Hindawi
Computational and Mathematical Methods in Medicine
Volume 2023, Article ID 5147399, 11 pages
https://doi.org/10.1155/2023/5147399

Research Article

Pathological Detection of Micro and Fuzzy Gastric Cancer Cells Based on Deep Learning

Qiuxia Guo,[1] Weiwei Yu,[2] Shasha Song,[3] Wenlin Wang,[4] Yufei Xie,[5] Lihua Huang,[1] Jing Wang,[1] Ying Jia,[6] and Sen Wang[7]

[1]*Wuhan Fourth Hospital, 430033 Wuhan, China*
[2]*Yantai Yuhuangding Hospital, 264001 Yantai, China*
[3]*College of Pharmacy, Shenzhen Technology University, 518118 Shenzhen, China*
[4]*Sino-German College of Intelligent Manufacturing, Shenzhen Technology University, 518118 Shenzhen, China*
[5]*School of Automation, Wuhan University of Technology, 430070 Wuhan, China*
[6]*Daqing Longnan Hospital, 163712 Daqing, China*
[7]*Shenzhen Gengfeng Technology Co., Ltd, Shenzhen, 518001 Guangdong, China*

Correspondence should be addressed to Shasha Song; songshasha@sztu.edu.cn and Wenlin Wang; wlwang0618@126.com

Received 3 September 2022; Revised 25 October 2022; Accepted 26 November 2022; Published 7 January 2023

Academic Editor: Ilias Elmouki

In recent years, with the increasing incidence of cancer, regular physical examination is an important way to find cancer. Nuclear screening is an important method for the diagnosis of gastrointestinal diseases, but it is challenging in the face of small and fuzzy gastrointestinal images. Different from traditional medical objects, pathological slice images are mostly blurry and tiny, which is somewhat difficult to detect and segment. The traditional diagnostic method lacks rapid quantitative analysis and has a certain delay in medical diagnosis, and traditional image processing uses morphological features and pixel distribution to extract features; it is often difficult to achieve the desired effect on small blurry images. This paper proposes a small, microfuzzy pathology detection algorithm based on the attention mechanism; the YOLOv5 is improved under small and micro fuzzy scenarios of the detection of cancer cells in the full field of digital pathology and tests it in the gastric cancer slice dataset. The network structure is improved, and the ability to learn features on small and micro targets is enhanced according to the law of feature distribution. Spatial and channel changes in network attention and attention weight distribution. In the deep blur scenario, the attention mechanism is added to optimize its recognition ability, and the test result shows F1_score is 0.616, and the mAP is 0.611, which can provide the decision support for clinical judgment.

1. Introduction

According to the 2020 global cancer statistics report, the number of cancer cases reached 19.2928 million. Gastric cancer is the fourth leading cause of death among all cancers. Stomach cancer is the third most common cancer in China, accounting for 10.5 percent of cancer patients. Therefore, gastric cancer is an urgent health problem [1].

Pathological image analysis is a common means of disease diagnosis, mainly through the whole section digital image (WSI) for pathological image screening, which is considered by the industry as one of the most effective methods to detect cancer. Full-section pathological images have the characteristics of high resolution, wide field of view, and so on. One image can cover thousands of cells. In the process of detection, doctors usually observe and analyze with naked eyes, which is inefficient, and the diagnostic process is subject to subjective experience constraints, so it cannot be widely applied [2]. How to quickly and professionally realize pathological image analysis is an important direction in cancer diagnosis.

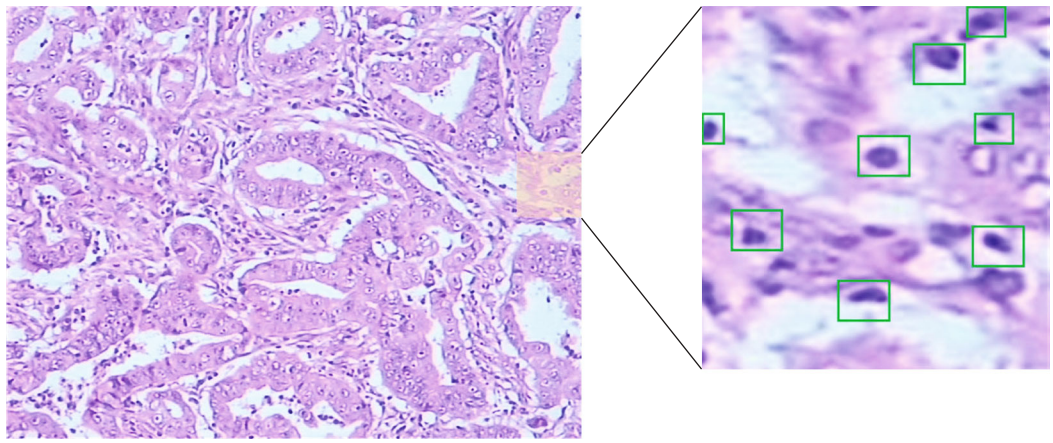


Figure 1: Example of pathological slices of patients with gastric cancer.

There are many ways to detect pathology; the traditional detection methods include the use of digital image processing technology to obtain the characteristic texture classification of cancer cells. Plissiti et al. [3] presented a fully automated detection method using morphological analysis on PAP smears for the first time and performed cluster analysis. Basavanhally et al. [4] proposed a method that combines regional growth and the Markov random field method to automatically detect lymphocytes when studying the pathological characterization of breast cancer. In 2011, Khurd et al. [5] proposed a SVM classifier based on texture features for the detection of prostate tumors. However, this method has certain limitations; it is difficult to quantitatively analyze, and it cannot be detected in large area and complex, small, and micro fuzzy scenes. With the development of deep learning technology, neural network algorithms have good detection effects on target recognition, such as YOLO algorithm [6–9], SSD [10], and RetinaNet [11]. Among them, the YOLO algorithm in convolutional neural network has certain advantages in operation speed and feature extraction.

To sum up, deep learning algorithms have achieved remarkable results, but the detection effect at the pathological cell level is not good, and there are mainly the following problems:

(1) In pathological image detection, cells change greatly, and their morphology will also change with the state of body fluids, making their distribution uneven and density uneven. At the same time, the differences between cells are small, which is difficult for feature extraction

(2) At the same time, the pathology image will also have the problem of manual labeling and counting. These will make the traditional object detection algorithm cannot be directly applied to WSI's image detection; its effect is poor

(3) The proportion of cell pixels is small, the scale is small, the target is dense, and the detection accuracy is still at a low level. In addition, the resolution of the device will also cause different degrees of blur, which will increase the difficulty of recognition

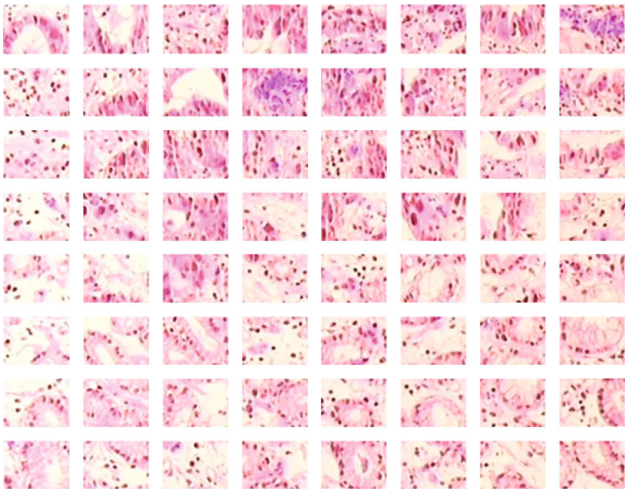


Figure 2: Gastric cancer pathology images are subdivided into 8 ∗ 8.

Aiming at the above problems, this paper realizes the complete process of case collection and collation, image analysis, database construction, and WSI dataset formation to algorithm design improvement and explores the recognition ability of neural network algorithms in small, micro-fuzzy pixels.

At the same time, an improved YOLOv5 algorithm with integrated attention mechanism is proposed to improve the algorithm's attention to key areas, so as to achieve the effect of feature enhancement recognition. Specific contributions are as follows:

(1) The original data images sampled by the hospital were selected, and the dataset was made according to the feature processing and clipping, which has general significance. The distribution of anchor frame is changed according to the improved K -means algorithm

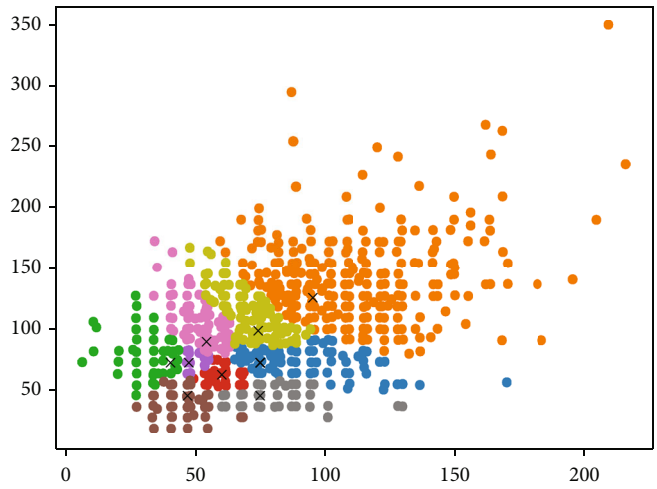


Figure 3: Anchor box clustering result visualization.

(2) The attention mechanism is introduced on the basis of YOLOv5 algorithm to improve its ability to learn small and fuzzy target data distribution. When the selected CBAM attention mechanism is added to backbone, the F1 score is increased by 4.2%

(3) In this research, the key to detection accuracy lies in the learning of scale distribution, and the difference between width and height of EIoU can be used as a correction to better replace the original aspect ratio. In this paper, the IoU basis in the loss function in the training process is modified, and EIoU is used as the cost loss function. With the CBAM attention mechanism, the F1 score showed almost no loss, but mAP increased by 0.7%

The rest of this paper is organized as follows: Section 2 provides a description of WSI image processing slicing method and the anchor frame scale clustering selection algorithm. Section 3 mainly presents the base algorithm and its improvement process and detection results. The calculation and analysis of detection index and the comparison results of ablation experiments are given in Section 4. Section 5 gives the conclusion of the research.

2. WSI Dataset Processing and Analysis

Digital pathological section datasets are not common. Compared with traditional section images, digital section image WSI has the characteristics of easy storage and easy analysis. Due to the different magnification of the microscope, the scope and scale of pathological sections obtained under the microscope are different. The gastric cancer data images provided by the hospital cannot be used directly. In order to obtain appropriate research data, the images need to be refined and sliced, such as the improved K-means algorithm.

2.1. Gastric Cancer Slice Dataset Processing. There are relatively few studies on the detection of gastric cancer cells at home and abroad. The research data in this paper were provided by the Fourth Hospital of Wuhan. The dataset contained a total of 147 partial tissue sections, all of which had backgrounds stained with chemical reagents, and the images in the dataset contained dense gastric cancer cells. The nuclei are clearly marked because of the staining. In the pathological section in Figure 1, an enlarged image of one of the regions shows gastric cancer cells distributed between the tissue fluid.

Table 1: Anchor box clustering results.

Serial points	Clustering results
1	(26.3, 54.0)
2	(31.2, 33.5)
3	(37.5, 47.0)
4	(38.2, 66.6)
5	(48.7, 36.9)
6	(50.4, 54.8)
7	(51.8, 77.4)
8	(69.8, 75.7)
9	(87.9, 113.12)

The yellow square coverage area represents the cut sample, and the green box represents the nuclei to be labeled. Due to the excessive range of slices, the pixel level is $756 * 568$ and $2592 * 1944$, and the number of cancer cells in a single sheet is too large to be used for supervised migration training, so this paper uses OpenCV (a computer vision library) to split the images in the dataset into $8 * 8$; that is, an image is divided into 64 blocks according to the row and column, and finally, 653 clear images are selected for training. The splitting result is shown in Figure 2. Labeling is clear in this state, and there is no tearing of the nucleus. Divide the dataset according to 8 : 2. In order to enable neural networks to fine-tune the location information and achieve supervised learning, an open source dataset tool, LabelImg, is used as a labeling tool to label the samples and obtain the final experimental gastric cancer dataset.

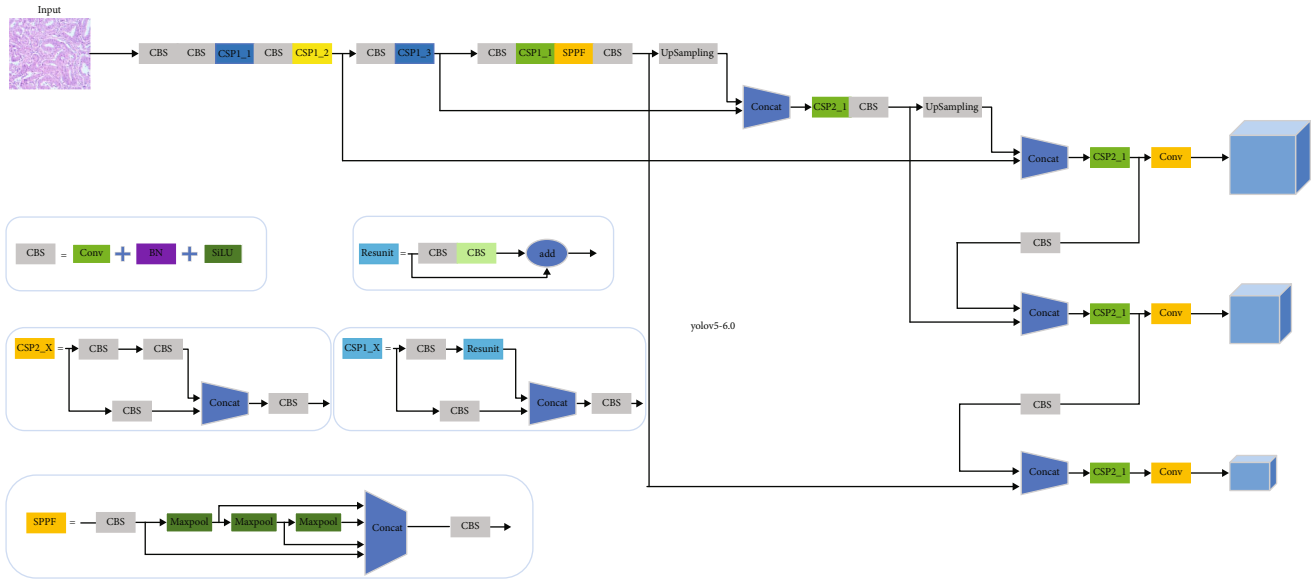


Figure 4: YOLOv5 network structure diagram.

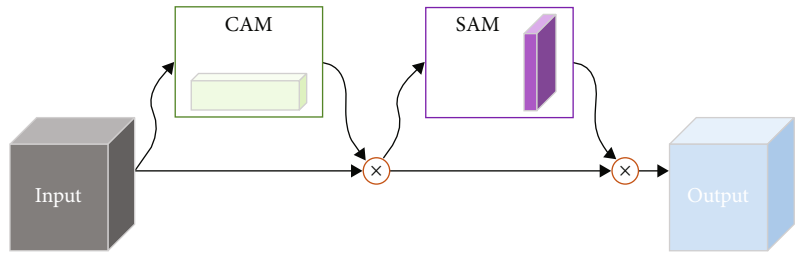


Figure 5: CBAM attention mechanism structure diagram.

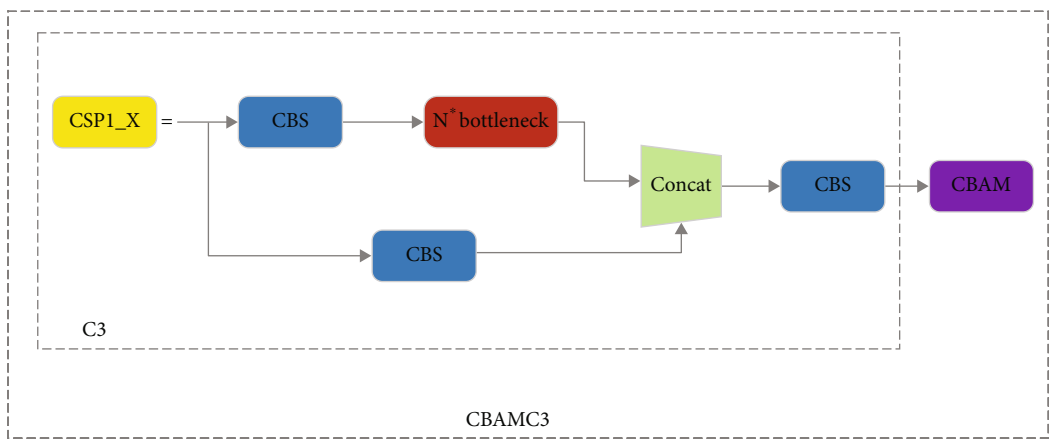


Figure 6: CBAMC3 module structure diagram.

2.2. Analysis of Recognition Characteristics of Small, Microblurry Images. Under the background of fuzzy small and micro, the size distribution of the binding boxes in the dataset is very different. Generally, the design of the anchor box is divided into 9, corresponding to 9 sizes from small to large. For the YOLO series algorithm, its detection heads are divided into 3, which detect the sensing fields of different scales, and each detection head is divided into three kinds of anchors, so that the binding box is usually divided into 9 categories.

In this experiment, the K-means algorithm is first used for clustering, which is divided into 9 categories. First, all anchor boxes are taken from the dataset as cluster sample sets, assuming that their center points are composed of w

Table 2: Optimizing hyperparameter settings.

Hyperparameters	Setpoint
Epoch	150
Batch_size	4
Init_learingrate	$1e-2$
Weight_decay	$5e-4$
Cos_lr	Auto
Mosaic_scalar	1
Optimizer	SGD

and h, and K points are randomly taken out as the center of the box cluster. The cluster center anchor set at this point can be represented as

$$C = \{c_1, c_2, c_3, \cdots, c_k\}. \tag{1}$$

Calculates the distance from all anchor boxes to the center of the cluster, assigns the closest distance to the corresponding class in this anchor box, and recalculates the points in the center of the cluster for each class.

$$c_i = \frac{1}{|\text{sum}_i|}\sum c_{\text{ij}}, \tag{2}$$

where the c_i is the cluster center of the i category, the sum_i represents all the number of points that are subordinate to the i cluster, and the c_{ij} represents the j anchor box that is subordinate to the i cluster.

Repeat the above calculations until the cluster center of the cluster no longer changes, which indicates the end of the cluster. The final output yields the anchor box required for this document. The result obtained by the above algorithm processing is shown in Figure 3.

As shown in Figure 3, the 9 colors represent the categories of anchor boxes. There are a total of 9 anchor centers, denoted by x. It can be seen that the aggregated anchor frame will be closer to the length and width distribution of the real frame of cancer cells.

Since the initial selection of the K-means algorithm is random, the first step is improved in this paper, using the IoU distance as the basis for calculation, so that the initial center point can be selected until the K center point is finally selected. After 1000 iterations, the cluster anchor box is shown in Table 1.

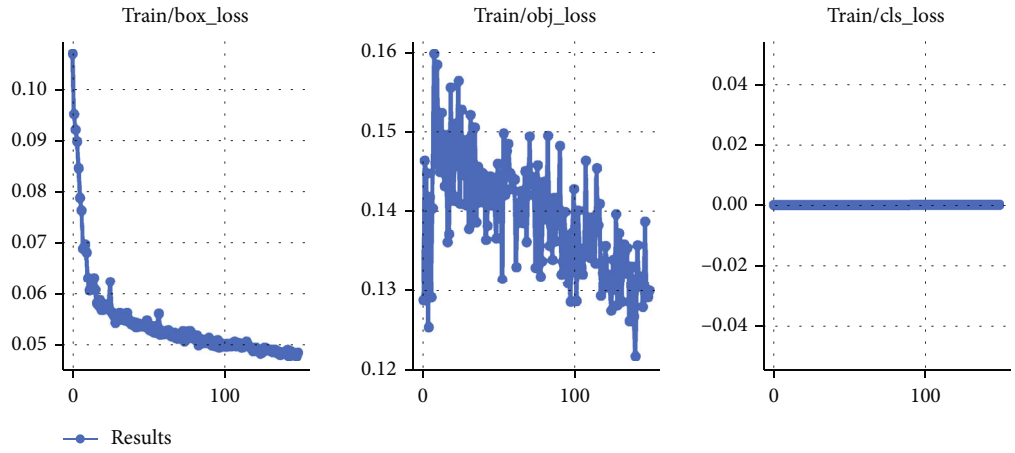
3. Pathological Detection Algorithm and Improvement

After the improvement and upgrade of YOLOv1~v4, the accuracy and speed of YOLO algorithm have been improved by leaps and bounds, and the effect of YOLOv5 target detection algorithm is better. In view of the excellent performance of YOLOv5 algorithm in target recognition, the feature extraction network is constructed on the basis of CSPDarkNet53 network based on YOLOv5 algorithm, and the attention mechanism is added to the C3 structure of the backbone to improve the channel and spatial weight of feature map, and the influence of various attention mechanisms on the results is compared. Finally, the improved algorithm is applied to the identification of small and micro fuzzy gastric cancer cells, and the quantitative analysis results are obtained.

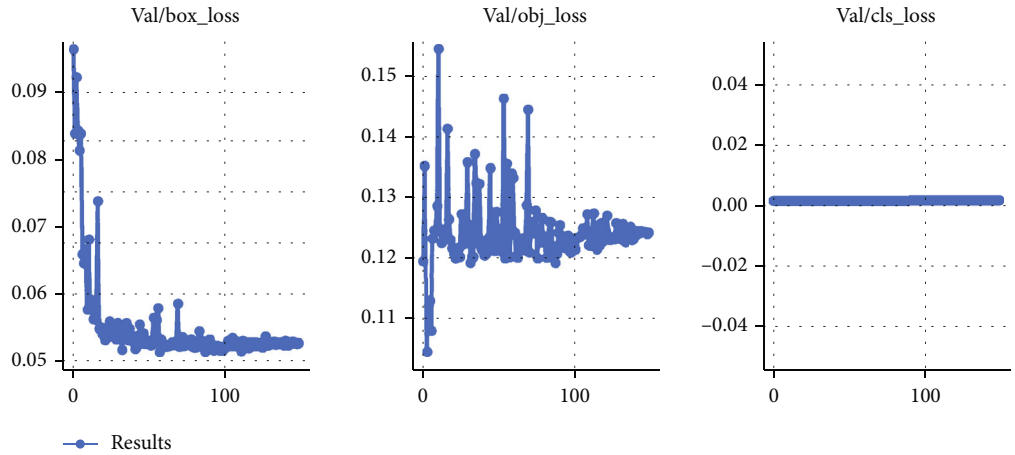
3.1. YOLOv5 Gastric Cancer Cell Pathology Detection. The backbone part of the network model consists mainly of CSPDarkNet53, with inputs through Mosaic data enhancement, MixUP, and the Focus network [12–14]. These algorithms are used to amplify the training capacity of the dataset to improve the detection effect of the algorithm. In the backbone network, a new CBS module is used as the basic convolutional block, which consists of a convolutional layer, a BN (batch normalization) layer, and a SiLU activation function. Residual structure Res-unit as a component of CSP structure, the network is spliced by CSPNet, CBS module, and several small residual structures. In this experiment, the imported slice image of the stomach cancer case will be extracted through the trunk network to obtain three useful feature layers; the input image resize is $640 * 640$ size, after the interval sampling of focus to obtain a feature map of $320 * 320 * 12$, and then through two CBS network blocks and CSP1_1 layers to obtain $160 * 160 * 128$ output and through CBS and CSP1_2 to obtain a characteristic map1. Similarly, you can obtain texturemap2 and texturemap3 and complete the trunk feature map extraction, which is 160, 160, and 128; (80, 80, and 256; and 20, 20, and 1024, respectively [15].

The neck part is the feature fusion layer of the network, which uses the FPN (feature pyramid network) structure to sample up and down and stitch the feature map, so as to realize the feature enhancement extraction of information at different scales. In the preceding trunk feature extraction network, three feature maps of different sizes and channels are output, representing the feature information of three different scale sizes. These effective feature layers will serve as a stitching basis when building the FPN. First, the channel adjustment is performed, the feature layer (20, 20, and 1024) is convoluted to obtain F5, and then, the F5 is fused with the feature layer after upsampling and the feature map2 to obtain 40, 40, and 512. Similarly, F4 can be obtained (80, 80, and 256). The feature map obtained in feature map1 is stitched and then downsampled to complete the downward feature fusion. The final output results of three different scale detectors, YOLO-head, are obtained. The specific network structure is shown in Figure 4, and it has excellent inference speed and precision.

3.2. Improved Algorithms Incorporating the Attention Mechanism. Cells in gastric cancer pathological sections are small and fuzzy samples, and the direct use of traditional neural networks has poor detection effect, because it cannot effectively extract the features of small parts. The nuclei of cancer cells studied in this paper occupy a small image area and are small targets. The attention mechanism is a process of weight allocation of target features to strengthen the attention to local effective features, so as to improve the



(a) Line chart of the training set loss function



(b) Verify the line chart of the set loss function

Figure 7: Loss function plot of the training process.

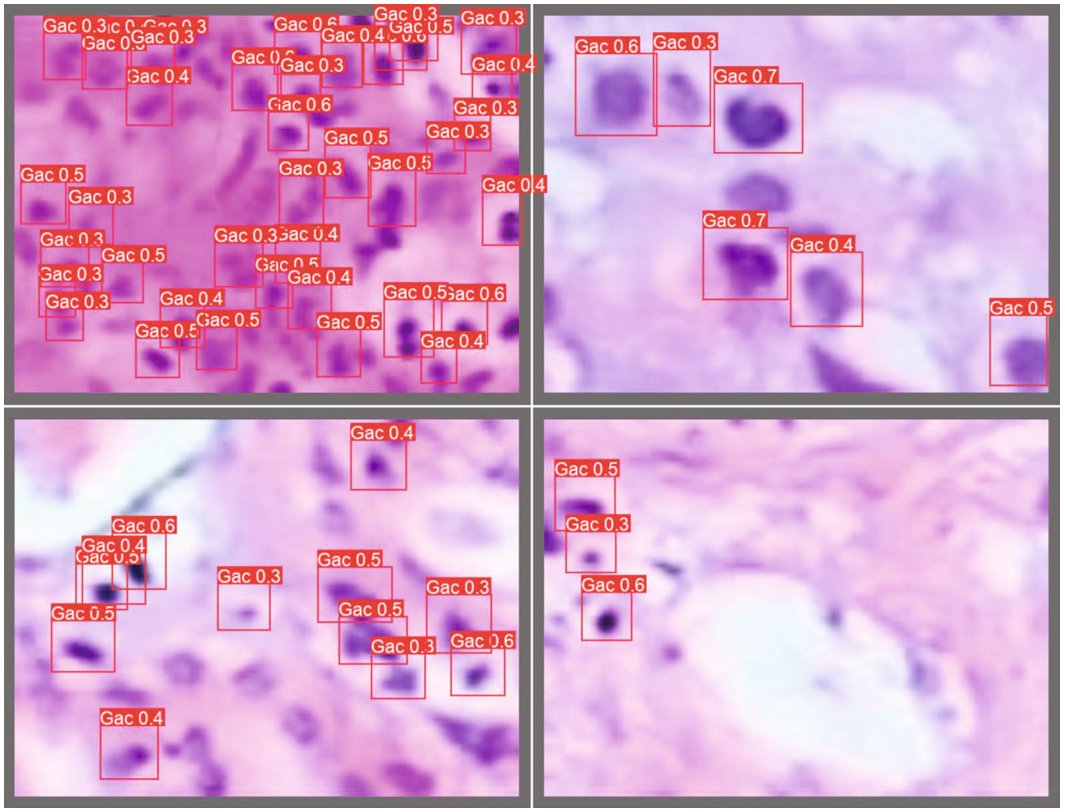
detection accuracy of gastric cancer nuclei under small and micro fuzzy background and optimize the effect of the algorithm.

The attention mechanism enables the weight reallocation of channels through a compression feature map [16]. The CBAM attention mechanism adds spatial weight allocation to the channel, using a large convolution to encode spatial features and strengthen feature representation [17]. The cam attention mechanism adopted in this paper has better calibration effect than other attention mechanisms. It mainly distributes the weight of small targets in space, as shown in Figure 5. Its specific implementation process is that the input is input size (h, w, and c), through the channel pooling becomes (h, w, and 2), and then through the convolution of $7 * 7$ and BN and sigmoid activation function to obtain the weight parameter matrix; the weight parameter matrix and input are multiplied to obtain the channel attention map, the feature map in the length and width direction of the global maximum pooling, and then through the multilayer perceptron and $1^* 1$ convolution to get $1^* 1^* C$ weight parameters, and finally through the sigmoid function to obtain a normalization result. Multiply to weigh the channel to the original input feature map.

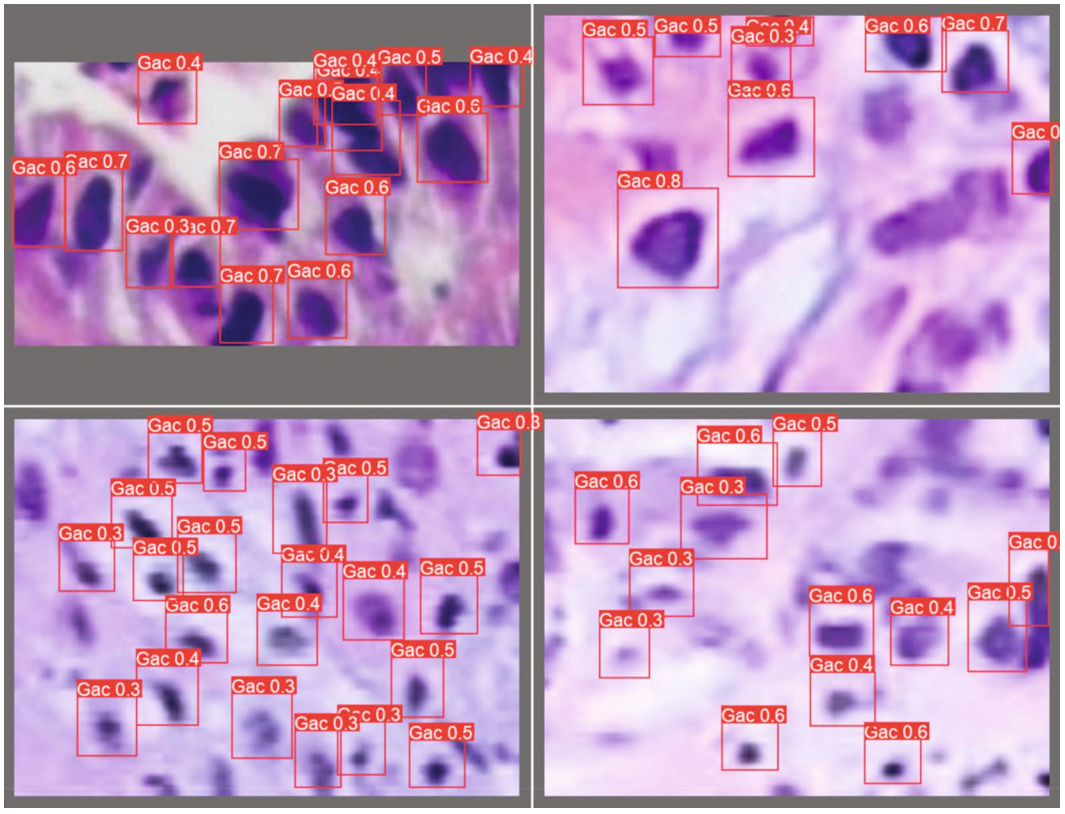
In order to realize the rescale and weight allocation of the small target features of the main trunk feature extraction, the C3 structure and the CBAM to C3 residual structure block is added to form a new CBMAC3 module, and the improved effect is shown in Figure 6.

3.3. Training Optimization and Loss Function Modification. The research device is a deep learning computer with an operating system of Windows 10, a CPU of Intel i7 10700 with a core frequency of 2.9 GHz and a running memory of 16 GB. It is equipped with an RTX Quadro 4000 GPU with 8 GB of video memory. A total of three sets of comparative experiments are involved, all of which are carried out on this configuration, and the parameters are set before the experiment begins. Depending on the dataset used and the scenario, the design hyperparameters are shown in Table 2.

In terms of the improvement of the loss function, GIoU [18] used the overlapping area as a basis to eliminate the loss zero problem when the boundary intersected with 0, but then, there was a problem of slow convergence, so later, researchers proposed DIoU [19] and added a center point distance on this basis. In this study, considering the difference of the cancer cell detection frame, the difference in



(a) Light-colored batch test results



(b) Dark nuclei test results

Figure 8: Test set test result graph.

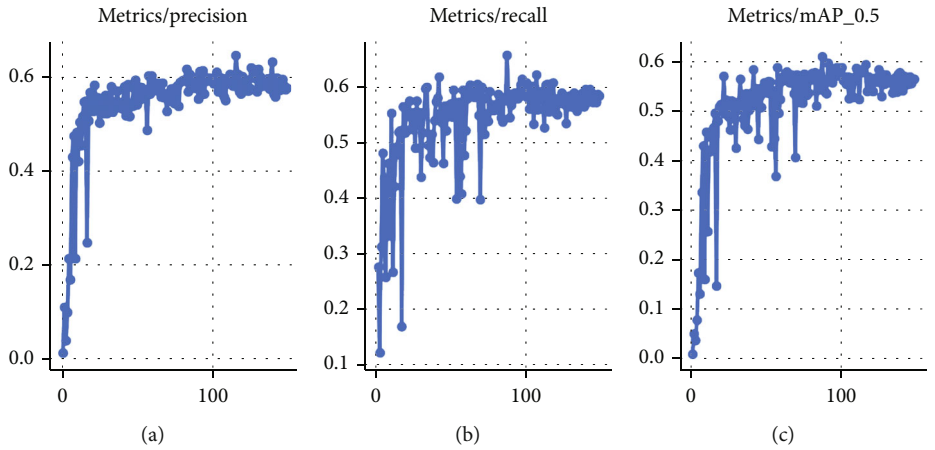


Figure 9: Test indicator result graph. (a) Precision result graph; (b) recall result graph; and (c) mAP@50 result graph.

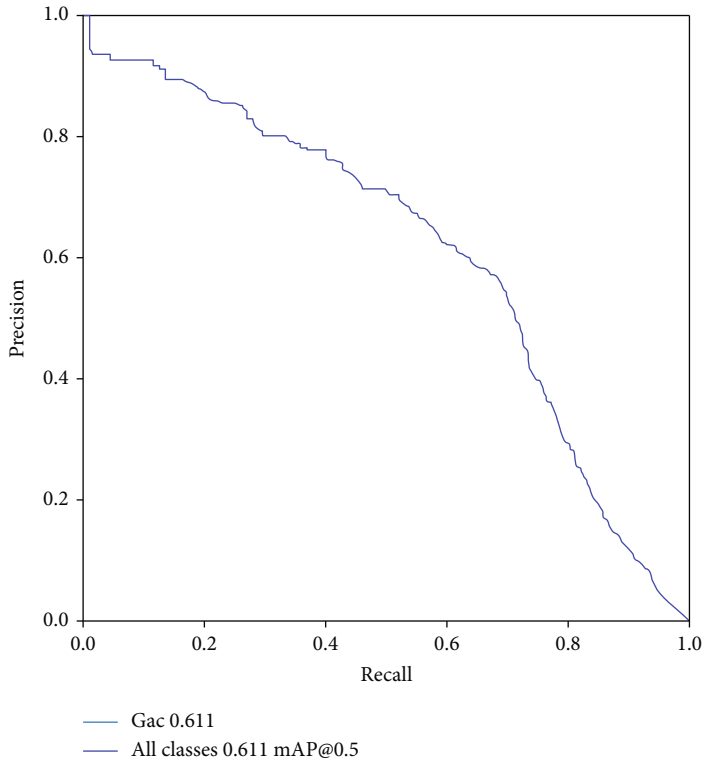


Figure 10: Test PR curve result plot.

width and height of EIoU [20] can be used as a correction to better replace the original aspect ratio. Therefore, the training process modifies GIoU and uses EIoU as the cost loss function. The EIoU loss consists of three parts: the overlap loss function L_{IoU}, the center distance loss L_{dis}, and the height-width loss function L_{asp}. Its expression can describe as

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - \mathrm{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_W{}^2} + \frac{\rho^2(h, h^{gt})}{C_h{}^2}, \quad (3)$$

where b and b^{gt} represent the center of the prediction box and the real box, respectively. ρ^2 represent the introduced Euclidean distance, and c represents the minimum closed region. Similarly, w and h are width and height, respectively. w^{gt} and h^{gt} are the width and height of the real box, respectively, and C_W and C_h represent the width and height of the minimum add-in box. The visualization results of the loss function during training are shown in Figure 7.

The loss function calculation consists of two parts: the first and second rows are represented as a line chart of the loss function of the training set and the validation set, respectively. It can be seen that its regression loss gradually decreases and stabilizes after 100 rounds, indicating that the model tends to

Table 3: Comparison of ablation experimental performance.

Model	CBAMC3	SEC3	EIoU	Precision	Recall	F1 score	mAP@50
YOLOv5	×	×	×	0.6	0.56	0.579	0.593
Improved model 1	×	√	×	0.637	0.592	0.614	0.592
Improved model 2	√	×	×	0.639	0.604	0.621	0.604
Improved model 3	√	×	√	0.581	0.656	0.616	0.611

converge. The confidence error generally shows a downward trend, and since only gastric cancer nuclei need to be detected in this experiment, the classification loss has been at a low level. In order to verify the real effect of its detection, after the training is completed, a weight file suitable for the detection of gastric cancer cells WSI scene is obtained, and the detection results of some images of the test set using this weight are shown in Figure 8.

As shown in Figure 8, in order to test the effect in different scenarios, the test included the following samples: sample 21 contained a scene with a darker staining background, sample 56 contained coarse and shallow stained objects, samples 126 and 63 had higher resolution, and the nuclei in the sample were more obvious.

4. Analysis and Comparison of Results

4.1. Analysis of Experimental Results. In order to quantitatively analyze the detection results, the commonly used indicators in pathological detection, including precision, recall, F1 comprehensive index (F1_score), and mean precision (mAP), are used as evaluation indicators to analyze the accuracy of the model. In the process of index calculation, the relevant concept of confusion matrix is involved, and this paper uniformly stipulates that TP indicates the allocation of correct positive samples; TN is assigned the correct negative sample; FP is a positive sample with an incorrect assignment; and the FN is a negative sample of the misassignment.

The accuracy of this study refers to the ratio of the samples assigned as positive and determined as gastric cancer cells to the correctly assigned samples. The expression is

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{4}$$

Recall rate refers to the proportion of all positive gastric cancer cell samples assigned correctly, and in pathology, the recall rate is more indicative of the quality of its test results. It can be expressed as

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{5}$$

The F1 is a comprehensive index, which is used to neutralize the contradiction between accuracy and recall. Specifically expressed as

$$\text{F1} = \frac{2 * P * R}{P + R}. \tag{6}$$

mAP represents the average detection accuracy, the most commonly used composite index in target detection, and in this study, it is mainly used to compare the effectiveness of ablation experiments.

$$\text{mAP} = \frac{1}{m} \sum_{i=1}^{m} \text{AP}_i, \tag{7}$$

where AP represents the average accuracy of detection for each class.

According to the formula modeling and by plotting the correlation curve, it can be seen that the index changes of the converged attention mechanism proposed in this paper and the improved model under EIoU are shown in Figure 9.

As can be seen from the values recorded by the algorithm, the accuracy is up to 0.62, the recall rate is up to 0.656, and the mAP value is up to 0.611. The change of index F1 is shown in Figure 10.

4.2. Comparison of Ablation Experiments. Pathological image detection of gastric cancer is an emerging research field. The scarcity of samples, difficulty in labeling, small detection target, and fuzzy scale have become important factors hindering its development. Therefore, this paper will demonstrate the effectiveness of the improved algorithm and make a deeper comparison between the detection results and the large datasets of other researchers.

In order to more fully compare the accuracy of the improved algorithm with that of the original algorithm, an ablation experiment is designed. The experimental results are shown in Table 3. Improved type 1 represents the addition of SE mechanism; that is, channel dimension adds attention mechanism, while improved type 2 adds cam attention mechanism. Compared with SE, CBAM has more obvious improvement and better results in average detection accuracy. Compared with the original algorithm of YOLOv5, the addition of CBAMC3 module improves the F1 index by 4.2%. Based on this result, after the original CIoU of YOLOv5 was modified to EIoU, the detection accuracy was improved again, reaching 0.611 on mAP, and the detection result mAP increased by 0.7% compared with the improved type 2 without EIoU.

5. Conclusion

Aiming at the current pathological detection, especially gastric cancer nuclear detection in small and micro fuzzy scenes, the accuracy of gastric cancer nuclei is not high, the error is large, and the model is unstable; a deep learning detection algorithm

based on improved YOLOv5 is proposed. The main improvement points of the algorithm are as follows:

(1) The anchor box clustering algorithm was modified. In the data preprocessing stage, in the process of anchor box clustering analysis, the initial point screening is improved to K-means ++ algorithm, and the random initial K-means clustering algorithm is improved to meet the requirements of the size and width ratio of anchor boxes under small and micro fuzzy detection

(2) The attention mechanism is integrated, so that the network can redistribute the weight, increase the attention of the network to small and micro target features, and greatly improve the accuracy of gastric cancer cell recognition. At the same time, two representative attention mechanisms are compared, and the cam attention mechanism which takes into account space and weight allocation is preferred

(3) In order to solve the problem that the length and width difference of small and micro fuzzy recognition scene is not big, the loss function is improved. The aspect ratio difference of EIoU was chosen as an alternative, which made the model evaluation more suitable for pathological cell section study and improved the average detection accuracy

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Qiuxia Guo and Weiwei Yu contributed equally to this work.

Acknowledgments

The authors appreciate the financial support from the National Natural Science Foundation of China (Grant No. 81700056) and the technical support of Shenzhen Gengfeng Technology Co., Ltd. (gengfengtechnology@126.com).

References

[1] Z. Liu, Z. Li, and Y. Zhang, "Interpretation of the 2020 global cancer statistics report," *Journal of Multidisciplinary Cancer Management*, vol. 7, no. 2, pp. 1–14, 2021.

[2] F. Chen, *Intelligent Identification of Endoscopic Images of Gastric Cancer and Diagnostic Technology of Fluorescent Peptides*, Zhejiang University, 2019.

[3] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, 2011.

[4] A. N. Basavanhally, S. Ganesan, S. Agner et al., "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 642–653, 2010.

[5] P. Khurd, L. Grady, A. Kamen, S. Gibbs-Strauss, E. M. Genega, and J. V. Frangioni, "Network cycle features: application to computer-aided Gleason grading of prostate cancer histopathological images," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1632–1636, Chicago, IL, 2011.

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

[7] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517–6525, Hawaii, USA, 2017.

[8] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 89–95, Salt Lake City, USA, 2018.

[9] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 2020.

[10] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Amsterdam, The Netherlands, 2016.

[11] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, Venice, Italy, 2017.

[12] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2778–2788, Montreal, Canada, 2021.

[13] Z. Chen, R. Wu, Y. Lin et al., "Plant disease recognition model based on improved YOLOv5," *Agronomy*, vol. 12, no. 2, pp. 365–379, 2022.

[14] Y. C. Xing and D. J. Li, "Remote sensing image target detection based on YOLOv5," *JiangXi Science*, vol. 39, no. 4, pp. 725–732, 2021.

[15] A. Kuznetsova, T. Maleva, and V. Soloviev, "Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images," in *International Symposium on Neural Networks*, pp. 233–243, Springer, Cham, 2020.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, USA, 2018.

[17] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolution block attention module," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3–19, Salt Lake City, USA, 2018.

[18] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition, pp. 658–666, California, USA, 2019.

[19] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: faster and better learning for bounding box regression," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[20] X. L i, W. Wang, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss v2: learning reliable localization quality estimation for dense object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11632–11641, 2021.