additionally, the novel polyp detection system DEEP<sup>2</sup> based on RetinaNet proposed by Livovsky et al. is also well-suited for small-size polyp detection. RetinaNet introduces a loss function called Focal Loss, which effectively addresses the issue of category imbalance and performs well, especially in detecting small targets [73]. Wan et al. proposed a YOLOv5 model for polyp detection based on a self-attention mechanism. This algorithm integrates an attention mechanism into the feature extraction process to enhance the contribution of information-rich feature channels while diminishing the interference from irrelevant channels. The experimental results showed that the method has a high accuracy in detecting small polyps and polyps with insignificant contrast [75].

*Improving reprocessing methods:* In the detection of colorectal polyps, post-processing methods can improve the accuracy of the detection model by eliminating false positives and reducing misses when detecting colorectal polyps. Lee et al. employed a median filter as a post-processing method to minimize false alarms, because the median filter removes impulse noise from the signal while retaining edge information well. The sensitivity of the algorithm for polyp detection was 96.7% and 90.2% on the two image datasets and 87.7% on the video dataset. Additionally, the false positive rate was reduced from 12.5% to 6.3% after using the median filter [50]. Nogueira et al. present a deep learning model for real-time polyp detection based on the pre-trained YOLOv3 architecture, which employs a deeper network structure (Darknet-53) to further improve feature extraction compared to the model of Lee et al. The model reduced false positives through a post-processing step based on a target tracking algorithm, achieving a frame rate of approximately 24 FPS [76]. Krenzer et al. proposed an automated polyp detection system called

**Table 2.** Comparison of deep learning algorithms for improving correctness

Categories	Method	Dataset	Result	Reference
Machine learning methods	Manual extraction of features	-	Accuracy: <93.2% AUC: <0.96	[14-23]
Early deep learning algorithms	CNN	-	Precision: <88.6%	[30-32]
Centralizing multiple models in a single algorithm	YOLOv4+SSD	Private	Accuracy: 92.7% Sensitivity: 87.96% (image) 92.31% (video)	[56]
	YOLOv5m	Public	Accuracy: 83.6%	[57]
	CNN	All	Accuracy: 98.3% Precision: 98.6%	[58]
Use of three-dimensional method	3D-FCN	Public	Precision: 88.1%	[65]
	3D-CNN	Private	CADe detected 94% of polyps tested (47/50)	[66]
Use of CNN-based models	CNN	Both	Accuracy: 96.4% (processing 1 frame in 10 ms)	[69]
	OptCNN	Public	Precision: 84.58	[72]
	RetinaNet	Private	Sensitivity: 99.8% (Polyp appearance time >30 s)	[73]
Detection of small-sized polyps	CNN	Both	Precision: 80.1% mAP@0.5: 81.7%	[74]
	MP-FSSD	Both	mAP@0.5: 91.56%	[54]
	YOLOv5	Private	Precision: 91.3%	[75]
Improving reprocessing methods	YOLOv2	Both	Sensitivity: 96.7%	[50]
	YOLOv3	Private	Precision: 89%	[76]
	CNN	Both	Precision: 99.06%	[78]
	YOLOv4	Both	Precision: 96.1%	[79]

CNN, convolutional neural network; YOLO, you only look once; SSD, single shot multibox detector; 3D-FCN, three-dimensional fully convolutional network; 3D-CNN, three-dimensional convolutional neural network; OptCNN, optical flow model combined with an on-the-fly trained convolutional neural network; SSDGPN, single shot multibox detector for gastric polyps; MP-FSSD, multiscale pyramidal fusion single-shot multibox detector network.

ENDOMIND-Advanced, which features a real-time post-processing method based on robust and efficient post-processing (REPP) [77]. This system connects bounding boxes across different frames using linking scores, discarding bounding frames that do not meet specific linking and prediction thresholds. The predicted frames detected in the past are used to adjust the current bounding box. Eventually, the system calculates and displays the filtered detection results. The algorithm can achieve 99.06% Precision on the CVC-VideoClinicDB dataset while maintaining real-time detection [78]. Zhang et al. introduced a novel post-processing method within the prediction phase of the YOLOv4 model, which utilizes neighboring frames to assess the detection accuracy of the current frame and integrates single-frame detection results with spatiotemporal information to make the final decision. The results indicate that the precision of this method can reach 96.1% on the CVC-ClinicVideoDB dataset, fulfilling real-time requirements [79]. In this section, we summarise several methods to improve the accuracy of the algorithms in the reviewed papers and present the main results in **Table 2**.

### Challenges and future directions

With the development of AI in recent years, the application of AI algorithms to real-time colonoscopy detection has emerged as a compelling area of cutting-edge research. Numerous studies have demonstrated that AI systems, particularly those utilizing Deep Learning techniques, are capable of efficiently analyzing vast amounts of colonoscopy video and image data, thereby aiding clinicians in the detection process during colonoscopies. However, there are still challenges in this area.

Firstly, among the many models developed, the sensitivity to image analysis is slightly higher than video analysis. Possible explanations for this observation include: 1) Many of the studies were based on image dataset studies, resulting in algorithms that do not perform as well on video datasets. 2) The quality of certain image frames in real-time colonoscopy videos is inferior to that of static images. This limitation can be addressed by increasing the training data and enhancing the quality of video frames. Researchers have proposed many solutions to improve the quality of video frames, such as exploiting the similarity between consecutive

frames to improve the quality of low-quality frames given neighboring high-quality frames [80, 81]. Another approach is video frame interpolation, which typically utilizes high-quality reference images to synthesize intermediate frames to generate high-frame-rate video from low-frame-rate video [82].

Secondly, polyps observed in colonoscopy images exhibit a variety of sizes, shapes, textures, colors, and orientations. This diversity and complexity make it difficult to generalize the model to different types of polyps. Furthermore, the ability of these models to generalize across different hospitals, devices, and patients presents an additional challenge. Several solutions have been proposed to address this issue: 1) Generating more diverse training data through data augmentation techniques such as rotation, scaling, and flipping. 2) Employing a transfer learning approach that utilizes models pre-trained on large-scale datasets and then fine-tuned on a specific polyp detection dataset. In addition, endoscopic video annotation is time-consuming and error-prone, which makes semi-supervised learning algorithms a growing trend.

## Conclusion

Colorectal cancer is the third most commonly diagnosed malignancy and the fourth leading cause of cancer deaths in the world. Colonoscopy detection algorithms can be effective in preventing colorectal cancer by improving the early detection of colorectal polyps. This paper reviews recent research on applying the application of conventional machine learning and deep learning algorithms for real-time colonoscopy detection. A detailed analysis and comparison of deep learning algorithms in colorectal polyp detection are presented, focusing on optimizations in both speed and accuracy. Additionally, the challenges associated with deep learning algorithms in colorectal polyp detection are discussed, along with potential solutions. The review may inform the development of AI algorithms in real-time detection of colorectal polyps, which will be helpful in the early detection and diagnosis of colorectal cancer.

## Acknowledgements

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS, grant No. 2021-I2M-1-010).

## Disclosure of conflict of interest

None.

**Address correspondence to:** Zheng Wang, Cancer Hospital Chinese Academy of Medical Sciences, 17 Panjiayuan South Lane, Chaoyang District, Beijing 100021, China. Tel: +86-18910160128; E-mail: 18910160128@163.com; Yan-Qiu Wang, Langfang Traditional Chinese Medicine Hospital, 108 Yinhe North Road, Langfang 065099, Hebei, China. Tel: +86-18031603558; E-mail: lfszywyq@163.com; Jia-Qi Lü, Hebei University of Technology, 5340 Xi Ping Road, Beichen District, Tianjin 300401, China. Tel: +86-13920691363; E-mail: lvjiaqi@hebut.edu.cn

## References

- [1] Bond JH. Colon polyps and cancer. *Endoscopy* 2003; 35: 27-35.
- [2] Simon K. Colorectal cancer development and advances in screening. *Clin Interv Aging* 2016; 11: 967-976.
- [3] Hassan C, Spadaccini M, Mori Y, Foroutan F, Facciorusso A, Gkolfakis P, Tziatzios G, Triantafyllou K, Antonelli G, Khalaf K, Rizkala T, Vandvik PO, Fugazza A, Rondonotti E, Glissen-Brown JR, Kamba S, Maida M, Correale L, Bhandari P, Jover R, Sharma P, Rex DK and Repici A. Real-time computer-aided detection of colorectal neoplasia during colonoscopy: a systematic review and meta-analysis. *Ann Intern Med* 2023; 176: 1209-1220.
- [4] Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, Zwierko M, Rupinski M, Nowacki MP and Butruk E. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010; 362: 1795-1803.
- [5] Rex DK, Schoenfeld PS, Cohen J, Pike IM, Adler DG, Fennerty MB, Lieb JG 2nd, Park WG, Rizk MK, Sawhney MS, Shaheen NJ, Wani S and Weinberg DS. Quality indicators for colonoscopy. *Am J Gastroenterol* 2015; 110: 72-90.
- [6] Ferlitsch M, Moss A, Hassan C, Bhandari P, Dumonceau JM, Paspatis G, Jover R, Langner C, Bronzwaer M, Nalankilli K, Fockens P, Hazzan R, Gralnek IM, Gschwantler M, Waldmann E, Jeschek P, Penz D, Heresbach D, Moons L, Lemmers A, Paraskeva K, Pohl J, Ponchon T, Regula J, Repici A, Rutter MD, Burgess NG and Bourke MJ. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) clinical guideline. *Endoscopy* 2017; 49: 270-297.
- [7] Yoshida H and Nappi J. Three-dimensional computer-aided diagnosis scheme for detec-

- tion of colonic polyps. *IEEE Trans Med Imaging* 2001; 20: 1261-1274.
- [8] Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, Ferrara E, Spadaccini M, Alkandari A, Fugazza A, Anderloni A, Galtieri PA, Pellegatta G, Carrara S, Di Leo M, Craviotto V, Lamona L, Lorenzetti R, Andrealli A, Antonelli G, Wallace M, Sharma P, Rosch T and Hassan C. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020; 159: 512-520, e517.
- [9] Kuiper T, Marsman WA, Jansen JM, van Soest EJ, Haan YC, Bakker GJ, Fockens P and Dekker E. Accuracy for optical diagnosis of small colorectal polyps in nonacademic settings. *Clin Gastroenterol Hepatol* 2012; 10: 1016-1020; quiz e79.
- [10] Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, Liu P, Li L, Song Y, Zhang D, Li Y, Xu G, Tu M and Liu X. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; 68: 1813-1819.
- [11] ELKarazle K, Raman V, Then P and Chua C. Detection of colorectal polyps from colonoscopy using machine learning: a survey on modern techniques. *Sensors (Basel)* 2023; 23: 1225.
- [12] LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521: 436-444.
- [13] Hassan C, Wallace MB, Sharma P, Maselli R, Craviotto V, Spadaccini M and Repici A. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. *Gut* 2020; 69: 799-800.
- [14] Kang J and Doraiswami R. Real-time image processing system for endoscopic applications. *CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No. 03CH37436)* 2003; 3: 1469-1472.
- [15] Hwang S, Oh J, Tavanapong W, Wong J and De Groen PC. Polyp detection in colonoscopy video using elliptical shape feature. *2007 IEEE International Conference on Image Processing* 2007; 2: II-465-II-468.
- [16] Li B and Meng MQH. Small bowel tumor detection for wireless capsule endoscopy images using textural features and support vector machine. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* 2009: 498-503.
- [17] Ameling S, Wirth S, Paulus D, Lacey G and Vilarino F. Texture-based polyp detection in colonoscopy. *Bildverarbeitung für die Medizin 2009: Algorithmen-Systeme-Anwendungen Proceedings des Workshops vom 22. bis 25. März 2009 in Heidelberg*. Springer Berlin Heidelberg 2009: 346-350.
- [18] Ševo I, Avramović A, Balasingham I, Elle OJ, Bergsland J and Aabakken L. Edge density based automatic detection of inflammation in colonoscopy videos. *Comput Biol Med* 2016; 72: 138-150.
- [19] Iakovidis DK, Maroulis DE, Karkanis SA and Brokos A. A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)* 2005: 575-580.
- [20] Häfner M, Liedlgruber M, Uhl A, Vécsei A and Wrba F. Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Med Image Anal* 2012; 16: 75-86.
- [21] Girtharan B, Yuan X, Liu J, Buckles B, Oh J and Tang SJ. Bleeding detection from capsule endoscopy videos. *Annu Int Conf IEEE Eng Med Biol Soc* 2008; 2008: 4780-4783.
- [22] Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, Tamaki T, Koide T, Kaneda K and Chayama K. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016; 83: 643-649.
- [23] Wang Y, Tavanapong W, Wong J, Oh JH and De Groen PC. Polyp-alert: near real-time feedback during colonoscopy. *Comput Methods Programs Biomed* 2015; 120: 164-179.
- [24] Karaman A, Karaboga D, Pacal I, Akay B, Basturk A, Nalbantoglu U, Coskun S and Sahin O. Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection. *Appl Intell* 2023; 53: 15603-15620.
- [25] Jiang X, Hu Z, Wang S and Zhang Y. Deep learning for medical image-based cancer diagnosis. *Cancers (Basel)* 2023; 15: 3608.
- [26] Chen J and Park C. A deep learning paradigm for medical imaging data. *Expert Syst Appl* 2024; 124480.
- [27] Tian L, Hunt B, Bell MAL, Yi J, Smith JT, Ochoa M, Intes X and Durr NJ. Deep learning in biomedical optics. *Lasers Surg Med* 2021; 53: 748-775.
- [28] Bokhorst JM, Nagtegaal ID, Fraggetta F, Vatrano S, Mesker W, Vieth M, van der Laak J and Ciompi F. Deep learning for multi-class semantic segmentation enables colorectal cancer detection and classification in digital pathology images. *Sci Rep* 2023; 13: 8398.
- [29] Bernal J, Tajbaksh N, Sanchez FJ, Matuszewski BJ, Hao Chen, Lequan Yu, Angermann Q, Romain O, Rustad B, Balasingham I, Pogorelov K, Sungbin Choi, Debard Q, Maier-Hein L,

- Speidel S, Stoyanov D, Brandao P, Cordova H, Sanchez-Montes C, Gurudu SR, Fernandez-Esparrach G, Dray X, Jianming Liang and Histace A. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. *IEEE Trans Med Imaging* 2017; 36: 1231-1249.
- [30] Tajbakhsh N, Gurudu SR and Liang J. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) 2015: 79-83.
- [31] Zhang R, Zheng Y, Poon CCY, Shen D and Lau JYW. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognit* 2018; 83: 209-219.
- [32] Poon CCY, Jiang Y, Zhang R, Lo WWY, Cheung MSH, Yu R, Zheng Y, Wong JCT, Liu Q, Wong SH, Mak TWC and Lau JYW. AI-doscopist: a real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices. *NPJ Digit Med* 2020; 3: 73.
- [33] Zand M, Etemad A and Greenspan M. Object-box: from centers to boxes for anchor-free object detection. *European Conference on Computer Vision* 2022: 390-406.
- [34] Tian Z, Shen C, Chen H and He T. FCOS: a simple and strong anchor-free object detector. *IEEE Trans Pattern Anal Mach Intell* 2020; 44: 1922-1933.
- [35] Cheng G, Wang J, Li K, Xie X, Lang C, Yao Y and Han J. Anchor-free oriented proposal generator for object detection. *IEEE Trans Geosci Remote Sens* 2022; 60: 1-11.
- [36] Yang X, Song E, Ma G, Zhu Y, Yu D, Ding B and Wang X. YOLO-OB: an improved anchor-free real-time multiscale colon polyp detector in colonoscopy. *arXiv.org* 2023: 2312.08628.
- [37] Duan K, Bai S, Xie L, Qi H, Huang Q and Tian Q. Centernet: keypoint triplets for object detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2019: 6569-6578.
- [38] Wang D, Zhang N, Sun X, Zhang P, Zhang C, Cao Y and Liu B. Afp-net: realtime anchor-free polyp detection in colonoscopy. 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI) 2019: 636-643.
- [39] Jeong SM, Lee SG, Seok CL, Lee EC and Lee JY. Lightweight deep learning model for real-time colorectal polyp segmentation. *Electronics* 2023; 12: 1962.
- [40] Ji Z, Li X, Liu J, Chen R, Liao Q, Lyu T and Zhao L. LightCF-Net: a lightweight long-range context fusion network for real-time polyp segmentation. *Bioengineering* 2024; 11: 545.
- [41] Xing B, Wang W, Qian J, Pan C and Le Q. A lightweight model for real-time monitoring of ships. *Electronics* 2023; 12: 3804.
- [42] Ou S, Gao Y, Zhang Z and Shi C. Polyp-yolov5-tiny: a lightweight model for real-time polyp detection. 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA) 2021: 1106-1111.
- [43] Yoo Y, Lee JY, Lee DJ, Jeon J and Kim J. Real-time polyp detection in colonoscopy using lightweight transformer. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2024: 7794-7804.
- [44] Yang Z, Gong B, Wang L, Huang W, Yu D and Luo J. A fast and accurate one-stage approach to visual grounding. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2019: 4683-4693.
- [45] Klamt S, Mahadevan R and Hädicke O. When do two-stage processes outperform one-stage processes? *Biotechnol J* 2018; 13: 1700539.
- [46] Redmon J, Divvala S, Girshick R and Farhadi A. You only look once: unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2016: 779-788.
- [47] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY and Berg AC. Ssd: single shot multibox detector. *Computer Vision-ECCV* 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing; 2016. pp. 21-37.
- [48] Lin TY, Goyal P, Girshick R, He K and Dollár P. Focal loss for dense object detection. *Proc IEEE Int Conf Comput Vis* 2017; 2980-2988.
- [49] Jha D, Ali S, Tomar NK, Johansen HD, Johansen D, Rittscher J, Riegler MA and Halvorsen P. Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *IEEE Access* 2021; 9: 40496-40510.
- [50] Lee JY, Jeong J, Song EM, Ha C, Lee HJ, Koo JE, Yang DH, Kim N and Byeon JS. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. *Sci Rep* 2020; 10: 8379.
- [51] Pacal I and Karaboga D. A robust real-time deep learning based automatic polyp detection system. *Comput Biol Med* 2021; 134: 104519.
- [52] Liu M, Jiang J and Wang Z. Colonic polyp detection in endoscopic videos with single shot detection based deep convolutional neural network. *IEEE Access* 2019; 7: 75058-75066.
- [53] Zhang X, Chen F, Yu T, An J, Huang Z, Liu J, Hu W, Wang L, Duan H and Si J. Real-time gastric polyp detection using convolutional neural networks. *PLoS One* 2019; 14: e0214133.

- [54] Souaidi M and El Ansari M. A new automated polyp detection network MP-FSSD in WCE and colonoscopy images based fusion single shot multibox detector and transfer learning. *IEEE Access* 2022; 10: 47124-47140.
- [55] Younas F, Usman M and Yan WQ. A deep ensemble learning method for colorectal polyp classification with optimized network parameters. *Appl Intell* 2023; 53: 2410-2433.
- [56] Zhao L, Wang N, Zhu X, Wu Z, Shen A, Zhang L, Wang R, Wang D and Zhang S. Establishment and validation of an artificial intelligence-based model for real-time detection and classification of colorectal adenoma. *Sci Rep* 2024; 14: 10750.
- [57] Ma C, Jiang H, Ma L and Chang Y. A real-time polyp detection framework for colonoscopy video. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV) 2022*: 267-278.
- [58] Sharma P, Balabantaray BK, Bora K, Mallik S, Kasugai K and Zhao Z. An ensemble-based deep convolutional neural network for computer-aided polyps identification from colonoscopy. *Front Genet* 2022; 13: 844391.
- [59] Arnold E, Al-Jarrah OY, Dianati M, Fallah S, Oxtoby D and Mouzakitis A. A survey on 3d object detection methods for autonomous driving applications. *IEEE Trans Intell Transp Syst* 2019; 20: 3782-3795.
- [60] Mao J, Shi S, Wang X and Li H. 3D object detection for autonomous driving: a comprehensive survey. *Int J Comput Vis* 2023; 131: 1909-1963.
- [61] Yousuf M, Harb S, Alkabbany I, Ali A, Elshazley S and Farag A. Colorectal polyps detection in virtual colonoscopy using 3D geometric features and deep learning. *2024 IEEE International Symposium on Biomedical Imaging (ISBI) 2024*: 1-4.
- [62] Puyal JGB, Bhatia KK, Brandao P, Ahmad OF, Toth D, Kader R, Lovat L, Mountney P and Stoyanov D. Endoscopic polyp segmentation using a hybrid 2D/3D CNN. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020: 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part VI* 23. Springer International Publishing; 2020. pp. 295-305.
- [63] González-Bueno Puyal J, Brandao P, Ahmad OF, Bhatia KK, Toth D, Kader R, Lovat L, Mountney P and Stoyanov D. Polyp detection on video colonoscopy using a hybrid 2D/3D CNN. *Med Image Anal* 2022; 82: 102625.
- [64] Feng L, Xu J, Ji X, Chen L, Xing S, Liu B, Han J, Zhao K, Li J, Xia S, Guan J, Yan C, Tong Q, Long H, Zhang J, Chen R, Tian D, Luo X, Xiao F and Liao J. Development and validation of a three-dimensional deep learning-based system for assessing bowel preparation on colonoscopy video. *Front Med (Lausanne)* 2023; 10: 1296249.
- [65] Yu L, Chen H, Dou Q, Qin J and Heng PA. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE J Biomed Health Inform* 2017; 21: 65-75.
- [66] Misawa M, Kudo S, Mori Y, Cho T, Kataoka S, Maeda Y, Ogawa Y, Takeda K, Nakamura H and Ichimasa K. Tu1990 Artificial intelligence-assisted polyp detection system for colonoscopy, based on the largest available collection of clinical video data for machine learning. *Gastrointest Endosc* 2019; 89: AB646-AB647.
- [67] Misawa M, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, Ogawa Y, Maeda Y, Takeda K, Ichimasa K, Nakamura H, Yagawa Y, Toyoshima N, Ogata N, Kudo T, Hisayuki T, Hayashi T, Wakamura K, Baba T, Ishida F, Itoh H, Roth H, Oda M and Mori K. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018; 154: 2027-2029, e2023.
- [68] Haj-Manouchehri A and Mohammadi HM. Polyp detection using CNNs in colonoscopy video. *IET Comput Vis* 2020; 14: 241-247.
- [69] Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W and Baldi P. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018; 155: 1069-1078, e1068.
- [70] Qadir HA, Balasingham I, Solhusvik J, Bergsland J, Aabakken L and Shin Y. Improving automatic polyp detection using CNN by exploiting temporal dependency in colonoscopy video. *IEEE J Biomed Health Inform* 2020; 24: 180-193.
- [71] Angermann Q, Bernal J, Sánchez-Montes C, Hammami M, Fernández-Esparrach G, Dray X, Romain O, Sánchez FJ and Histace A. Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures: 4th International Workshop, CARE 2017, and 6th International Workshop, CLIP 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 4* 2017: 29-41.
- [72] Zheng H, Chen H, Huang J, Li X, Han X and Yao J. Polyp tracking in video colonoscopy using optical flow with an on-the-fly trained CNN. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 2019; 79-82.
- [73] Livovsky DM, Veikherman D, Golany T, Aides A, Dashinsky V, Rabani N, Ben Shimol D, Blau Y, Katzir L, Shimshoni I, Liu Y, Segol O, Goldin E,

## Artificial intelligence algorithms for real-time detection of colorectal polyps

- Corrado G, Lachter J, Matias Y, Rivlin E and Freedman D. Detection of elusive polyps using a large-scale artificial intelligence system (with videos). *Gastrointest Endosc* 2021; 94: 1099-1109, e1010.
- [74] Fu J, Gao Y, Zhou P, Huang Y, Jiao J, Lin S, Wang Y and Guo Y. D2polyp-Net: a cross-modal space-guided network for real-time colorectal polyp detection and diagnosis. *Biomed Signal Process Control* 2024; 91: 105934.
- [75] Wan J, Chen B and Yu Y. Polyp detection from colorectum images by using attentive YOLOv5. *Diagnostics* 2021; 11: 2264.
- [76] Nogueira-Rodríguez A, Domínguez-Carbajales R, Campos-Tato F, Herrero J, Puga M, Remedios D, Rivas L, Sánchez E, Iglesias A and Cubiella J. Real-time polyp detection model using convolutional neural networks. *Neural Comput Appl* 2022; 34: 10375-10396.
- [77] Sabater A, Montesano L and Murillo AC. Robust and efficient post-processing for video object detection. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2020: 10536-10542.
- [78] Krenzer A, Banck M, Makowski K, Hekalo A, Fitting D, Troya J, Sudarevic B, Zoller WG, Hann A and Puppe F. A real-time polyp-detection system with clinical application in colonoscopy using deep convolutional neural networks. *J Imaging* 2023; 9: 26.
- [79] Zhang Z, Ma L, Chana Y, Xiao L, He Q, Ma C and Jiang H. A practical polyp detecting model in colonoscopy video by post-processing. 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). *IEEE* 2021; 1-6.
- [80] Guan Z, Xing Q, Xu M, Yang R, Liu T and Wang Z. MFQE 2.0: a new approach for multi-frame quality enhancement on compressed video. *IEEE Trans Pattern Anal Mach Intell* 2021; 43: 949-963.
- [81] Yang R, Xu M, Wang Z and Li T. Multi-frame quality enhancement for compressed video. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2018: 6664-6673.
- [82] Shen W, Bao W, Zhai G, Chen L, Min X and Gao Z. Video frame interpolation and enhancement via pyramid recurrent framework. *IEEE Trans Image Process* 2020; 30: 277-292.