ISSN:3068-9376

Frontiers in Preventive Medicine

DOI: 10.64904/fpm25018

Original Research

A Gastrointestinal Polyp Detection and Treatment-Assistance Method Based on an Improved YOLOv5n Network

Ran Chen [1,2], Jingkun Liang [,1,2], Cong Chen [2]*

[1] Yazhou Bay Innovation Institute of Hainan Tropical Ocean University, China

[2] School of Marine Information Engineering, Hainan Tropical Ocean University, Sanya 572022, China

Abstract

Gastrointestinal polyps are common lesions in the digestive system, and some lesions carry a risk of malignant transformation. Timely and accurate detection is of great clinical significance. However, due to factors such as weak polyp texture, low contrast, and complex backgrounds, traditional endoscopic examinations still have a certain detection missed rate. To address this, this paper proposes an improved network for gastrointestinal polyp detection based on the lightweight object detection model YOLOv5n. The proposed method introduces the ContextGuidedC3 module to enhance multi-scale contextual information modeling. Additionally, the CARAFE (Content-Aware Re-Assembly of Feature Elements) operator is used for content-aware upsampling, improving the reconstruction quality of high-resolution features and the expression of small target edge details. Experiments were conducted on the polyps dataset and compared with lightweight models such as YOLOv5n, YOLOv8n, YOLOv10n, and YOLOv11n. The results show that, compared to the original YOLOv5n, the proposed method improved the Recall from 85.7% to 90.3%, mAP@0.5 from 92.4% to 94.2%, and mAP@0.5-0.95 from 67.3% to 69.6%. The method also demonstrates better stability and robustness in terms of confusion matrix, visualized detection results, and PR curves. This indicates that the proposed method effectively improves the detection accuracy and small target recognition capability of gastrointestinal polyps while maintaining low computational overhead, and has potential clinical application value.

Keywords: Gastrointestinal polyps; Medical imaging; YOLOv5n; CARAFE upsampling operator; ContextGuidedC3 module

1. Introduction

Gastrointestinal polyps are common lesions in the digestive system, some of which carry a potential risk of malignant transformation, serving as important precursor lesions for colorectal cancer and other gastrointestinal tumors. Clinically, gastrointestinal endoscopy remains the gold standard for detecting and removing polyps. However, in practice, due to factors such as small polyp size, the similarity in color with the surrounding mucosa, diverse shapes, and differences in endoscopist experience, polyp miss rates still occur [1]. Therefore, improving the sensitivity and accuracy of gastrointestinal polyp detection holds significant clinical importance.

In recent years, with breakthroughs in deep learning, especially Convolutional Neural Networks (CNNs) in image recognition [2], computer-aided diagnostic

* *Corresponding author: 451758104@qq.com*

systems based on artificial intelligence have shown great potential in medical image analysis [3]. For the automated analysis of endoscopic images, deep learning methods, with their strong feature representation ability, can automatically learn the shape, texture, and edge features of polyps from complex backgrounds, enabling efficient detection and localization of polyps [3, 4]. Numerous studies have shown that deep learning-based gastrointestinal polyp detection systems have achieved accuracy, sensitivity, and real-time performance close to or even meeting clinical requirements [2, 3], offering feasible solutions to assist endoscopists in reducing miss rates and improving examination efficiency.

Despite significant progress in existing studies, many challenges remain in practical applications. For example, Rostami et al. [5] conducted a systematic review of the performance of YOLOv7 in colonoscopic polyp detection and evaluated its feasibility for transfer to hysteroscopic polyp detection based on technical similarity. However, direct experimental validation based on hysteroscopic images is still lacking, and conclusions are mainly drawn from indirect inferences from colonoscopy research results. Haider et al. [6] proposed four variant models of YOLOv9 (Gelan-c, Gelan-e, YOLOv9-c, YOLOv9-e) for training and evaluation on colorectal polyp detection tasks, validating the model's performance through data augmentation and multiple evaluation metrics. However, on the large-scale dataset LDPolypVideo, the best model's mAP@50 was only 55.56, indicating that its performance was relatively average and that the model's generalization ability in complex real-world scenarios remained limited. Sun et al. [7] proposed a lightweight polyp detection model EP-1. YOLO based on YOLOv10. This model introduced the GBottleneck module, designed a lightweight detection head GHead, added small target detection layers, improved the SE_SPPF attention module, and adopted the Wise-IoU loss function, achieving2. high precision and efficiency on multiple public polyp datasets. However, the model still experienced false negatives and false positives in extreme high-brightness or low-resolution scenarios, and its generalization ability needs further improvement. Keshavarz et al. [8] proposed a lightweight deep learning model based on pre-trained ResNet50, combining transfer learning and multi-task learning to perform both polyp classification and bounding box detection, balancing the two tasks with a weighted loss function. However, non-polyp image data still relied on external supplementation, and the dataset remained limited in terms of diversity and scale. Viet et al. [9] applied the YOLOv8 object detection algorithm to train and validate a dataset containing 50 colonoscopy videos and 20,616 images, evaluating its accuracy in detecting polyps in real videos through recall rate, precision, and F1 score. However, at higher IoU thresholds, the model's precision was low, and false positives were still present. Chen et al. [10] proposed the YOLO-MF model, which embedded a dual-feature perception module (DFP) in the YOLOv11 backbone network to enhance multi-scale feature extraction ability and introduced a fine-grained feature calibration module (FGFC) in the neck network to alleviate the loss of small targets and detail information, effectively improving polyp detection accuracy in complex scenes. However, this model has not been fully validated for computational efficiency and real-time performance, potentially facing challenges in inference speed under high computational load scenarios.

To address the shortcomings of existing gastrointestinal polyp detection methods in complex scenarios with weak texture, blurry boundaries, and strong background noise, this paper proposes structural improvements based on YOLOv5n. The main contributions of this paper are as follows:

To address the weak texture and unclear boundaries of polyps, we have made targeted improvements to the YOLOv5n structure, enhancing its feature representation ability and detection accuracy while maintaining model lightness.

We propose the ContextGuidedC3 module to strengthen multi-scale contextual information modeling. In the C3 module, a context-guided mechanism is introduced, and multiple ContextGuided

units are connected to effectively integrate local details and global semantics, thereby enhancing the model's discriminative ability in complex backgrounds.

We introduce the CARAFE upsampling operator. By dynamically predicting reassembly kernels, it can adaptively restore edge and texture information based on feature content, significantly improving the localization accuracy and boundary fitting of small-scale polyps.

2. Method design

To enhance the feature representation ability and object detection performance of lightweight detection networks in gastrointestinal polyp recognition tasks, this paper is based on an improved YOLOv5n [11] network structure. The overall method is shown in Figure 1. In the backbone network and feature fusion module, the original C3 module is replaced with the ContextGuidedC3 module proposed in this paper to strengthen the model's multi-scale semantic representation and contextual dependency modeling ability. This enables the model to more effectively distinguish polyp regions with weak textures, blurry boundaries, and strong background noise. Additionally, the CARAFE content-aware upsampling operator is introduced in the upsampling path to dynamically predict the upsampling kernel, thereby improving the reconstruction quality of high-resolution features and providing higher edge clarity and localization accuracy when processing small-scale polyps.

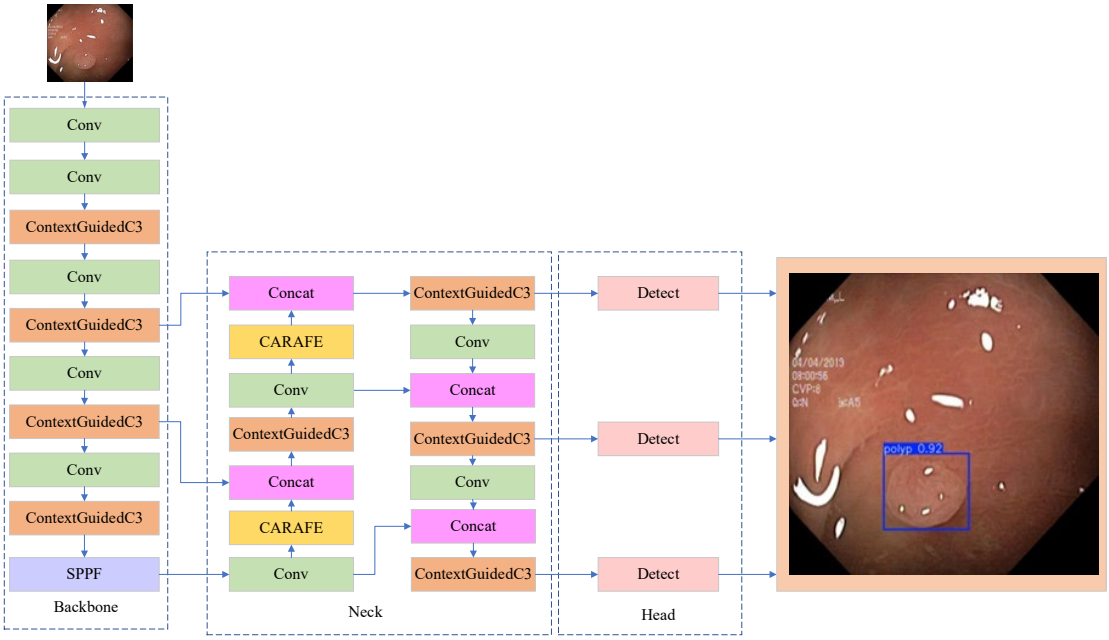


Figure 1. Improved YOLOv5n Network Architecture.

2.1 ContextGuidedC3 Module

In lightweight detection networks, efficiently integrating local detail features with global contextual information is crucial for improving object detection performance. The traditional C3 module has a simple structure and low computational cost but is insufficient in utilizing long-range dependencies and contextual information. To address this, this paper introduces a context-guided mechanism based on the C3 module, proposing the ContextGuidedC3 module. By concatenating multiple ContextGuided units in the bottleneck section, it strengthens the multi-scale contextual modeling ability. This improves the representation of polyp edges and textures while adding minimal parameters, thus enhancing overall detection performance.

As shown in Figure 2, let the input feature be X. First, the module performs initial feature extraction and channel adjustment on the input through a convolutional layer, resulting in the feature F_0.

$$F_0 = Conv_1(X), \tag{1}$$

Subsequently, the input F_0 is passed through the main branch, which is composed of multiple ContextGuided [12] units, to model multi-scale contextual information layer by layer and enhance semantic representation, resulting in the context-enhanced feature F_c

$$F_c = CG^{(n)}(F_0), \tag{2}$$

Here, $CG^{(n)}(\cdot)$ represents the concatenation of n ContextGuided modules. Meanwhile, a bypass is introduced from the input X, where a separate convolution is applied to obtain the featureF_s which is used to preserve the original structural information and improve the stability of gradient propagation.

$$F_s = Conv_2(X), \tag{3}$$

Next, the output feature F_c from the main branch and the feature F_s from the bypass are concatenated along the channel dimension to obtain the fused feature F_{Concat}.

$$F_{concat} = Concat(F_c, F_s), \tag{4}$$

Finally, a convolutional layer is applied to integrate and re-adjust the channels of the fused feature, resulting in the output feature $X'$of the module.

$$X' = Conv_3(F_{concat}), \tag{5}$$

This structure effectively fuses deep contextual information F_c with shallow local structural features F_s enhancing the feature representation ability while maintaining a low parameter overhead.

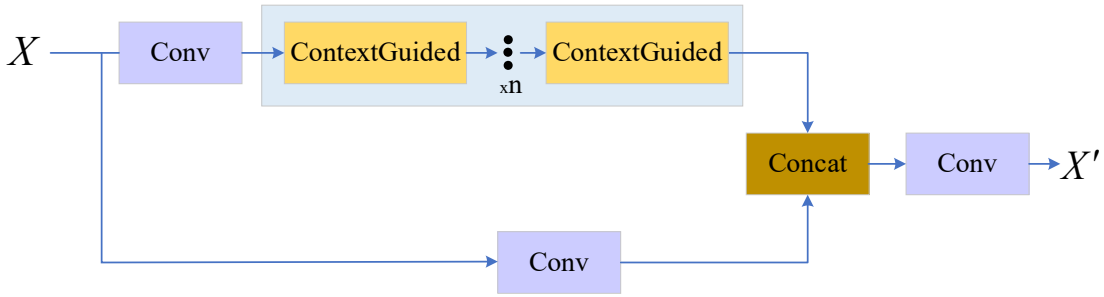


Figure 2. ContextGuidedC3 module.

2.2 Sampling Operator on CARAFE

In object detection networks, the upsampling process has a crucial impact on the reconstruction quality of high-resolution features. Traditional upsampling methods, although computationally lightweight, lack adaptability to image content and are prone to causing blurriness in edge regions, which affects the localization accuracy of small objects. To address this issue, this paper introduces the CARAFE upsampling operator [13], which regenerates the feature map in a "content-aware" manner. It adaptively predicts the reassembly weights based on the semantic information of the input features, thereby better restoring detailed structures and enhancing the representation of polyp edges and small-scale regions.

As shown in Figure 3, CARAFE (Content-Aware Reassembly of Features) is an innovative feature upsampling module. Its core idea is to redefine the upsampling process as "content-aware reassembly" rather than simple interpolation. Traditional methods (such as bilinear interpolation) apply fixed computation rules to all positions, whereas CARAFE generates a dynamic, position-specific upsampling kernel based on the semantic information of each position in the input feature map. Specifically, it first compresses the channel dimension to reduce computational cost, then encodes the local context, and predicts normalized reassembly weights in real-time. This allows the generation of kernels that preserve sharpness in object edges while smoothing the background to generate kernels that emphasize consistency. Subsequently, the content-aware reassembly module uses these dynamic kernels to extract the corresponding local receptive fields from the input features and perform weighted fusion, outputting a refined upsampling result.

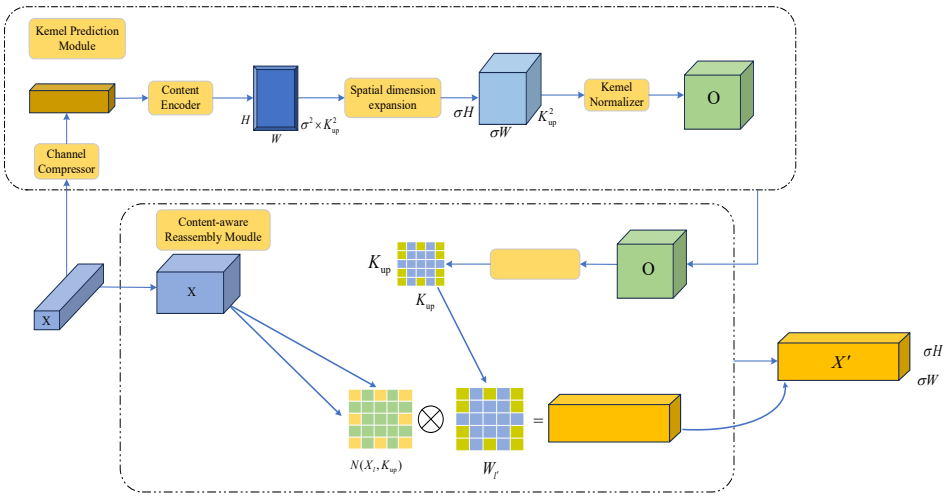


Figure 3. Sampling operator on CARAFE.

3. Experiment

3.1 Dataset

The dataset used in this study contains two types of endoscopic images: wireless capsule endoscopy (WCE) samples [14] and the publicly available colonoscopy dataset Kvasir-SEG [15]. During the data preparation process, the raw images were first filtered and cleaned to remove low-quality samples (such as blurry images, strong reflections, and duplicate frames). Then, annotations were standardized using the minimum bounding rectangle and consistency checks were performed to ensure the reliability of the annotations. After the dataset was divided, data augmentation techniques (such as rotation, translation, scaling, flipping, etc.) were applied to expand the training set. A total of 1611 valid samples were retained, including 1288 images for training and 323 images for testing. All images were color normalized and cropped, and resized to a consistent resolution to eliminate pixel differences. The test set was carefully selected to cover different types of polyps, lighting conditions, and endoscopic viewpoints, and included some challenging images (such as blurry edges, partial occlusion, or high-noise images) to evaluate the model's stability and adaptability under complex conditions. Figure 4 shows image samples from the polyp dataset.

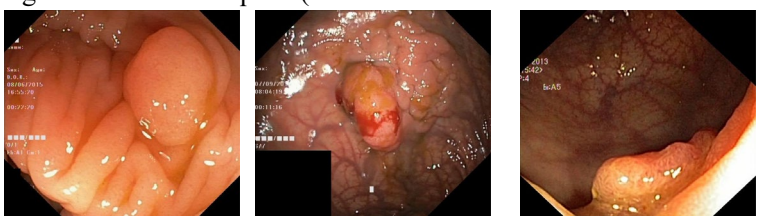


Figure 4. Image samples from the polyps dataset.

3.2 Experimental Platform and Hyperparameter Setting

This experiment was conducted on a high-performance computing environment, using PyTorch 1.10.0 as the deep learning framework, running on a software platform consisting of Python 3.8 and Ubuntu 20.04, combined with CUDA 11.3 to fully leverage the parallel computing capabilities of the GPU. The hardware configuration includes an NVIDIA RTX 4090 graphics card (24 GB VRAM), an AMD EPYC 7T83 64-core processor, and 90 GB of memory, providing ample computational support for large-scale data loading and model training. For the training hyperparameters, the input image size was set to 640×640, the batch size was set to 64, and the total number of training epochs was 200. The initial learning rate was set to 0.01, and the number of data loading processes was 8. The optimizer's momentum factor was set to 0.937, and the weight decay coefficient was set to 0.0005. This

configuration not only ensured stable convergence during the training process but also contributed to improving the overall detection performance of the model.

3.3 Evaluation Indicators

The experiments in this paper use the F1 score, Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP) as evaluation metrics [16], with their calculation formulas as shown below:

$$\text{Precision} = \frac{T_{\text{p}}}{T_{\text{p}} + F_{\text{p}}}, \tag{6}$$

$$\text{Recall} = \frac{T_{\text{p}}}{T_{\text{p}} + F_{\text{N}}}, \tag{7}$$

$$\text{AP} = \int_0^1 P(R)\text{d}R, \tag{8}$$

$$\text{mAP} = \frac{1}{n}\sum_{\text{i=0}}^{n} AP(i), \tag{9}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{10}$$

Where T_{p} represents the number of correctly detected objects; Fp represents the number of incorrectly detected objects; FN represents the number of missed detections; n represents the number of classes; and $AP(i)$ represents the average precision of the i-th object class.

4. Experimental analysis

4.1 Comparison of Results from Multiple Different models

To validate the effectiveness of the proposed method in object detection tasks, this paper selects commonly used lightweight detection models YOLOv5n, YOLOv8n [17], YOLOv10n [18], and YOLOv11n [19] as baseline comparisons, and performs comparative experiments under the same experimental settings. The evaluation metrics used include Recall, mAP@0.5, and mAP@0.5-0.95, to comprehensively assess the detection performance of the models from different perspectives. The quantitative results of each model on the test set are shown in Table 1.

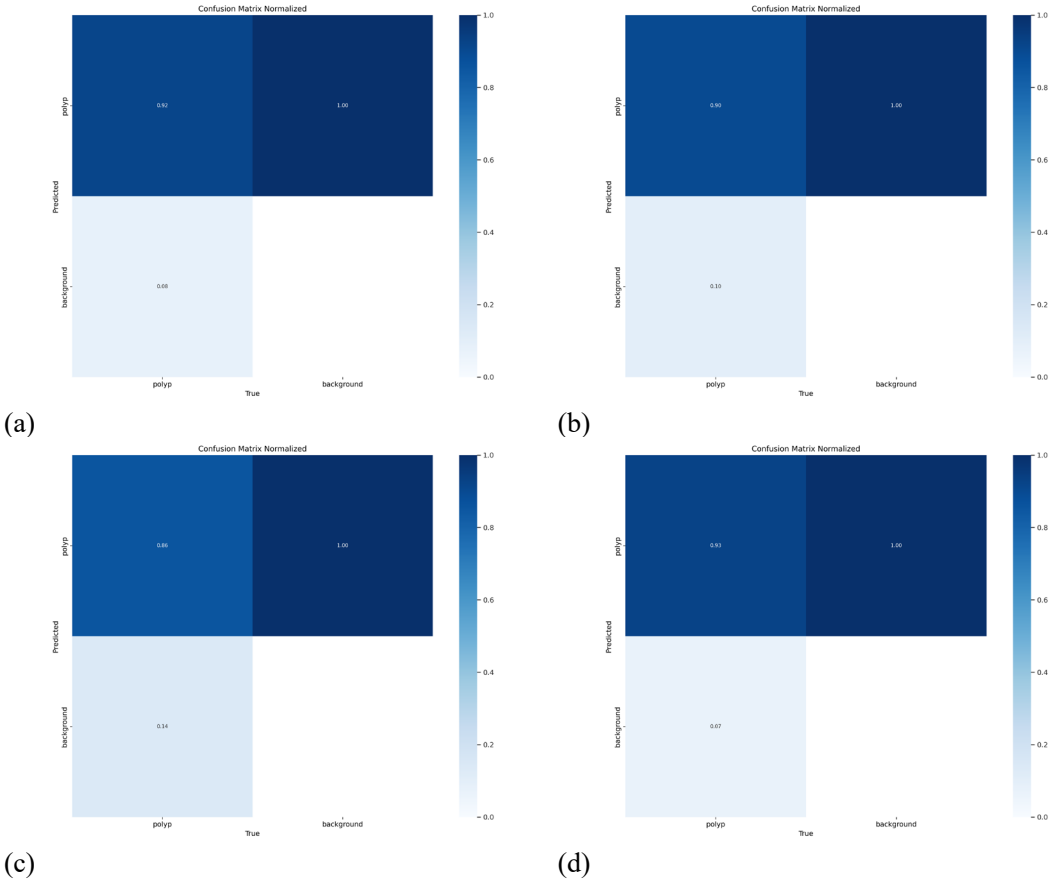
Table 1. Comparison of Results Across Different Models.

Algorithms	Recall/%	mAP@0.5/%	mAP@0.5-0.95/%
YOLOv5n	85.7	92.4	67.3
YOLOv8n	88.2	92.5	67.1
YOLOv10n	83.2	89.5	62.2
YOLOv11n	87.8	92.8	69.5
Ours	90.3	94.2	69.6

Table 1 presents a comparison of the detection performance of different models on this dataset. Overall, Ours method achieves the best results across all three metrics. The Recall reaches 90.3%, which is a 2.1% improvement over the second-best YOLOv8n (88.2%), effectively reducing the miss detection rate. The mAP@0.5 improves to 94.2%, surpassing YOLOv11n's 92.8%, indicating that the proposed method has an advantage in both object localization and classification accuracy at lower IoU thresholds. In the more stringent mAP@0.5-0.95 metric, Ours also slightly outperforms YOLOv11n with 69.6% versus 69.5%, demonstrating its robust performance across different IoU thresholds. In summary, the experimental results fully validate the comprehensive superiority of our method in terms of detection accuracy and stability compared to various YOLO lightweight baseline models.

To further analyze the false positive and false negative cases of each model in the polyp detection task from a classification perspective, relying solely on the global metrics in Table 1 is insufficient. Therefore, this paper computes the normalized confusion matrix for the predictions of different models on the test set, as shown in Figure 5. By comparing the confusion matrices of YOLOv5n, YOLOv8n, YOLOv10n, and Ours, we can more intuitively observe the models' ability to distinguish between

polyps and background samples, as well as the differences in error classification patterns across the different methods.



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

Figure 5. Normalized Confusion Matrix: (a) YOLOv5n; (b) YOLOv8n; (c) YOLOv10n; (d) Ours.

As shown in Figure 5, YOLOv5n, YOLOv8n, and YOLOv10n all have relatively high accuracy in correctly identifying the polyp class. Specifically, YOLOv5n achieves a correct detection rate of 0.92 and a miss detection rate of 0.08, YOLOv8n achieves a correct detection rate of 0.90 and a miss detection rate of 0.10, while YOLOv10n's detection rate drops to 0.86 and its miss detection rate rises to 0.14, indicating relatively weaker performance. In contrast, the Ours model achieves the highest correct detection rate of 0.93 for the polyp class and reduces the miss detection rate to 0.07, making it the best among the four models. Additionally, it has the lowest false positive rate for classifying background areas as polyps, demonstrating that the proposed method performs better in both reducing false positives and false negatives. Overall, the confusion matrix results are consistent with the conclusions from the previous quantitative metrics, further validating that the proposed method offers higher detection accuracy and better robustness in the polyp detection task.

To further evaluate the detection performance of each model in actual endoscopic scenarios from a visual perspective, relying solely on the aforementioned quantitative metrics and confusion matrices is insufficient to fully reflect bounding box regression accuracy and confidence distribution. Therefore, this paper selects representative polyp image samples and visualizes a comparison of the detection results from YOLOv5n, YOLOv8n, YOLOv10n, YOLOv11n, and the proposed method Ours, as shown in Figure 6. By examining the positions, scales, and confidence levels of the detection boxes output by different models, a more intuitive analysis can be made of the differences in bounding box alignment with the lesion boundaries, suppression of redundant boxes, and robustness to complex polyp shapes across the methods.

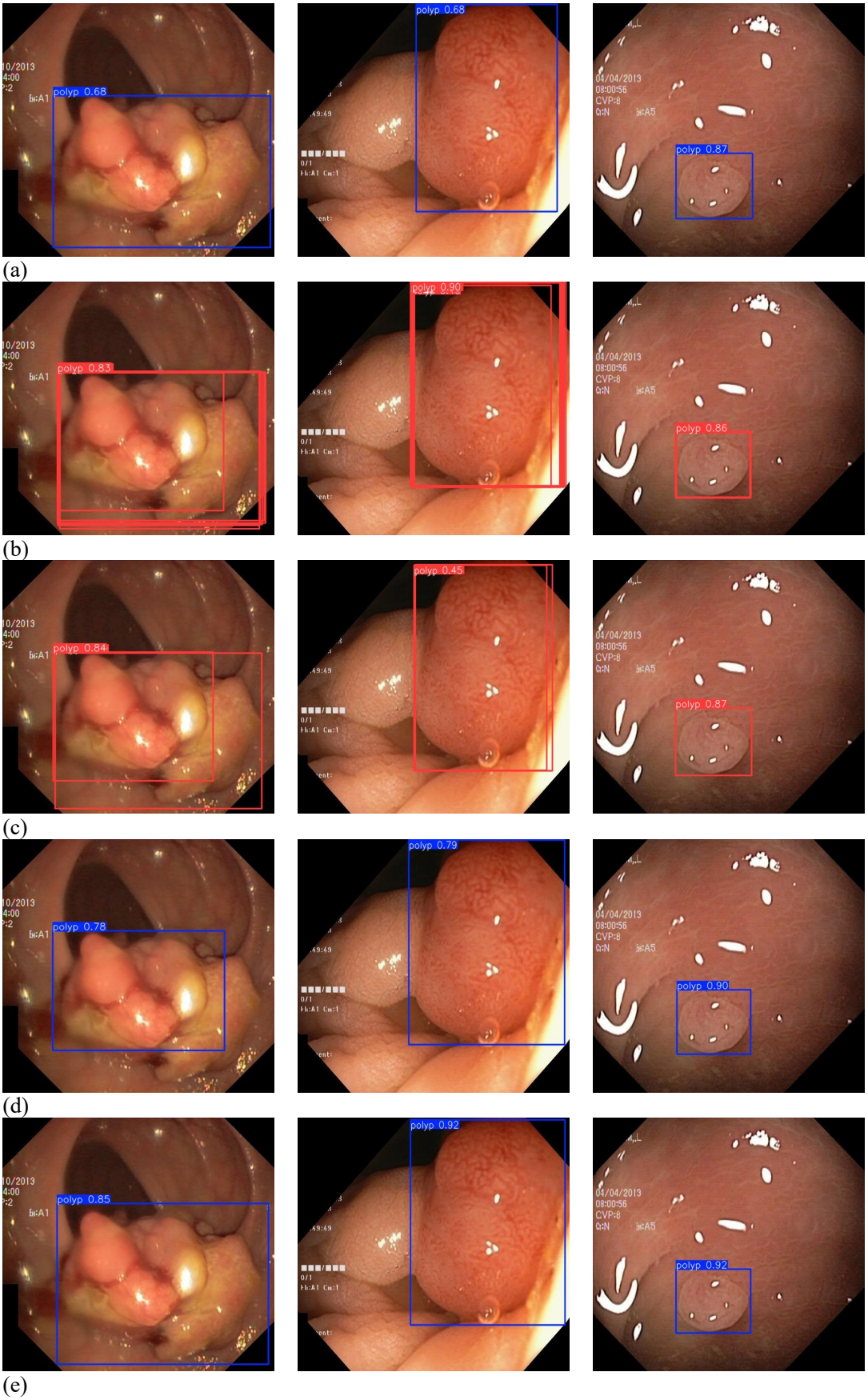
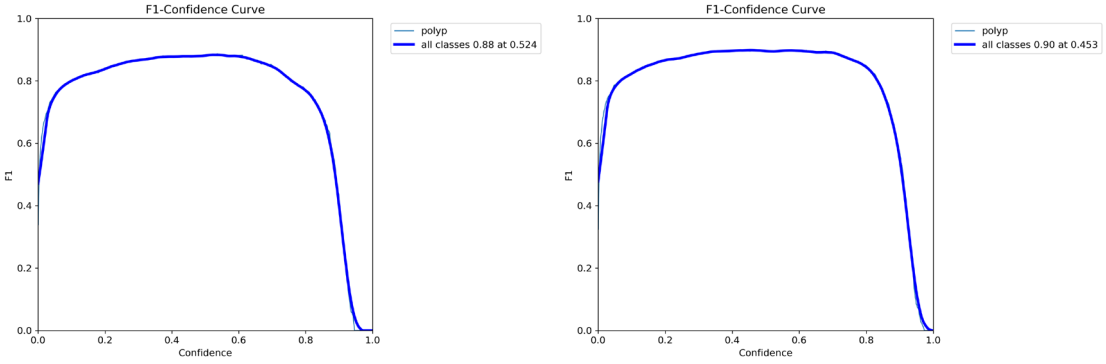


Figure 6. Comparison of Detection Results from Different Models on Gastrointestinal Polyp Images: (a) YOLOv5n; (b) YOLOv8n; (c) YOLOv10n; (d) YOLOv11n; (e) Ours.

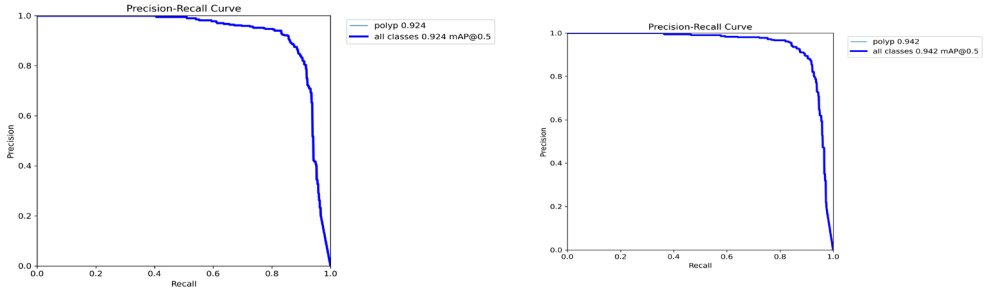
Figure 6 shows that although YOLOv5n is able to locate the lesion area, the predicted bounding box positions are significantly off, and the overall confidence is relatively low, ranging from about 0.68 to 0.87. This indicates that its stability under complex textures or lighting interference is limited. YOLOv8n shows an improvement in confidence for some samples, but multiple highly overlapping candidate boxes can be seen in the image, resulting in obvious redundant confidence boxes, which is not ideal for intuitive interpretation in clinical settings. YOLOv10n's detection results are more volatile, especially in the second image where the confidence is only 0.45, accompanied by box displacement, indicating poor robustness in detecting polyps with weak textures or blurry boundaries. In contrast, YOLOv11n's predicted boxes are more compact and aligned with the target area, with confidence maintained between 0.78 and 0.90, though slight boundary offsets remain for some samples. Notably, Ours method demonstrates higher detection stability and localization accuracy across all images. The predicted confidence significantly improves to between 0.85 and 0.92, and only a single, well-fitted bounding box is output, effectively avoiding the redundant box issue. These visual results are consistent with the quantitative evaluation metrics, further proving the robustness and practical value of the proposed method in complex endoscopic images.



(a) (b)

Figure 7. F1-Confidence Curve: (a) YOLOv5n; (b) Ours.

As shown in Figure 7, both the F1-Confidence curves of YOLOv5n and Ours exhibit a relatively high and stable "platform" structure in the medium confidence range. However, there are differences between the two in terms of peak height and corresponding threshold values: YOLOv5n achieves its highest F1 value of 0.88, with the peak occurring at a confidence of around 0.524, while Ours reaches a peak F1 of 0.90, with the corresponding optimal confidence threshold at approximately 0.453. In comparison, Ours not only outperforms YOLOv5n in terms of peak F1 but also shows a flatter and wider high-platform region. This indicates that Ours maintains a better Precision-Recall balance across a larger confidence range, is less sensitive to threshold selection, and exhibits more stable and robust overall detection performance.



(a) (b)

Figure 8. Precision-Recall Curve: (a) YOLOv5n; (b) Ours.

As shown in Figure 8, both the Precision-Recall curves of YOLOv5n and Ours exhibit the typical characteristic of maintaining high Precision in the high Recall range. However, there are noticeable differences in overall performance between the two. YOLOv5n's PR curve starts to decline significantly after Recall reaches around 0.8, with the corresponding mAP@0.5 of 0.924. In contrast, Ours maintains a smoother and slower decline in the high Recall region, with overall higher Precision, resulting in an improved mAP@0.5 of 0.942. This indicates that Ours can recall more targets while maintaining a higher accuracy, thus achieving a better balance between false positives and false negatives. Overall, Ours demonstrates stronger stability and detection reliability in complex scenarios, further validating its effectiveness in model optimization.

4.2 Ablation Experiment

To systematically evaluate the contribution of each improved module to the model's performance enhancement, this paper constructs multiple sets of control experiments based on YOLOv5n and progressively introduces different structural components for verification. Under the same training strategy, the original model, the model with the CARAFE upsampling module, and the complete scheme with both the ContextGuidedC3 module and CARAFE were tested. The performance of each method on metrics such as Recall, mAP@0.5, and mAP@0.5-0.95 is shown in Table 2.

Table 2. Results of Ablation Experiment.

number	Experiments	Recall/%	mAP@0.5/%	mAP@0.5-0.95/%
1	YOLOv5n	85.7	92.4	67.3
2	YOLOv5n+CARAFE	88.8	94.1	70.1
3	YOLOv5n+ContextGuidedC3+CARAFE	90.3	94.2	69.6

Table 2 presents the ablation experimental results with the progressive introduction of different modules based on YOLOv5n. It can be seen that the baseline model YOLOv5n has Recall, mAP@0.5, and mAP@0.5-0.95 values of 85.7%, 92.4%, and 67.3%, respectively. When only the CARAFE feature upsampling module is added (Experiment 2), Recall increases to 88.8%, mAP@0.5 improves to 94.1%, and mAP@0.5-0.95 rises to 70.1%, indicating that CARAFE helps enhance the recovery of high-level semantic information, thereby improving overall detection accuracy. On this basis, the further introduction of the ContextGuidedC3 module (Experiment 3) leads to a continued increase in Recall to 90.3%, with a slight improvement in mAP@0.5 as well, showing that context-guided feature modeling can further enhance the model's receptive field and discriminative ability for polyp targets. Overall, both improvement modules play a positive role in performance enhancement, validating the effectiveness of the designed network structure.

5. Conclusion

This paper addresses the detection challenges of gastrointestinal polyps in clinical endoscopic scenarios, such as weak textures, blurry boundaries, and complex background noise. We propose an improved lightweight object detection model based on YOLOv5n. By introducing the ContextGuidedC3 module into the network backbone and neck structure, the model's ability to capture multi-scale contextual semantics is effectively enhanced, allowing it to more accurately distinguish polyps from surrounding mucosa in complex backgrounds. Additionally, the CARAFE (Content-Aware Re-Assembly of Feature Elements) operator is introduced during the upsampling stage. By dynamically predicting reassembly kernels, it improves the reconstruction quality of high-resolution features, especially excelling in the recovery of small-scale polyp edges and details.

Experimental results on the polyps dataset demonstrate that, compared to the original YOLOv5n, the proposed method significantly improves key

metrics such as Recall, mAP@0.5, and mAP@0.5-0.95. The model also exhibits better stability and robustness in multi-dimensional evaluations, including confusion matrices, visual detection results, F1-Confidence curves, and PR curves. Moreover, ablation experiments validate the effectiveness of the ContextGuidedC3 module and CARAFE upsampling operator in enhancing model detection performance.

Despite the promising results, this study has certain limitations. The current validation is based on a dataset of limited scale and diversity, which may not fully capture the variability encountered in real-world clinical practice. Furthermore, the model's performance under extreme conditions (e.g., severe occlusion or poor image quality) requires further investigation. Finally, a more comprehensive evaluation of the model's computational efficiency and real-time performance on embedded devices is necessary to assess its practical deployment potential. Future work will focus on addressing these limitations by expanding the dataset with multi-center clinical data and conducting rigorous benchmarks on portable hardware platforms.

Received: November 14, 2025;
Accepted: November 19, 2025

References

[1] Jahn, B., Bundo, M., Arvandi, M., Schaffner, M., Todorovic, J., Sroczynski, G., ... & Siebert, U. (2025). One in three adenomas could be missed by white-light colonoscopy–findings from a systematic review and meta-analysis. BMC gastroenterology, 25(1), 170.

[2] Taghiakbari, M., Mori, Y., & von Renteln, D. (2021). Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. World journal of gastroenterology, 27(47), 8103.

[3] Nie, M. Y., An, X. W., Xing, Y. C., Wang, Z., Wang, Y. Q., & Lü, J. Q. (2024). Artificial intelligence algorithms for real-time detection of colorectal polyps during colonoscopy: a review. American Journal of Cancer Research, 14(11), 5456.

[4] Dougherty, K. E., Melkonian, V. J., & Montenegro, G. A. (2021). Artificial intelligence in polyp detection-where are we and where are we headed?. Artificial Intelligence in Gastrointestinal Endoscopy, 2(6), 211-219.

[5] Rostami, G., Hosseini Berneti, S. H., Habibzadeh, N., & Bazir, M. (2025). From Colon to Uterus: Potential of YOLOv7 for Real-Time Polyp Detection in Hysteroscopy. *InfoScience Trends*, *2*(4), 48-57.

[6] Haider, Z., Azar, A. T., Fati, S. M., & Ibraheem, I. K. (2025, February). Deep Learning and AI for Superior Colorectal Polyp Detection with YOLOv9 Variants. In *2025 8th International Conference on Data Science and Machine Learning Applications (CDMA)* (pp. 150-155). IEEE.

[7] Sun, X., Ma, J., & Li, Y. (2025). Efficient polyp detection algorithm based on deep learning. *Scandinavian Journal of Gastroenterology*, *60*(6), 502-515.

[8] Keshavarz, H., Ansari, Z., Abootalebian, H., Sabet, B., & Momenzadeh, M. (2025). Introducing a Deep Neural Network Model with Practical Implementation for Polyp Detection in Colonoscopy Videos. *Journal of Medical Signals & Sensors*, *15*(6), 17.

[9] Viet, H. D., Nguyen, T. T., Lam, H. N., Nguyen, B. P., Vu, T. Q., Nguyen, H. M., ... & Nguyen, T. T. (2025). Validation of YOLOv8 algorithm in detecting colon polyps in endoscopy videos. *Journal of Medical Artificial Intelligence*, *8*, 35.

[10] Chen, S., Chen, T., Chen, J., Chen, H., Chen, Y., & Zhang, J. (2025, June). YOLO-MF: A Multi-Branch Feature Sensing and Fine-Grained Calibration Network for Gastroenteroscopic Polyp Detection. In *2025 5th International Conference on Computer Graphics, Image and Virtualization (ICCGIV)* (pp. 78-82). IEEE.

[11] Jocher, G., Stoken, A., Chaurasia, A., Borovec, J., Kwon, Y., Michael, K., ... & Thanh Minh,

M. (2021). ultralytics/yolov5: v6. 0-YOLOv5n'Nano'models, Roboflow integration, TensorFlow export, OpenCV DNN support. Zenodo.

[12] Wu, T., Tang, S., Zhang, R., Cao, J., & Zhang, Y. (2020). CGNet: A light-weight context guided network for semantic segmentation. IEEE Transactions on Image Processing, 30, 1169-1179.

[13] Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C. C., & Lin, D. (2019). Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3007-3016).

[14] Silva, Juan, et al. "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer." International journal of computer assisted radiology and surgery 9.2 (2014): 283-293.

[15] Jha, Debesh, et al. "Kvasir-seg: A segmented polyp dataset." International conference on multimedia modeling. Cham: Springer International Publishing, 2019.

[16] Ming Chen, Tingting Chen, Yukang Lou, Yan Li, Jiyang Yu; Remote sensing image detection method based on context-aware mechanism and transformer architecture. AIP Advances 1 July 2025; 15 (7): 075326. https://doi.org/10.1063/5.0283520

[17] Wang, Z., Hua, Z., Wen, Y., Zhang, S., Xu, X., & Song, H. (2024). E-YOLO: Recognition of estrus cow based on improved YOLOv8n model. Expert Systems with Applications, 238, 122212.

[18] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., & Han, J. (2024). Yolov10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems, 37, 107984-108011.

[19] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725.