applied sciences

MDPI

Article

UY-NET: A Two-Stage Network to Improve the Result of Detection in Colonoscopy Images

Cheng-Si He *, Chen-Ji Wang, Jhong-Wei Wang and Yuan-Chen Liu *

Department of Computer Science, College of Science, National Taipei University of Education,
Taipei City 10671, Taiwan; gemini910610@gmail.com (C.-J.W.); ericwang850325@gmail.com (J.-W.W.)
* Correspondence: hesisi05@gmail.com (C.-S.H.); liu@tea.ntue.edu.tw (Y.-C.L.)

Abstract: The human digestive system is susceptible to various viruses and bacteria, which can lead to the development of lesions, disorders, and even cancer. According to statistics, colorectal cancer has been a leading cause of death in Taiwan for years. To reduce its mortality rate, clinicians must detect and remove polyps during gastrointestinal (GI) tract examinations. Recently, colonoscopies have been conducted to examine patients' colons. Even so, polyps sometimes remain undetected. To help medical professionals better identify abnormalities, advanced deep learning algorithms that can accurately detect colorectal polyps from images should be developed. Prompted by this proposition, the present study combined U-Net and YOLOv4 to create a two-stage network algorithm called UY-Net. This new algorithm was tested using colonoscopy images from the Kvasir-SEG dataset. Results showed that UY-Net was significantly accurate in detecting polyps. It also outperformed YOLOv4, YOLOv3-spp, Faster R-CNN, and RetinaNet by achieving higher spatial accuracy and overall accuracy of object detection. As the empirical evidence suggests, two-stage network algorithms like UY-Net will be a reliable and promising aid to image detection in healthcare.

Keywords: colorectal polyp; colonoscopy; deep learning; image detection; object detection algorithm; two-stage network algorithm

check for updates

Citation: He, C.-S.; Wang, C.-J.; Wang, J.-W.; Liu, Y.-C. UY-NET: A Two-Stage Network to Improve the Result of Detection in Colonoscopy Images. *Appl. Sci.* **2023**, *13*, 10800. https://doi.org/10.3390/app131910800

Academic Editors: Teen-Hang Meen and Cosimo Nardi

Received: 29 July 2023
Revised: 19 September 2023
Accepted: 22 September 2023
Published: 28 September 2023

1. Introduction

Cancer is a life-threatening disease that seriously affects human health, accounting for many deaths in Taiwan annually. For instance, approximately 28.0% of deaths in 2021 were cancer-related (51,656 deaths) [1]. Among all types of cancer, colorectal cancer has the second-highest incidence rate and the third-highest mortality rate [2]. Because its early symptoms are often not obvious, regular screening tests are needed to detect them. As statistics show, the earlier colorectal cancer is accurately diagnosed and properly treated, the higher the survival rate. In some cases, the survival rates may even exceed 90% [2]. Since colorectal cancer develops from polyps in the colon, early detection and removal of such polyps at the treatable stage can halt their progression and reduce associated death rates.

As mentioned above, the most effective approach to prevent colorectal cancer is for individuals to undergo regular screening. Among the tools used to achieve this purpose is colonoscopy. It is a highly patient-centered and minimally invasive procedure that enables medical professionals to observe, diagnose, and treat colon abnormalities. Nevertheless, the rates of misdiagnosis for colorectal cancer after a colonoscopy can range from 5% to 27% [3]. At least four reasons can explain these high error rates: (1) inexperienced endoscopists who are not familiar with the appearances of polyps may encounter difficulty in detecting them; (2) polyps smaller than one centimeter can be very smooth, flat and easily overlooked [4]; (3) polyps may exist beyond the field of view of endoscopes and (4) abnormalities may remain unnoticed due to rapid movements of endoscopes during examinations [5].

However, with the assistance of advanced technologies, clinical data such as colonoscopy images can be stored appropriately for instant and subsequent analysis. This advantage benefits patients, medical practitioners, and the healthcare system. Doctors can save as many images as they need for an immediate diagnosis. These digitalized images can also be scrutinized for further justification and assessment. Moreover, such images can be applied to train computer-aided diagnosis (CADx) algorithms (e.g., image detection algorithms) to assist medical professionals in making correct diagnoses [6]. For instance, busy and exhausted proctologists, particularly those working in understaffed medical institutions, can employ CADx tools to help detect abnormalities from images and reduce the miss rate for colorectal polyps.

Only a limited number of image detection algorithms have been specifically designed to analyze medical images. One such algorithm is U-Net, which can extract information from a large number of images. Another algorithm that excels over others in detection accuracy is YOLOv4. However, no study has yet integrated the two algorithms to test whether the resulting model could accurately detect colorectal polyps from colonoscopy images. To bridge this research gap, the present study first combined U-Net and YOLOv4 to create a two-stage, deep-learning network algorithm called UY-Net. Its accuracy was then evaluated. Based on the evidence collected through this study, two contributions are noted:

1. UY-Net displays higher spatial accuracy and overall accuracy of polyp detection than other single detection algorithms such as YOLO3-spp, YOLOv4, RetinaNet, and Faster R-CNN. The development and utilization of the two-stage deep learning network, instead of the one-stage or two-stage detection algorithms, can significantly reduce misdiagnosis for colorectal polyps. Patients only receive the most suitable treatment when colorectal polyps are accurately detected. UY-Net can support clinicians in accomplishing this goal.
2. For the two-stage network algorithm, the sequence of performing image segmentation followed by image detection assumes a critical role in enhancing its accuracy. Precise segmentation of objects in advance can improve the performance of subsequent detection algorithms. This accounts for why the detection accuracy of UY-Net reaches a significantly high level.

2. Literature Review

Most of the early algorithms used for detecting colorectal polyps involved analyzing edge shapes [7], textures [8], colors [9], or a combination of these factors [10]. For example, Hwang et al. [7], who had observed that most polyps have an elliptical shape, proposed a new model to detect colorectal polyps. They applied the marker-controlled watershed algorithm, along with other techniques, to conduct region segmentation, ellipse fitting, and ellipse filtering by computing curve direction, curvature, edge distance, and intensity. In contrast, Ameling et al. [8] chose texture features, like grayscale intensity and local binary patterns, to distinguish colorectal polyps. Tajbakhsh, Gurudu, and Liang [10] employed shapes and texture features to recognize polyps. They differentiated regions with polyps from polyp-free areas by analyzing texture features such as local binary patterns (LBP: a texture descriptor used to represent the local texture of a computer image or vision by comparing the intensity of a pixel to those of its neighboring pixels), distribution of intensity values, and frequency content of a local neighborhood. They also examined shape features by considering boundary curves to enhance the reliability of localization. However, the applicability of these traditional models is limited and restricted because they can only recognize typical polyps but not those with non-typical shapes or textures.

With the recent development of deep learning techniques, algorithms based on Convolutional Neural Networks (CNNs) have gained considerable attention. Take Bernal et al.'s study [11] as an example. They used WM-DOVA energy maps to localize the positions of colorectal polyps without considering the sizes or types of such polyps. Pozdeev, Obukhova, and Motyko [12] advanced a fully automated system to segment colorectal polyps using a Fully Convolutional Network (FCN) for pixel-level prediction. Likewise, Bernal et al. [13] adopted CNNs and achieved state-of-the-art (SOTA) performance in a competition to detect colorectal polyps in colonoscopy videos automatically. Shin et al. [14] also employed a region-based CNN for the automated detection of colorectal polyps in colonoscopy. They chose Inception ResNet for feature learning and incorporated post-processing techniques to reach more reliable detection. Another study by Shin et al. [15] used Generative Adversarial Networks (GAN) [16] to generate images of colorectal polyps. In their study, image generation was unsatisfactory, but image detection was still significantly improved.

Moreover, a study by Wang et al. [17] showed that using the SegNet architecture [18] to detect colorectal polyps achieved a detection speed of 25 frames per second. It also demonstrated high sensitivity, specificity, and memory efficiency. Poorneshwaran et al. [19] selected GAN to segment colorectal polyps from images. In their model, GAN comprised the generator and discriminator. The generator was responsible for generating polyp segmentation masks, while the discriminator distinguished real masks from fake ones. Since the generator and discriminator were incorporated, high segmentation precision was observed on a challenging dataset. Similarly, Guo and Matuszewski adopted the Fully Convolutional Neural Network (FCNN) architecture, reporting that their proposed algorithm effectively segmented polyps from images [20,21]. Along with these researchers, Kang and Gwak [22] trained and fine-tuned two Mask R-CNN models where ResNet50 and ResNet100 were used as backbone architectures, respectively. By combining the two models using an ensemble method, their resulting framework significantly outperformed other SOTA methods in segmenting colorectal polyps. Lee et al. [23] utilized the YOLOv2 algorithm [24] for the localization and detection of colorectal polyps. They contended that YOLOv2 yielded high sensitivity and near real-time computational performance, with great potential to compensate for the limited visual field of an endoscopist.

As prior literature suggests, successfully recognizing colorectal polyps from images primarily relies on fulfilling three major functions: segmentation, localization, and detection. In deep learning, effective image segmentation involves the precise classification of individual pixels and the delineation of boundaries. To achieve the localization function, the coordinates of the bounding box must be calculated correctly. Finally, the detection function can be satisfied by accurately predicting the classification of target objects. Therefore, a deep learning algorithm that aims to attain the three functions, concurrently or separately, must consist of four components: Input, Backbone, Neck, and Head. These components are explained as follows:

- Input can be an inputted image, a patch, or a processed and sampled image;
- Backbone is responsible for pre-training, and a network based on CNNs such as ResNet, CSPDarkNet, AlexNet, DarkNet, or VGGNet is commonly adopted;
- Neck is to extract features at different levels, and another network such as Feature Pyramid Network (FPN), PANet, or Bi-FPN can be chosen to attain this objective;
- Head is responsible for predicting bounding boxes, and a one-stage network (e.g., Region Proposal Network: RPN, YOLO, or RetinaNet [25]) or a two-stage network (e.g., Faster R-CNN [26] or R-FCN) can be selected for this purpose.

Of the current deep learning algorithms, YOLOv4 [27] has gained great popularity among researchers. It is an updated version of YOLOv3 [28]. The older algorithm calls for revision because CNNs can encounter the problem of gradient vanishing when the number of network layers is increased. This leads to information loss at each network layer during the training phase and deteriorates the efficiency of layer learning. For instance, if the information is propagated by copying, as in the case of ResNet, it will demand more computational resources to process. To solve this problem, researchers develop or choose innovative networks as the backbones of YOLOv4. For instance, DarkNet or other DarkNet-based networks such as Cross Stage Partial Network (CSPNet) [29] are commonly selected. With these new backbone architectures, information is split and combined in the propagation process through the use of additional transition layers. This allows certain information to be directly merged with the convolution results, thereby reducing computational complexity, facilitating the network's learning capacity, and increasing the utilization of layer parameters. Accordingly, YOLOv4 achieves higher accuracy but demands lower hardware requirements than the older algorithm while maintaining the same speed.

Recently, significant progress has been witnessed in image segmentation following the introduction of CNN-based architectures. By repeatedly downsampling the inputted images, low-level features (also known as feature maps) can be effectively extracted. Subsequently, upsampling can be performed to enable pixel-level prediction and image segmentation. To illustrate, after Long, Shelhamer, and Darrell [30] had advanced FCN (the first end-to-end trainable image segmentation algorithm), Ronneberger, Fischer, and Brox [31] modified it to create U-Net. U-Net is a deep-learning algorithm specifically designed for medical image segmentation [32]. Its architecture consists of a U-shaped network that enables the capture of both contextual and positional information. It also includes a pathway between an encoder and a decoder (i.e., the skip connection). The encoder comprises multiple convolutional and pooling layers and is responsible for feature extraction. The decoder in U-Net uses deconvolution to restore localization information. With the established skip connections, high-level features learned in the encoder can be transmitted to the decoder. This helps reduce information loss during the upsampling process. Figure 1 illustrates the architecture of U-Net.

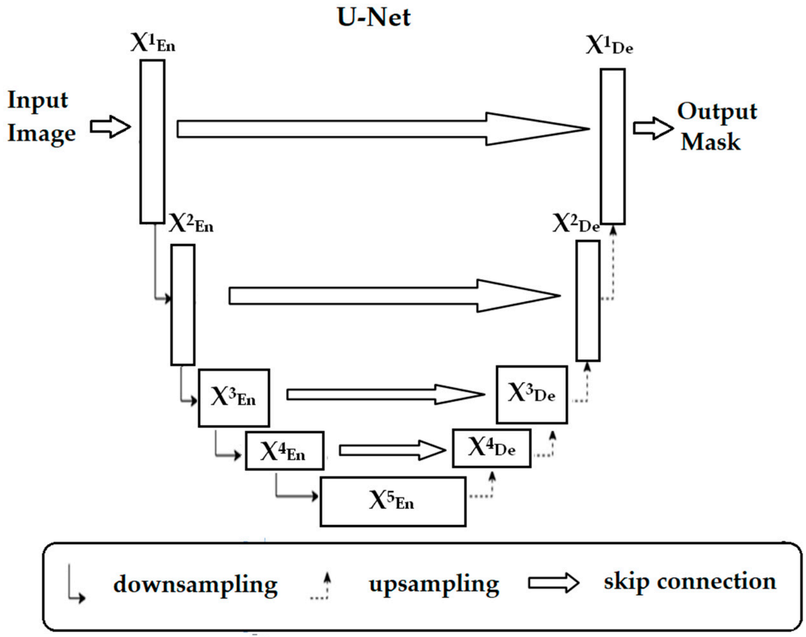


Figure 1. Architecture of U-Net.

It is worth noting that medical images typically exhibit relatively simple semantic features, fixed structures, and less irrelevant information. In other words, most features extracted from these images convey plain yet sufficient information, making the skip connections in the U-shaped structure relatively effective. Since U-Net utilizes a U-shaped structure, its application for segmenting medical images holds great promise.

As discussed earlier, both U-Net and YOLOv4 are highly suitable for detecting abnormalities from medical images, including those obtained by endoscopes. U-Net is also known for its relatively simple structure, while YOLOv4 is renowned for its widespread usage. However, no study has ever combined the two algorithms to establish a new two-stage network model, let alone explore its accuracy in detecting polyps from colonoscopy images. To gather evidence to answer the unknown question, the present study combined U-Net and YOLOv4 to create UY-Net. The accuracy of UY-Net was estimated and compared to the performance of the four individual object detection algorithms (i.e., YOLO3-spp, YOLOv4, RetinaNet, and Faster R-CNN). To be specific, two hypotheses formulated for testing were presented as follows:

1. Performing U-Net first would result in precise segmentation of abnormalities from the colonoscopy images; after abnormalities were precisely segmented, the subsequent application of YOLOv4 would result in accurate detection of colorectal polyps;
2. UY-Net would achieve higher accuracy of polyp detection than the four detectors.

3. Method

3.1. UY-Net

UY-Net consists of two main components: (1) image segmentation and (2) object localization and detection. In the present study, image segmentation was first performed, followed by object detection. In the first stage, U-Net (with the Adam optimizer and ResNet as the backbone) was applied to segment images. Figure 2 presents sample images of segmentation.

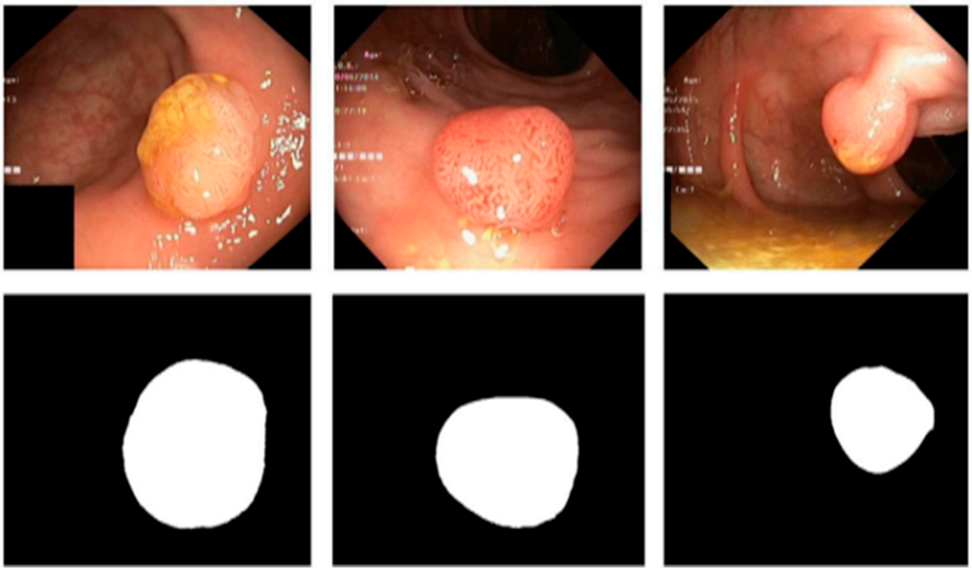


Figure 2. Sample images of segmentation.

In the second stage, YOLOv4 (DarkNet as the backbone) was utilized to detect colorectal polyps. Its application resulted in the bounding box localization information of polyps, including x_{center}, y_{center}, $yolo_w$, and $yolo_h$. Figure 3 illustrates the framework of UY-Net for image segmentation, localization, and detection.

As shown, x_{center} represents the proportion of the center x-coordinate of the bounding box relative to the length of the entire image's x-axis. Similarly, y_{center} represents the proportion of the center y-coordinate of the bounding box relative to the length of the entire image's y-axis. On the other hand, $yolo_w$ represents the proportion of the width of the bounding box relative to the width of the entire image, while $yolo_h$ represents the proportion of the height of the bounding box relative to the height of the entire image.

3.2. Dataset

This study analyzed the images obtained from the Kvasir-SEG dataset [33]. This dataset comprises 1000 images of colorectal polyps, along with data of corresponding Mask and Bounding Box Ground Truth. The resolution of these images varies, ranging from 332 × 487 to 1920 × 1072 pixels. The ground truth has been manually annotated by medical experts using the Labelbox software. The dataset contains a total of 1071 colorectal polyps, including 700 large polyps (larger than 160 × 160 pixels), 323 medium-sized polyps (between 160 × 160 and 64 × 64 pixels), and 48 small polyps (smaller than 64 × 64 pixels).

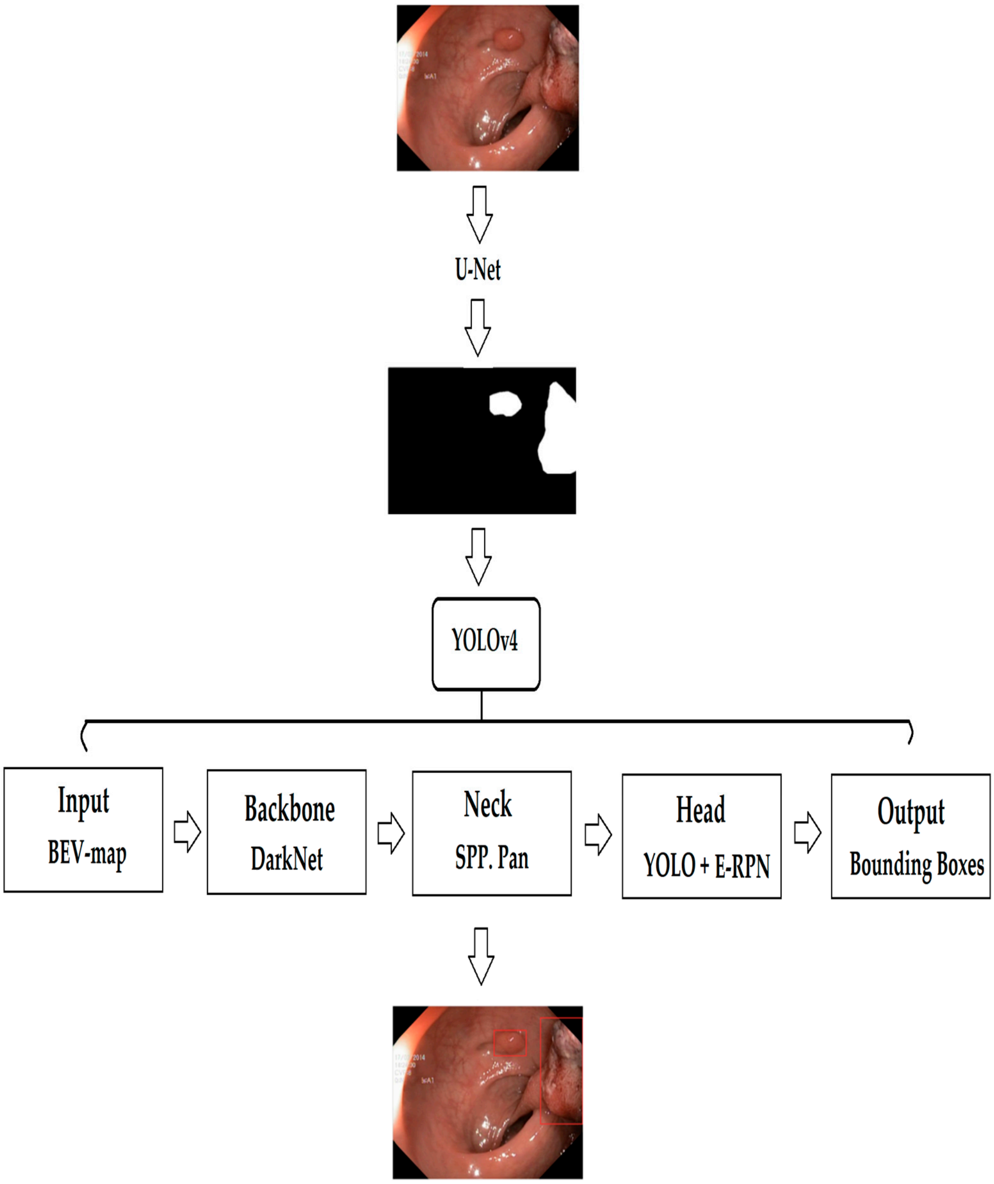


Figure 3. Framework of UY-Net.

3.3. Intersection over Union (IoU) and Average Precision (AP)

The two types of metrics commonly used to assess object detection and localization are IoU and AP. In the present study, IoU was calculated by dividing the intersection of the ground truth and predicted regions by the union of the ground truth and predicted regions. In other words, it measured the overlap ratio between the two regions. The Equation (1) is shown below: (GT stands for the ground truth region, and PD stands for the predicted region).

$$\mathrm{IoU} = \frac{\mathrm{GT} \cap \mathrm{PD}}{\mathrm{GT} \cup \mathrm{PD}} \tag{1}$$

Figure 4 depicts the intersection and union between the ground truth and predicted regions. The red area represents the ground truth, while the yellow area represents the predicted region.

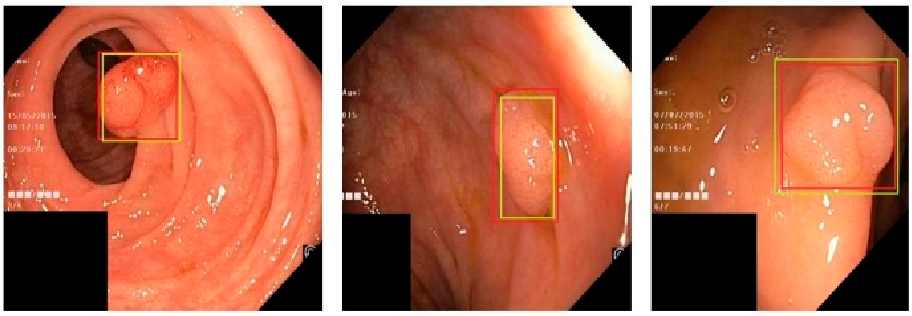
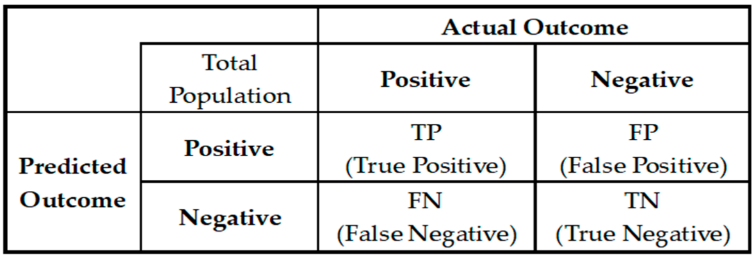


Figure 4. Illustration of GT and PD.

AP was calculated as the area under the precision-recall curve. The predicted targets were evaluated based on IoU calculations. If the value of IoU was greater than a predefined threshold, then the target would be considered a true positive (TP). If the value of IoU was below the threshold, the target would be considered a false positive (FP). Both TP and FP represented the states in the confusion matrix (see Figure 5). In the present study, the threshold for AP was set within a specified range. For example, the IoU threshold was set from 0.25 to 0.75 with an interval of 0.05, denoted as (AP@[0.25:0.05:0.75]). If the IoU threshold was 0.50, it would be referred to as AP50.



		Actual Outcome	
	Total Population	**Positive**	**Negative**
Predicted Outcome	**Positive**	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

Figure 5. Confusion matrix.

Precision was the proportion of predicted targets that were true targets (also known as the Positive Predictive Value: PPV). Recall was the proportion of targets that were correctly predicted as targets (also known as Sensitivity). Precision (2) and Recall (3) equations are shown below.

$$\text{Precision } = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall } = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

The calculation of AP involved selecting the maximum precision corresponding to each change in Recall. Then, these recalls were considered as calculation points. The Equation of AP (4) is shown below:

$$\text{AP} = \sum_{\text{k}=0}^{\text{k}=\text{n}-1} [\text{Recalls}(\text{k}) - \text{Recalls}(\text{k}+1)] \times \text{Precisions}(\text{k}) \tag{4}$$

where Recalls (n) = 0, Precisions(n) = 1, n = Number of Thresholds

3.4. Settings and Procedures

In the present study, Faster R-CNN, RetinaNet, YOLOv3-spp, YOLOv4, and UY-Net were tested. By incorporating these algorithms into the experiment, it became possible to assess whether the proposed network could outperform one-stage or two-stage detectors in accurately detecting polyps.

The training was conducted using Google Colab with an NVIDIA Tesla P100 GPU and the PyTorch machine learning library. The dataset was divided into 880 training images and 120 validation images. Since the image sizes were not fixed, they were uniformly resized to 512 × 512 for training. Because UY-Net is a combination of the two different algorithms, its training was conducted separately. U-Net was trained using both the images of colorectal polyps and their corresponding masks. YOLOv4, on the other hand, was trained using the colorectal polyps and the ground truth bounding boxes.

The configuration of hyper-parameters is crucial for the training of deep learning models. For U-Net, the backbone was ResNet, the learning rate was set to 1×10^{-5}, the optimizer was Adam, the batch size was 8, the loss function was cross-entropy, and the decay rate was 1×10^{-4}. The respective hyper-parameter settings for all algorithms are presented in Table 1.

Table 1. Hyper-parameters of YOLOv4 and UY-Net.

Algorithm	Learning Rate	Optimizer	Batch Size	Anchors	Loss	Threshold
Faster R-CNN	2.5×10^{-4}	Adam	8	256	$L1^{smooth}$ log loss	0.4
RetinaNet	1×10^{-5}	SGD	8	15	$L1^{smooth}$ focal loss	0.3
YOLOv3-spp	1×10^{-3}	SGD	16	8	MSE, CE	0.25
YOLOv4	1×10^{-3}	SGD	16	8	CioU, CE	0.25
UY-Net	1×10^{-3}	SGD	32	18	CioU, CE	0.25

4. Results and Discussion

The values of AP and IoU were computed and used as indexes to estimate the accuracy of object detection. Table 2 presents these results.

Table 2. AP and IoU for different algorithms.

Algorithm	Backbone	AP	IoU
Faster R-CNN	ResNet	0.7866	0.5621
RetinaNet	ResNet	0.8697	0.7313
YOLOv3-spp	ResNet	0.8105	0.8248
YOLOv4	DarkNet	0.8513	0.8205
UY-Net	DarkNet, U-Net	0.9915	0.9395

The table shows that the AP and IoU values for YOLOv4 and YOLOv3-spp are all above 0.81, indicating that the two YOLO models detect polyps to an adequate level. This finding aligns with what previous research has reported [34]. For instance, Doniyorjon et al. [35] tested five YOLO algorithms (i.e., YOLOv3, YOLOv3-tiny, YOLOv4, YOLOv4-tiny, and YOLOv4-tiny with the Inception-ResNet-A block), and all models were found to achieve at least 89% training accuracy and 85% testing accuracy. In other words, they effectively detected polyps by drawing bounding boxes around these detected objects. As Doniyorjon et al.'s study and the present study suggest, YOLO algorithms can aid medical practitioners in detecting abnormalities from endoscopic images. However, UY-Net achieves a significantly higher accuracy level (AP = 0.9915; IoU = 0.9395), exceeding that of YOLOv3-spp or YOLOv4 by at least 10%. Based on this finding, the first hypothesis of this study can be substantially corroborated:

1. Applying U-Net followed by YOLOv4 results in considerably higher accuracy in detecting colorectal polyps from colonoscopy images.

The proposed two-stage network displays the highest levels of spatial accuracy and overall accuracy of object detection. This suggests that image detection should not be carried out alone but coupled with image segmentation. For example, de Moura Lima et al. [36] proposed a two-stage design that used transformers to detect polyps in colonoscopy images. In the segmentation stage, they first used the Dense Prediction Transformer (DPT) model to extract depth maps of salient objects. Then, they used the Visual Saliency Transformer (VST) architecture to extract depth geometric information of regions associated with these suspicious objects. In the second stage, DEtection TRansformer (DETR) architecture was applied to detect the polyps. de Moura Lima et al.'s model achieved an AP of 0.92 in the Kvasir-SEG dataset. Like UY-Net, it can also accurately detect colorectal polyps in medical images. Therefore, a design with two stages, first for image segmentation and/or extraction, followed by image detection, may be a promising framework for facilitating polyp detection accuracy.

Moreover, UY-Net surpasses RetinaNet by at least 12% and Faster R-CNN by 20% in accuracy. RetinaNet is a detector that combines region proposal generation and object classification into one stage. By simplifying its architecture and incorporating FPN to create a feature pyramid, RetinaNet may perform well on detection accuracy and speed [25]. Faster R-CNN [37], on the other hand, is a two-stage detection algorithm. In its first stage, RPN is used to output a set of regional proposals. In the second stage, these regional proposals are used for object detection and classification. Regardless of their one-stage or two-stage detection design, both RetinaNet and Faster R-CNN do not achieve the same level of accuracy in polyp detection as UY-Net. This finding lends strong support to the second hypothesis:

2. The two-stage network UY-Net would be more accurate in detecting colorectal polyps than the one-stage or two-stage detection algorithms.

It also highlights the need to experiment with a segmentation architecture and a detection algorithm to design an innovative two-stage network. To illustrate, in the present study, we hypothesized that the more precisely a region with abnormalities could be segmented in advance, the more likely it was for these abnormalities to be accurately detected thereafter. Therefore, U-Net was trained first to precisely extract and obtain regions of interest (ROI) from images. Then, YOLOv4 underwent training, but it was not applied to analyze the well-segmented regions until its accuracy was elevated. The exceptional performance of UY-Net in polyp detection validates our hypothesis, implying that the procedural sequence should factor into the improved accuracy of object detection. The two algorithms of a two-stage model should be trained independently, with the segmentation algorithm being trained first, followed by the training of the detection algorithm and its application.

To the best of our knowledge, this study may be the first attempt to create a two-stage network by combining U-Net and YOLOv4. As ELKarazle et al. ([34], p. 10) argued, "the YOLO architecture has been the preferred go-to solution for real-time detection tasks as it can process 45 frames per second". This feature makes it popular among researchers and one of the most used methods for polyp detection. Yang and Yu [38] also emphasized that U-Net is distinguished from other segmentation algorithms by its relatively simple structure with few parameters. This simplicity helps to avoid overfitting and improves the accuracy of image segmentation. U-Net is, therefore, one of the most preferred image segmentation methods in the medical domain, especially for small datasets such as the Kvasir-SEG dataset. Furthermore, the common adoption of YOLOv4, U-Net, and their revised versions has led to the availability of several open-source libraries for executing these algorithms [39]. Encouraged by the promising results of the present study and the availability of the source codes, some researchers in the medical field may choose to develop and evaluate their own two-stage network models using the recently improved versions of YOLO and U-Net. Other researchers may be motivated to incorporate different CNN-based segmentation and detection architectures to develop novel two-stage frameworks. Either way, these researchers can generate new models to improve the accuracy of detecting colorectal polyps from endoscopic images. In this respect, the present study significantly contributes to the medical research community by creating a new and promising pathway and protocol for advancing medical image research.

5. Conclusions

The novelty of the present study lies in the incorporation of a segmentation algorithm and a detection algorithm into a two-stage network. While some researchers continue to focus on improving single algorithms, the present study explores a novel and promising alternative by developing and evaluating the two-stage model for polyp detection. The present study also takes an innovative approach to model training. The segmentation algorithm is trained first, followed by the detection algorithm. The sequence of training may assume a significant role in the high accuracy of the proposed model. Taken together, the colorectal polyps in colonoscopy images can be computationally quantifiable and identifiable through object localization and detection after being precisely segmented. UY-Net, therefore, can outperform any of the single detection algorithms in the accuracy of colorectal polyp detection. This sheds light on the potential of a two-stage network model for improving the detection and diagnosis of abnormalities in medical images.

It is important to note that the present study deliberately lowers resolutions of certain original input images (e.g., from 1920 × 1072 to 512 × 512) to reduce memory complexity and time required to run the algorithms. Additionally, all the algorithms are executed on a GPU instead of a CPU. This would improve runtime. Moreover, U-Net is trained with the Adam optimizer in order to reduce memory usage and increase inference speed. However, UY-Net contains YOLOv4, a deep-learning algorithm known to be memory-intensive. It also needs to run U-Net. Accordingly, UY-Net excels at polyp detection but incurs greater memory complexity and longer runtime when contrasted with a single algorithm. Researchers, therefore, should continue to delve deeper into the use of deep learning in bioengineering to develop fast, reliable, and efficient algorithms for image detection.

Although the findings are promising, caution should be exercised before the results can be appropriately generalized. First, colorectal polyps may develop into cancer, so failing to detect them will pose a life-threatening danger to patients. Researchers still need to improve the UY-Net algorithm to reduce the miss rates. To this end, we plan to replicate the present study and incorporate the relatively recent U-Net3+ and YOLOv7 into a two-stage model. We will then compare the accuracy of this new model with that of UY-Net to assess if it can better detect colorectal polyps than the older model. Second, calculating the bounding box based on the edges seems intuitively simple. Nevertheless, the edges obtained by U-Net tend to be less smooth. These minor irregularities in the edges may or may not impact the precision of image segmentation. More effort is needed to clarify this concern.

Author Contributions: Conceptualization, C.-S.H. and Y.-C.L.; Software, J.-W.W.; Validation, C.-J.W.; Writing—original draft, J.-W.W.; Writing—review and editing, C.-S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 2021 Cause of Death Statistics in Taiwan. Available online: https://www.mohw.gov.tw/cp-16-70314-1.html (accessed on 22 May 2023).
2. Overview of Colorectal Cancer Prevention and Control in Taiwan. Available online: https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=615&pid=1126 (accessed on 22 May 2023).
3. Ahn, S.B.; Han, D.S.; Bae, J.H.; Byun, T.J.; Kim, J.P.; Eun, C.S. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. *Gut Liver* **2012**, *6*, 64–70. [CrossRef]

4. Heresbach, D.; Barrioz, T.; Lapalus, M.G.; Coumaros, D.; Bauret, P.; Potier, P.; Sautereau, D.; Boustière, C.; Grimaud, J.C.; Barthélémy, C.; et al. Miss rate for colorectal neoplastic polyps: A prospective multicenter study of back-to-back video colonoscopies. *Endoscopy* **2008**, *40*, 284–290. [CrossRef]
5. Vázquez, D.; Bernal, J.; Sánchez, F.J.; Fernández-Esparrach, G.; López, A.M.; Romero, A.; Drozdzal, M.; Courville, A. A benchmark for endoluminal scene segmentation of colonoscopy images. *J. Healthc. Eng.* **2017**, *2017*, 4037190. [CrossRef]
6. de Lange, T.; Halvorsen, P.; Riegler, M. Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy. *World J. Gastroenterol.* **2018**, *24*, 5057–5062. [CrossRef] [PubMed]
7. Hwang, S.; Oh, J.; Tavanapong, W.; Wong, J.; de Groen, P.C. Polyp detection in colonoscopy video using elliptical shape feature. In Proceedings of the 2007 the IEEE International Conference on Image Processing (IEEE-ICIP), San Antonio, TX, USA, 16–19 September 2007; pp. 465–468. [CrossRef]
8. Ameling, S.; Wirth, S.; Paulus, D.; Lacey, G.; Vilarino, F. Texture-based polyp detection in colonoscopy. In Proceedings of the Bildverarbeitung für die Medizin: Algorithmen-Systeme-Anwendungen, Proceedings of BVM Workshop 2009, Heidelberg, Germany, 22–25 March 2009; Meinzer, H., Deserno, T.M., Handels, H., Tolxdorff, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 346–350.
9. Karkanis, S.A.; Iakovidis, D.K.; Maroulis, D.E.; Karras, D.A.; Tzivras, M. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans. Inf. Technol. Biomed.* **2003**, *7*, 141–152. [CrossRef] [PubMed]
10. Tajbakhsh, N.; Gurudu, S.R.; Liang, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans. Med. Imaging* **2016**, *35*, 630–644. [CrossRef] [PubMed]
11. Bernal, J.; Sánchez, F.J.; Fernández-Esparrach, G.; Gil, D.; Rodríguez, C.; Vilariño, F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Comput. Med. Imaging Graph.* **2015**, *43*, 99–111. [CrossRef] [PubMed]
12. Pozdeev, A.A.; Obukhova, N.A.; Motyko, A.A. Automatic analysis of endoscopic images for polyps detection and segmentation. In Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (IEEE EIConRus), Saint Petersburg/Moscow, Russia, 28–31 January 2019; pp. 1216–1220. [CrossRef]
13. Bernal, J.; Tajkbaksh, N.; Sanchez, F.J.; Matuszewski, B.J.; Chen, H.; Yu, L.; Angermann, Q.; Romain, O.; Rustad, B.; Balasingham, I.; et al. 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. [CrossRef]
14. Shin, Y.; Qadir, H.A.; 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. [CrossRef]
15. Shin, Y.; Qadir, H.A.; Balasingham, I. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. *IEEE Access* **2018**, *6*, 56007–56017. [CrossRef]
16. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27, Proceedings of Annual Conference on Neural Information Processing Systems 2014, Montreal, Quebec, Canada, 8–13 December 2014*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Boston, MA, USA, 2014; pp. 2672–2680.
17. Wang, P.; Xiao, X.; Brown, G., Jr.; Berzin, T.M.; Tu, M.; Xiong, F.; Hu, X.; Liu, P.; Song, Y.; Zhang, D.; et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat. Biomed. Eng.* **2018**, *2*, 741–748. [CrossRef] [PubMed]
18. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
19. Poorneshwaran, J.M.; Kumar, S.S.; Ram, K.; Joseph, J.; Sivaprakasam, M. Polyp segmentation using generative adversarial network. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 7201–7204. [CrossRef]
20. Guo, Y.; Matuszewski, B.J. GIANA polyp segmentation with fully convolutional dilation neural networks. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 25–27 February 2019; pp. 632–641.
21. Guo, Y.; Bernal, J.; Matuszewski, B.J. Polyp segmentation with fully convolutional deep neural networks-Extended evaluation study. *J. Imaging* **2020**, *6*, 69. [CrossRef] [PubMed]
22. Kang, J.; Gwak, J. Ensemble of instance segmentation models for polyp segmentation in colonoscopy images. *IEEE Access* **2019**, *7*, 26440–26447. [CrossRef]
23. Lee, J.Y.; Jeong, J.; Song, E.M.; Ha, C.; Lee, H.J.; Koo, J.E.; Yang, D.; Kim, N.; Byeon, J. Real-time detection of colon polyps during colonoscopy using deep learning: Systematic validation with four independent datasets. *Sci. Rep.* **2020**, *10*, 8379. [CrossRef]
24. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
25. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
26. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
27. Bochkovskiy, A.; Wang, C.; Liao, H.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.

28. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
29. Wang, C.; Liao, H.M.; Wu, Y.; Chen, P.; Hsieh, J.; Yeh, I. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580. [CrossRef]
30. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
31. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer Cham: Basel, Switzerland, 2015; pp. 234–241.
32. Sharma, N.; Gupta, S.; Koundal, D.; Alyami, S.; Alshahrani, H.; Asiri, Y.; Shaikh, A. U-Net model with transfer learning model as a backbone for segmentation of gastrointestinal tract. *Bioengineering* **2023**, *10*, 119. [CrossRef]
33. Jha, D.; Smedsrud, P.H.; Riegler, M.A.; Halvorsen, P.; de Lange, T.; Johansen, D.; Johansen, H.D. Kvasir-seg: A segmented polyp dataset. In *Multimedia Modeling, Proceedings of the 26th International Conference on Multimedia Modeling (MMM 2020), Daejeon, Korea, 5–8 January 2020*; Ro, Y.M., Cheng, W., Kim, J., Chu, W., Cui, P., Choi, J., Hu, M., de Neve, W., Eds.; Springer Cham: Basel, Switzerland, 2020; pp. 451–462.
34. ELKarazle, K.; Raman, V.; Then, P.; Chua, C. Detection of colorectal polyps from colonoscopy using machine learning: A survey on modern techniques. *Sensors* **2023**, *10*, 1225. [CrossRef]
35. Doniyorjon, M.; Madinakhon, R.; Shakhnoza, M.; Cho, Y. An improved method of polyp detection using custom YOLOv4-tiny. *Appl. Sci.* **2022**, *12*, 10856. [CrossRef]
36. de Moura Lima, A.C.; de Paiva, L.F.; Bráz, G., Jr.; de Almeida, J.D.S.; Silva, A.C.; Coimbra, M.T.; de Paiva, A.C. A two–stage method for polyp detection in colonoscopy images based on saliency object extraction and transformers. *IEEE Access* **2023**, *11*, 76108–76119. [CrossRef]
37. Li, J.; Zhang, J.; Chang, D.; Hu, Y. Computer-assisted detection of colonic polyps using improved faster R-CNN. *Chin. J. Electron.* **2019**, *28*, 718–724. [CrossRef]
38. Yang, R.; Yu, Y. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. *Front. Oncol.* **2021**, *11*, 638182. [CrossRef]
39. Yan, T.; Qin, Y.Y.; Wong, P.K.; Ren, H.; Wong, C.H.; Yao, L.; Hu, Y.; Chan, C.I.; Gao, S.; Chan, P.P. Semantic segmentation of gastric polyps in endoscopic images based on convolutional neural networks and an integrated evaluation approach. *Bioengineering* **2023**, *10*, 806. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.