2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Meta-learning and Residual Block Enhanced YOLO for Accurate Detection of Gastrointestinal Pathology Lesions

Xiangyu Xue[1] Yakun Wang[2]

[1] School of Engineering Medicine, Beihang University, Beijing, 100191, China

Abstract. Early identification and accurate diagnosis of gastrointestinal diseases, particularly gastric cancer, are paramount for enhancing patient survival rates and treatment outcomes. However, diagnosing these diseases can be challenging, especially when symptoms are mild or absent. Endoscopy, a standard diagnostic tool, relies heavily on the endoscopist's expertise. Integrating artificial intelligence (AD) with endoscopic imaging has the potential to assist in diagnosis, reduce missed cases, and expedite timely treatment. Previous studies have focused on refining disease classification and and improving diagnostic accuracy, often neglecting issues of data reliability and imbalance. This study proposes a novel approach utilizing model-agnostic meta-learning (MAML) strategies to address the challenges posed by sparse and imbalanced medical image data. We introduce the YOLO-MR model, which incorporates meta-recognition mechanisms and residual blocks into the YOLO framework. Experimental results demonstrate that the traditional YOLO model achieves an average precision (mAP) of only 41.7% on imbalanced data, highlighting the negative impact of data imbalance. Traditional data augmentation techniques improve the mAP to 65.2%. whereas our proposed YOLO-MR model achieves an impressive mAP of 96%, representing a significant improvement of 54.3% over the traditional model. This enhancement effectively reduces the diagnostic accuracy gap between different disease categories and mitigates the issue of data imbalance. Furthermore, our research validates the strong potential of advanced techniques such as MAML and residual blocks in resource-limited medical image recognition tasks. These findings provide valuable insights into addressing the challenges of limited and imbalanced medical data in the healthcare field.

Keywords: Gastrointestinal endoscopy, Medical image, Meta-learning, YOLO, Lesions

1 Introduction

Image-based diagnostics are crucial in medicine, with gastrointestinal (GI) imaging playing a key role in evaluating the digestive system [1]. GI endoscopy is the primary method for directly detecting anomalies like tumors, ulcers, and bleeding, complemented by techniques such as X-rays, CT, and MRI. Early and accurate diagnosis of GI conditions, such as gastric cancer, is vital for patient survival but often challenging due to subtle early symptoms [2, 3].

A major hurdle in developing automated detection systems using computer vision (e.g., object detection, classification) is the difficulty in obtaining large, well-structured medical datasets, particularly endoscopic images [4-6]. Class imbalance is a common issue, where certain conditions are underrepresented. Data augmentation techniques (e.g., oversampling, SMOTE, geometric transformations) are frequently used to mitigate this [7-9]. However, while methods like sampling [7, 8], selective transformations [9], transfer learning [10], and transformer-based models [11] have shown promise in improving detection on imbalanced medical data, traditional augmentation involving data alteration can raise reliability concerns critical in the medical domain.

Meta-learning offers a compelling approach to address data scarcity and imbalance by enabling models to adapt quickly from limited examples, potentially without altering the original data [13, 14]. Existing methods like Meta-SSD [13] and Meta-YOLO [14] demonstrate its potential in object detection. Inspired by this, and building upon advancements like EAD-YOLO [15], we propose YOLO-MR. This model integrates meta-learning, specifically the model-agnostic meta-learning (MAML) algorithm, with the YOLO object detection framework. We further incorporate Residual Blocks to enhance feature extraction for challenging lesion identification tasks. Meta-learning is particularly suitable due to its effectiveness with limited data and its ability to learn from domain-specific medical data characteristics, bypassing potential issues with mismatched pre-training data.

The contributions of this study are summarized as follows:

- Development of YOLO-MR, an automated object detection algorithm for identifying gastric lesions (cancer, adenoma, ulcer).
- Integration of meta-learning for optimized weight initialization and Residual Blocks into the YOLO architecture, improving lesion identification performance over existing methods.
- Experimental investigation into the impact of data imbalance and validation of the proposed YOLO-MR's effectiveness for robust real-time lesion detection.

2 Basic knowledge of proposed baseline

This section provides explanations of important theories behind the proposed algorithm. First, it describes the object detection and YOLO model, which serves as the basic framework. It then elaborates on meta-learning and the MAML algorithm, which are the key concepts driving the algorithm. Finally, we discuss residual blocks.

2.1 Object detection and YOLO

Object detection is a computer vision problem that involves the simultaneous identification of the location and class of objects in images or videos [16]. It has various applications in fields, such as autonomous driving, medical image analysis, and security. Several algorithms have been developed for object detection, including region-based convolutional neural network (R-CNN) [17], Fast R-CNN [18], Faster R-CNN [19], you only looking once (YOLO) [20], and single-shot multibox detector (SSD) [21].

YOLO has evolved through multiple versions [22-28] and is structured with three components for object detection: backbone, neck, and head. The backbone is responsible for extracting essential features from the input image and is typically composed of convolutional neural networks [29]. It processes images of various scales and resolutions to generate feature maps used to capture object shapes and visual features. Second, the neck collected and combined feature maps with different resolutions and scales from the backbone to create a feature pyramid. This feature pyramid allows detection of objects of all sizes, from small to large. Third, the head is where the final detection results are the output. It comprises output layers for class prediction and bounding-box regression. Class predictions indicate the probability of an object's class within a grid cell, whereas bounding box predictions provide information about the object's location and size. Multiple anchor boxes are used in each grid cell to predict multiple bounding boxes, thereby enabling adaptability to various sizes and aspect ratios.

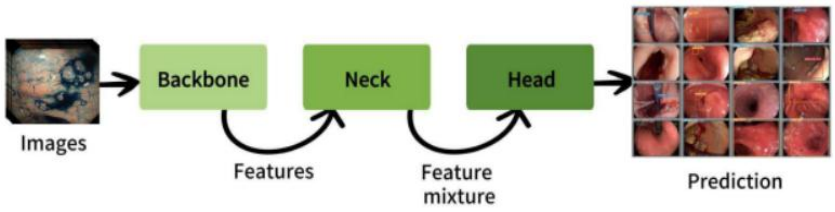

Fig.1. YOLO structure flow.

Figure 1 illustrates the overall structure of YOLO, which combines three components: the backbone, neck, and head. This structure provides excellent performance and speed for real-time objects.

2.2 Meta-learning and model agonistic meta learning

Meta-learning [30] is a method that enhances the ability of a machine learning model to adapt quickly to new tasks. It can be broadly categorized into three main perspectives: research on adjusting a model's hyperparameters to achieve optimal performance; exploring model structures or initial parameters that can quickly adapt to new tasks using knowledge and experience from various tasks or domains; and utilizing information on relationships and similarities between datasets to improve generalization performance. One meta-learning technique that can be applied regardless of the model is MAML [31].

Figure 2 shows a structure of the MAML. MAML focuses on training an algorithm to quickly modify a deep learning model, addressing the problem of finding model structures or initial parameters that can rapidly adapt to new tasks. This approach is effective even with a small amount of new data, and is suitable for novice-level learning. MAML fine-tunes initial model parameters using example data and the adjusted initial parameters are used for adaptation to different tasks or domains. This process was repeated several times, with each iteration aimed at improving the generalization

of the initial parameters across various tasks. Therefore, the goal of MAML is to create models that can quickly learn and adapt to various tasks by adjusting initial parameters. It is a versatile meta-learning algorithm that can be applied to various fields.

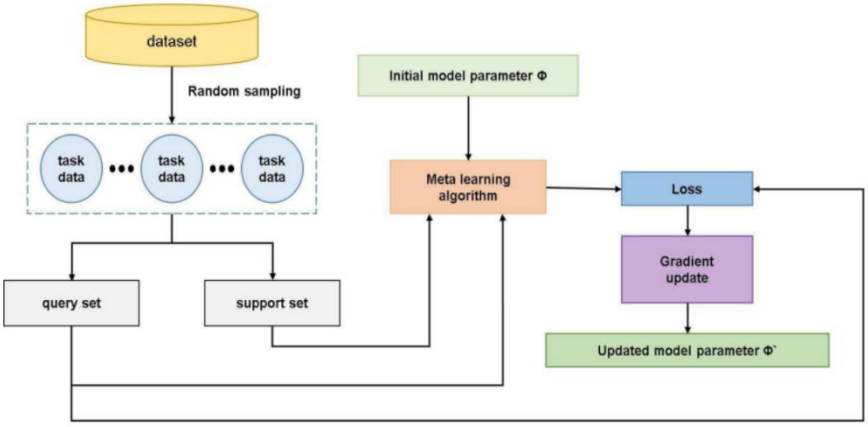

Fig.2. MAML structure.

2.3 Residual block

The residual block [32] is a crucial component of deep learning networks that helps mitigate the vanishing gradient problem and improves learning performance while increasing the depth of the network. Residual blocks decompose the input data into their original values and residuals (the difference between the input and output). This was achieved by adding a residual connection to the output of the previous layer, allowing the neural network to obtain additional learned representations of the input data. These residual connections facilitate the smooth propagation of gradients throughout the neural network, thereby alleviating the vanishing gradient problem that can occur as the network depth increases.

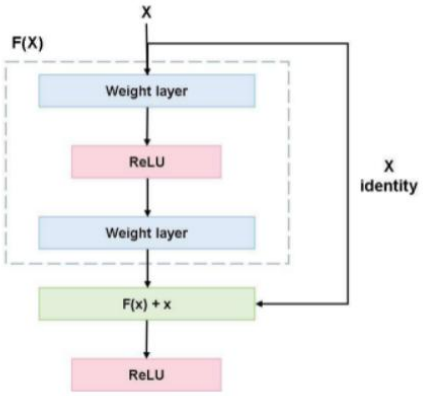

Fig.3. Residual block structure.

As shown in Figure 3, x represents the input, and F(x) represents the transformation function for the input (e.g., convolutional layers and feedforward neural networks). F(x) transforms input x to create a new representation y, and the residual connection combines the transformed y with the original input x to obtain the final output. This process can be expressed by the following equation:

$$y=F(x)+x \tag{1}$$

where y is the final output, F(x)is the transformed representation, and x is the original input. The addition operation combines the transformed representation with the original input, yielding the final output.

3 Proposed method

First,we investigate the impact of class imbalances resulting from differences in data quantity on the accuracy of endoscopic image classification. To address this issue, we performed the following experiments. In all the experiments, the training and testing data were set at a ratio of 9:1, and each experiment was run for 1,000 epochs.

3.1 Correlation between data imbalance and accuracy

Experiments were conducted under the assumption that all classes had an equal number of data samples. Considering that the ulcer class contained approximately 4,000 samples, we constructed an experimental dataset based on this class. In other words, we extract 4,000 samples for each class or the experiments, which was the same as the number of samples in the ulcer class. Additionally, we conducted experiments using a smaller dataset comprising 400 samples, which accounted for 10% of the dataset.

Table 1. Test accuracy for balanced data using 400 samples.

	Num(train/test)	**p**	**r**	**mAP**
Cancer	360/40	0.897	0.9	0.933
Ulcer	360/40	0.927	0.905	0.97
Adenoma	360/40	0.744	0.756	0.814
all	1080/120	0.856	0.854	0.906

Table 2. Test accuracy for balanced data using 4000 samples.

	Num(train/test)	p	r	mAP
Cancer	3600/400	0.776	0.723	0.751
Ulcer	3600/400	0.653	0.558	0.61
Adenoma	3600/400	0.645	0.695	0.661
all	10800/1200	0.691	0.695	0.674

First, we conducted experiments using 400 and 4,000 balanced data samples, respectively. Table 2 presents the results of training with 4,000 balanced data samples for each class using pre-trained weights, showing that the cancer class achieves a relatively high accuracy of approximately 0.75 compared to the other classes, with ulcer at 0.61 and adenoma at 0.661. This suggests that while the ulcer and adenoma classes exhibit slightly lower accuracy, the overall difference between the classes is not substantial. However, referring to Table 1, when the data quantity is limited to 400 samples, the accuracy is significantly high at approximately 0.9. Particularly, the ulcer class demonstrates higher accuracy than cancer. These results are likely attributed to overfitting and insufficient test data, highlighting the challenge of distinguishing between cancer and ulcer. Secondly, we conducted experiments considering the imbalances in the number of data samples for each class.

Table 3. 1/10th of imbalanced data test accuracy.

	Num(train/test)	p	r	mAP
Cancer	728/104	0.539	0.51	0.473
Ulcer	308/44	0.516	0.257	0.314
Adenoma	951/136	0.475	0.499	0.463
all	1987/284	0.51	0.422	0.417

Table 4. Imbalanced data test accuracy.

	Num(train/test)	p	r	mAP
Cancer	7289/1043	0.785	0.769	0.81
Ulcer	3090/442	0.68	0.57	0.62
Adenoma	9522/1361	0.738	0.683	0.714
all	19901/2846	0.734	0.674	0.715

The following presents the results of experiments conducted using an imbalanced dataset. According to Table 3, the accuracies for the cancer, ulcer, and adenoma classes are 0.473, 0.314, and 0.463, respectively, with an average accuracy of approximately 0.417. These results demonstrate a decrease in accuracy for classes with relatively fewer data samples. Table 4 illustrates the results of experiments conducted by increasing the dataset size. The accuracy for the cancer class improved slightly to 0.81. The ulcer class still recorded a low accuracy of 0.62, while the adenoma class showed a slight increase to approximately 0.714. Therefore, the overall average accuracy increased by approximately 29.8% with the increase in dataset size.

However,low accuracy persists in cases with relatively few data samples, such as the ulcer class. This confirms the difficulty of distinguishing objects belonging to specific classes during the learning process. However, a more fundamental problem lies in the limited number of data samples and the learning approach used.

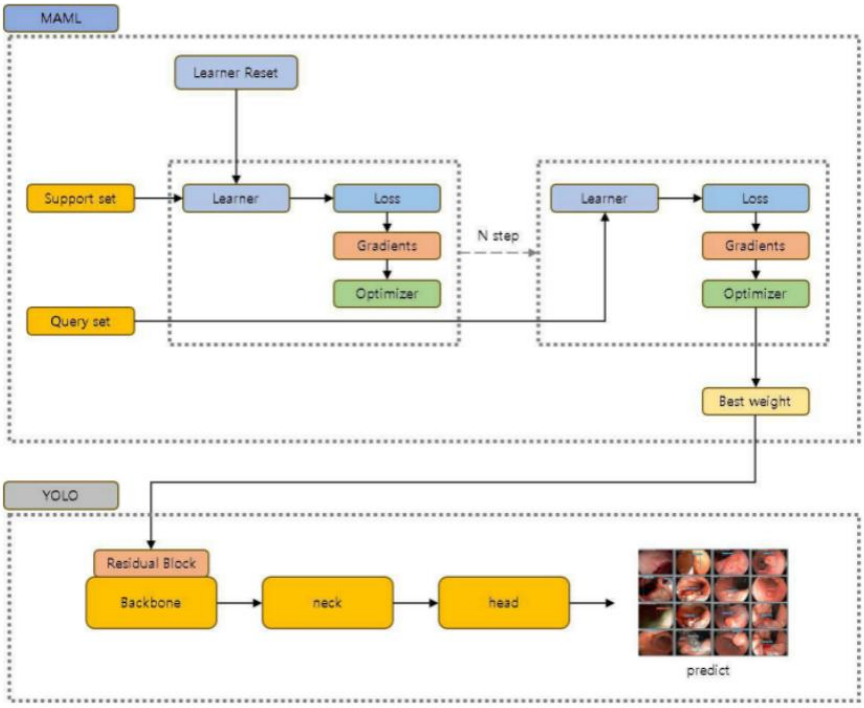

Fig.4. Proposed model structure.

3.2 Proposed structure

In this paper, we propose a method based on the YOLOv7 model to address object detection considering the imbalanced nature and class characteristics of medical data. We utilize MAML (Meta-learning Adaptive Model) to learn the optimal weights and

apply them to a YOLO model with residual blocks. The model trained using this approach is named YOLO-MR (YOLO with meta-learning and residual blocks), and its structure is illustrated in Figure 4.

Model-agnostic meta-learning module

Model-agnostic meta-learning is a meta-learning algorithm utilized to rapidly adapt model parameters to various tasks. On the other hand, residual blocks act as a network structure, establishing a direct pathway between the input and output, thereby mitigating the issue of gradient vanishing that can arise in complex networks and enabling deeper network training.

The primary objective of our proposed YOLO-MR model is to combine the advantages of MAML and residual blocks to achieve high-performance object detection. MAML leverages a network architecture that combines a convolution-based backbone with the YOLO object detection head and utilizes gradient-based optimization algorithms to determine optimal weights. MAML consists of two main steps. In the first step, the initial model parameters are updated using the support set data, while in the second step, the performance of the updated initial parameters is evaluated and optimized using the query set data. By iteratively performing these steps, the initial parameter values are finely adjusted, resulting in model weights capable of adapting to diverse tasks. Consequently, the derived optimal weights are utilized in object detection tasks within the YOLO model, enabling MAML to maintain high performance while being adaptable to a variety of image object detection tasks.

YOLO with residual block module

In the YOLO framework, residual blocks play a crucial role in facilitating seamless information transfer from ConvModules to subsequent layers. ConvModules consist of a combination of layers, including convolution, batch normalization, and activation functions, which perform transformations on the input data.

Importantly,even after the post-ConvModule processing, the input data are directly transmitted through a skip connection, establishing a residual association between the input and output data. Within the residual block, an additive summation occurs between the input and output data, resulting in the generation of a residual value between the output and input of the ConvModule. This residual value is utilized during the learning process and can contribute to improved accuracy. By keeping the residual value concise and compact compared to the previous pathway, information is transmitted without loss, enabling deeper learning within the network. Consequently, residual blocks work collaboratively wit ConvModules to enhance the accuracy of object detection tasks.

Therefore,an algorithmic structure that leverages MAML and residual blocks has been proposed, ensuring high performance even in the presence of data imbalances.

4 Experimental results

The experiments compared the conventional YOLO model with a YOLO model that uses data augmentation techniques and a YOLO model that applies both MAML and residual blocks (YOLO-MR).

4.1 Data and experimental environment

The dataset used in this study comprised endoscopy data collected from patients who underwent upper gastrointestinal endoscopy at Gachon University Gil Medical Center's outpatient and inpatient departments from 2008 to October 2022. The data included patients'medical records, excluding cases with unclear diagnoses, and consisted of patient records stored in EMR and image databases. The dataset (IRB Number: GBIRB2021-383) included 61,734 cases classified into four classes, as shown in Table 5. The dataset exhibits variations in the number of data samples per class, with the ulcer class having notably fewer data samples compared to the other classes. As demonstrated in Section III-A, such dataset imbalances were shown to lead to accuracy degradation. This experiment is conducted with a dataset about one-tenth the size of all datasets except for the normal class. The details of the equipment utilized for this study are provided in Table 6.

Table 5. Dataset.

Class	Count
Cancer	10414
Ulcer	4415
Adenoma	13603
Normal	33302

Table 6. Experimental environment.

CPU	AMD Ryzen Threadripper 3960X 24-Core Processor 3.79 GHz
GPU	NVIDIA Geforce RTX 2080 Ti
RAM	64GB
OS	Windows11,64bit OS

4.2 YOLO

Figure 5 presents the prediction results obtained using the conventional YOLO model.

Fig.5. YOLO results (L:GroundTruth, R:Prediction).

As shown in Table 7, the average precision (AP) for the cancer, ulcer, and adenoma classes were 0.473, 0.314, and 0.463, respectively, resulting in a mean average precision (mAP) of 0.417. This experiment was conducted using only 10% of the dataset, which led to issues related to data imbalance and a limited amount of data. As observed in Table 7, there was a significant disparity in accuracy between each class, and the overall accuracy was also low. Comparing the average accuracy obtained in this experiment with the 0.7 accuracy achieved when training on the full dataset, the results obtained in this experiment were approximately half. This suggests that the model struggled to learn properly due to the limited amount of data and resulting data imbalance.

Table 7. Data test results using the YOLO model.

	Num(train/test)	**p**	**r**	**mAP**
Cancer	728/104	0.539	0.51	0.473
Ulcer	308/44	0.516	0.257	0.314
Adenoma	951/136	0.475	0.499	0.463
all	1987/284	0.51	0.422	0.417

4.3 Data augmentation YOLO

Data augmentation involves transforming existing data in various ways to expand a dataset. To compare it with the proposed model, data augmentation was used to increase the amount of data during training and to compare it with the previous YOLO model. Among the various data augmentation methods, a basic approach was employed to perform data augmentation on existing image data within a range that did not significantly distort the data. The data augmentation methods applied include illumination changes, contrast adjustments, saturation changes, and Gaussian blurring, as shown in Figure 6.

Fig.6. Data augmentation example.

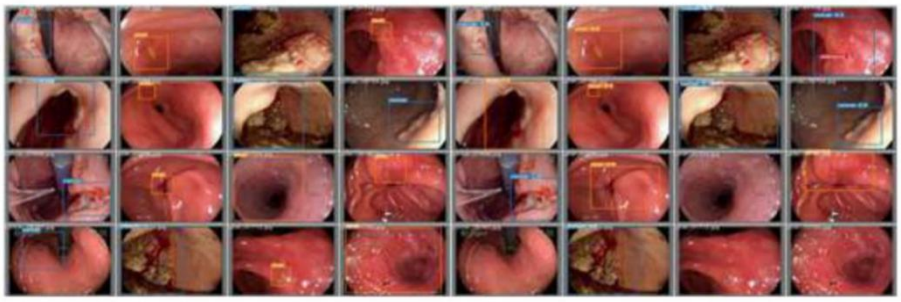

Fig.7. Presents the prediction results corresponding to data augmentation.

In this study,a dataset was generated based on 10% of the dataset for data augmentation.However,the augmentation techniques used increased the number of original data samples to approximately 19,870. Data augmentation techniques that minimally altered the data were employed to maintain the reliability of the medical data. As shown in Table 8, the average precision (AP) for the cancer, ulcer, and adenoma classes were 0.697, 0.602, and 0.657, respectively. The mean average precision (mAP) across these classes wa approximately 0.652, representing a significant improvement of approximately 23.5% compared to the YOLO baseline results. This confirms that, as previously observed, increasing the amount of data has an impact on performance.

Table 8. Test results of YOLO model with data augmentation.

	Num(train/test)	**p**	**r**	**mAP**
Cancer	728(7280)/104	0.739	0.641	0.697
Ulcer	308(3080)/44	0.646	0.579	0.602
Adenoma	951(9510)/136	0.652	0.666	0.657
all	1987(19870)/284	0.679	0.692	0.652

However,it is worth noting that these results are sill approximately 6.3% lower than those obtained from training with the original dataset. Additionally, when analyzing the

differences in accuracy between classes, it can be observed that the issue of data imbalance persists, although to a lesser extent than in the initial experiments.

4.4 Results of the proposed model

Figure 8 presents the prediction results using YOLO-MR.

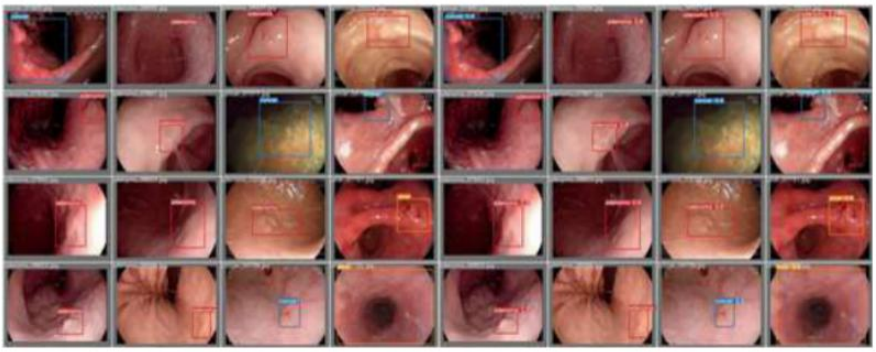

Fig.8. YOLO-MR results (L:GroundTruth,R:Prediction).

In the final experiment,the Meta-Learning Adaptive Model (MAML) was employed to learn the optimal weights, which were subsequently set as the initial weights for the YOLO model. Following that, the YOLO-MR architecture with residual blocks was applied. According to Table 9,the average precision (AP) for the cancer, ulcer, and adenoma classes were 0.984, 0.919, and 0.976, respectively. The overall mean average precision (mAP) was approximately 0.96, signifying significant improvements of approximately 54.3% and 30.8% compared to the previous YOLO model and the experiment utilizing data augmentation techniques, respectively. These findings are particularly noteworthy considering that the amount of data utilized was only one-tenth of that used in the data augmentation experiment. This indicates a substantial contribution of MAML and residual blocks to the enhancement of performance.

Table 9. Test results learned on YOLO-MR model.

	Num(train/test)	**p**	**r**	**mAP**
Cancer	364/364/104	0.947	0.974	0.984
Ulcer	154/154/44	0.928	0.878	0.919
Adenoma	475/475/136	0.936	0.951	0.976
all	993/993/284	0.937	0.934	0.96

Moreover,the disparity between the highest and lowest accuracies was approximately 0.07, significantly lower than in previous experiments. This implies that YOLO-MR contributed to a reduction in performance disparities among classes, thereby alleviating the issue of data imbalance.

Table 10. Below provides an overview of the average precision for each model across all the experiments.

Model	cancer	ulcer	adenoma	mAP
YOLO	0.473	0.314	0.463	0.417
Data augmentation YOLO	0.697	0.602	0.657	0.652
Our YOLO-MR	0.984	0.919	0.976	0.96

Table 10 presents a summary of the previous three experiments. The results obtained from the YOLO baseline experiment were unsatisfactory, and the method of increasing the dataset by tenfold through data augmentation did not achieve comparable performance for our model. Our approach exhibited the best performance despite being trained on a smaller dataset compared to data augmentation. Furthermore, we successfully minimized the performance gap between classes.

Figure 9 shows the precision-recall curves of all evaluated models. The larger the area under the curve, the better the performance. The proposed method showed the highest precision and reproducibility curves compared to the existing YOLO model and the data-enhanced YOLO model, which shows the superior performance.

Figure 10 depicts the training and validation loss curves for the entire experiment. Upon closer examination, the graph exhibits periods of stagnation. but overall, it demonstrates a decreasing trend. Furthermore, it can be observed that the proposed model exhibits a small gap between the training and validation loss.

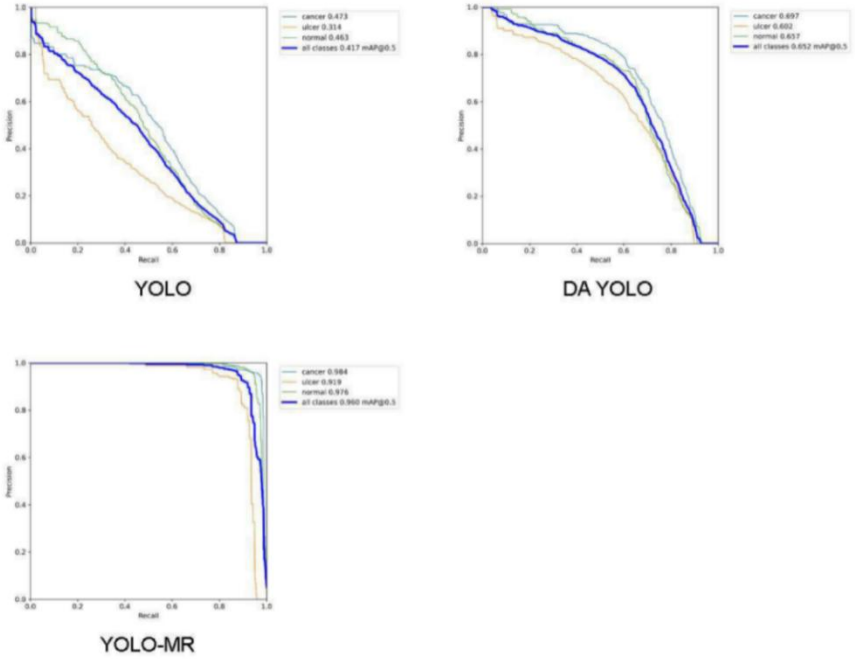

Fig.9. Precision-recall curves of YOLO baseline and data augmentation YOLO,YOLO-MR.

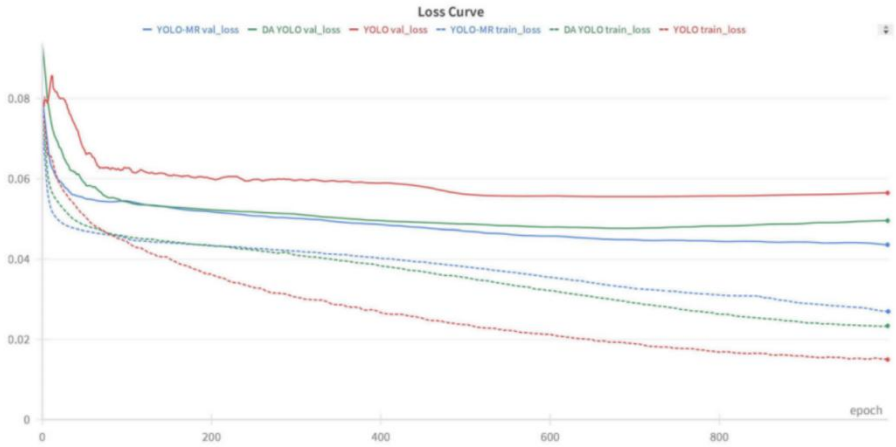

Fig.10. Train loss and val loss graph.

Table 11. Overview of the accuracy for each model across all the experiments and related research indicators.

Model	cancer	ulcer	adenoma	Accurary or mAP
InvNorm				0.842
ASSD-GPNet				0.942
DenseNet121				0.9868
Mask R-CNN+BiFPN				0.9333
Original YOLOv7	0.473	0.314	0.463	0.417
Data augmentation YOLO	0.697	0.602	0.657	0.652
Our YOLO-MR	0.984	0.919	0.976	0.96

To summarize once again, when dealing with imbalanced data, the conventional YOLO model showed a relatively low mAP of 0.417. However, by augmenting the data and increasing the dataset size, the mAP improved to 0.625, representing a 30.8% increase. In contrast, the proposed YOLO-MR approach achieved a higher mAP of 0.96, indicating a 54.3% increase in accuracy compared to the traditional YOLO model. When comparing the differences between classes, the YOLO model exhibited the largest difference of approximately 0.16, while the data-augmented YOLO model showed a difference of approximately 0.09. The proposed YOLO-MR model significantly reduced the class imbalance to approximately 0.07.

These results emphasize the effectiveness of techniques such as meta-learning and residual blocks in addressing data imbalance in image recognition tasks. This holds practical potential for addressing imbalanced data in the field of medical image analysis,

and it is expected that these methodologies can be applied to various problem-solving scenarios beyond the medical domain.

Furthermore,when comparing the performance of recent research papers on object detection in gastrointestinal endoscopy, a study conducted by using the InVNorm model [33] achieved an mAP of 84.2% by applying interpretable style normalization, without compromising the reliability of medical data augmentation. Another study proposed the ASSD-GPNet model [34], which achieved an mAP of 94.2% for gastrointestinal endoscopy videos and 76.9% for the Pascal VOC dataset. This model demonstrated outstanding performance by generating intricate feature maps that focus on specific information, aiding in the detection of small polyps. A study introducing the DenseNet121 model, used for histopathological image analysis achieved a top accuracy of 98.68% and an AUC of 98.58%. Lastly, a study proposed the Mask R-CNN+BiFPN model, which combined the object detection method with endoscopic images, improved feature fusion, and enhanced early detection of gastrointestinal lesions, achieving an mAP of 93.33%. Our model exhibited high accuracy and achieved a commendable mAP of 96%, compared to recent research. Models based on the SSD model, which utilized refined map blocks (RMB) and attention cascades to improve accuracy, outperformed our study.

5 Conclusion

In this paper, we emphasize the importance of early detection and accurate diagnosis of gastrointestinal diseases, including gastric cancer, through gastrointestinal endoscopy. However, the accuracy of disease identification in this field varies depending on the endoscopist, and there is a possibility of missed diagnoses. To address these challenges, the application of artificial intelligence as an assistive tool has shown promising results in reducing missed diagnoses and improving patient survival rates by enabling early treatment. However, previous studies have mainly focused on disease classification and improving classification accuracy, overlooking the practical difficulties in medical data collection and the handling of imbalanced datasets.

In this study, we implemented-meta learning using the MAML algorithm and proposed the YOLO-MR model by combining the YOLO object detection algorithm with Residual Blocks. The YOLO-MR model significantly improved the object detection accuracy compared to the baseline YOLO model. The object detection mAP of the baseline YOLO model was relatively low,with detection AP for cancerous tumors, ulcers, and adenomatous tumors being 0.473, 0.314, and 0.463, respectively, resulting in an mAP of 0.417. By augmenting the dataset and increasing its size by 10 times, the accuracy improved, with detection AP for cancerous tumors, ulcers, and adenomatous tumors being 0.697, 0.602, and 0.657, respectively, resulting in an mAP of 0.652. However, these results were still lower compared to training with the original data without data augmentation. The proposed YOLO-MR method, utilizing MAML and residual blocks, achieved detection AP for cancerous tumors, ulcers, and adenomatous tumors of 0.984, 0.919, and 0.976, respectively, resulting in an mAP of 0.96. Furthermore, the proposed approach significantly reduced the accuracy gap between different classes and contributed to addressing the issue of data imbalance.

In summary,when dealing with imbalanced data, using only the conventional YOLO model leads to relatively low mAP. Data augmentation can greatly improve mAP, but the YOLO-MR method surpasses such improvements, achieving 54.3% increase in mAP. In particular, it successfully reduces the accuracy gap between classes and effectively addresses the issue of data imbalance. These results highlight the effectiveness of techniques such as meta-learning and residual blocks in addressing the challenge of data imbalance in image recognition tasks and emphasize the performance of medical image object detection. However, both the data augmentation technique and the proposed YOLO-MR model have the limitation of long execution times in the data augmentation module and meta-learning module. Nevertheless, once the best model is trained using the YOLO model, quick test results can be obtained.

Therefore, training with imbalanced class datasets using the proposed model can achieve good performance and comparable performance to state-of-the-art research. Additionally, it would be worthwhile to explore the applicability of these methods in other fields and rare diseases.

Disclosure of Interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. M. Moor et al. "Foundation models for generalist medical artificial intelligence," vol. 616, no. 7956, pp. 259-265, 2023.
2. J. P. Griffin-Sobel, "Gastrointestinal cancers: screening and early detection," in Seminars in Oncology Nursing, 2017, vol. 33, no. 2, pp. 165-171: Elsevier.
3. X. Li, J. Lu, J. Zhou, W. Liu, K. J. C. A. Zhang, and V. Worlds, "Multi — temporal scale aggregation refinement graph convolutional for skeleton — based action recognition," vol. 35, no. 1, p. e2221, 2024.
4. M. Khushi et al., "A comparative performance analysis of data resampling methods on imbalance medical data," vol. 9, pp. 109960-109975, 2021.
5. G. Yue et al., "Automated endoscopic image classification via deep neural with class imbalance loss," vol. 72, pp. 1-11, 2023.
6. S. G. Ali et al., "Egdnet: an efficient glomerular detection network for multiple anomalous pathological feature in glomerulonephritis," pp. 1-18: 2024.
7. A. Bria, C. Marrocco, F. J. C. i. b. Tortorella, and medicine, "Addressing class imbalance in deep learning for small lesion detection on medical images," vol. 120, p. 103735, 2020.
8. F. Deeba, S. K. Mohammed, F. M. Bui, and K. A. Wahid, "Learning from imbalanced data: A comprehensive comparison of classifier performance for bleeding detection in endoscopic video," in 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016, pp. 1006-1009: IEEE.
9. H. Wang, Q. Wang, F. Yang, W. Zhang, and W. J. a. p. a. Zuo, "Data augmentation for object detection via progressive and selective instance-switching," 2019.
10. X. Liz, H. Zhang, X. Zhang, H. Liu, and G. Xie, "Exploring transfer learning for gastrointestinal bleeding detection on small-size imbalanced endoscopy images," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (ENOC), 2017, pp. 1994-1997: IEEE.

11. L. Bai, L. Wang, T. Chen, Y. Zhao, and H. J. E. Ren, "Transformer-based disease identification for small-scale imbalanced capsule endoscopy dataset," vol. 11, no. 17, p. 2747, 2022.
12. X. Thu et al., "TMSDNet: Transformer with multi — scale dense network for single and multi — view 3D reconstruction," vol. 35, no. 1, p. e2201, 2024.
13. K. Fu et al., "Meta-SSD: Towards fast adaptation for few-shot object detection with meta-learning," vol. 7, pp. 77597-77606, 2019.
14. X. Ren, W. Zhang, M. Wu, C. Li, and X. J. A. S. Wang, "Meta-Yolo: Meta-learning for few-shot traffic sign detection via decoupling dependencies," vol. 12, no. 11, p. 5543, 2022.
15. Z. Yuan, J. Ye, C. Qian, and X. Liz, "EAD-YOLO: Improved YOLOv5 for Endoscopic Artefact Detection," in 2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI), 2023, pp. 151-158: IEEE.
16. Y. Chen, B. Chen, W. Xian, J. Wang, Y. Huang, and M. J. T. V. C. Chen, "LGFDR: local and global feature denoising reconstruction for unsupervised anomaly detection," pp. 1-14, 2024.
17. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.
18. R. J. a. p. a. Girshick, "Fast r-cnn," 2015.
19. S. Ren, K. He, R, Girshick, J. J. I. t. o. p. a. Sun, and m. intelligence, "Faster R-CNN: Towards real-time object detection with region proposal vol. 39, no. 6, pp. 1137-1149, 2016.
20. J. Redmon, "You only look once: Unified: real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
21. W. Liu et al., "Ssd: Single shot multibox detector," in Computer Vision — ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part 114, 2016, pp. 21-37: Springer.
22. J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263-7271.
23. J. J. a. p. a. Redmon, "Yolov3: An incremental improvement," 2018.
24. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. J. a. p. a. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.
25. G. Zheng, L. Songtao, W. Feng, L. Zeming, and S. J. a. p. a. Jim, "YOLOX: Exceeding YOLO series in 2021," 2021.
26. S. Xu et al., "PP-YOLOE: An evolved version of YOLO," 2022.
27. C. Li et al., "YOLOv6: A single-stage object detection framework for industrial applications," 2022.
28. C-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-aft for real-time object detectors," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7464-7475.
29. Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," vol. 1, no. 4, pp. 541-551, 1989.
30. Z. M. Baum, Y. Hu, and D. C. J. I. t. o. m. i. Barratt, "Meta-learning initializations for interactive medical image registration," vol. 42, no. 3, pp. 823-833, 2022.
31. C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in International conference on machine learning, 2017, pp. 1126-1135: PMLR.
32. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
33. W. Fan, Y. Yang, K. Qiu, S. Wang, and Y. J. a. p. a. Guo, "InvNorm: Domain generalization for object detection in gastrointestinal endoscopy," 2022.

34. D. Mushtaq, T. M. Madni, U. I . Janjua, F. Anwar, A. J. I. J. o. I. S. Kakakhail, and Technology, "An automatic gastric polyp detection technique using deep learning," vol. 33, no. 3, pp. 866-880, 2023.