ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Optimized Transfer Learning for Polyp Detection

Noppakun Boonsim[1] and Saranya Kanjaruek[2]

ABSTRACT

Early diagnosis of colorectal cancer focuses on detecting polyps in the colon as early as possible so that patients can have the best chances for successful treatment. This research presents the optimized parameters for polyp detection using a deep learning technique. Polyp and non-polyp images are trained on the InceptionResnetV2 model by the Faster Region Convolutional Neural Networks (Faster R-CNN) framework to identify polyps within the colon images. The proposed method revealed more remarkable results than previous works, precision: 92.9 %, recall: 82.3%, F1-Measure: 87.3%, and F2-Measure: 54.6% on public ETIS-LARIB data set. This detection technique can reduce the chances of missing polyps during a prolonged clinical inspection and can improve the chances of detecting multiple polyps in colon images.

Article information:

Keywords: Polyp Detection, Transfer Learning, InceptionRestnetV2

Article history:
Received: November 29, 2022
Revised: January 6, 2023
Accepted: January 21, 2023
Published: February 18, 2023 (Online)

DOI: 10.37936/ecti-cit.2023171.250910

1. INTRODUCTION

According to the International Agency for Research on cancer's report on cancer in 2020 [1], colorectal cancer was the third (female 0.86 million cases) and fourth (male 1.06 million cases) cause of death worldwide. Medical imaging is a crucial technique that uses for early cancer detection. If any malignancy is detected, it can be diagnosed and cured at an early stage, leading to higher rates of successful treatment and extending life. Physicians can carry out early detection via examinations, screening symptoms, or by using medical imaging (X-ray, Computed Tomography, and Magnetic Resonance Imaging). Medical screening is suitable for both colorectal and cervical cancer [1]. This process, screening medical images, is ideal for colorectal and cervical cancer [1].

Colorectal cancer can early detect by putting a camera through a patient's colon to find polyps, excising them, and curtail the cancer's growth. Colorectal cancer causes by abnormal polyps and can be detected early by inserting a camera to find evidence of polyps in the intestine. The abnormal polyps are initially benign, but they might become malignant over time. Colonoscopy screening can help with early detection by putting a camera through the colon and will find the suspicious polyp from obtained images. The camera can be flexible (sigmoidoscopy), colonoscopy, and wireless camera (wireless capsule endoscopy). Sigmoidoscopy is a flexible camera that physicians will put through a patient's anus to inspect the colon to look for polyps. However, the camera has limitations because it can only examine from the rectum to the sigmoid. Next, colonoscopy is like sigmoidoscopy but not the same. Colonoscopy can inspect almost any part of the colon, about 1,200-1,500 mm, while a sigmoidoscopy can only examine the distal portion of about 600 mm of the colon [3].

On the other hand, Wireless Capsule Endoscopy (WCE) [4] is a small camera (capsule size) including wireless communication, which a patient can swallow to capture the entire colon. After a patient swallows a WCE, it will traverse through the digestive tract for about 8 hours, producing about 50,000 images (2-3 images per second). Next, a physician will inspect all very time-consuming and labor-intensive images. This study aims to develop a deep learning technique to accurately detect polyp positions in colonoscopy images from the public dataset. The following three significant contributions represent the benefits of this method:

Firstly, an automated detection process can be beneficial in assisting physicians in discovering polyps better and reducing their workloads. Furthermore, this technique can help inexperienced physicians or medical staff members by offering support.

[1,2] The authors are with Faculty of Interdisciplinary Studies, Khon Kaen University, Thailand., E-mail: boonsim@kku.ac.th and kanjaruek@kku.ac.th

Secondly, a proposed deep learning strategy can solve the problem of variations in the appearance of the polyps for detecting polyps in the colon images.

Finally, this approach achieved superior performance on public datasets for detecting polyps compared to the state-of-the-art methods.

The remainder of this paper organizes as follows: Section 2 presents work related to polyp detection and the deep learning technique. Section 3 proposes the methodology in detail. Section 4 describes the experimental results and comparisons. Finally, the conclusion and future works summarize in Section 5.

2. LITERATURE REVIEWS

Many researchers have presented polyp detection in colon images. The early work was proposed by Karkanis et al. [5]. Color wavelet transformation was used to extract polyp information. Linear Discrimination Analysis (LDA) was utilized to classify the polyps and the non-polyp images. The experimental results were tested on 1,200 randomly selected images from 66 patient videos, and the Precision and Recall values of 99.3% and 93.6% were respectively reported. However, with large variations in the polyps' colors, the Karkanis' method can only detect some polyp colors. Due to this, the color-based method is sensitive to lighting conditions. In different light intensities, the colors of the polyps can truly change. In work [6], the texture features, the Grey-Level-Co-occurrence, and the Local Binary Pattern (LBP) were presented to obtain information on the polyps.

Moreover, the Support Vector Machine (SVM) was used to classify those features. The study was carried out with 1,736 polyp images, and the area under the ROV curve reported classification results, which were found to be 0.96. An edge-based technique was proposed to detect polyps [7]. The detection algorithm began with applying the Log Gabor filter to extract the polyp regions. The Susan edge detector was applied to extract the features of the polyps' edges. The Log Gabor output and the edge features were combined as polyp features. The SVM technique classified these features, and the detection Precision and Recall were reported at 96.7% and 72.5%, respectively after 50 polyp images had been evaluated. Li et al. [8] presented a combined wavelet transform feature and the LBP feature. Polyp images were classified by SVM. The technique was evaluated on a dataset with 1,200 images (600 normal and 600 polyp images). The classification accuracy was described at approximately 91.6%. Many features were integrated into the polyp detection technique to improve the performance. In work [9], a combined technique of many features, such as the Scale Invariant Feature Transform, LBP, and a Histogram of the Oriented Gradient, was presented. These features were grouped using the k-mean technique, which can be called a "bag of features" and classified by SVM and Fisher's LDA. Therefore, Li et al. [8] reported the experimental results were to outperform the method with 93.20% accuracy, which examined 2,500 images consisting of 500 polyp images and 2,000 typical colon images. The deep learning technique was presented in several works [9, 10, 11] to manage various polyp information. Deep feature learning is a sub-type of machine learning that excludes the manual step of extracting features from the images. The technique learns features directly from images that load to the learning algorithm. The advantages of this method are avoided in the feature extraction step, and classification accuracy outperforms the previous methods. However, achieving the performance of the deep learning techniques depends upon suitable architecture, parameters, and quality images to place into the networks [12]. Convolutional Neural Network (CNN) was presented to identify the polyp images in work [9]. The network architecture consisted of three Convolutional layers, three max-pool layers, one feature layer, and one output layer. The study tested 62 images of the CVC-CLINIC database [12]. Precision and recall were reported at 65.7% and 82.7%, respectively. In work [10], CNN based on AlexNet [13] was proposed. The appearances of the polyps (i.e., their colors, shapes, and temporalities) were separately extracted and sent into the proposed network.

Regarding the experimental results, significant improvements were reported over the state-of-the-art techniques, and the number of false alarms was reduced. The work of Brandao et al. [14] presented a fully convolutional neural network of the VGG network [15]. The experimental results reported a detection Precision and Recall of 73.61% and 83.31%, respectively. The technique was tested on 4,664 polyp images in three standard datasets: CVC-CLINIC [12], ETIS-LARIB [16], and ASU-Mayo [17]. With large variations in the appearance of the polyps, the textures, shapes, sizes, colors, and environments, including light intensity and image quality, are problematic concerning polyp detection. Some previous studies have only been able to detect some polyp types. Moreover, in most studies, the researchers implemented their datasets.

Recently, Shin et al. [18] presented CNN and the learning process to detect polyps. The research used image augmentation to place more images into the training set. The experiments revealed that the technique had outperformed the previous deep learning techniques.

Several object detection frameworks exist, such as the region-based approach, the Region CNN (R-CNN) [19], the You Only Look Once, and the Single Shot Detection. Regarding object detection accuracy, the region-based method has been shown to outperform the other techniques, but its detection time is the slowest [20]. The method has two sub-processes: 1) extracting possible object regions (region propos-

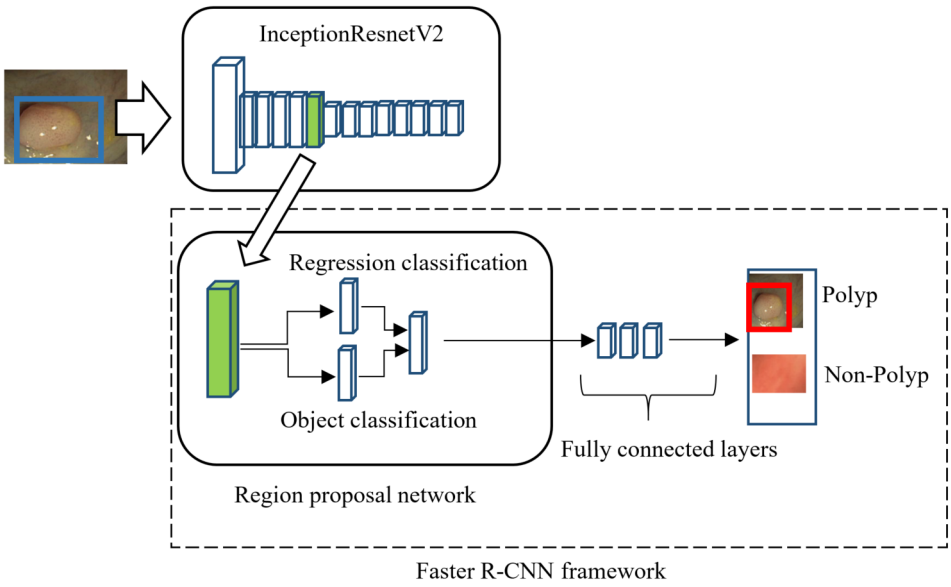

Fig.1: *The proposed polyp detection architecture.*

als) and 2) the detection sub-process. R-CNN has been modified to reduce detection time by utilizing Fast R-CNN [21] and Faster R-CNN [22]. Faster R-CNN, which uses a region proposal network to extract region proposals, can reduce the object detection time from 49 seconds to 0.2 seconds per image.

The You Only Look Once (YOLO) [23] technique separates the image into grid cells. Each grid cell is classified into an object class with a confidence score. Any objects with confidence scores that were less than the threshold were removed. Single Shot Detection (SSD) [20] is the fastest object detection framework, and its detection time is 59 frames per second (fps). In contrast, YOLO and R-CNN spend 45 and 7 fps to detect object location, respectively. The technique implements a single deep network by combining the region proposals and the feature extraction process. After that, the extracted features are sent through a classification network to classify the object's class.

All frameworks can be modified by using another pre-trained network such as Resnet101 [24], MobileNetV2 [25], GoogleNet [26], InceptionV3 [27], InceptionResnetV2 [28], EfficientNetB0 [29], Densenet201 [30] and so on to be a based network. After that, the framework must choose a feature extraction layer for transfer learning features to solve a new problem and might have better outcomes. Therefore, the work of polyp detection still offers a challenging problem.

This research presents the most accurate method for detecting polyps using the Fast R-CNN framework, which is considered suitable for the job in the medical field that requires high accuracy in diagnosing the discovery polyp. The research uses InceptionResnetV2 as a based network because it was found that the network gave the best results in experiments.

3. METHODOLOGY

3.1 Proposed methodology

The proposed technique is based on the Faster R-CNN framework, as shown in Fig. 1. The architecture receives the image into the network, which performs resizing to 400 × 400 pixels, and then the image is sent to the region proposal network (RPN) to generate the region proposals (RP) which may contain polyps. In this study, 256 region proposals were defined in processing to find polyps in the image. Subsequently, the RPs are dispatched to the classification layer to distinguish whether the RP is an object.

The RPs are also sent to the regression classification layer to predict the location of the found object. Finally, objects found will be reclassified at the fully connected layer to classify again as polyps or non-polyps. This research utilized the InceptionResnetV2 transfer learning model, while the Faster Region Convolution Neural Network (Faster R-CNN) [22] framework was used to modify the network to detect polyps.

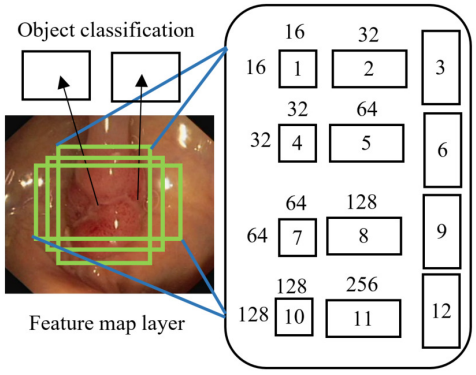

Fig.2: *Examples of anchor boxes.*

From the introduction section, many pre-trained networks can be used as a based network. Some experiments were conducted to determine the best layer of the InceptionRestnetV2 network [28] for the feature extraction layer. InceptionRestnetV2 network has a total layer of 164 layers. The experiments were randomly implemented on those layers such as "block1_1_ac", "block2_2_ac", "block3_3_ac," and "block4_4_ac" layers and found that layer "block17_20_ac" showed the best accuracy. Subsequently, the layer was used to fine-tunning for the best results.

After obtaining the feature extraction layer, anchor boxes are defined to determine the possible size of the polyp. The selection of effective anchor boxes is important in the training process, which can improve accuracy and spend less time on parameter tuning. Anchor boxes are bounding boxes that define object size, width, and height to be detected. During training, predefined anchor boxes are placed all over the image. It then predicts the object found in the position of the anchor box. The predictions will be refined to the ground truth object position as much as possible.

This study experimentally determined the number of anchor boxes, for example, 3, 6, 9, 12, and 15, for each feature map in the feature extraction layer. The number of anchor boxes at 12 reported the best performance. There are anchor boxes of sizes 16, 32, 64, and 128. Each box has other box aspect ratios of 2 and 0.5 for 12 boxes, as shown in Fig. 2.

3.2 Model learning and Implement

In the proposed object detection algorithms, the loss function is based on [21] consisting of the sum of the object localization (loc) and object classification (cls). Most object detection algorithms use the Intersection of Union (IoU) to determine the degree of overlap between the predicted box and the ground target box. The loss function is formulated as follows:

$$L(\{p_i\},\{t_i\}) = L_{cls} + \lambda * L_{loc} \tag{1}$$

$$L_{cls} = \frac{1}{N_{cls}} \sum_{i=1}^{N} L_{clsloss}(p_i, p_i^*) \tag{2}$$

$$L_{loc} = \frac{1}{N_{loc}} \sum_{i=1}^{N} p_i^* L_{reg}(ti, t_i^*) \tag{3}$$

Where λ denotes the weight term to balancing between L_{cls} and L_{loc}, N_{cls} is the number of classes. N_{loc} is the number of object locations, p_i, and p_i^* represent the classes confidences of anchor i and the label of target box i, respectively, t_i and t_i^* denote the localization vectors of the prediction box and target box, consisting of central point coordinate (cx, cy) and box's width and height (w, h). The classification loss $L_{clsloss}$ is log loss over two classes (object vs. not object). The term p_i^* L_{reg} means the regression loss is activated only for the positive anchor (p_i^*=1) and disabled otherwise (p_i^*=0). The outputs of the cls and reg layers consist of $\{p_i\}$ and $\{t_i\}$, respectively.

As mentioned above, in this work, the state-of-the-art CNN network, InceptionResnetV2, pre-trained on ImageNet [13], is transferred to learn polyp feature representations from the labeled training data. For the network architecture, all the layers were copied from the InceptionResnetV2 model to a target network except the last fully connected layer. Then the last fully connected layer was modified for adapting the CNN model to the polyp classification task by replacing the last fully connected layer (intended for 1000 classes) with a new fully connected layer for the two classes, polyps, and non-polyps. Next, the initial CNN filter weights derived from the natural images were then fine-tuned (optimized) using the training data, i.e., the polyp images and corresponding bounding boxes through back-propagation.

To minimize the loss function, the stochastic gradient descent (SGD) method, which calculates a random subset of training examples, was used to estimate the mean gradient for all the training examples. The learning rate was set to 0.001, and the momentum was 0.9. In general, the early layers of a CNN learn low-level image features, which apply to most vision tasks, but the late layers learn high-level features specific to the application at hand. Therefore, fine-tuning the last few layers is usually sufficient for transfer learning.

However, because the distance between images in ImageNet [13] and polyps are significant, therefore in our work, fine-tuned early layers have been employed as well. The algorithm starts from the last layer and then incrementally includes more layers in the updating process until the desired performance is reached. The study set the overlap value between the polyp prediction location and target box (ground truth) to more than 0.6, and if the overlap value is less than 0.3, it will be defined as non-polyps.

4. EXPERIMENTAL RESULTS

The experiments were conducted on publicly available polyp datasets. The research used ETIS-LARIR

Table 1: *Information about public data sets used in this study.*

Dataset	Polyp Image	Resolution	Purpose	Download link
CVC-Clinic	612	384 × 288	Training	https://polyp.grand-challenge.org/CVCClinicDB/
Kvasir	1000	720 × 576	Training	https://datasets.simula.no/kvasir/
ETIS-LARIR	196	1255 × 966	Testing	https://polyp.grand-challenge.org/EtisLarib/

[16] for testing the model, while the training of the model was performed on the CVC-CLINIC [12] and the Kvasir [31]. Detailed information for each data set, such as the number of images, image resolutions, and the purpose of use, is given in Table 1.

Due to the different image sizes in the dataset, images were resized into fixed dimensions with a spatial size of 400 × 400 before feeding the network.

The research applied image augmentation to oversampling the data set to reduce the chance of overfitting on our model and increase robustness.

Training images were applied augmentation techniques: random rotation between -30 degrees and 30 degrees, zooming in of 1 to 1.2x for small polyps and zooming out of 0.8x to 1 for large polyps, translation in x, y direction between 0 and 20 pixels and shearing in x- and y-axis between 0 and 15 pixels followed by centering the polyp.

All experiments in this study were done on a Window 10 machine. The features of the computer used for training and testing the deep learning model are an Intel Core i7-10700 (2.9 GHz) processor, 32 GB DDR4 RAM, and a single NVIDIA GeForce GTX 1080 graphic card. MATLAB R2022b (Trial version) Experiments have been conducted with image processing and deep learning toolboxes.

Training a network with a pre-trained network is more advantageous than training from scratch. Moreover, a deep learning model needs a large data set to train and gain good classification accuracy. Transfer learning is a machine learning method that is used to transfer the knowledge gained during training in one type of problem to a different area or related task. Fine-tuning, a concept of transfer learning, also includes improvements to the hyper-parameters to ensure the network has the best performance.

Optimization algorithms, also known as optimizers, are the main approach to minimizing the error rate in training deep learning models. Two basic approaches are used to determine the effectiveness of optimizers: speed of convergence, the process of reaching a global optimum for gradient descent, and the ability to generalize the model's performance on new test data. This research used two optimizers, Adam [32] and SGD [33]. In the experiment results, The SGD optimizer performs better than the Adam optimizer in both training and generalization criteria in the data set (CVC-ClinicDB and Kvasir). In addition, by providing hyper-parameter optimization with the help of an optimal selection of values, such as learning rate and momentum, the SGD optimizer was used in this work due to its better performance.

Precision and recall are used to evaluate the effectiveness of the proposed model. F1-Measure and F2-Measure were employed to balance missed polyps and false alarms. Precision is used to measure the predicted positive observation of the predicted observations in a positive class and is formulated as:

$$Precision = \frac{TP}{TP + FP}, \tag{4}$$

True positive (TP) indicates the predicted bounding box falls into the ground truth of the polyp. False positive (FP) denotes the predicted bounding box falling outside the ground truth of the polyp.

A recall is evaluated the proposition of positive observations that are correctly classified, which is formulated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

False negative (FN) indicates no predicted bounding box, but the frame contains a polyp.

The F-Measure (F1, F2) value calculates the harmonic weight of precision and recall, which is formulated as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

$$F1 = \frac{5 \times Precision \times Recall}{4 \times Precision + Recall} \tag{7}$$

Fig. 3 shows examples of the polyp detection results in the ETIS-LARIB images. Polyps are detected with yellow boxes and are labeled as polyps. The evaluation methods used were precision and recall,

Table 2: *A comparison of the proposed method as compared to other pre-trained networks.*

Method	TP	FP	FN	Precision (%)	Recall (%)	F1 (%)	F2 (%)	Speed (ms.)
faster r-cnn + VGG16	107	17	85	86.3	55.7	67.7	42.3	9377
faster r-cnn + VGG19	112	11	80	91.1	58.3	71.1	44.4	1421
faster r-cnn + Squeeznet	72	71	120	50.3	37.5	43.0	26.9	379
faster r-cnn + Resnet50	102	26	90	79.7	53.1	63.8	39.8	2722
faster r-cnn + Resnet101	83	18	109	82.2	43.2	56.7	35.4	6985
faster r-cnn + GoogleNet	87	30	105	74.4	45.3	56.3	35.2	1488
faster r-cnn + MobileNetV2	100	17	92	85.5	52.1	64.7	40.5	2331
faster r-cnn + InceptionV3	145	22	47	86.8	75.5	80.8	50.5	2012
faster r-cnn + InceptionResnetV2 (**Proposed method**)	158	12	34	92.9	82.3	87.3	54.6	410

F1-Measure, and F2 Measures, as shown in equations 4, 5, 6, and 7, respectively.

The proposed method was compared with the Faster R-CNN framework but changed the pre-trained networks such as VGG19, GoogleNet, MobileNetV2, and InceptionV3 to evaluate the performance, as shown in Table 2. All methods were trained with the CVC-ClinicDB and Kvasir to make a fair performance comparison.

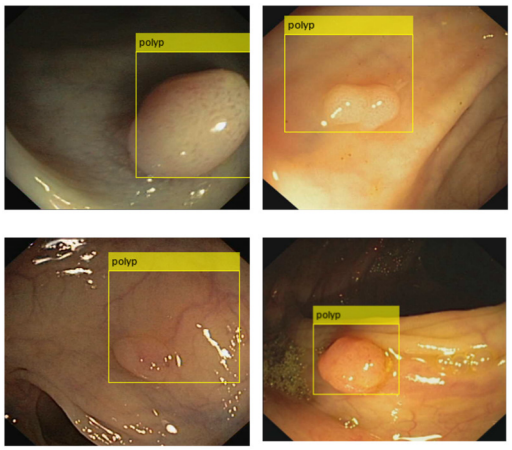

Fig.3: *Examples of the polyp detection results.*

For evaluation, ETIS-LARIR was used for testing. Furthermore, other parameters used in the training process were used the same such as activation function, loss function, augmentation techniques, number of epochs, learning rate, and optimizer. The model evaluation used the IOU threshold of 0.6, batch size of 1, and confidence threshold of 0.5. The proposed method located at the end of the table shows the best polyp detection result, in which the precision, recall, F1- Measure, and F2-Measure were 92.9%, 82.3%, 87.3%, and 54.6%, respectively.

Table 3 shows a comparison of the proposed method with other state-of-the-art methods. All methods were tested on the same dataset, ETIS-LARIR, which has 196 images. Some methods were reviewed in the 2015 MICCAI challenge [12]. The top three results were selected: CUMED and OUS, and these teams used CNN-based end-to-end learning for the polyp detection task. CUMED (Department of Computer Science and Engineering, Chinese University of Hong Kong) employed a CNN-based segmentation strategy [16] where pixel-wise classification was performed with ground-truth polyp masks. The OUS (Oslo University Hospital, OUS Norway, University of Oslo) team adopted the AlexNet CNN model [12] along with the traditional sliding window approach for patch-based classification [12]. The UNS-UCLAN (School of Engineering, University of Central Lancashire, Preston, UK, and University of Nice-Sophia Antipolis, Nice, France) team utilized three CNNs for feature extraction of different spatial scales and adopted one independent Multi-Layer Perceptron (MLP) network for classification [12]. In addition, other research using the same framework (Faster R-CNN, R-CNN) was compared, for example, Shin et al. [18], Kang et al., 2019 [34] and Sornapudi et al., 2019 [35].

From the results presented in Table 3, the proposed method outperforms other methods in all evaluation terms (precision, recall, F1, and F2-Measure) except the speed, which takes detection time than any other methods of 450 milliseconds per image. There is far from success in real-time performance because the study is a two-stage approach.

Although computational time is expensive and cannot be used in real-time processing, polyp detection work is medical work that requires high accuracy. A processing time of about a half second per image is still acceptable. Although the proposed method shows a more accurate polyp detection rate than other methods, there are still some errors. It was found that the localizations of the sessile polyp, a type of polyp that protrudes to a lesser degree out of the colon, are often missed, as shown in Fig. 4.

Another problem was that the technique had located polyps in bigger regions than polyp size and had located some parts of the polyps, as shown in Fig. 4.

Table 3: *A comparison of the proposed method as compared to other state-of-the-art methods.*

Method	Method	TP	FP	FN	Precision (%)	Recall (%)	F1 (%)	F2 (%)	Speed (ms.)
CUMED [12]	CNN	144	55	64	72.3	69.2	70.7	44.2	200
OUS [12]	CNN	131	57	77	69.7	63.0	66.1	41.4	5000
Shin et al. [18]	Faster R-CNN	148	14	60	91.4	71.2	80	50.0	390
Kang et al. [34]	Mask R-CNN	n/a	n/a	n/a	73.8	73.4	74.1	n/a	n/a
Sornapudi et al. [35]	R-CNN	167	62	41	72.9	80.3	76.4	47.8	317
Qadir et al., 2019 [36]	Mask R-CNN	n/a	n/a	n/a	80.0	72.6	76.1	n/a	430
Jia et al. [37]	Faster R-CNN	170	96	38	63.9	81.7	71.7	44.8	n/a
Qadir et al., 2021 [38]	F-CNN	180	28	28	86.5	86.1	86.3	54.1	39
Proposed Method	Faster R-CNN	158	12	34	92.9	82.3	87.3	54.6	410

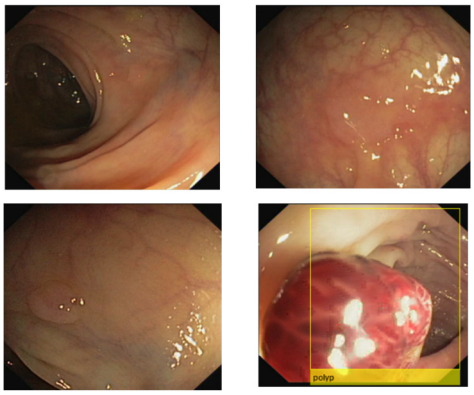

Fig.4: *Some examples of polyps that missed detection.*

5. CONCLUSIONS

In this research, the Faster R-CNN framework and InceptionResnetV2 network, the two-step approach, were used for automatic polyp detection. The experimental results showed a higher accurate detection of polyps than previous methods. This research presented optimized parameters such as activation function, loss function, augmentation techniques, and training model options to have a high polyp detection accuracy.

The future work might be changed to the large pre-trained networks, such as EffienceNet, DenseNet, and NASNet that the accuracy may improve. In addition, the pre-trained networks used are not related to the medical image (ImageNet). If used pre-trained on medical images may improve the polyp detection accuracy.

The Faster R-CNN technique is time-consuming (17 fps) for locating polyps. Future research might employ other techniques, such as SSD and YOLO, to improve the localization speed.

References

[1] International Agency for Research on Cancer, "New Global Cancer Data," accessed January 22, 2021. [online]. Available: `https://www.uicc.org/news/globocan-2020-new-global-cancer-data`

[2] National Institute of Diabetes and Digestive and Kidney Diseases, "Flexible Sigmoidoscopy," accessed May 22, 2018. [online]. Available: `https://www.niddk.nih.gov/health-information/diagnostic-tests/flexible-sigmoidoscopy`

[3] N. N. Baxter, M. A. Goldwasser, L. F. Paszat, R. Saskin, D. R. Urbach and L. Rabeneck, "Association of colonoscopy and death from colorectal cancer," *Annals of Internal Medicine*, vol. 150, no. 1, pp.1–8, 2009.

[4] G. Arkady, "Wireless capsule endoscopy," *Sensor Review*, vol. 23, pp.128–133, 2003.

[5] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras and M. Tzivras, M, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE transactions on information technology in biomedicine*, vol. 7, pp.141–152, 2003.

[6] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung Für Die Medizin, Springer*, pp.346–350, 2009.

[7] A. Karargyris and N. Bourbakis, "Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos," *IEEE Transactions on biomedical engineering*, vol. 58, pp.2777–2786, 2011.

[8] B. Li and M.Q.H. Meng, " Automatic polyp detection for wireless capsule endoscopy images," *Expert Systems with Applications*, vol. 39, pp.10952–10958, 2021.

[9] S. Park, M. Lee and N. Kwak, *Polyp detection in colonoscopy videos using deeply-learned hierarchical features*, Seoul National University, 2015.

[10] N. Tajbakhsh, S. R. Gurudu and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," *Proceeding of 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp.79-83, 2015.

[11] J. Bernal, N. Tajkbaksh, F. J. Sánchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham and K. Pogorelov, "Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge," *IEEE transactions on medical imaging*, vol. 36, no.6, pp.1231–1249, 2017.

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

[13] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no.6, pp.84-90, 2017.

[14] P. Brandao, E. Mazomenos, G. Ciuti, R. Caliò, F. Bianchi, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo and D. Stoyanov, "Fully convolutional neural networks for polyp segmentation in colonoscopy," *Proceeding of Medical Imaging 2017*, vol. 10134, pp.101-107, 2017.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.

[16] J. Silva, A. Histace, O. Romain, X. Dray and B. Granado, "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer," *International Journal of Com-*

puter Assisted Radiology and Surgery, vol.9, pp.283–293, 2014.

[17] N. Tajbakhsh, S. R. Gurudu and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE transactions on medical imaging*, vol.35, pp.630–644, 2016.

[18] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland and I. Balasingham, "Automatic colon polyp detection using region based deep cnn and post learning approaches," *IEEE Access*, vol.6, pp.40950-40962, 2018.

[19] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.580-587, 2014.

[20] W. Liu, D. Anguelov, D. Erhan, C, Szegedy, S. Reed, C. Y. Fu and A. C. Berg, "Ssd: Single shot multibox detector," *Proceedings of the European conference on computer vision*, pp. 21–37, 2016.

[21] R. Girshick, "Fast r-cnn," *Proceeding of the IEEE international conference on computer vision*, pp.1440-1448, 2015.

[22] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp.91-99, 2015.

[23] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.779-788, 2016.

[24] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770-778, 2016.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.4510-4520, 2018.

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1-9, 2015.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.2818-2826, 2016.

[28] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *Proceedings of the Thirty-first AAAI conference on artificial intelligence*, 2017.

[29] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *Proceedings of the International conference on machine learning*, pp.6105-6114, 2019.

[30] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.4700-4708, 2017.

[31] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D. T. Dang-Nguyen, M. Lux, P. T. Schmidt and M. Riegler, "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," *Proceedings of the 8th ACM on Multimedia Systems Conference*, pp.164–169, 2017.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[33] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv:1609.04747*, 2016.

[34] J. Kang and J. Gwak, "Ensemble of instance segmentation models for polyp segmentation in colonoscopy images," *IEEE Access*, vol.7, pp.26440-26447, 2019.

[35] S. Sornapudi, F. Meng and S. Yi, "Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps," *Applied Sciences*, vol.9, no.12, pp.2404-2418, 2019.

[36] H. A. Qadir, Y. Shin, J. Solhusvik, J. Bergsland, L. Aabakken and I. Balasingham, "Polyp detection and segmentation using mask R-CNN: Does a deeper feature extractor CNN always perform better?," *Proceedings of the 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, pp.1-6, 2019.

[37] X. Jia, X. Mai, Y. Cui, Y. Yuan, X. Xing, H. Seo, L. Xing and M. Q. H., Meng, "Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction," *IEEE Transactions on Automation Science and Engineering*, vol.17, no.3, pp.1570-1584, 2020.

[38] H. A. Qadir, Y. Shin, J. Solhusvik, J. Bergsland, L. Aabakken and I. Balasingham, "Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction," *Medical Image Analysis*, vol.68, pp.101897-101906, 2021.

Noppakun Boonsim received B.Sc., and M.Sc. degrees in computer science from Khon Kaen University, Khon Kaen, Thailand, and a Ph.D. degree in computer science from the University of Bedfordshire. He is currently an Assistant Professor at the Faculty of Interdisciplinary Studies, Khon Kaen University, Thailand. His current research interests include image processing computer vision and image pattern recognition with applications.

Saranya Kanjaruek received B.Sc., and M.Sc. degrees in computer science from the Khon Kaen University, Khon Kaen, Thailand, and a Ph.D. degree in information systems from the University of Bedfordshire. She is currently an Assistant Professor at the Faculty of Interdisciplinary Studies, Khon Kaen University, Thailand. Her current research interests include information systems and data analysis.