Article

Colorectal Polyp Detection Model by Using Super-Resolution Reconstruction and YOLO

Shaofang Wang *,† , **Jun Xie** † , **Yanrong Cui** and **Zhongju Chen**

School of Computer Science, Yangtze University, Jingzhou 434023, China; 2022710617@yangtzeu.edu.cn (J.X.); cyanr@yangtzeu.edu.cn (Y.C.); chenzj@yangtzeu.edu.cn (Z.C.)
* Correspondence: wangshaofang2022@yangtzeu.edu.cn
† These authors contributed equally to this work.

Abstract: Colorectal cancer (CRC) is the second leading cause of cancer-related deaths worldwide. Colonoscopy is the primary method to prevent CRC. However, traditional polyp detection methods face problems such as low image resolution and the possibility of missing polyps. In recent years, deep learning techniques have been extensively employed in the detection of colorectal polyps. However, these algorithms have not yet addressed the issue of detection in low-resolution images. In this study, we propose a novel YOLO-SRPD model by integrating SRGAN and YOLO to address the issue of low-resolution colonoscopy images. Firstly, the SRGAN with integrated ACmix is used to convert low-resolution images to high-resolution images. The generated high-resolution images are then used as the training set for polyp detection. Then, the C3_Res2Net is integrated into the YOLOv5 backbone to enhance multiscale feature extraction. Finally, CBAM modules are added before the prediction head to enhance attention to polyp information. The experimental results indicate that YOLO-SRPD achieves a mean average precision (mAP) of 94.2% and a precision of 95.2%. Compared to the original model (YOLOv5), the average accuracy increased by 1.8% and the recall rate increased by 5.6%. These experimental results confirm that YOLO-SRPD can address the low-resolution problem during colorectal polyp detection and exhibit exceptional robustness.

Keywords: polyp detection; colonoscopy; medical image processing; deep learning; SRGAN; YOLO; ACmix; Res2net

1. Introduction

Colorectal cancer (CRC) is one of the top three leading causes of cancer death in the United States [1]. Adenomatous polyps carry a significantly higher probability of malignant transformation at a quicker rate than hyperplastic polyps [2]. Early detection of these colorectal poslyps through colonoscopies can significantly curtail the risk of progression to cancer [1]. Computer-aided diagnosis (CAD) can help physicians accurately detect polyps, decide treatment plans, and predict patient prognosis. Hence, early enteroscopies enable the excision of nascent polyps to prevent advancement towards cancer and subsequent degeneration [3].

Traditional intestinal endoscopy encounters limitations due to factors impacting the field of view, structure, and image resolution. Detecting polyps in low-resolution images undoubtedly poses a greater challenge for physicians. Furthermore, the process of detecting colorectal polyps necessitates a high degree of physician expertise and is thwarted by physician subjectivity [4], resulting in an estimated manual oversight rate of colorectal polyps at approximately one-quarter [5].

Deep learning can be applied in many fields. In the field of medical imaging, the tasks of object detection, recognition, and classification of medical images have become easier and play an important role in practical medical detection [6–8]. Texture, color, and shape features of images are commonly employed in traditional detection of colorectal polyps, such as scale-invariant feature transform (SIFT) [9], support vector machines (SVMs) [10],

and Gaussian mixture models [11] for classification. However, these traditional methods heavily rely on the given data features and limit their ability to capture the complex variations in polyp images. This leads to poor performance in real-world scenarios. Therefore, these methods are prone to overfitting and exhibit poor generalization.

In the field of medical image processing and cancer detection, super-resolution (SR) reconstruction plays a distinctive role [12,13]. SR reconstructs low-resolution images into high-resolution images, helping to improve image resolution and enhance image details. Pavlou et al. [14] proposed using SRGAN to enhance the resolution of OCT images, distinguishing between BCC lesions and scar tissue in cryoimmunotherapy. Shi et al. [15] proposed utilizing an improved SRGAN to reconstruct phase contrast polarimetry (PCP) images with enhanced resolution, and they incorporated a counting network for added functionality. These studies demonstrate the potential of SR in enhancing diagnostic quality in medical image processing.

Currently, researchers are increasingly employing artificial intelligence (AI) methods to assist in colonoscopy for detecting polyps. Zhu et al. [16] improved polyp diagnosis accuracy by introducing the PAM-Net and GWD loss functions. Ghose et al. [17] proposed a method that uses data augmentation and fine-tunes parameters to improve polyp detection performance. Yasmin et al. [18] proposed the GastroNet to detect and classify gastrointestinal polyps and abnormalities and achieved high accuracy through hyperparameter tuning. These approaches employ detection algorithms for polyp identification. However, several challenges still exist in using these methods to assist in detecting polyps.

- Low-resolution images leads to insufficient accuracy in polyp detection.
- During colorectal polyp detection, the intestinal environment is complex. Some polyps are small, making them more difficult to identify.

To solve the problems of low-resolution image in polyp detection, this paper proposes a polyp detection method based on super-resolution reconstruction. Super-Resolution Generative Adversarial Network (SRGAN) [19] is used to increase the resolution of colonoscopy images and decrease manual missed detection due to low-resolution images. The improved You Only Look Once (YOLO) then uses the reconstructed images as input to help physicians identify polyps.

The main contributions of this study are listed below:

(1) We propose YOLO-SRPD (Super-Resolution Reconstruction for Polyp Detection) to address the problems of low-resolution polyp images during the detection by combining super-resolution reconstruction and YOLOv5.

(2) To enhance partial texture and details of polyps, we introduce attention-based mixed convolution modules (ACmix) in the generator and discriminator of SRGAN.

(3) We propose the improved YOLOv5 algorithm by incorporating the Res2net-based C3 module in the backbone. The Res2net-based C3 module can enlarge the convolutional receptive fields and enhance multiscale feature extraction within the backbone network.

(4) This paper also incorporates the CBAM attention mechanism into head layers of YOLO. CBAM can focus on polyp information and enhance overall detection capability.

This paper proposes an algorithm for polyp detection that involves the use of SRGAN to reconstruct low-resolution images, followed by detection algorithm. Section 2 reviews the pertinent research on deep learning in medical diagnosis, with a focus on polyp detection. In Section 3, we describe the algorithms and framework employed in this study. Section 4 presents experimental results and algorithm comparisons. Finally, a summary of the paper is provided in Section 5.

2. Related Work

Deep learning has made substantial advancements in cancer detection, particularly in the areas of lesion detection and segmentation. Many studies have shown that employing deep learning models can significantly enhance the accuracy of cancer detection [20–22], enabling effective differentiation between normal and cancerous tissue. Tan et al. [23] proposed a small target breast mass detection network, introducing an adaptive positive

sample selection algorithm to automatically select positive samples. This method significantly improved the detection accuracy of small masses in breast mass detection. However, there may be issues with missed detections in the edge regions during the breast cancer detection process. Deep learning models heavily rely on a large and accurate dataset for training, and insufficient data can potentially compromise their generalization capabilities.

The majority of traditional colorectal polyp detection relies on physicians' medical expertise. Therefore, problems like misdetection and missing detection may arise during the detection process due to inexperience and manual mistakes of physicians. Accurate and quick diagnosis of colorectal polyps using AI technology has become possible in the field of medical detection [24–26].

With the application of computer vision in the medical field, convolutional neural networks (CNNs) play a crucial role in segmentation [27–30] and detection of colorectal polyps. Ozawa et al. [31] demonstrated the viability of CNN as a polyp detection support system by proposing a polyp classification architecture based on a single-shot multibox detector (SSD) on a private dataset. In 2020, Kayser et al. [32] employed the RetinaNet network for intestinal polyp diagnosis to lessen the impact of image artifacts. In 2020, Kayser et al. [32] employed the RetinaNet network for polyp detection in datasets such as EAD2019 [33], CVC-Clinic [34], ETIS-Larib [35], and Kvasir-SEG [36], aiming to mitigate image artifacts and achieved a precision of 53.7% and a recall of 72.6%. Zeng et al. [37] proposed the RetinaNet model, which employed CNNs to capture structural patterns in human colon optical coherence tomography (OCT) images.

One of the critical challenges in gastric polyp detection is the wide range in size and shape of gastric polyps. In order to solve this issue, Laddha et al. [38] developed a feature fusion module based on deep learning to identify target problems on the CLV-14SL [39] dataset, achieving a precision of 93%, a recall of 91%, and a mean average precision (mAP) of 91%. Zhang et al. [40] proposed a ResYOLO model which was trained on nonmedical data and fine-tuned using colonoscopy images. Tang et al. [41] used GAN to generate polyp images for YOLO training. The accuracy was improved by using Gaussian blur to simulate blurred images and deblur the images. Carrinho et al. [42] utilized the YOLOv4 and achieved real-time detection through optimization and quantization with NVIDIA TensorRT. However, this optimization may sacrifice the generalization ability on different types of images. Tang et al. [43] proposed narrow-band imaging (NBI) technology on a private dataset to enhance polyps' contrast and vascular patterns, and this method positively impacts polyp identification and classification tasks. Chou et al. [44] employed discrete wavelet transform (DWT) and GAN2 (presumably referring to StyleGAN2) to enhance the discriminative characteristics of polyps. Chen et al. [45] proposed an accelerated R-CNN architecture that leverages self-attention mechanisms for polyp detection. They achieved a precision of 94.3%, a recall of 92.5%, and F1-score of 93.4% on a private dataset.

The deep-learning-based intestinal polyp detection framework offers practical applications in the detection of intestinal polyps and the reduction of missed detection rates. The YOLOv5 framework is the primary basic framework discussed and employed in this study. YOLOv5 satisfies the demands for high accuracy and frame rate in colonoscopy detection scenarios.

3. Materials and Methods

In this paper, we propose a model for intestinal polyp detection that combines a fused SR and an improved YOLOv5 algorithm. The pertinent structure is depicted in Figure 1. The process starts with the low-resolution image (I_{LR}) being reconstructed into a super-resolution image (I_{SR}) by using SRGAN. I_{LR} are obtained by applying a Gaussian filter to high-resolution images (I_{HR}) and then downsampling. Subsequently, the reconstructed images are input into YOLOv5 to detect colon polyps.
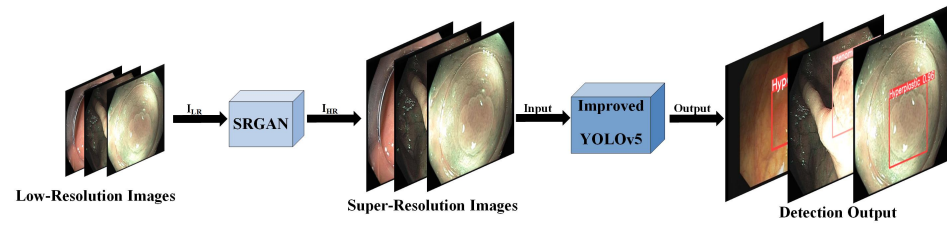
Figure 1. Illustration of the polyp detection architecture.

3.1. Super-Resolution Reconstruction Using SRGAN

Compared to traditional image processing algorithms, SRGAN can generate high-quality images with enhanced details and textures. Furthermore, it achieves visually more realistic results by leveraging deep learning techniques. The SRGAN algorithm is employed to generate high-resolution images from low-resolution inputs. The approach comprises a generator network and a discriminator network. The generator network incorporates multiple residual block structures, followed by sub-pixel convolution layers.

The discriminator network module consists of seven convolutional layers with the LeakyReLU activation function. To differentiate between I_{HR} and I_{SR}, two fully connected layers and a sigmoid activation function are added after the convolutional layer. The perceptual loss function of SRGAN [19] is shown in Equation (1). The perceptual loss is composed of two parts: content loss (l_X^{SR}) and adversarial loss (l_{Gen}^{SR}).

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR}. \tag{1}$$

This study employs the VGG19 network for content loss, where the content loss is represented in Equation (2).

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y} \right)^2, \tag{2}$$

where W and H are the width and height of the feature map, i and j denote the j-th convolutional layer before the i-th max pooling layer, ϕ denotes the obtained feature map, and $\phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y}$ represents the pixel value (x, y) in the feature map extracted from I_{LR}.

The adversarial loss function is expressed in Equation (3).

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{LR})), \tag{3}$$

where $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ denotes the discriminator that judges the image generated by the generator $G_{\theta_G}(I^{LR})$ as a natural image.

Although SRGAN could improve the image quality and increase pixel count, wrong detections may still occur. It is vitally important to preserve texture and detail in medical image reconstruction. This paper adds the self-attention convolution module ACmix to both the generator and discriminator networks of SRGAN. This addition aims to enhance colorectal polyp detection and facilitate the training of super-resolution images that closely resemble real images.

The ACmix module effectively combines the advantages of traditional convolution and the self-attention mechanism to improve the network's focus on details. The structure of this module is illustrated in Figure 2.

Figure 3 presents the ACmix architecture, which operates in two stages. First, the input $H \times W \times C$ feature map is broken into N features by three 1×1 convolutions, obtaining rich intermediate features. The feature information generated in stage I is sent into the convolution and self-attention branches in stage II. In the convolution branch, the features first travel through a convolution kernel of size K, after which the feature data are divided

into K^2 subset feature maps through the dense layer, and, ultimately, a new feature map is generated. The features in the self-attention branch are separated into N groups, and three 1×1 convolutions are employed for self-attention multiplication. The two learnable parameters α and β are then used to add the features of the two branches to the channel. To obtain the final feature map, as illustrated in Equation (4), the feature output of the two branches is finally sent through the ACmix module.

$$F_{out} = \alpha F_{att} + \beta F_{conv} \, , \tag{4}$$

where F_{out} represents the final output features, F_{att} represents the self-attention branch, and F_{conv} denotes the convolutional branch, while α and β are parameters that can be learned.
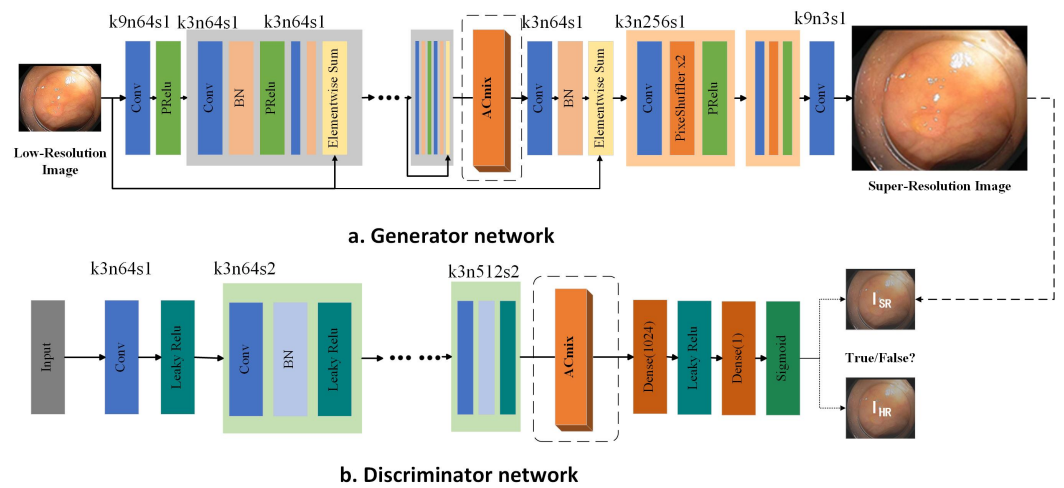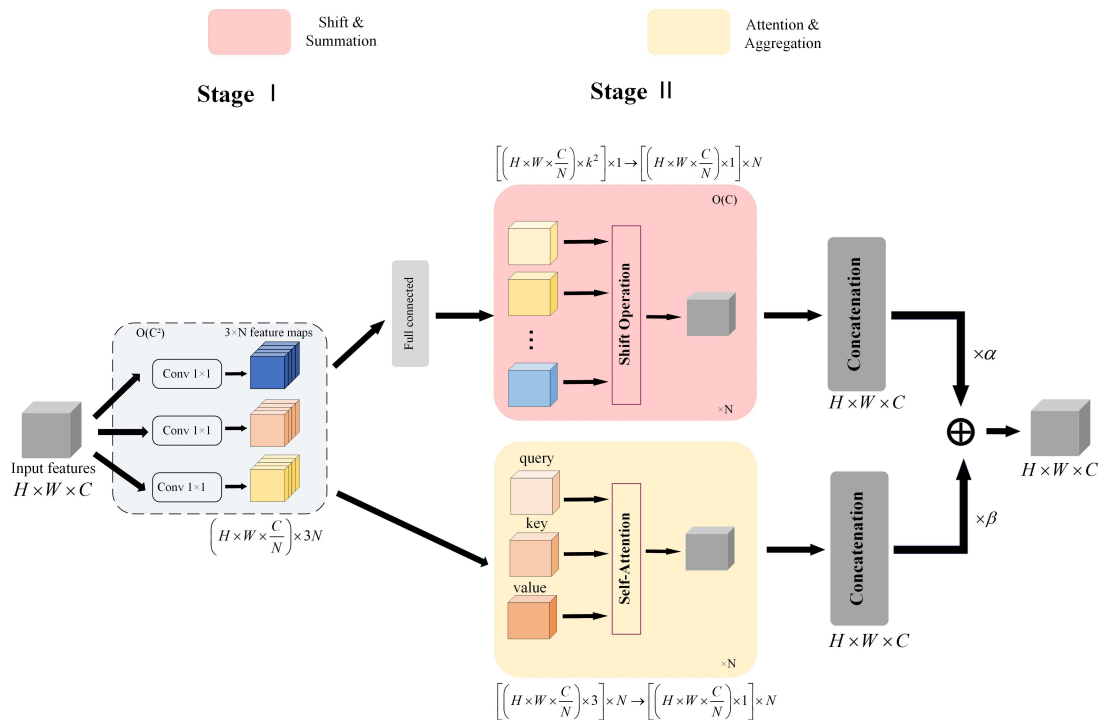

Figure 2. Structure of improved SRGAN.

Figure 3. Structure of ACmix.

3.2. Improved YOLOv5 Polyp Detection Algorithm

Aiming to enhance the efficiency of the conventional YOLO detection algorithm, this paper proposes improvements to YOLOv5, specifically, a C3 fusion feature extraction module based on Res2net [46] within the backbone network. Additionally, we incorporate an attention mechanism, CBAM [47], into the detection layer. As a result, an intestinal polyp detection framework based on YOLOv5s is proposed. The precise configuration is exhibited in Figure 4.
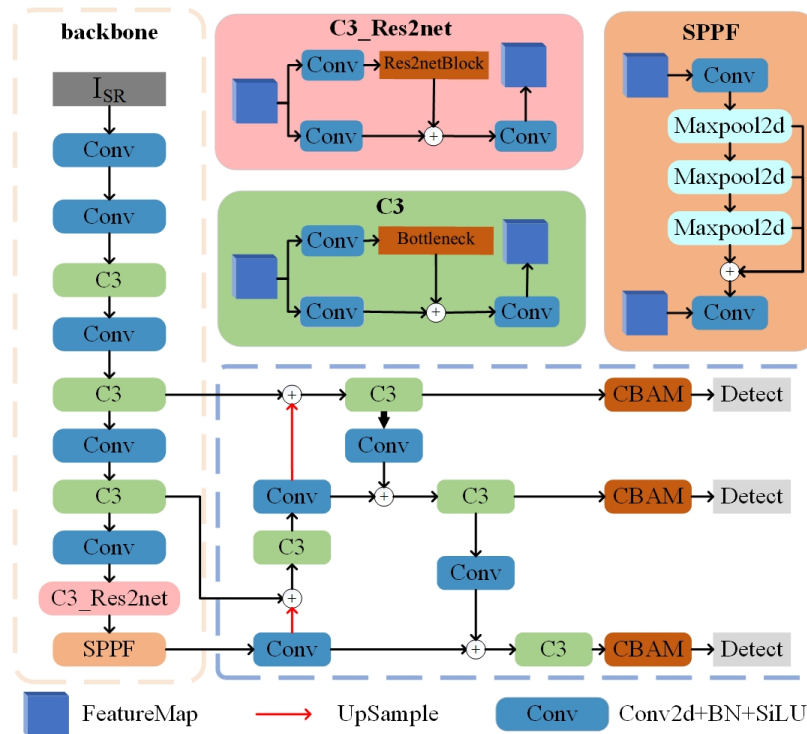

Figure 4. Improved YOLOv5-based polyp detection framework.

3.2.1. C3 Module Fused with Res2Net

The conventional C3 module has limited feature extraction capabilities, as it only employs three convolution layers. The main innovation of Res2Net lies in employing hierarchical cascaded feature group convolutions, which facilitate the enlargement of receptive fields. The finer-grained multibranch structure is used to achieve more effective feature extraction. We propose a new module, the C3_Res2Net module, which combines the C3 module and Res2Net. This module improves the accuracy of the YOLOv5, better extracts features of different scales, and broadens the receptive field. Consequently, it allows for more comprehensive capture of intestinal image feature information. The final C3 convolution module in the backbone network is replaced with the C3_Res2Net module in this article.

The primary architecture of C3_Res2Net is depicted in Figure 5. Initially, the feature map undergoes a 1×1 convolution, partitioning the features into s subsets, with the parameter s set to 4 in this study. Except for x_1 and x_2, the other subgroups can accept features from the left branch and perform element-wise addition. Then, a corresponding 3×3 convolution is applied, denoted by g_i, thereby expanding the receptive field of the feature convolution. Subsequently, the final output y_i of the split features is recombined and passed through a 1×1 convolution to produce the resulting feature. Consequently, within the fused C3_Res2Net module, features are extracted at a finer granularity, allowing

for more effective handling of global and local features, thereby improving recognition accuracy. The corresponding mathematical expression can be represented by Equation (5):

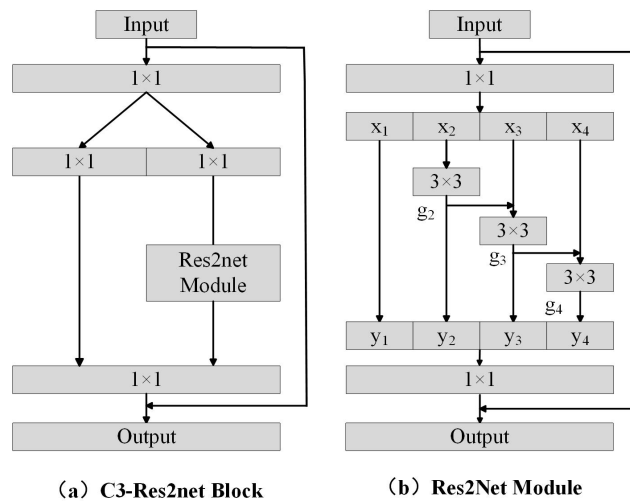$$y_i = \begin{cases} x_i, & i = 1 , \\ g_i(x_i), & i = 2 , \\ g_i(x_i + y_{i-1}), & 2 < i \leq s . \end{cases} \tag{5}$$

（a）**C3-Res2net Block**   （b）**Res2Net Module**

Figure 5. C3_Res2net Block (**a**) and Res2Net Module (s = 4) (**b**).

3.2.2. CBAM Attention Mechanism Module

The attention mechanism is commonly used in machine learning. The accuracy of the network can be affected by the presence of both small polyps and blocked intestinal polyps. To solve this problem, we introduce the CBAM before the prediction head. By incorporating the CBAM, we aim to enhance the accuracy of polyp detection and minimize the impact of the complex intestinal environment. The CBAM is composed of the spatial attention module (SAM) and the channel attention module (CAM). It simultaneously generates channel and spatial attention feature map information before performing adaptive calibration on the input feature map. Figure 6 depicts its primary structure.
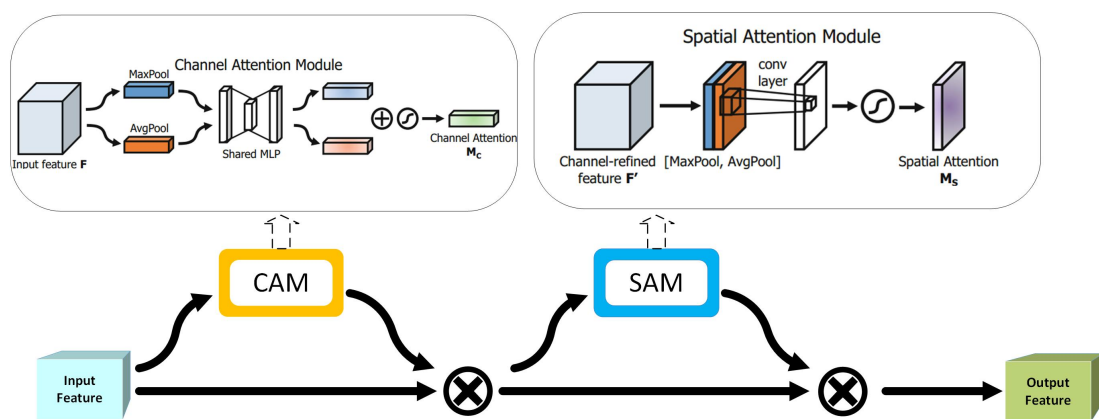

Figure 6. CBAM attention mechanism module.

The CBAM attention mechanism initiates by directing input feature values to CAM, where weighted calculations occur. Subsequently, these processed values are directed to SAM, where weighted calculations are once again performed. The specific calculations are defined by Equations (6) and (7).

$$F' = M_c(F) \otimes F , \tag{6}$$

$$F'' = M_S(F') \otimes F', \tag{7}$$

where F represents input features, F' denotes the one-dimensional channel attention module, $M_S(\cdot)$ denotes the two-dimensional spatial attention module, F'' denotes the output eigenvalue, and $\otimes$ represents element-wise multiplication.

3.3. Evaluation Indexes

Evaluation indexes in DL are crucial tools for assessing algorithm performance. This study primarily focuses on *precision*, *recall*, and *mAP*, which facilitate the assessment of algorithmic effectiveness.

3.3.1. SR Evaluation Index

Peak signal-to-noise ratio (PSNR) and structure similarity index measure (SSIM) serve as standard metrics for assessing image reconstruction quality. PSNR measures the fidelity of image reconstruction, while SSIM quantifies the similarity between the reconstructed I_{SR} and I_{HR}. The mathematical expressions for PSNR and SSIM are provided in Equations (8) and (9).

$$PSNR = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right), \tag{8}$$

where MSE represents the mean square error between the two images and MAX is the maximum value that can be calculated from the image pixels. The PSNR value is generally within the range of 20 to 50 dB, with higher PSNR values indicating better image quality.

$$SSIM = \frac{(2\mu_{HR}\mu_{SR} + C_1)(2\sigma_{HR,SR} + C_2)}{(\mu_{HR}^2 + \mu_{SR}^2 + C_1)(\sigma_{HR}^2 + \sigma_{SR}^2 + C_2)}, \tag{9}$$

where μ represents the gray mean, σ represents the variance, and C_1 and C_2 are constants that keep the equation valid. The SSIM value ranges from -1 to 1. In practical applications, it is typically between 0 and 1. SSIM can measure the similarity between two polyp images. The greater their structural similarity, the higher the $SSIM$ value.

3.3.2. Indices for Object Detection Evaluation

In YOLO detection, evaluation metrics such as *precision*, *recall*, *mAP* (mean average precision), and F1-score are commonly employed. The specific formulas for these metrics [48] are provided in Equations (10)–(13).

$$Precision = \frac{TP}{TP + FP}, \tag{10}$$

$$Recall = \frac{TP}{TP + FN}, \tag{11}$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{12}$$

where TP (true positive) represents the number of correct predictions, FP (false positive) denotes the number of incorrect positive predictions, FN (false negative) represents the number of positive instances that the model failed to predict correctly. *Precision* represents the proportion of correct positive predictions, and *recall* represents the proportion of all correct predictions. The F1-score ranges from 0% to 100%. The F1-score represents the weighted average of *precision* and *recall*. The F1-score of 100% represents the best possible classification performance. The higher the F1-score, the better the model performance.

$$mAP = \frac{AP}{N} = \frac{\int_0^1 p(r)dr}{N}, \tag{13}$$

where mAP represents the average AP values, N is the variety of classes contained, p is the value of *precision*, and r is the value of *recall*.

4. Results

4.1. Experimental Platform and Parameters

The model experimentation was conducted on Windows 11, with a system comprising 16 GB of memory, an Intel i7-12700F 2.10 GHz CPU, and an NVIDIA GeForce RTX3070 graphics card. The relevant environmental configuration is listed in Table 1.

Table 1. Experiment-related environment configuration.

Environmental Configuration	Specification
CUDA	11.7
CuDnn	8.7.0
Pytorch	2.0.0

4.2. Data Preparation

The dataset was obtained from a public dataset (the 8th National Student Biomedical Engineering Innovation Design Competition held in China, 2023), and has been anonymized to protect personal privacy. The dataset contains approximately 28,000 images of colorectal polyps and is used both for training the super-resolution reconstruction process and conducting detection experiments. The dataset consists of a variety of images (unstructured data) and corresponding annotation files (structured data). Using super-resolution reconstruction technology, we have improved the resolution of the images and enhance the ability to perceive fine structures. High-resolution images provide more detail and help the model to more accurately identify and locate colorectal polyps.

This dataset includes two main types: hyperplastic polyps and adenomatous polyps, which aids the model in distinguishing between these two different types of polyps during detection, thereby providing more information for clinical decision making.

4.3. Experimental Results

To further substantiate the effectiveness of the model proposed in this study for colon polyp detection, comparative analyses were conducted with various diagnostic algorithms. A series of ablation experiments were employed to evaluate the impact of different modules on the algorithm.

4.3.1. SR Experiment Results

In the super-resolution reconstruction experiment, a total of 28,000 colonoscopy images were processed. This reconstruction technique improves the resolution quality of the captured images, resulting in enhanced accuracy for polyp identification. The achieved $PSNR$ value of 30.14 and $SSIM$ measure of 0.79 demonstrate significant improvement in the SR reconstruction, providing essential and more detailed information for future detection. Figure 7 presents images from enteroscopy, both before (a) and after (b) SR reconstruction. The clearer edges of the polyps after SR reconstruction contribute to improved subsequent polyp detection. Figure 8 shows the details of some intestinal polyp images before and after reconstruction. The reconstructed images contain more texture, structure, and other detailed information.

Table 2 presents the $PSNR$ and $SSIM$ values for both the original SRGAN and the improved model. The incorporation of the ACmix module effectively enhances the quality of the generated image.

Table 2. SR comparison experiment.

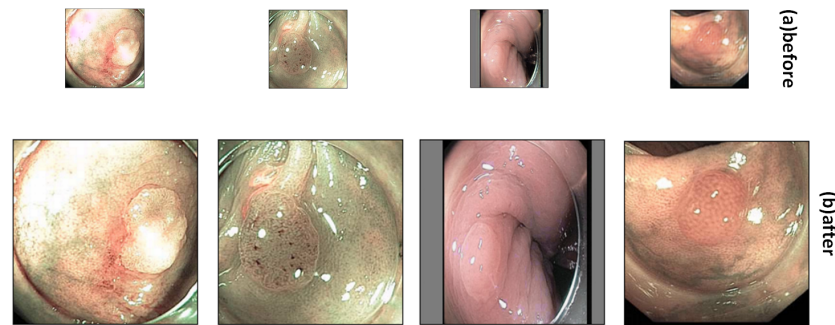| Methods	*PNSR*/%	*SSIM*/%
SRGAN	28.5	0.77
SRGAN + ACmix	30.14	0.79

Figure 7. Comparison of images before (**a**) and after (**b**) SR reconstruction.
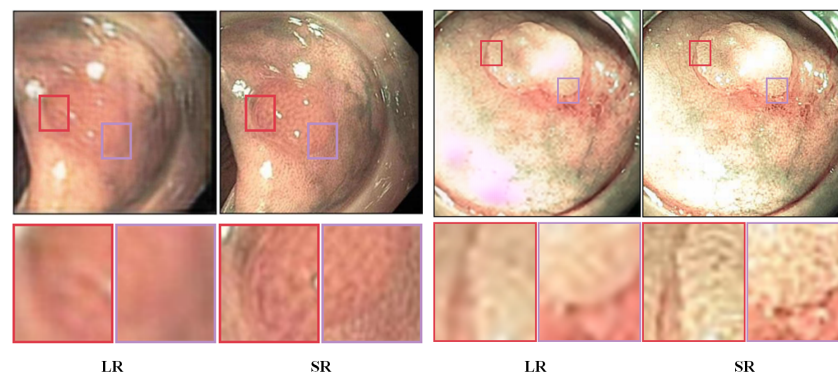

Figure 8. Visualization results before and after colon polyp image reconstruction.

With the increase in experimental epochs, the curves of PNSR and SSIM are shown in Figures 9 and 10. The table illustrates that the SRGAN that introduces the ACmix self-attention convolution module improves the SSIM by 0.02 compared with the original model. In comparison to the original SRGAN, the *PSNR* in this article increased by 1.64, thereby improving the visual effect of intestinal polyp data.
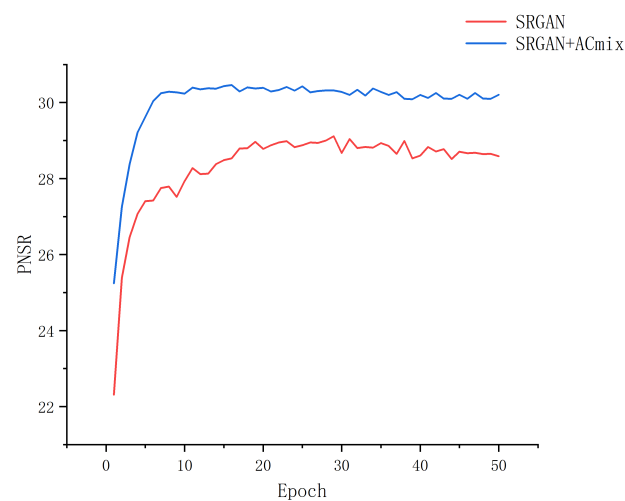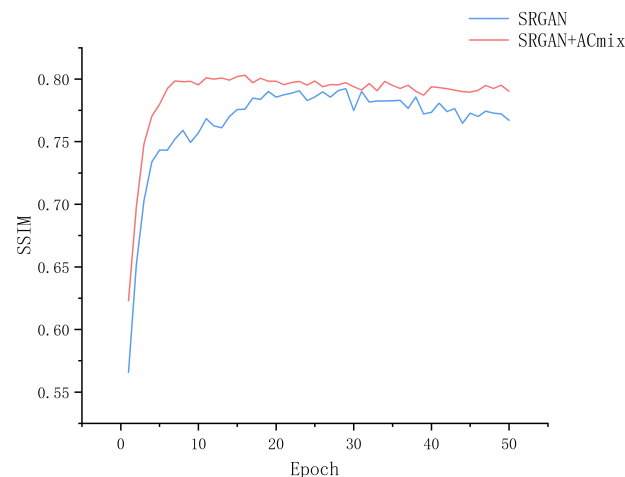

Figure 9. PNSR training.

Figure 10. SSIM training curve.

4.3.2. Experimental Results of Polyp Detection Modeling

The polyp detection model was trained using super-resolution (SR) reconstructed images. The training process utilized a batch size of 16 and spanned over 150 training rounds. Due to limitations in computational resources, we opted to train the polyp detection model using 8000 images. This approach allows for a balance between computational resources and the training of the polyp detection model. The dataset was divided into an 8:1:1 ratio. In the polyp detection model, we used 3000 images of hyperplastic polyps and 3400 images of hyperplastic polyps as the training set to ensure data balance during the training process. The validation set and test set each contain 800 polyp images for evaluating the training results. Table 3 depicts the comparative outcome of this model with other experimental methods.

Table 3. Comparison experiment.

Methods	*Precision*/%	*Recall*/%	$mAP_{@0.5}$/%	*F1-Score*/%
Faster R-CNN [49]	75.6	73.2	79.1	78.2
Faster R-CNN + SR	80.5	75.6	82.1	81.0
EfficientDet [50]	91.2	87.8	92.5	91.0
EfficientDet + SR	92.4	89.1	93.0	90.8
YOLOv5s	92.9	85.9	92.4	91.8
YOLOv5s + SR	94.1	88.6	92.7	91.8
YOLOv7-tiny [51]	79.2	68.1	73.8	80.7
YOLOv7-tiny + SR	82.5	72.4	78.1	82.6
YOLOv7 [51]	92.1	85.1	91.4	90.2
YOLOv7 + SR	93.6	87.5	91.4	91.8
YOLOv9 [52]	92.7	89.0	93.9	90.0
YOLOv9 + SR	94.2	91.4	92.6	93.5
Ours	95.2	91.5	94.2	94.1

In comparison to other models [49–51], the model proposed in this study exhibits improved performance across three key indicators. The detection model presented here surpasses the YOLOv5s benchmark by enhancing mean average precision (mAP) by 1.8%, precision by 2.3%, and recall by 5.6%. When comparing YOLOv7-tiny and YOLOv7, we observe distinct improvements in precision, recall, and mAP. Specifically, precision increased by 16% and 3.1% for YOLOv7-tiny and YOLOv7, respectively. Additionally, recall exhibited significant enhancements of 23.4% and 6.4% for the two models, while mAP experienced notable improvements of 20.4% and 2.8%, respectively. Compared to the Faster R-CNN model, the enhanced model demonstrates a notable 19.6% improvement in precision, an 18.3% increase in recall, and a significant 15.1% enhancement in mAP.

Utilizing EfficientDet resulted in achieving an mAP of 92.5%. These findings collectively demonstrate the enhanced efficiency of detecting colorectal polyps using the proposed model in this investigation.

Figure 11 illustrates the comparison of the mAP@0.5 curves between YOLO-SRPD and other models. As depicted in the graph, it is clear that our proposed method outperforms other approaches as the number of training iterations increases, resulting in higher mAP values. This improvement is particularly noticeable after around 60 training iterations. However, the method proposed in this study performs slightly worse than the standalone YOLOv5 in the first 40 epochs, mainly due to the introduction of SRGAN, which makes the training process more complex and requires more time to stabilize and realize its advantages. However, as training progresses, the proposed method outperforms the original YOLOv5 model. This indicates that it provides higher detection accuracy after sufficient training. The experimental results demonstrate the effectiveness of our suggested method in detecting colon polyps. The experimental results demonstrate the effectiveness of our suggested method in detecting colon polyps.
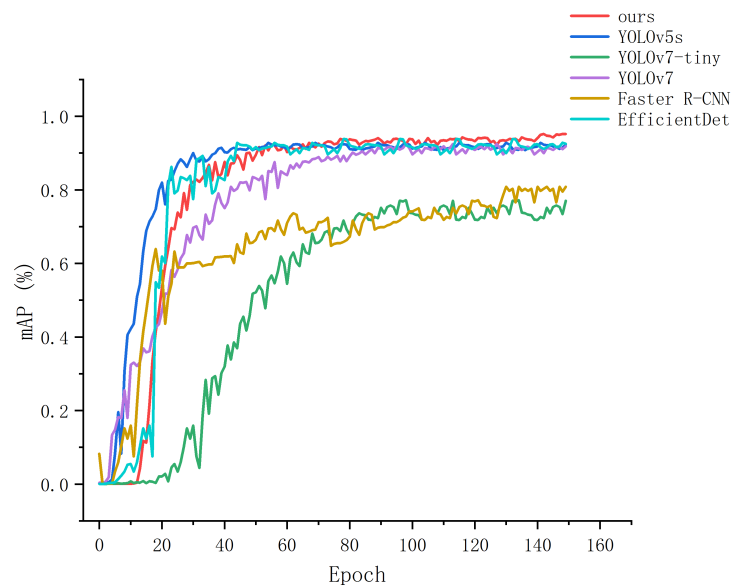

Figure 11. Comparison curve of different models.

This study conducted experiments to investigate the impact of replacing different C3 modules in the backbone of YOLOv5 with C3-Res2Net on the extraction of polyp feature information. The experimental results are presented in Table 4.

The replacement location of the C3 module is displayed in Figure 12. In this study, the C3 modules in the backbone are replaced with C3-res2net. There are a total of four C3 modules in the backbone. Starting from the input of the backbone, they are numbered sequentially as 1–4 (first, second, third, and fourth, as depicted in Figure 12). We investigate the impact of replacing different modules on the accuracy of polyp detection.

Table 4. Comparison of the replacement positions by C3-Res2Net.

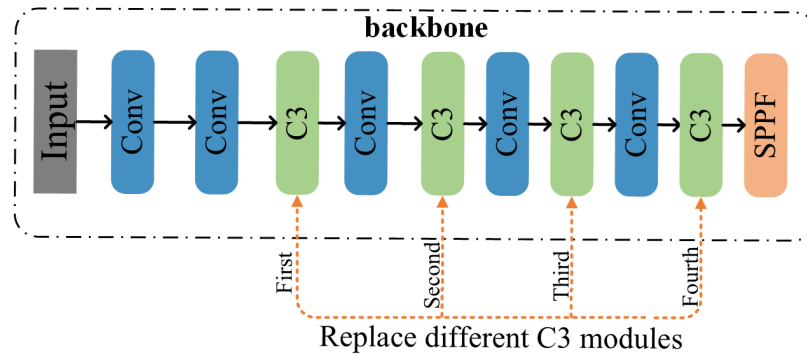| The Replacement Positions	*Precision*/%	*Recall*/%	$mAP_{@0.5}$/%
First	93.7	88.2	92.8
Second	93.8	91	92.6
Third	94.7	92	94
ALL	95.4	91.5	93.5

Figure 12. The replacement location of the C3 module.

Based on the experimental data, it is evident that incorporating Res2Net into the last layer of the C3 module is beneficial and improves the model's detection accuracy.

In addition, Figure 13 shows the training results curve of YOLO-SRPD. Figure 13 can demonstrate that there were no issues such as overfitting during the training process, proving that the model training results are satisfactory. During the training process, the training loss (Figure 13a–c) rapidly decreases and then stabilizes as the number of iterations increases, indicating that the model's performance in detecting polyps is gradually improving and the classification task is converging. In the validation loss (Figure 13f–h), it gradually decreases and stabilizes as training progresses, showing that the model is also learning better polyp predictions on the validation set. The precision (Figure 13d) rapidly increases at the beginning of training and then stabilizes, indicating that the model can more accurately identify positive samples. The mAP (Figure 13i) curve rises rapidly at the beginning and quickly converges, indicating that the overall performance of the model in the detection task is gradually improving. Furthermore, the similar trends of the training loss and validation loss curves indicate that the model does not exhibit significant overfitting. This demonstrates that our model performs well during training and gradually achieves high detection performance.
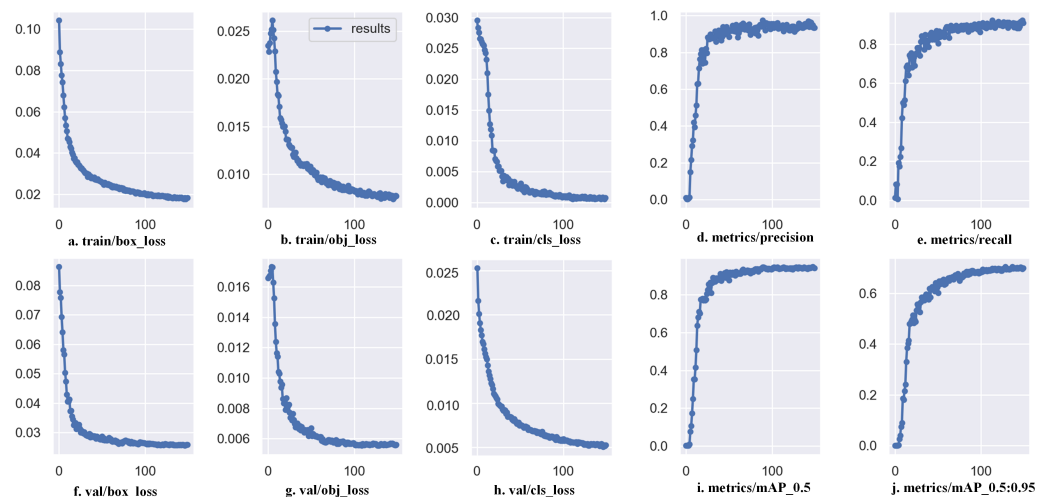

Figure 13. Training results curve of YOLO-SRPD.

4.3.3. Ablation Experiment

The results of the ablation experiment are presented in Table 5. These data demonstrate that the fusion of SR reconstruction, Res2Net module fused with the C3 module, and the inclusion of CBAM improve the accuracy of intestinal polyp detection.

Table 5. Ablation experiments.

*Yolov*5	SR	C3-Res2net	CBAM	*Precision*/%	*Recall*/%	*mAP*$_{@0.5}$/%	*F1-Score*/%
√				92.9	85.9	92.4	91.4
√	√			94.1	88.6	92.7	91.8
√	√	√		94.8	88.2	93.8	92.7
√	√	√	√	95.2	91.5	94.2	94.1

To verify the effectiveness of all modules in YOLO-SRPD, the ablation experiments were conducted to investigate the impact of each module. After incorporating the SR module, precision increased by 1.2%, indicating its efficacy in recovering information from low-resolution polyp images. Further addition of the C3_Res2net module led to a 0.7% increase in precision, a 1% rise in recall, and a 0.8% increased in mAP@0.5, demonstrating the strong multiscale feature extraction capability of the C3 module with residual structure, as it effectively extracts features from polyps of different scales. Moreover, incorporating SR, C3_Res2net, and CBAM together resulted in significant improvements across these three metrics. In summary, the improved model exhibits significant enhancements in both accuracy and precision.

4.4. Cross-Dataset Validation

To further verify the robustness of the model proposed in this paper, we conducted experiments on different public datasets. Table 6 presents the experimental results of YOLO-SRPD on various datasets. Meanwhile, the experimental results in Table 6 show significant differences compared to the results on other publicly available datasets. The experimental outcomes demonstrate that the detection performance of the model decreases significantly during cross-dataset evaluation. The information island effect in medical images greatly impacts the model's generalization ability.

Table 6. Experiments on different public datasets.

Dataset	*Precision*/%	*Recall*/%	*mAP*$_{@0.5}$/%
ETIS-Larib	89.6	87.4	88.8
CVC-ClinicDB	90.2	82.7	89.0
Kvasir-SEG	86.7	88.6	89.4

4.5. Visualization Detection of Colon Polyps

Figure 14 presents the results of polyp detection, depicting results for two distinct polyp types. The model demonstrates proficiency in detecting various polyp types.
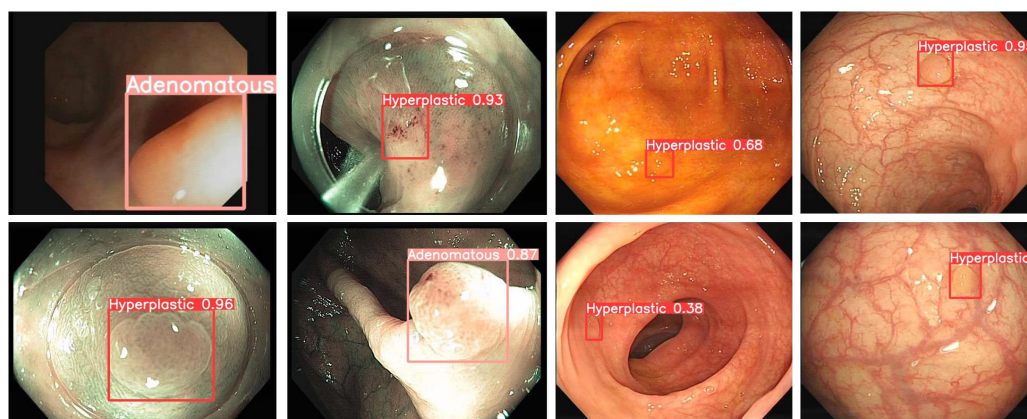

Figure 14. Examples of polyp detection.

5. Conclusions

Some polyps may progress to colorectal cancer tumors. Early colonoscopy can detect and remove some polyps. However, the low resolution of colonoscopy images and the small size of some polyps pose a diagnostic challenge. By using SRGAN for super-resolution reconstruction of images, the pixel size was increased by a factor of 4. This process led to an improvement in the clarity and texture of polyp images, thereby enhancing their visibility.

To address the challenge of misdiagnosis in colorectal polyps attributed to low-resolution images during colonoscopy, this study presents a model integrating an enhanced SRGAN for image super-resolution and an improved YOLOv5s model for polyp detection. Firstly, the study addresses the problem of insufficient resolution in colorectal polyp images by performing super-resolution reconstruction. An improved SRGAN model is employed, incorporating mixed self-attention mechanisms and convolutional modules (ACmix) in both the generator and discriminator of SRGAN. This enhancement bolsters subsequent convolutional neural networks in effectively extracting features. Secondly, the YOLOv5s model is improved by integrating the Res2Net module into the C3 module, resulting in the proposed C3_Res2net fusion module. This modification increases the receptive fields of convolutional kernels, thereby enhancing the detection rate of polyps of varying sizes. Additionally, a CBAM attention mechanism is incorporated to augment the model's focus on colorectal polyps. The experimental results indicate that the model proposed in this paper exhibits high accuracy in detecting colorectal polyps. With an mAP of 94.2% and a precision of 95.2%, the model effectively localizes polyps in the colon. Consequently, employing this proposed model facilitates efficient detection of polyp locations. In the future, we will further investigate lightweight models for polyp detection to achieve even faster screening capabilities.

Compared to other detection models, the method proposed in this study demonstrates higher accuracy, making it a practical tool for assisting medical professionals in colorectal polyp detection and reducing the rate of missed diagnoses. However, it is crucial to acknowledge that this heightened accuracy comes with an associated increase in computational complexity. Future research will explore lightweight target detection models to address these computational challenges. In recent years, more effective super-resolution reconstruction algorithms have been proposed. Therefore, our next research direction will focus on the reconstruction and detection of colorectal polyp images using the latest algorithms. In addition, we also aim to combine the advantages of super-resolution algorithms with the YOLO series detection algorithms, making the integration of these into a hybrid framework for comprehensive detection a future research direction.

Author Contributions: Conceptualization, Y.C. and Z.C.; methodology, S.W. and J.X.; software, J.X.; validation, S.W. and Z.C.; formal analysis, S.W.; investigation, S.W. and J.X.; writing—original draft preparation, J.X.; writing—review and editing, S.W.; visualization, Z.C.; supervision, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The experimental dataset can be accessed at the following: https://aistudio.baidu.com/datasetdetail/216022 (accessed on 26 May 2023).

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

References

1. Siegel, R.L. Cancer Statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [CrossRef]
2. Jass, J.R. Hyperplastic polyps and colorectal cancer: Is there a link? *Clin. Gastroenterol. Hepatol.* **2004**, *2*, 1–8. [CrossRef]
3. Glover, B.; Teare, J.; Patel, N. The Status of Advanced Imaging Techniques for Optical Biopsy of Colonic Polyps. *Clin. Transl. Gastroenterol.* **2020**, *11*, e00130. [CrossRef] [PubMed]
4. Abadir, A.P.; Ali, M.F.; Karnes, W.; Samarasena, J.B. Artificial intelligence in gastrointestinal endoscopy. *Clin. Endosc.* **2020**, *53*, 132–141. [CrossRef]

5. Le Clercq, C.M.; Bouwens, M.W.; Rondagh, E.J.; Bakker, C.M.; Keulen, E.T.; de Ridder, R.J.; Winkens, B.; Masclee, A.A.; Sanduleanu, S. Postcolonoscopy Colorectal Cancers Are Preventable: A Population-Based Study. *Gut* **2014**, *63*, 957–963. [CrossRef]

6. Shi, Y. Study of Machine Learning Techniques and Applications in Med-Ical Image Analysis. Ph.D. Thesis, Nanjing University, Nanjing, China, 2013.

7. Jie, L.; Liang, P.; Zhao, Z.; Chen, J.; Chang, Q.; Zeng, Z. Adan: An adversarial domain adaptation neural network for early gastric cancer prediction. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022; pp. 2169–2172.

8. ELKarazle, K.; Raman, V.; Then, P.; Chua, C. Detection of colorectal polyps from colonoscopy using machine learning: A survey on modern techniques. *Sensors* **2023**, *23*, 1225. [CrossRef]

9. Kominami, Y.; Yoshida, S.; Tanaka, S.; Sanomura, Y.; Hirakawa, T.; Raytchev, B.; Tamaki, T.; Koide, T.; Kaneda, K.; Chayama, K. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest. Endosc.* **2016**, *83*, 643–649. [CrossRef] [PubMed]

10. Tamaki, T.; Yoshimuta, J.; Kawakami, M.; Raytchev, B.; Kaneda, K.; Yoshida, S.; Takemura, Y.; Onji, K.; Miyaki, R.; Tanaka, S. Computer-aided colorectal tumor classification in NBI endoscopy using local features. *Med. Image Anal.* **2013**, *17*, 78–100. [CrossRef] [PubMed]

11. Min, M.; Su, S.; He, W.; Bi, Y.; Ma, Z.; Liu, Y. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. *Sci. Rep.* **2019**, *9*, 2881. [CrossRef]

12. Hegazy, M.A.; Cho, M.H.; Lee, S.Y. Half-scan artifact correction using generative adversarial network for dental CT. *Comput. Biol. Med.* **2021**, *132*, 104313. [CrossRef]

13. Yoshimura, T.; Nishioka, K.; Hashimoto, T.; Kogame, S.; Seki, K.; Sugimori, H.; Yamashina, H.; Kato, F.; Aoyama, H.; Kudo, K.; et al. Evaluation of Visualizing the Prostatic Urinary Tract in MRI with a Super Resolution Deep Learning Model for Urethra Sparing Radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2021**, *111*, e121–e122. [CrossRef]

14. Pavlou, E.; Gaitanis, G.; Bassukas, I.D.; Kourkoumelis, N. BCC and Immunocryosurgery scar differentiation through computa-tional resolution-enhanced OCT images and skin optical attenuation: A proof-of-concept study. *Exp. Dermatol.* **2024**, *33*, e15019. [CrossRef] [PubMed]

15. Shi, J.; Ye, Y.; Liu, H.; Zhu, D.; Su, L.; Chen, Y.; Huang, Y.; Huang, J. Super-resolution reconstruction of pneumocystis carinii pneumonia images based on generative confrontation network. *Comput. Methods Programs Biomed.* **2022**, *215*, 106578. [CrossRef]

16. Zhu, P.C.; Wan, J.J.; Shao, W.; Meng, X.C.; Chen, B.L. Colorectal image analysis for polyp diagnosis. *Front. Comput. Neurosci.* **2024**, *18*, 1356447. [CrossRef] [PubMed]

17. Ghose, P.; Ghose, A.; Sadhukhan, D.; Pal, S.; Mitra, M. Improved polyp detection from colonoscopy images using finetuned YOLO-v5. *Multimed. Tools Appl.* **2024**, *83*, 42929–42954. [CrossRef]

18. Yasmin, F.; Hassan, M.M.; Hasan, M.; Zaman, S.; Bairagi, A.K.; El-Shafai, W.; Fouad, H.; Chun, Y.C. GastroNet: Gastrointestinal polyp and abnormal feature detection and classification with deep learning approach. *IEEE Access* **2023**, *11*, 97605–97624. [CrossRef]

19. Ledig, C. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 June 2017.

20. Mahmood, T.; Saba, T.; Rehman, A.; Alamri, F.S. Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. *Expert Syst. Appl.* **2024**, *249*, 123747. [CrossRef]

21. Zhang, S.; Yuan, Z.; Zhou, X.; Wang, H.; Chen, B.; Wang, Y. VENet: Variational energy network for gland segmentation of pathological images and early gastric cancer diagnosis of whole slide images. *Comput. Methods Programs Biomed.* **2024**, *250*, 108178. [CrossRef] [PubMed]

22. Kim, B.S.; Kim, B.; Cho, M.; Chung, H.; Ryu, J.K.; Kim, S. Enhanced multi-class pathology lesion detection in gastric neoplasms using deep learning-based approach and validation. *Sci. Rep.* **2024**, *14*, 11527. [CrossRef]

23. Tan, L.; Liang, Y.; Xia, J.; Wu, H.; Zhu, J. Detection and Diagnosis of Small Target Breast Masses Based on Convolutional Neural Networks. *Tsinghua Sci. Technol.* **2024**, *29*, 1524–1539. [CrossRef]

24. Vinsard, D.G.; Mori, Y.; Misawa, M.; Kudo, S.E.; Rastogi, A.; Bagci, U.; Rex, D.K.; Wallace, M.B. Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointest. Endosc.* **2019**, *90*, 55–63. [CrossRef]

25. Sanchez-Peralta, L.F.; Bote-Curiel, L.; Picon, A.; Sanchez-Margallo, F.M.; Pagador, J.B. Deep learning to find colorectal polyps in colonoscopy: A systematic literature review. *Artif. Intell. Med.* **2020**, *108*, 101923. [CrossRef]

26. Kavitha, M.S.; Gangadaran, P.; Jackson, A.; Venmathi Maran, B.A.; Kurita, T.; Ahn, B.C. Deep Neural Network Models for Colon Cancer Screening. *Cancers* **2022**, *14*, 3707. [CrossRef] [PubMed]

27. Yang, K.; Chang, S.; Tian, Z.; Gao, C.; Du, Y.; Zhang, X.; Liu, K.; Meng, J.; Xue, L. Automatic polyp detection and segmentation using shuffle efficient channel attention network. *Alex. Eng. J.* **2022**, *61*, 917–926. [CrossRef]

28. Su, Y.; Cheng, J.; Zhong, C.; Zhang, Y.; Ye, J.; He, J.; Liu, J. FeDNet: Feature Decoupled Network for polyp segmentation from endoscopy images. *Biomed. Signal Process. Control* **2023**, *83*, 104699. [CrossRef]

29. Yu, T.; Wu, Q. HarDNet-CPS: Colorectal polyp segmentation based on Harmonic Densely United Network. *Biomed. Signal Process. Control* **2023**, *85*, 104953. [CrossRef]

30. Su, Y.; Cheng, J.; Yi, M.; Liu, H. FAPN: Feature augmented pyramid network for polyp segmentation. *Biomed. Signal Process. Control* **2022**, *78*, 103903. [CrossRef]

31. Ozawa, T.; Ishihara, S.; Fujishiro, M.; Kumagai, Y.; Shichijo, S.; Tada, T. Automated Endoscopic Detection and Classification of Colorectal Polyps Using Convolutional Neural Networks. *Ther. Adv. Gastroenterol.* **2020**, *13*, 1756284820910659. [CrossRef]

32. Kayser, M.; Soberanis-Mukul, R.D.; Zvereva, A.M.; Klare, P.; Navab, N.; Albarqouni, S. Understanding the Effects of Artifacts on Automated Polyp Detection and Incorporating That Knowledge via Learning without Forgetting. *arXiv* **2020**, arXiv:2002.02883. [CrossRef].

33. Ali, S.; Zhou, F.; Daul, C.; Braden, B.; Bailey, A.; Realdon, S.; East, J.; Wagnières, G.; Loschenov, V.; Grisan, E.; et al. Endoscopy Artifact Detection (EAD 2019) Challenge Dataset. *arXiv* **2019**, arXiv:1905.03209. [CrossRef].

34. 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]

35. Silva, J.; Histace, A.; Romain, O.; Dray, X.; Granado, B. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *Int. J. Comput. Assist. Radiol. Surg.* **2014**, *9*, 283–293. [CrossRef] [PubMed]

36. 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 Proceedings of the MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, Republic of Korea, 5–8 January 2020; Proceedings, Part II 26, 2020; pp. 451–462.

37. Zeng, Y.; Xu, S.; Chapman, W.C.; Li, S.; Alipour, Z.; Abdelal, H.; Chatterjee, D.; Mutch, M.; Zhu, Q. Real-Time Colorectal Cancer Diagnosis Using PR-OCT with Deep Learning. *Theranostics* **2020**, *10*, 2587–2596. [CrossRef] [PubMed]

38. Laddha, M.; Jindal, S.; Wojciechowski, J. Gastric Polyp Detection Using Deep Convolutional Neural Network. In Proceedings of the 2019 4th International Conference on Biomedical Imaging, Signal Processing, New York, NY, USA, 27–29 September 2020; ICBSP '19; pp. 55–59. [CrossRef]

39. Zhang, X.; Chen, F.; Yu, T.; An, J.; Huang, Z.; Liu, J.; Hu, W.; Wang, L.; Duan, H.; Si, J. Real-Time Gastric Polyp Detection Using Convolutional Neural Networks. *PLoS ONE* **2019**, *14*, e0214133. [CrossRef] [PubMed]

40. Zhang, R.; Zheng, Y.; Poon, C.C.Y.; Shen, D.; Lau, J.Y.W. Polyp Detection during Colonoscopy Using a Regression-Based Convolutional Neural Network with a Tracker. *Pattern Recognit.* **2018**, *83*, 209–219. [CrossRef] [PubMed]

41. Tang, C.P.; Chang, H.Y.; Wang, W.C.; Hu, W.X. A Novel Computer-Aided Detection/Diagnosis System for Detection and Classification of Polyps in Colonoscopy. *Diagnostics* **2023**, *13*, 170. [CrossRef] [PubMed]

42. Carrinho, P.; Falcao, G. Highly Accurate and Fast YOLOv4-based Polyp Detection. *Expert Syst. Appl.* **2023**, *232*, 120834. [CrossRef]

43. Tang, C.P.; Hsieh, C.H.; Lin, T.L. Computer-Aided Image Enhanced Endoscopy Automated System to Boost Polyp and Adenoma Detection Accuracy. *Diagnostics* **2022**, *12*, 968. [CrossRef] [PubMed]

44. Chou, Y.C.; Chen, C.C. Improving deep learning-based polyp detection using feature extraction and data augmentation. *Multimed. Tools Appl.* **2023**, *82*, 16817–16837. [CrossRef]

45. Chen, B.L.; Wan, J.J.; Chen, T.Y.; Yu, Y.T.; Ji, M. A self-attention based faster R-CNN for polyp detection from colonoscopy images. *Biomed. Signal Process. Control* **2021**, *70*, 103019. [CrossRef]

46. Gao, S.H. Res2net: A New Multi-Scale Backbone Architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [CrossRef] [PubMed]

47. Woo, S. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

48. Bishop, C.M. Pattern recognition and machine learning. *Springer Google Sch.* **2006**, *2*, 1122–1128.

49. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9. [CrossRef] [PubMed]

50. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10781–10790.

51. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.

52. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.