OKU Fen Bilimleri Enstitüsü Dergisi
8(3): 1240-1257, 2025

OKU Journal of The Institute of Science and Technology,
8(3): 1240-1257, 2025

Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Osmaniye Korkut Ata University Journal of The Institute of Science and Technology

Gerçek Zamanlı Polip Tespiti: Yolov5 ve Yolov6'nın Hız ve Performans Analizi

Semih DEMİREL[1*], Azer ÇELİKTEN[2], Andaç AKPULAT[3], Kerem DEMİR[4], Ece BİNGÖL[5], İdris GÜLTEKİN[6], Abdulkadir BUDAK[7], Hakan KARATAŞ[8]

[1,2,3,4,5,6,7,8]Akgun Computer Inc., Department of Artificial Intelligence and Image Processing, Ankara, Türkiye
[1]Gazi University, Graduate School of Informatics, Department of Information Systems, Ankara, Türkiye
[4]Atılım University, Faculty of Engineering, Department of Industrial Engineering, Ankara, Türkiye
[5]Atılım University, Faculty of Engineering, Department of Software Engineering, Ankara, Türkiye

[1] https://orcid.org/0000-0002-3454-3631
[2] https://orcid.org/0000-0002-6804-737X
[3] https://orcid.org/0000-0001-9611-9038
[4] https://orcid.org/0009-0001-4737-1496
[5] https://orcid.org/0009-0006-7615-1392
[6] https://orcid.org/0000-0002-3240-606X
[7] https://orcid.org/0000-0002-0328-6783
[8] https://orcid.org/0000-0002-9497-5444
*Sorumlu yazar: semih.demirel@akgun.com.tr

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 06.09.2024
Kabul tarihi: 20.02.2025
Online Yayınlanma:16.06.2025

Anahtar Kelimeler:
Gerçek zamanlı polip tespiti
Yolov5
Yolov6
Bilgisayar destekli tanı
Nesne tespiti

ÖZ

Kolorektal kanser, kolonoskopi sırasında gözden kaçan poliplerin bilgisayar destekli tanı sistemi ile tespit edilmesiyle potansiyel olarak önlenebilir. Endoskopi uzmanlarına yardımcı olmak amacıyla, gerçek zamanlı polip tespiti için yolov5 ve yolov6 modelleri kullanılarak performans ve hız analizi yapılmıştır. Polip tespiti için you look only once v5 (yolov5) ve you look only once v6 (yolov6) modelleri kullanılmıştır. Nesne tespiti modellerini eğitmek için açık kaynaklı bir veri setine yeni bir özel veri seti eklenmiştir. Bulgular, yolov5x ve yolov6l modelleri için ortalama doğruluk (mAP50) oranlarının sırasıyla 0.896 ve 0.913 olduğunu göstermiştir. Yolov5x ve yolov6l modelleri karşılaştırıldığında, yolov5x'in daha yüksek bir doğruluk oranına, yolov6l'nin ise daha yüksek bir duyarlılık oranına sahip olduğu belirlenmiştir. Modeller, literatürdeki diğer araştırmalarla karşılaştırıldığında, yolov5x ve yolov6l'nin sırasıyla duyarlılık (0.893) ve f1-skoru (0.876) açısından önceki çalışmalardan daha başarılı olduğu görülmüştür.

Real-Time Polyp Detection: A Speed and Performance Analysis of Yolov5 and Yolov6

Research Article

Article History:
Received: 06.09.2024
Accepted: 20.02.2025
Published online:16.06.2025

Keywords:
Real-time polyp detection
Yolov5
Yolov6
Computer aided diagnosis
Object detection

ABSTRACT

Colorectal cancer can potentially be prevented by detecting polyps missed during colonoscopy using a computer aided diagnosis system. In order to help endoscopy specialists, a performance and speed analysis were carried out utilizing the yolov5 and yolov6 models for real-time polyp detection. You look only once v5 (yolov5) and you look only once v6 (yolov6) models were used for polyp detection. To train object detection models, a new private dataset was added to the open-source dataset. The findings showed that the mean average precision 50 (mAP50) rates for yolov5x and yolov6l were 0.896 and 0.913, respectively. After comparing yolov5x and yolov6l, it was determined that yolov5x had superior precision while yolov6l had superior recall. When models were compared to other research in the literature, yolov5x and yolov6l succeeded better than previous studies in terms of recall (0.893) and f1-score (0.876), respectively.

To Cite: Demirel S., Çelikten A., Akpulat A., Demir K., Bingöl E., Gültekin İ., Budak A., Karataş H. Real-Time Polyp Detection: A Speed and Performance Analysis of Yolov5 and Yolov6. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2025; 8(3): 1240-1257.

1. Introduction

According to data from the American Cancer Society, colorectal cancer is the second leading cause of cancer-related deaths worldwide (Rahim et al., 2021). Colorectal cancer is the term used to describe the gradual development of malignant tumors from benign polyps in the colon and rectum (Pacal and Karaboga, 2021). Polyps can be detected by colonoscopy (Pacal et al., 2022). Research suggests that a considerable proportion of polyps, as many as 28%, may be missed during a colonoscopy (Wang et al., 2021). In order to combat colorectal cancer, professionals may find it useful to use an automated diagnosis system that was created for polyp identification.

In recent years, object detection models have advanced significantly and are now widely employed in a variety of industries, including healthcare. Using two-stage models yields good results for problems involving object detection. Furthermore, one-stage versions have gained popularity recently (Lin et al., 2017). Two-stage models are exemplified by the region-based convolutional neural networks (R-CNN) series. A convolutional neural network is used to extract characteristics from generated region proposals in a region-based convolutional neural network (Girshick et al., 2014). A support vector machine is used to classify the regions based on the features that were extracted (Girshick et al., 2014). Fast region-based convolutional neural network (Girshick, 2015) is a single-stage object identification model that arranges locations according to convolution-based assumptions. It synchronizes region proposals with the pooling of regions of interest. The region proposal network is used to create region proposals at various sizes on feature maps in the faster region-based convolutional neural network (Ren et al., 2017), which is the more sophisticated variant of the fast region-based convolutional neural network. The region proposal network helps to improve the quality of region proposals and produce more accurate outcomes. One example of a single-stage model is yolo. The single shot detector model is an additional single-stage model (Liu et al., 2016). According to Redmon et al. (2016), yolov1 predicts items in a single network and considers object identification tasks to be a regression problem. To increase the accuracy performance of yolov1, yolov2 uses anchor boxes instead of a fully connected network (Redmon and Farhadi, 2017). The feature extractor network's residual connections (He et al., 2016) enable the extraction of superior features in yolov3 (Redmon and Farhadi, 2018). Building on the success of earlier iterations, yolov4 (Bochkovskiy et al., 2020) is an upgraded version that incorporates a number of improvements to achieve greater speed and accuracy. The main concept of a single shot detector is to detect objects immediately on a single network pass, eliminating the need for post-processing or extra proposal creation (Liu et al., 2016). Despite their high accuracy, two-stage models are not suitable for real-time problems due to their speed (Lin et al., 2017). Yolo's novel network architecture enables high accuracy performance free from speed fluctuations. As a result, the yolo series can be used for tasks involving object recognition in real time.

Significant research on polyp detection has recently been carried out in the literature with the introduction of object identification models based on convolutional neural networks. Rahim et al. (2021) improved feature extraction and polyp detection on colonoscopy images by using convolution kernels

of different sizes. Following a variety of augmentation methods, images were entered into a 16-convolution model. Two fully linked layers and a softmax function were then used to make predictions. Their findings showed that the suggested convolutional neural network model has a f1-score of 88.30%, 82.92% sensitivity, and 94.44% precision.

In their work, Pacal and Karaboga (2021) improved the yolov4 model for real-time polyp recognition. They accomplished this by using a number of models as the feature extractor, with distance intersection over union (Zheng et al., 2020) serving as the loss and mish (Misra, 2019) serving as the activation. They claimed that the accuracy was increased by employing input of varying sizes. Yolo models were said to be effective at detecting polyps in real time.

Chen et al. (2021) used faster r-cnn to detect polyps in their investigation. In order to make the polyps more noticeable, they preprocessed the colonoscopy images by increasing the contrast. They added a self-attention module to the feature extractor network in order to improve the quality of the feature map. Their findings showed that the suggested model had a 0.934 f1-score, 0.925 recall, and 0.943 precision.

In order to achieve excellent performance in polyp identification, Pacal et al. (2022) employed the cross stage partial network (Wang et al., 2020) in the neck and head networks of yolov3 and in the backbone network of yolov4. It was shown that the false positive rate is decreased by include negative samples. Other networks came out worse than the suggested yolov4 cross stage partial (yolov4-csp) model.

In order to perform data augmentation on image data for polyp detection in their experiment, Qian et al. (2022) used a conditional generative adversarial network (Mirza and Osindero, 2014) because they lacked colonoscopy images. By using skip connection and dilated convolution, they enhanced the yolov4's architecture. Based on the results, the suggested model produced 17.2 frames per second and an average precision of 92.37%.

Karaman et al. (2022) adjusted the yolo models' hyperparameters using the artificial bee colony algorithm. Following the recommended modification, the scaled-yolov4 (Wang et al., 2021) model's mean average precision value and f1-score both increased by 2% and 3%, respectively. Tang et al. (2022) employed image processing techniques to simulate the narrow band imaging approach used during colonoscopy. When compared to the original images, the multi-scale retinex approach was more successful in identifying polyps.

In this work, the most recent yolo models, yolov5 and yolov6 (Li et al., 2022), were used to create an automated diagnostic solution for real-time polyp identification. This study's primary goal is to evaluate the yolov5 and yolov6 model variants' performance and speed. Open-source data and a new private dataset were used as the datasets. The private dataset only included a tiny amount of data, thus data augmentation techniques such motion blur, rotation 90 degrees, and vertical and horizontal flips were applied. To lower the amount of false positives, negative samples were added to the training set of data. Precision, recall, f1-score, f2-score, mAP50, and mAP50:95 metrics were used to compare the models' performance. The models were also put through a speed test. The yolov5x and yolov6l models were also compared with previous studies in the literature.

This study offers the following contributions that are different from other studies for real-time polyp detection:

- Performance comparison of the accuracy and speed of yolov5 and yolov6 models.
- A new dataset for polyp detection was created and validated by an endoscopy specialist.
- The performance in polyp detection was significantly improved by using data augmentation techniques.

The rest of the content of this research is divided into the following sections: In Section 2, the materials and methods of this study are presented. Section 3 consists of performance metrics, experimental results of the models, and discussions about the results. Section 4 gives the conclusion and future research.

2. Materials and Method

In this study, real-time polyp detection was carried out using yolov5 and yolov6 models.

2.1. Dataset

In this study, open source datasets and private dataset were used for polyp detection. Table 1 provides comprehensive details about the dataset.

Table 1. Detailed information about the dataset.

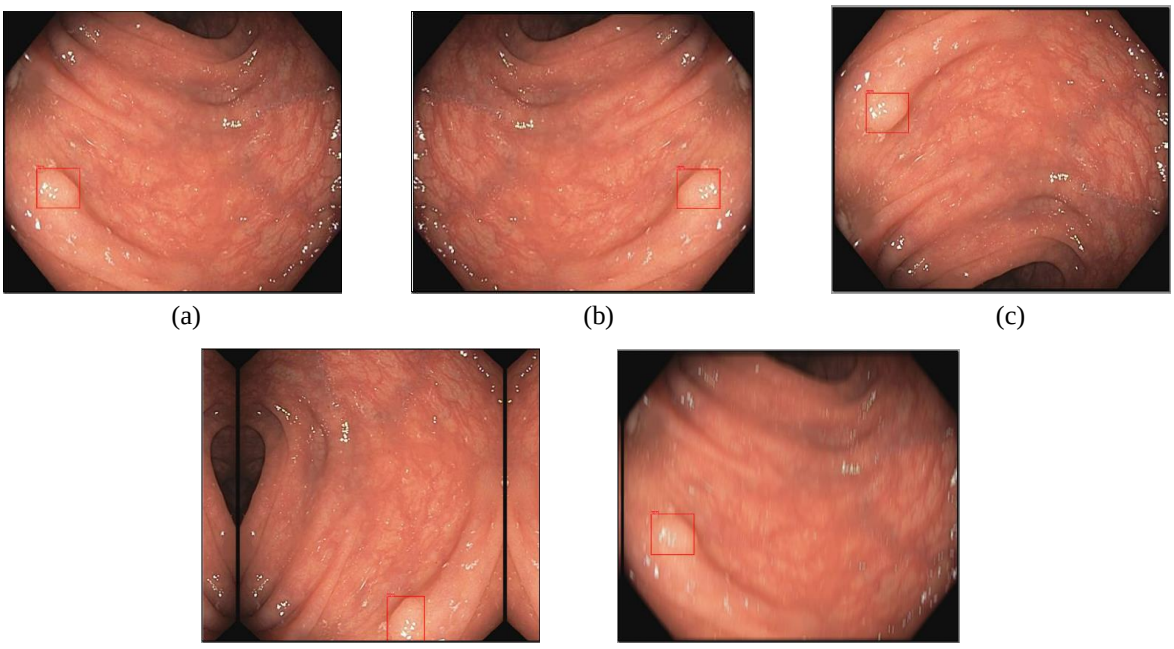
	Training	Testing	Total
CVC-ColonDB (Bernal et al., 2012)	244	56	300
ETIS (Silva et al., 2014)	157	39	196
Kvasir-SEG (Jha et al., 2020)	800	200	1000
CVC-ClinicDB (Bernal et al., 2015)	500	112	612
Private dataset	516	122	638
Total	2217	529	2746

There are 2217 and 529 data points for training and test, respectively, in Table 1. A new private dataset was created in this study. The private dataset was obtained from a private healthcare facility in Turkey, while the open-source datasets were sourced via online sites. Video recordings were gathered following the acquisition of the required ethical permissions. An endoscopic specialist recognized and annotated frames in the videos that included polyps. Data augmentation was used because the private dataset only contained a small number of images. For the private dataset in this investigation, data augmentation techniques such as vertical flip, horizontal flip, rotate 90, and motion blur were used. The albumentations (Buslaev et al., 2020) library was used for data augmentation. In addition, it was stated in the study by (Pacal et al., 2022) that adding negative sample to the training data reduces the false positive rate. Therefore, 765 negative sample images were added to the training data in this study. The added negative samples are images that do not contain polyps. In addition to negative sample processing and data augmentation, no other preprocessing was used. Following these methods, Table 2 provides the dataset information.

Table 2. Relevant details about the dataset after taking data augmentation methods.

	Training	Testing	Total
CVC-ColonDB	244	56	300
ETIS	157	39	196
Kvasir-SEG	800	200	1000
CVC-ClinicDB	500	112	612
Private dataset	516	122	638
Private dataset(Horizontal flip)	516	-	516
Private dataset(Vertical flip)	516	-	516
Private dataset(Rotate 90)	516	-	516
Private dataset(Motion blur)	516	-	516
Negative sample	765	-	765
Total	5046	529	5575

Table 2 shows that while the quantity of data for test stays the same, the number of data for training expands from 2217 instances to 5046 instances. In Figure 1, some examples from the private dataset are shown. The red bounding box represents the ground truth of a polyp.



(a) (b) (c) (d) (e)

Figure 1. Example of images from the dataset. (a) Original image, (b) Horizontal flip, (c) Vertical flip, (d) Rotate 90, (e) Motion blur

2.2. Yolov5

Cross stage partial network is used as a backbone in yolov5. Cross stage partial network consists of cross stage partial modules from which information is extracted using convolution filters. Cross stage partial modules learn features using convolution, gain stability using batch normalization, and perform smooth optimization using mish activation (Sun et al., 2022). Information from the upper layers is transmitted to the lower layers using feature pyramid network to enhance the fusion feature in the neck network (Mushtag et al., 2023). In addition, spatial information of feature pyramids is aggregated and sent to the head via path aggregation network (Liu et al., 2018). Similar to the previous series, the head

network predicts the bounding box, object accuracy score, and object category. From the fewest to the most in terms of parameters, the models in yolov5 are N, S, M, L, and X. Architecture of the yolov5 models are given in Figure 2.

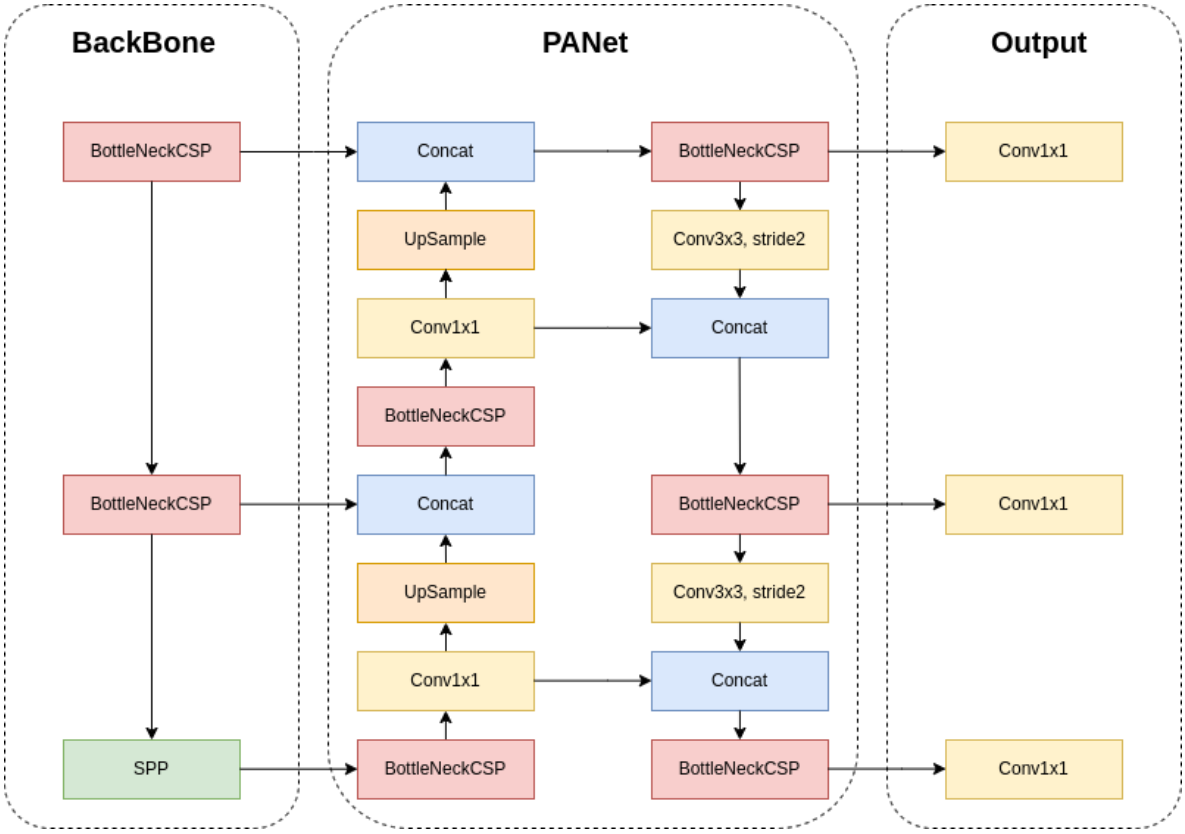


Figure 2. Architecture of the yolov5 models (Kurniawan et al., 2024).

2.3. Yolov6

Yolov6 proposes small and large models for different purposes as a one-stage model. While large models yield better results with the increase in parameters, the single path architecture creates a cost for large models. For small models, the single path architecture provides a consistency in terms of cost. Yolov6 shows high performance without sacrificing inference with domain-specific methods and cutting-edge quantization methods (Li et al., 2022). The models in yolov6 range in size from the smallest to the largest with respect to the quantity of parameters: N, T, S, M, and L. Figure 3 shows the architecture of the yolov6 N and S models.

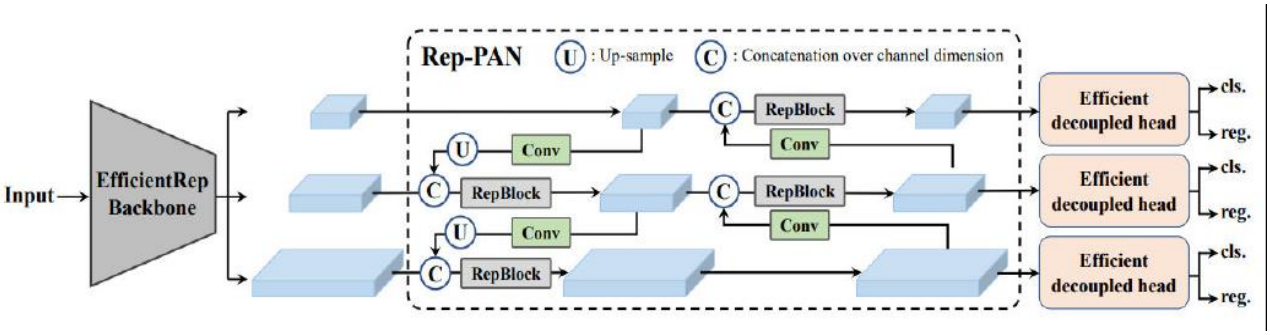


Figure 3. Architecture of the yolov6 N and S models (Li et al., 2022).

A backbone called efficientrep, which includes cross stage partial (Wang et al., 2020) blocks representing strong feature characteristics, is introduced for the backbone network. Thanks to the structural design of the multi-branch methodology and the reparameterization technique used during the training, repblocks (Ding et al., 2021) offers an optimal compromise between speed and accuracy.

A network architecture called a path aggregation network is intended to increase information flow and feature fusion. By allowing both top-down and bottom-up paths to aggregate features across several levels of a neural network, it expands on the concept of feature pyramids. In the neck network, the path aggregation network strengthens the information feature by performing fusion integration with feature aggregation at the layers (Xiao et al., 2022). The residual module located in the cross stage partial block provides a deep network, and the cross stage partial modules implement cross-stage hierarchy. Yolov6 uses the efficient decoupled head module for the head network. Hybrid-channel integration was provided for better performance. Yolov6 uses focal loss based varifocal loss (Zhang et al., 2021) to minimize classification loss. Varifocal loss is given in Equation 1.

$$loss_{varifocal} = \begin{cases} -q(qlog(p) + (1-q)\log(1-p)), & q > 0 \\ -\alpha p^{\gamma}\log(1-p), & q = 0 \end{cases} \quad (1)$$

where, p is the predicted classification score based on localization-aware, and q is the actual value. γ is adjustable parameter and defined as $\gamma \geq 0$.

Yolov6 uses scylla intersection over union (Gevorgyan, 2022) and generalized intersection over union (GIoU) (Rezatofighi, 2019) to localize bounding boxes. Equation 2 gives the generalized intersection over union loss function.

$$loss_{GIoU} = \frac{|A \cap B|}{|A \cup B|} - \frac{C/(A \cup B)}{|C|} \quad (2)$$

where, A represents the predicted bounding box, whereas B represents the actual bounding box. The smallest convex object that encloses a given shape or set of points is referred to as C.

Distribution focal loss (DFL) (Xiang, 2020) is used as probability loss for the M and L models of yolov6. Distribution focal loss is given in Equation 3.

$$loss_{DFL}(S_i, S_{i+1}) = -((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1})) \quad (3)$$

where, y is the target value, S_i and S_{i+1} are defined as $S_i = \frac{y_{i+1}-y}{y_{i+1}-y_i}$ and $S_{i+1} = \frac{y-y_i}{y_{i+1}-y_i}$, respectively.

3. Findings and Discussion

In this study, real-time polyp detection was performed using the yolov5 and yolov6 object detection models. Different models of yolov5 and yolov6 were compared with each other. Additionally, the best yolov5 and yolov6 models in terms of mAP50 were also compared with previous studies in the literature. Precision, recall, f1-score, f2-score, mAP50 and mAP50:95 were used as evaluation metrics.

In this study, training and tests were conducted on ubuntu 20.04 operating system. Computer hardware specifications are as follows: Intel Core i7-10700 CPU (Central Processing Unit) 2.90GHz, Nvidia GeForce 3060 12GB GPU (Graphic Processing Unit), 16GB RAM (Random Access Memory).

3.1. Performance Evaluation Metrics

The intersection over union (IoU) measure is the fundamental performance evaluation information for identifying objects. In Equation 4, the intersection over union metric is presented.

$$IoU = \frac{area(B' \cap B)}{area(B' \cup B)} \tag{4}$$

where, the intersection region between the region of interests is represented by the numerator. The union area between the region of interests is represented by the denominator. The intersection over union value, similar to the classification, can potentially used to generate a confusion matrix. Using the components of the confusion matrix, metrics can be calculated. Calculated metrics are given in Equations 5, 6, 7 and 8, respectively.

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

$$f1score = 2 \times \frac{precision \times recall}{precision + recall} \tag{7}$$

$$f2score = \frac{5 \times precision \times recall}{4 \times precision + recall} \tag{8}$$

Precision-recall curve can be obtained by using different intersection over union threshold values. The precision-recall curve is commonly used to compute the mean average precision given in Equation 9.

$$mAP = \frac{1}{k}\sum_{l=1}^{k} AP_l \tag{9}$$

where, k is the number of classes, AP is the average precision value for class l.

3.2. Yolov5 Test Results

In this study, yolov5 models were trained at 400 epochs. During training, the batch size and image size of the models were set to 32 and 512, respectively, and the optimization algorithm was set to adam (Diederik, 2014). The learning rate and momentum hyperparameters for adam optimization were set to 0.0032 and 0.843, respectively. Models were evaluated using a 0.5 confidence and intersection over union threshold. Table 3 provides the performance results of the yolov5 models.

Table 3. Outcomes of the yolov5 models.

Models	Precision	Recall	F1-Score	F2-Score	mAP50	mAP50-95
Yolov5n	0.865	0.825	0.844	0.832	0.860	0.694
Yolov5s	0.897	0.821	0.857	0.835	0.878	0.744
Yolov5m	0.920	0.807	0.859	0.827	0.882	0.772
Yolov5l	0.901	0.858	0.878	0.866	0.891	0.767
Yolov5x	0.909	0.847	0.876	0.858	0.896	0.783

The performance outcomes of the models utilizing the accuracy, recall, f1-score, f2-score, mAP50, and mAP50:95 metrics are given in Table 3. According to the results obtained, the yolov5m model achieved the highest precision value of 0.92. This means that the yolov5m model achieved the lowest false positive value in the detection of polyps. The yolov5l model achieved the highest recall value of 0.858. In the case of not missing polyp identification, the yolov5l model performs the best. The yolov5l model achieved the best score in regarding the f1-score with 0.878. This suggests that the L model offers the optimum precision and recall arrangement. The yolov5l model outperformed in terms of f2-score with a value of 0.866 compared to other models. Since f2-score gives more weight to the recall metric, this indicates that the yolov5l performs the greatest performance with respect to the recall when considering the f2-score metric. According to the mAP50 and mAP50:95 metrics, the yolov5x model achieved superior performance compared to other studies with values of 0.896 and 0.783, respectively. Therefore, the yolov5x model outperformed other models in polyp detection at different intersection over union thresholds. Table 4 provides the speed test outcomes of the yolov5 models.

Table 4. The speed test outcomes of the yolov5 models.

Models	(B, C, W, H)	Preprocess(ms)	Inference(ms)	NMS(ms)	Total(ms)
Yolov5n	(32, 3, 512, 512)	0.1	1.1	0.9	2.1
Yolov5s	(32, 3, 512, 512)	0.1	2.4	0.9	3.4
Yolov5m	(32, 3, 512, 512)	0.1	5.6	1.3	7
Yolov5l	(32, 3, 512, 512)	0.1	10.2	2.0	12.3
Yolov5x	(32, 3, 512, 512)	0.1	17.4	1.8	19.3

In Table 4, batch size is 32, images are 3-channel and image size is 512. In the table, the preprocess time of all models is 0.1 ms. In terms of NMS(Non Maximum Suppression) time, the fastest model is yolov5n and yolov5s with 0.9 ms, while the slowest model is yolov5l with 2.0 ms. Models with more parameters have a higher inference time. The yolov5n model with the fewest parameters, which has an inference time of 2.1 ms, is the fastest model, while the yolov5x model with the most parameters, which has an inference time of 19.3 ms, is the slowest. The yolov5n is the most efficient when thinking of overall time, while the slowest model is the yolov5x model. The yolov5m, with its 0.772 mAP50:95 value and 7 ms speed, provides the best accuracy-to-speed balancing. A few instances of the yolov5x outcomes for object detection are shown in Figure 4.

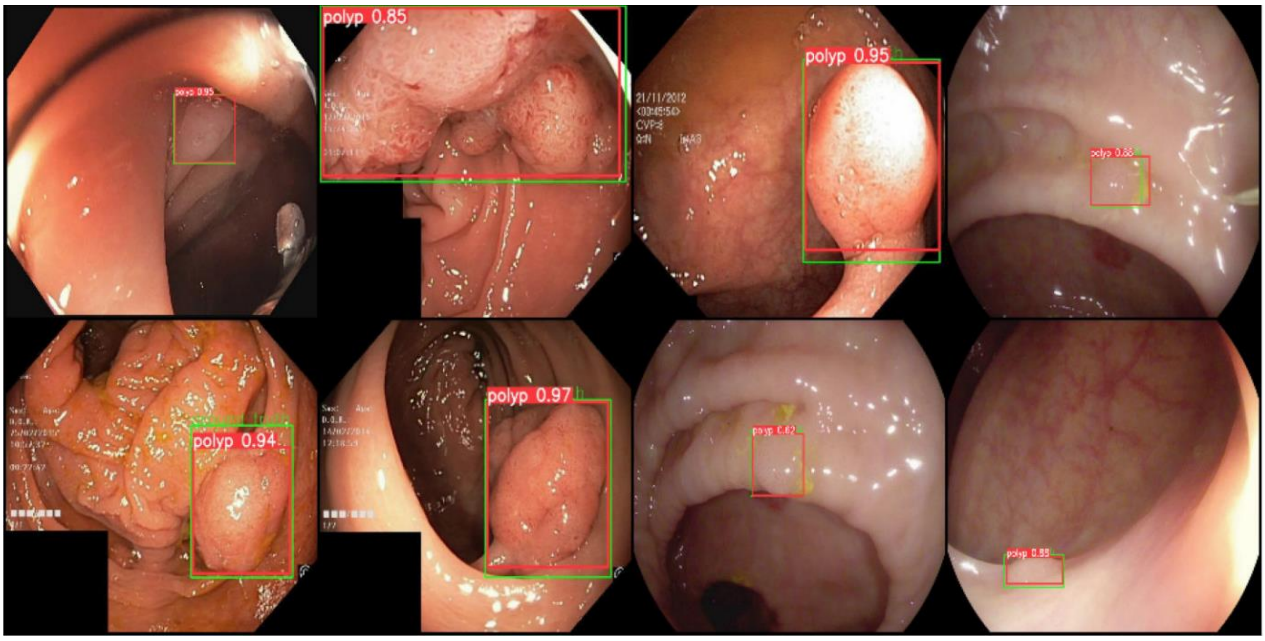


Figure 4. Examples of images containing predictions of the yolov5x model for polyp detection.

The bounding boxes shown in Figure 4 correspond to the true regions, indicated by the green boxes, and the detected regions, indicated by the red boxes.

3.3. Yolov6 Test Results

In this study, yolov6 models were trained at 400 epochs. During training, the batch size and image size of the models were set to 32 and 512, respectively and the optimization algorithm was set to adam. The learning rate and momentum hyperparameters for adam optimization were set to 0.0032 and 0.843, respectively. The intersection over union and confidence thresholds were both set to 0.5 to evaluate models. Table 5 provides the performance outcomes of yolov6 models.

Table 5. Outcomes of the yolov6 models.

Models	Precision	Recall	F1-Score	F2-Score	mAP50	mAP50-95
Yolov6n	0.854	0.777	0.814	0.791	0.873	0.632
Yolov6t	0.761	0.843	0.800	0.825	0.864	0.637
Yolov6s	0.785	0.840	0.811	0.828	0.857	0.632
Yolov6m	0.832	0.875	0.853	0.866	0.881	0.678
Yolov6l	0.842	0.893	0.867	0.882	0.913	0.710

Table 5 shows that the yolov6n model, with a rate of 0.854, offers the greatest precision value. This suggests that the N model performs the best at avoiding false-positive polyp detection. The yolov6l model, with a ratio of 0.893, offers the greatest recall value. The L model has the least number of missed detections in terms of detecting polyps compared to other models. The yolov6l model, with a ratio of 0.867, achieves the greatest f1-score. This implies that the L model has the highest precision to recall ratio when compared to other models. With a rate of 0.882, the yolov6l model has the highest f2-score. This suggests that, when using the f2-score metric, the yolov6l model performs the best with respect to the recall. The yolov6l model has the highest mAP50 and mAP50:95 metrics with rates of 0.913 and 0.71, respectively. This indicates that, when it comes to recognizing polyps at various intersection over union thresholds, the yolov6l model outperforms the other models. Table 6 provides the speed test outcomes of yolov6 models.

Table 6. The speed test outcomes of the yolov6 models.

Models	(B, C, W, H)	Preprocess(ms)	Inference(ms)	NMS(ms)	Total(ms)
Yolov6n	(32, 3, 512, 512)	0.12	1.24	0.56	1.92
Yolov6t	(32, 3, 512, 512)	0.11	2.26	0.56	2.93
Yolov6s	(32, 3, 512, 512)	0.09	3.07	0.69	3.85
Yolov6m	(32, 3, 512, 512)	0.09	6.40	0.64	7.13
Yolov6l	(32, 3, 512, 512)	0.09	10.08	0.61	10.78

In Table 6, batch size is 32, images are 3-channel and image size is 512. According to the results obtained, the fastest models in terms of preprocessing time are yolov6s, yolov6m, and yolov6l with 0.09 ms, while the slowest model is yolov6n with 0.12 ms. In terms of NMS, the fastest models are yolov6n and yolov6t with 0.56 ms. The slowest model in terms of NMS is yolov6s with 0.69 ms. Due to its significantly larger parameter count, the yolov6l model demonstrates the slowest inference time, taking approximately 10.08 milliseconds. With the smallest parameter count among the models, the yolov6n model achieves the fastest inference time, completing in just 1.24 milliseconds.

In terms of the overall time elapsed, it can be observed that the yolov6l model exhibits a comparatively slower performance, while the yolov6n model demonstrates superior speed. In terms of speed, the models from the fastest to the slowest are the N, T, S, M, and L models. The yolov6t model presents an optimal compromise with regard to accuracy and speed, offering a commendable mean average precision (mAP50:95) rate of 0.637 alongside a swift inference time of 2.93 milliseconds. Figure 5 shows example images containing the polyp predictions of the yolov6l model.

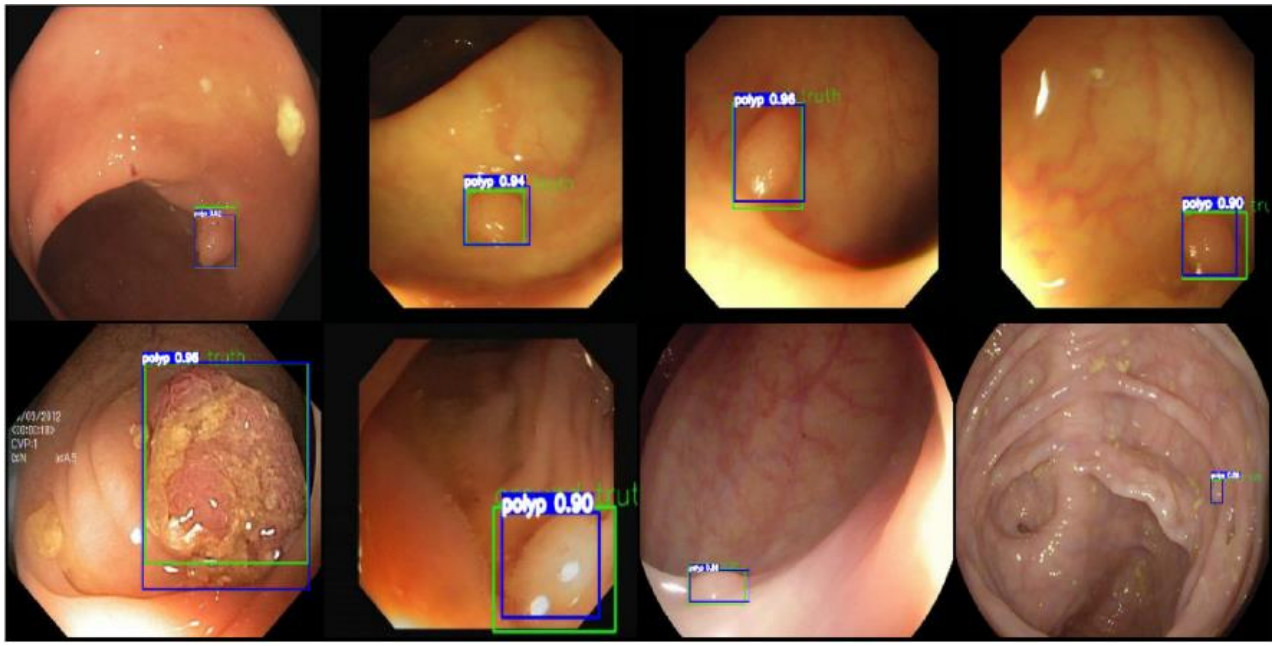


Figure 5. Examples of images containing predictions for polyp detection by the yolov6l model.

Figure 5 depicts the actual box, which is shown in green, and the detected box, which is shown in blue. When yolov5 and yolov6 were compared, the yolov5m model obtained the highest precision value with 0.92. In terms of precision, the other models of the yolov5 also outperformed all the models of the yolov6. From this, it can be inferred that the false positive rate of yolov5 models is smaller compared to yolov6 models. In terms of recall, the yolov6l and yolov6m models outperformed all models of yolov5 with a ratio of 0.893 and 0.875, respectively. This indicates that the ability to not miss polyps is better in yolov6. The yolov5l model was particularly better than all of the yolov6 models, with a f1-score ratio of 0.878. It is clear from comparing the yolov6 and yolov5 models that the yolov5 model provides the best balance between precision and recall. The yolov6l model, which obtained a mAP50 score of 0.913, performed better than any of the yolov5 models. This suggests that the yolov6l model is the best option in regard to mAP50 when using an intersection over union threshold of 0.5. With a mAP50:95 score of 0.783, the yolov5x model performed better than any of the yolov6 variants. In terms of mAP50:95 at various intersections over union thresholds, this suggests that yolov5x performed better than the other models of yolov6.

The slowest model among the yolov5 models is yolov5x, which has a speed of 19.3 ms, while the slowest model among the yolov6 models is yolov6l, which has a speed of 10.78 ms. Compared to the yolov5x model, the yolov6l model exhibits an 8.52 ms speed advantage. With an impressive speed of 2.1 ms, the yolov5n model is the fastest of the examined yolov5 versions. At 1.92 ms, the yolov6n model is the quickest of the yolov6 variants. Yolov6n model is faster than yolov5n model. The evaluation revealed that the yolov6 models exhibit superior speed in the context of polyp detection compared to the yolov5

models. The yolov5x and yolov6l models, which show the best performance in terms of the mAP50:95 metric, are compared with previous studies in the literature that perform polyp detection in Table 7.

Table 7. Comparison of yolov5x and yolov6l models with previous studies in the literature.

Paper	Method	Precision	Recall	F1-Score	F2-Score	mAP50	mAP50-95
Proposed models	Yolov5x	0.909	0.847	0.876	0.858	0.896	0.783
	Yolov6l	0.842	0.893	0.867	0.882	0.913	0.710
(Shin et al., 2018)	Faster R-CNN	0.815	0.803	0.809	-	-	-
(Pacal and Karaboga, 2021)	Yolov4-CSP	0.916	0.825	0.868	-	-	-
(Karaman et al., 2022)	Yolov4-CSP+ABC	0.860	0.840	0.850	-	0.890	-
(Taş and Yılmaz, 2021)	Faster R-CNN	0.710	0.844	0.770	0.810	-	-
(Xu et al., 2021)	Yolov3	0.832	0.716	0.770	0.736	-	-
(Sornapudi et al., 2019)	R-CNN	0.729	0.802	0.764	-	-	-

The performance results of studies that carried out polyp detection in the literature, along with the proposed yolov5x and yolov6l models, are given in Table 7. Based on the acquired results, the yolov4-CSP model, as suggested by (Pacal and Karaboga, 2021), exhibited the highest precision performance, attaining a value of 0.916. This indicates that the yolov4-CSP model proposed by (Pacal and Karaboga, 2021) has the least false detections. The precision value of the yolov5x model is 0.909. The yolov5x model exhibits the highest precision performance compared to the other studies, with a precision value very similar to that of the yolov4-CSP model proposed by (Pacal and Karaboga, 2021). Compared to other studies, the yolov6l and yolov5x models exhibit the best performance with recall values of 0.893 and 0.847, respectively. This indicates that the yolov6l and yolov5x models have the best capacity to identify polyps. The yolov5x model displayed the best performance in terms of the f1-score metric, with a value of 0.876. After the yolov5x model, the yolov4-CSP model put forward by Pacal and Karaboga (2021) obtained the greatest f1-score value of 0.868. The yolov5x and yolov6l models we suggested operate differently from the yolov4-CSP model put forward by Pacal and Karaboga (2021). The yolov4-CSP model performed the best in terms of precision. Our suggested yolov6l model, however, performed better than the yolov4-CSP model in terms of recall. The yolov5x model offered the best results in terms of the f1-score. The yolov5x model performed better than the yolov6l and yolov4-CSP models because the f1-score is the harmonic mean of recall and precision. At a value of 0.882, the yolov6l model had the best performance in terms of the f2-score metric. Furthermore, the yolov5x model trailed the yolov6l model with a value of 0.858, achieving the second-best performance. The yolov3 proposed by Xu et al. (2021) and the quicker R-CNN proposed by Taş and Yılmaz (2021) both achieved f2-score values of 0.736 and 0.81, respectively. Our suggested models performed better than the models in the references

(Taş and Yılmaz, 2021) and (Xu et al., 2021), especially when it came to the f2-score metric. Furthermore, the yolov6l model showed the best performance when taking the mAP50 measure into account, with a value of 0.913. This suggests that the yolov6l model obtained the highest average precision value when the intersection over union threshold was set to 0.5. The yolov5x model has the highest mAP50 value (0.896), followed by yolov6l. The yolov4-CSP+ABC model put forward by Karaman et al. (2022) has a mAP50 value of 0.89. The yolov5x and yolov6l models performed better than the yolov4-CSP+ABC model put forward by Karaman et al. (2022) in terms of mAP50. There was no information regarding mAP50:95 in other studies. As a result, yolov5x performed better than yolov6l when our suggested models were evaluated, with a mAP50:95 score of 0.783.

The quicker R-CNN model achieved precision, recall, and f1-score values of 0.815, 0.803, and 0.809, respectively, according to the research conducted by Shin (2018). Our suggested models outperformed the quicker R-CNN model put forward by Taş and Yılmaz (2021) in terms of precision, recall, and f1-score measures. Taş and Yılmaz (2021) reported that the quicker R-CNN model achieved a recall value of 0.844. This figure is poorer than our models, but it is better than other studies. The precision value of the yolov3 proposed by Xu et al. (2021) was 0.832. This result is worse than yolov5x and yolov6l, but better than the models put forward by Taş and Yılmaz (2021) and Sornapudi et al. (2019). With a ratio of 0.802, the R-CNN model put forward by Sornapudi et al. (2019) performs better than the yolov3 model put forward by Xu et al. (2021), but it is still inferior to our models and previous research.

3. Conclusions

Through the identification of polyps, computerized detection has the potential to reduce the death rate associated with colorectal cancer. The death rate from colorectal cancer can be decreased by using computer-aided diagnosis methods to detect polyps. By identifying polyps, computer-aided diagnostic tools can lower the death rate from colorectal cancer. Consequently, within the framework of this investigation, real-time polyp detection was carried out using the yolov5 and yolov6 object detection models. To enhance the performance of object detection, this study incorporated data augmentation techniques and introduced negative samples into the training data to mitigate the false positive rate.

According to the obtained results, the yolov6l model outperformed the yolov5x model, which had a mAP50 value of 0.896, with a mAP50 value of 0.913. When considering the mAP50-95 measure, the yolov5x model scored better than the yolov6l model, with a mAP50-95 score of 0.783 versus 0.71. The matching yolov5 models were all outperformed by the yolov6 in terms of speed. Additionally, with f1-score values of 0.876 and 0.867, respectively, the yolov5x and yolov6l models outperformed earlier models documented in the literature. In future studies, real-time multi-class polyp detection is going to be carried out as a computerized diagnostic system.

Acknowledgment

This study was supported by AKGUN Computer Incorporated Company and TUBITAK(Scientific and Technological Research Council of Turkey). We would like to thank AKGUN Computer Inc. and TUBITAK for providing all the necessary resources and funding for the execution of this study.

References

Bernal J., Sanchez J., Vilarino F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition 2012; 45(9). 3166-3182.

Bernal J., Sanchez FJ., Esparrach GF., Gil D., Rodriguez C., Vilarino F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics 2015; 43: 99-111.

Bochkovskiy A., Wang CY., Liao HYM. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Buslaev A., Iglovikov V., Khvedchenya E., Parinov A., Druzhinin A., Kalinin M., Alexandr A. Albumentations: fast and flexible image augmentations. Information 2020; 11(2).

Chen BL., Wan JJ., Chen TY., Yu YT., Ji M. A self-attention based faster R-CNN for polyp detection from colonoscopy images. Biomedical Signal Processing and Control 2021; 70: 103019.

Diederik PK. Adam: A method for stochastic optimization. 2014.

Ding X., Zhang X., Ma N., Han J., Ding G., Sun J. RepVGG: making vgg-style convnets great again. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, 13728-13737, Nashville, TN, USA.

Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression. arXiv preprint arXiv:2205.12740.

Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580-587, Columbus, OH, USA.

Girshick R. Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), 2015, 1440-1448, Santiago, Chile.

He K., Zhang S., Ren S., Sun J. Deep residual learning for ımage recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770-778, Las Vegas, NV, USA.

Jha D., Smedsrud PH., Riegler MA., Halvorsen P., Lange TD., Johansen D., Johansen HD. Kvasir-SEG: A segmented polyp dataset. 2020 Multimedia Modeling, 2020, 451-462, Daejeon, South Korea.

Karaman A., Karaboga D., Pacal I., Akay B., Basturk A., Nalbantoglu U., Coskun S., 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. Applied Intelligence 2022; 53(12): 15603-15620.

Kurniawan H., Arief MAA., Manggala B., Lee S., Kim H., Cho BK. Advanced detection of foreign objects in fresh-cut vegetables using YOLOv5. LWT-Food Science and Technology 2024; 212: 116989.

Lin TY., Goyal P., Girshick R., He K., Dollar P. Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision(ICCV), 2017, 2999-3007, Venice, Italy.

Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu CY., Berg AC. SSD: Single shot multibox detector. 2016 European Conference on Computer Vision(ECCV), 2016, 21-37, Amsterdam, Netherlands.

Li C., Li L., Jiang H., Weng K., Geng Y., Li L., Ke Z., Li Q., Cheng M., Nie W., Li Y., Zhang B., Liang Y., Zhou L., Xu X., Chu X., Wei X., W Xiaolin. Yolov6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976, 55.

Liu S., Qi L., Qin H., Shi J., Jia J. Path aggregation network for instance segmentation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018, 8759-8768, Salt Lake City, UT, USA.

Mirza M., Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Misra D. Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681.

Mushtag F., Ramesh K., Deshmukh S., Ray T., Parimi C., Tandon P., Jha PK. Nuts and bolts: yolo-v5 and image processing based component identification system. Engineering Applications of Artificial Intelligence 2023; 118: 105665.

Pacal I., Karaboga D. A robust real-time deep learning based automatic polyp detection system. Computers in Biology and Medicine 2021; 134: 104519.

Pacal I., Karaman A., Karaboga D., Akay B., Basturk A., Nalbantoglu U., Coskun S. An efficient real-time colonic polyp detection with yolo algorithms trained by using negative samples and large datasets. Computers in Biology and Medicine 2022; 141: 105031.

Qian Z., Jing W., Lv Y., Zhang W. Automatic polyp detection by combining conditional generative adversarial network and modified you-only-look-once. IEEE Sensors Journal 2022; 22(11): 10841-10849.

Rahim T., Hassan SA., Shin SY. A deep convolutional neural network for the detection of polyps in colonoscopy images. Biomedical Signal Processing and Control 2021; 68: 102654.

Ren S., He K., Girshick R., Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017; 39(6): 1137-1149.

Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; 779-788, Las Vegas, NV, USA.

Redmon J., Farhadi A. Yolo9000: better, faster, stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 6517-6525, Honolulu, HI, USA.

Redmon J., Farhadi A. Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767.

Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized intersection over union: a metric and a loss for bounding box regression. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2019; 658-666, Long Beach, CA, USA.

Silva J., Histace A., Romain O., Dray X., Granado B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. International Journal of Computer Assisted Radiology and Surgery 2014; 9: 283-293.

Shin Y., Qadir HA., Aabakken L., Bergsland J., Balasingham I. Automatic colon polyp detection using region based deep cnn and post learning approaches. IEEE Access 2018; 6: 40950-40962.

Sornapudi S., Meng F., Yi S. Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps. Applied Sciences 2019; 9(12).

Sun X., Jia X., Liang Y., Wang M., Chi X. A defect detection method for a boiler inner wall based on an improved yolo-v5 network and data augmentation technologies. IEEE Access 2022; 10: 93845-93853.

Tang CP., Hsieh CH., Lin TL. Computer-aided image enhanced endoscopy automated system to boost polyp and adenoma detection accuracy. Diagnostics 2022; 12(4).

Taş M., Yılmaz B. Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images. Computers and Electrical Engineering 2021; 90: 106959.

Wang CY., Liao HYM., Wu YH., Chen PY., Hsieh JW. Cspnet: A new backbone that can enhance learning capability of cnn. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, 1571-1580, Seattle, WA, USA.

Wang CY., Bochkovskiy A., Liao HYM. Scaled-yolov4: scaling cross stage partial network. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2021; 13024-13033, Nashville, TN, USA.

Wang S., Yin Y., Wang D., Lv Z., Wang Y., Jin Y. An interpretable deep neural network for colorectal polyp diagnosis under colonoscopy. Knowledge-Based Systems 2021; 234: 107568.

Xiang L., Wang W., Wu L., Chen S., Hu X., Li J., Tang J., Yang J. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. Advances in Neural Information Processing Systems 2020; 33: 21002-21012.

Xiao Y., Chang A., Wang Y., Huang Y., Yu J., Huo L. Real-time object detection for substation security early-warning with deep neural network based on yolo-V5. 2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET), 2022, 45-50, Arad, Romania.

Xu J., Zhao R., Yu Y., Zhang Q., Bian X., Wang J., Ge Z., Qian D. Real-time automatic polyp detection in colonoscopy using feature enhancement module and spatiotemporal similarity correlation unit. Biomedical Signal Processing and Control 2021; 66: 102503.

Zhang H., Wang Y., Dayoub F., Sünderhauf N. VarifocalNet: an iou-aware dense object detector. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2021; 8510-8519, Nashville, TN, USA.

Zheng Z., Wang P., Liu W., Li J., Ye R. Distance-IoU loss: faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence 2020; 34(7): 12993-13000.