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

An Optimized Object Detection Approach on Medical Image using Feature Enhancement and Dynamic Loss

Yingjun Liu [1 2][0000-0001-7852-4013] and Fuchun Liu [1] and Yingbin Huang [2]

[1] Guangdong University of Technology, School of Computers, 510006, Guangzhou , China

[2] Guangzhou City University of Technology, School of Robot Engineering, 510800, Guangzhou , China
liuyj@gcu.edu.cn

Abstract. Detection based on medical images is crucial for improving disease cure rates and patient prognosis. However, existing methods have limitations in feature extraction and computational efficiency. This paper presents an optimized method for medical image object detection using feature enhancement and dynamic loss (FENet-UIoU). It combines receptive field attention convolution (RFAConv), efficient up-sampling convolution block (EUCB), large separable kernel attention (LSKA), and a dynamic loss function to address the shortcomings of convolutional neural network (CNN) in medical image detection. RFAConv highlights tumor features through spatial attention mechanisms, EUCB improves feature map resolution and computational efficiency, LSKA enhances feature capture and expression, and the unified intersection over union (UIoU) dynamic loss function uses dynamic weight allocation to optimize the prediction of box focus. According to the experimental evaluation, the proposed method yields a maximum 7.8% improvement on mean average precision (MAP) over the existing approach. Meanwhile, ablation experiments verify the synergistic effect of each module, indicating that this study provides a high-precision and high-efficiency solution for medical image object detection.

Keywords: Medical Image, Object Detection, Feature Enhancement, Dynamic Loss.

1 Introduction

Early and accurate detection on disease is of great clinical significance for improving patient survival rates and optimizing treatment plans [1][2]. Deep learning-based medical image analysis techniques have witnessed rapid development in recent years [3][4][5]. However, the parameter sharing mechanism of convolution kernels in traditional CNN models limits the flexibility of feature extraction [3]. To address the limitations of feature extraction, researchers have proposed various improvement schemes. For example, squeeze and excitation networks enhance key features through channel attention mechanisms [6], but their dynamic adaptability in the spatial dimension is insufficient [7][8]. MobileNet reduces computational costs using depth-wise separable convolutions [9], but it performs poorly in multi-scale feature fusion for medical images [10][11]. In terms of loss function design, the researcher group optimizes bounding box regression by introducing center point distances [12], but it still cannot dynamically balance the weight allocation of high or low-quality prediction boxes [13][14]. Tumor edges and texture features were focused on in [15] [16]and long-range dependencies

between tumors and surrounding tissues through separable dilated convolutions were captured in [16][17]. These issues indicate that existing methods urgently need to improve the synergy of feature enhancement, computational efficiency, and dynamic optimization in medical image detection tasks.

This paper proposes a feature enhancement and dynamic loss-based object detection framework (FENet-UIoU) to break through the above technical bottlenecks. First, the RFAConv module introduces a spatial attention mechanism to dynamically allocate convolution kernel parameters and focus on edges and texture features. Second, the EUCB module is designed to combine depth-wise separable convolutions with channel shuffle strategies to achieve high-resolution feature reconstruction and computational efficiency balance. Furthermore, the LSKA module is adopted to capture long-range dependencies through separable dilated convolutions. At the loss function level, we introduce the UIoU loss, which integrates dynamic bounding box scaling to enhance the responsiveness to high-quality prediction boxes. According to the experimental evaluation, the proposed method yields a maximum 7.8% improvement on mean average precision (MAP) over the existing approach.

2 Methods

To tackle the fundamental problems of inadequate feature extraction, constrained computational efficiency, and diminished sensitivity in boundary regression within medical image object detection, this paper introduces a feature enhancement and dynamic loss-based object detection framework, named FENet-UIoU. The architecture of the FENet-UIoU model is visualized in Fig. 1 with illustrating the complete processing pipeline from input image to final detection outcomes.

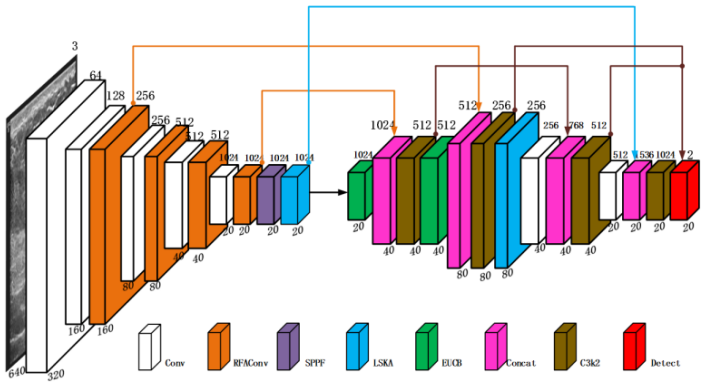

Fig. 1. Architecture of the FENet-UIoU Model.

The architecture achieves optimization by integrating the design of the feature extraction network with the loss function. The model accepts medical images of size 640×640×3 as input and progressively decreases the spatial resolution of the feature

maps. The RFAConv module dynamically allocates convolution kernel parameters using a spatial attention mechanism to improve the key features. The Spatial Pyramid Pooling with Feature Fusion (SPPF) captures multi-scale features, while the LSKA module captures global contextual information through separable dilated convolutions. The EUCB module combines depth-wise separable convolutions with channel shuffle strategies to enhance the resolution and computational efficiency of the feature maps. The Concat module fuses information from different levels, and the C3k2 module optimizes the feature fusion process. Finally, the detection head outputs the detection results, which include bounding boxes and class probabilities.

2.1 Feature Enhancement

RFAConv Model: RFAConv introduces a receptive field attention mechanism that customizes convolution kernel parameters for different regions to focus on the key features such as edges and shapes, while suppressing irrelevant background interference. Specifically, RFAConv dynamically generates attention weights based on different regions of the input feature map and combines these weights with the convolution kernel parameters to produce customized kernels for each region. The implementation process of RFAConv is shown in Fig. 2.

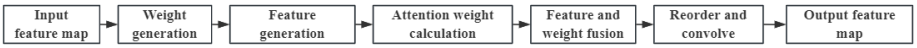

Fig. 2 . Flowchart of RFAConv Implementation

Weight generation: The input feature X shapes as (b,c,h,w), where b: batch size, c: channels, h×w: spatial resolution. Average pooling (AvgPool2d) is applied to X to get the global information in Equation 1:

$$W_{pool} = AvgPool2d(X, kernel\_size = k, padding = p, stride = s) \tag{1}$$

Then, a 1×1 convolution (Conv2d) is performed to generate the weight tensor W, as shown in Equation 2 :

$$W = Conv2d(W_{pool}, out\_channels = c \times k^2, kernel\_size = 1, groups = c, bias = False) \tag{2}$$

Rearrangement and convolution: Use rearrange to reorganize the weighted feature map, adjusting its shape to fit subsequent convolution operations. Perform convolution on the rearranged data to obtain the final output feature map Y with the shape (b, c, h_{out}, w_{out}), as shown in Equation 3 :

$$conv\_data = rearrange(F_{weighted}, bc(n1n2)hw -> bc(hn1)(wn2)', n1 = k, n2 = k)$$

$$Y = Conv2d(conv\_data, out\_channels = c_{out}, kernel\_size = k, stride = k) \tag{3}$$

After completing the above steps, RFAConv produces the output feature map Y, where the features have been adjusted according to the attention weights, highlighting key regions and suppressing unimportant information.

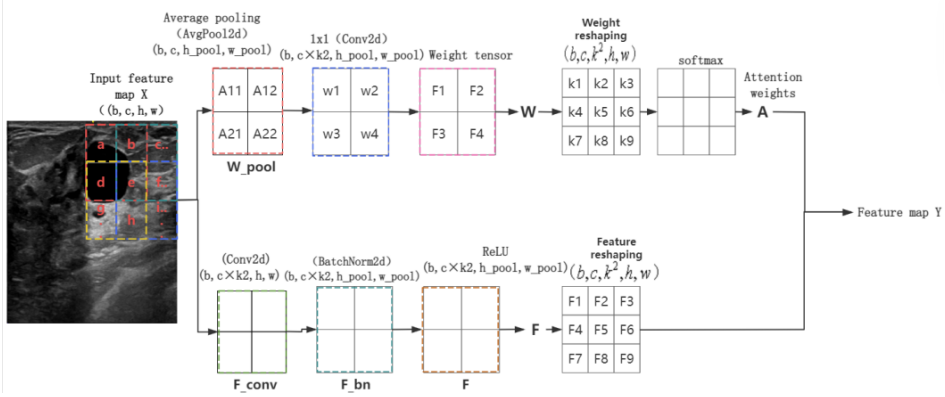

Fig. 3 . RFAConv Neural Network Diagram

The detailed structure of RFAConvis shown in Fig. 3. The input feature map is divided into receptive field blocks (e.g., A11 denotes the block in row 1, column 1). Each block is assigned weights via a dynamic attention map, generating an attention weight map with the same shape as the input. This emphasizes key areas and reduces background noise. The feature map is then element-wise dotted with the attention weights to create a weighted feature map, enhancing key features while weakening non-key ones. Finally, the RFAConv module reorganizes this weighted map to produce the output feature map.

Efficient Up-sampling Convolution Block (EUCB): EUCB module combines operations such as up-sampling, depth-wise separable convolution, batch normalization, activation functions, channel shuffling, and point-wise convolution to achieve efficient up-sampling and feature enhancement. The implementation process of EUCB is designed in Fig. 4.

Fig. 4 . EUCB Process Diagram

The UpSampling operation effectively restores low-resolution feature maps to match the dimensions and resolution of the next skip connection feature map, without adding too much computational load. This process accurately captures feature information at different levels, providing a rich data basis for subsequent precise analysis. The output map X_{up} is designed in Equation 4:

$$X_{up} = \text{UpSampling}(X) \tag{4}$$

Depth-wise convolution performs convolution operations on each input channel separately, using a K×K convolution kernel for each channel. The depth-wise convolution operation is expressed in Equation 5:

$$(f_c * k_c)(i,j) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} f_c(i+m, j+n) \bullet k_c(m,n) \tag{5}$$

Here, f_c is the c-th channel of the input feature map, k_c is the corresponding convolution kernel. Depth-wise convolution is applied to the up-sampled feature map X_{up} as shown in the Equation 6 :

$$X_{dw} = \text{Depth-wise Convolution}(X_{up}) \tag{6}$$

Then batch normalization is applied to the feature map X_{dw} after convolution, which accelerates training and enhances model stability,as shown in Equation 7:

$$X_{bn} = BatchNorm2d(X_{dw}) \tag{7}$$

The ReLU activation function is adopted to the batch-normalized feature X_{bn} to introduce non-linearity, enhancing the expressive capability as shown in Equation 8:

$$X_{act} = \text{ReLU}(X_{bn}) \tag{8}$$

The feature map X_{act} is subjected to channel shuffle operation, which exchanges channel information to enrich the features, as shown in the following Equation:

$$X_{shuffle} = \text{Channel\_shuffle}(X_{act}) \tag{9}$$

Finally, A 1×1 convolution kernel is used to perform point-wise convolution on the feature map $x_{shuffle}$ as $X_{out} = \text{Conv2d}(X_{shuffle})$ adjusting the number of channels to ensure the output meets the requirements. Throughout the entire network architecture, the smooth flow of data between different modules is essential, and matching the number of channels is one of the key factors to ensure accurate data transfer and effective integration. This operation helps stabilize the training of the entire network, ensuring that each module can fully exert its function and ultimately achieve accurate prediction and analysis of medical images.

LSKA Model: The neural network diagram of LSKA is designed as Fig. 5. It decomposes the two-dimensional convolution kernel into cascaded one-dimensional convolution kernels and attention mechanisms to lower computational costs while improving feature extraction and representation. Horizontal one-dimensional Convolution, as shown in the following Equation:

$$Z_C^h = W_C^h * F_C \tag{10}$$

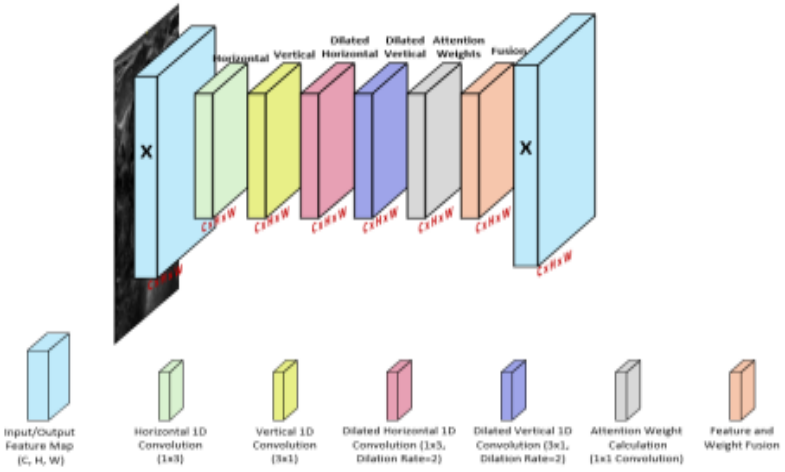

Fig. 5 . Neural Network Diagram of the LSKA Module

where W_C^h is the convolution kernel for the horizontal direction, with a shape of 1×k. F_c is the C_{th} channel of the input feature. * denotes the operation of convolution . Z_C^h is the feature map after horizontal convolution. Vertical one-dimensional convolution, as shown in the following Equation:

$$Z_C^v = W_C^v * Z_C^h \tag{11}$$

where W_C^v is the convolution kernel for the vertical direction, with a shape of k×1, Z_C^v is the feature map after vertical convolution. Dilated convolution for horizontal one-dimensional convolution, as shown in the following Equation:

$$Z_C^{h\_dilated} = W_C^{h\_dilated} *^d Z_C^v \tag{12}$$

where $W_C^{h\_dilated}$ is the horizontally dilated convolution kernel with a shape of 1×k. $*^d$ denotes the dilated convolution operation, and d is the dilation rate. $Z_C^{h\_dilated}$ is the feature map achieved by horizontal convolution with dilation. Vertical one-dimensional dilated convolution is shown in the following Equation:

$$Z_C^{v\_dilated} = W_C^{v\_dilated} *^d Z_C^{h\_dilated} \tag{13}$$

where $W_C^{v\_dilated}$ is the vertically dilated convolution kernel with a shape of k×1. $Z_C^{v\_dilated}$ is the feature map after vertical convolution with dilation. Attention weights are computed via a 1×1 convolution layer to dynamically re-weight the feature map which is shown in the following Equation:

$$A_C = W_{1\times1} * Z_C^{v\_dilated} \tag{14}$$

where $W_{1\times1}$ is the 1×1 convolution kernel. Ac is the generated attention weight. By performing an element-wise multiplication between the attention weights and the input feature map, the final output feature map is obtained, as expressed in the following equation:

$$\overline{F_C} = A_C \otimes F_C \tag{15}$$

where A_c is the generated attention weight. F_c is the C-th channel of the input feature map. $\otimes$ denotes element-wise multiplication, $\overline{F_C}$ is the final output feature map.

2.2 Dynamic Loss

The conventional Intersection over Union (IoU) metric serves to quantify the overlap between predicted and ground truth bounding boxes. The same weight allocation is given to low-quality and high-quality prediction boxes, which fails to effectively address the problem of sample imbalance.

$$L_{IoU} = 1 - \left| \frac{B \cap B_{gt}}{B \cup Bgt} \right| \tag{16}$$

Let B be the predicted box and B_{gt} the ground truth box. $B \cap B_{gt}$ represents the overlapping area between the predicted and ground truth boxes. $B \cup B_{gt}$ denotes the total area covered by both boxes. The L_{IoU} is the IoU loss function; a smaller value indicates closer alignment, with $L_{IoU} = 0$ when they perfectly overlap. The formula given is for the L_{IoU} loss function, which assesses the overlap between B and B_{gt}.

The dynamic box weighting strategy dynamically adjusts the size of the predicted and ground truth boxes, assigning different weights to prediction boxes of varying quality. This strategy allows the model to focus more on high-quality prediction boxes during the training process, thereby improving detection accuracy and convergence speed. The dynamic box scaling ratio is adjusted dynamically based on the number of training epochs, specifically through a linearly decreasing strategy, as shown in Equation 17 :

$$\text{ratio} = -0.005 \times Num_e + 2 \tag{17}$$

where Num_e is the training epochs, and ratio=2, the model significantly enlarges the boxes to focus on a large number of low-quality boxes, accelerating convergence; as training progresses, ratio decreases linearly, and the model gradually focuses on high-quality boxes to improve accuracy.

$$\begin{aligned} ww1 &= w1 \times \text{ratio}; \quad hh1 = h1 \times \text{ratio} \\ ww2 &= w2 \times \text{ratio}; \quad hh2 = h2 \times \text{ratio} \end{aligned} \tag{18}$$

Here, ww1 and hh1 represent the dimensions of the scaled detection box. They are obtained by multiplying the original width and height (w1 and h1) by a scaling factor

(ratio). To calculate the boundary coordinates of the predicted box, the Equation is as follows:

$$bb1\_x1 = bb1\_xc - \frac{ww1}{2};\quad bb1\_x2 = bb1\_xc + \frac{ww1}{2};$$

$$bb1\_y1 = bb1\_yc - \frac{hh1}{2};\quad bb1\_y2 = bb1\_yc + \frac{hh1}{2} \tag{19}$$

where $bb1_{xc}$, $bb1_{yc}$ are the center coordinates of the predicted box, Calculate the boundary coordinates of the ground truth box, the Equation is as follows:

$$bb2\_x1 = bb2\_xc - \frac{ww2}{2};\quad bb2\_x2 = bb2\_xc + \frac{ww2}{2}$$

$$bb2\_y1 = bb2\_yc - \frac{hh2}{2};\quad bb2\_y2 = bb2\_yc + \frac{hh2}{2} \tag{20}$$

Where the predicted box's center is given by (bb2xc, bb2yc), and the ground truth box 's scaled width and height are ww2 and hh2, respectively.After obtaining the scaled bo unding box coordinates, the next step is to calculate the intersection area (inter) and th e union area (union), and finally compute the IoU value. The Equations for calculating the intersection area inter and the union area, and computing IoU, are as follows:

$$union = w1 \times h1 + w2 \times h2 - inter + \varepsilon$$

$$iou = \frac{inter}{union} \tag{21}$$

The intersection (inter) is the area of overlap between the predicted bounding box B and the ground truth bounding box B_{gt}. Here, the predicted box has dimensions (w1, h1), and the ground truth box has dimensions (w2, h2). Based on the IoU value, adjust the weights dynamically to make the model focus more on high-quality prediction boxes.

3 Experimental Results and Analysis

The study assesses the performance of the proposed model in breast tumor detection tasks. The breast tumor dataset is sourced from the publicly available Breast Ultrasound Dataset (BUSI), which was constructed and open-sourced by Al-Dhabyani et al. in 2020 [19]. This dataset comprises 3172 ultrasound images of breast tumor cases confirmed by pathology with benign tumors (1717 images) and malignant tumors (1455 images). All images were annotated with tumor bounding boxes and class labels by professional radiologists, and the annotation consistency was verified using the Kappa coefficient (k=0.92) to ensure reliability. The dataset was divided into training (2,051 images),

validation (335 images), and test sets (335 images) using random sampling, ensuring balanced distribution across categories.

3.1 Comparative Experiments

The comparison of the training loss between the proposed model and other models (YOLOv11, YOLOv8, YOLOv5, Faster-rcnn, SSD) is shown in Fig. 6. The proposed model demonstrates significant advantages during the training process. In the first 10 epochs, the loss of the proposed model rapidly decreases from 5.5 to approximately 1.5, while the loss of YOLOv11 decreases from 5.5 to about 2.0, YOLOv8's loss drops from 5.5 to about 2.5, YOLOv5's loss falls from 5.5 to around 3.0. The loss of Faster-rcnn decreases from 5.5 to about 3.5. The loss of SSD drops from 5.5 to about 4.0. At the final stage (Epoch 100), the loss of the proposed model further decreases to about 0.5, the loss of YOLOv11 stabilizes at about 1.2, YOLOv8's loss stabilizes at about 1.8, YOLOv5's loss stabilizes at about 2.2, Faster-rcnn's loss stabilizes at about 2.8, and SSD's loss stabilizes at about 3.2. The data indicate that the proposed model not only converges faster during the training process but also achieves a lower final loss value.

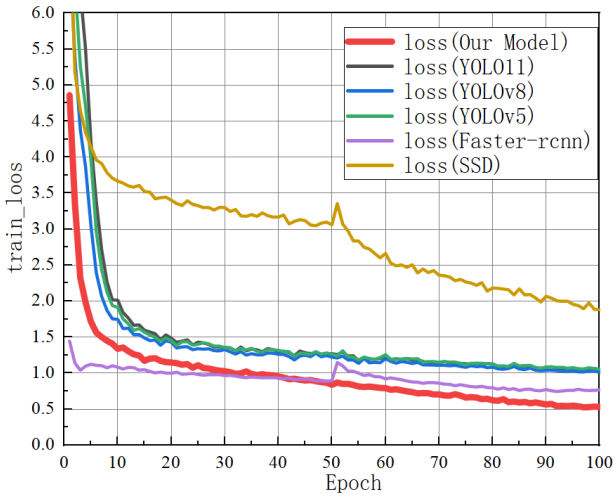

Fig. 6 Loss Metric Performance

The plot of the Map50 metric for the proposed model and other models (YOLOv11, YOLOv8, YOLOv5, Faster-rcnn, SSD) can be seen in Fig. 7. The proposed model demonstrates significant advantages during the training process. In the first 10 epochs, the Map50 value of the proposed model quickly rises to about 0.75, YOLOv11's Map50 value increases to about 0.70, YOLOv8's Map50 value rises to about 0.65, YOLOv5's Map50 value increases to about 0.60, Faster-rcnn's Map50 value increases to about 0.55, and SSD's Map50 value rises to about 0.50. At the final stage (Epoch 100), the proposed model's Map50 value stabilizes at about 0.88, YOLOv11's Map50 value stabilizes at about 0.82, YOLOv8's Map50 value stabilizes at about 0.78, YOLOv5's Map50 value stabilizes at about 0.75, Faster-rcnn's Map50 value stabilizes at about

0.70, and SSD's Map50 value stabilizes at about 0.65. Judging from the shape and numerical changes of the curves, the proposed model not only shows a faster convergence speed at the beginning of training but also maintains a higher Map50 value throughout the entire training process.

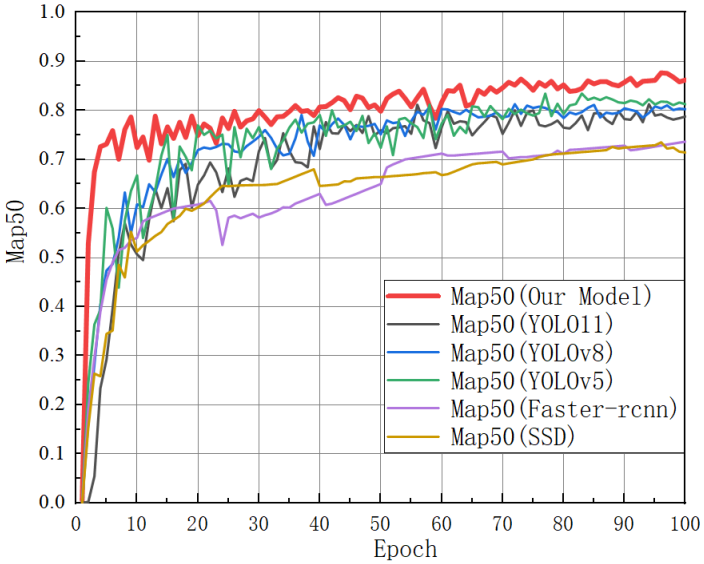

Fig. 7 Map50 Metric Performance

The proposed model's Map50 value rises quickly within the first 10 epochs, much faster than other models, indicating its ability to quickly learn effective feature representations. At the mid-stage, the proposed model's Map50 value has already reached 0.85 and stabilizes at 0.88 at the final stage, while other models' Map50 values are all below this level. This shows that the proposed model can continuously optimize during the training process, ultimately achieving higher detection accuracy. In summary, the proposed model demonstrates a faster convergence speed, higher Map50 value, and a more stable training process during the training, which fully proves its advantages in feature extraction, parameter optimization, and computational efficiency.

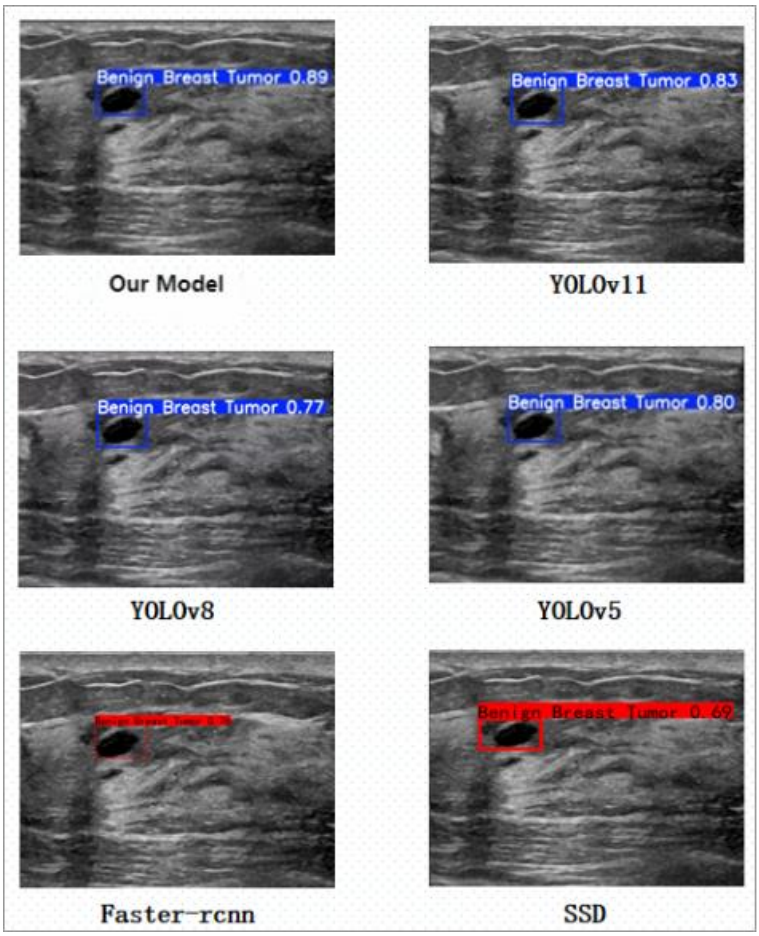

Fig. 8 Comparative Experiment Identification Effect Diagram

The inference results of the proposed model and models such as (YOLOv11, YOLOv8, YOLOv5, Faster-rcnn, SSD) in breast tumor detection tasks are shown in Fig. 8. The proposed model significantly outperforms other models in terms of detection accuracy and confidence. Confidence of the proposed model for tumor detection reaches 0.89, while the confidence levels for YOLOv11, YOLOv8, YOLOv5, Faster-rcnn, and SSD are 0.83, 0.77, 0.80, 0.75, and 0.69, respectively.

Table 1 presents total comparison between the proposed model and other models in breast tumor detection tasks. The proposed model shows a marked advantage in breast tumor detection tasks, providing more reliable and accurate detection results for clinical diagnosis.

Table 1. Model Performance Results Comparison Table

Models	Map50	Map50.90	FPS	Accuracy	Recall
Faster.rcnn	0.816	0.336	12.86	0.7390	0.7559
SSD	0.797	0.324	54.61	0.6208	0.6598
YOLOv5	0.804	0.528	40.90	0.8341	0.8415
YOLOv8	0.812	0.524	42.30	0.8095	0.8190
YOLO11	0.824	0.532	32.20	0.8354	0.8484
Our Model	**0.876**	**0.584**	**28.40**	**0.8736**	**0.8812**

This indicates that the proposed model captures tumor features more precisely and can identify tumor types with greater confidence. The detection boxes of the proposed model are closer to the actual tumor boundaries, effectively reducing false positives and false negatives.

3.2 Ablation Study

Ablation experiments were carried out to analyze the individual contributions of RFAConv, EUCB, LSKA to the proposed model. The plot of the Map50 metric during the training process for the proposed model and models without some modules can be seen in Fig. 9.

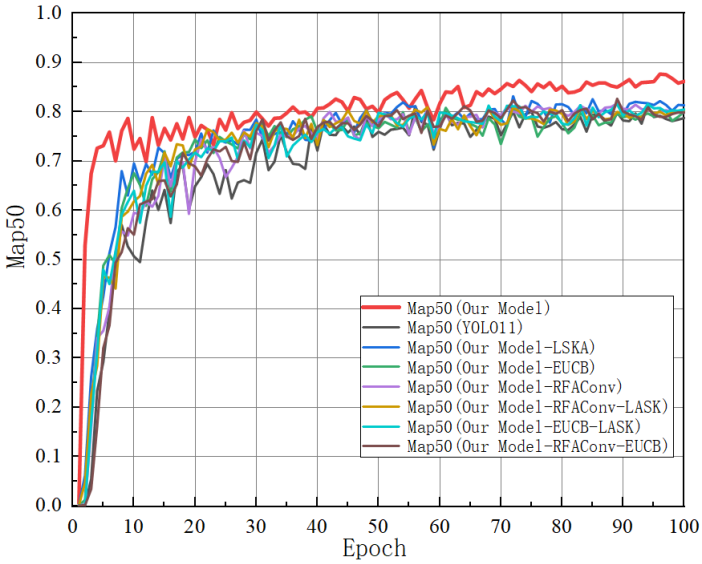

Fig. 9 .Map50 Metric Performance

In the first 10 epochs, the Map50 value of the proposed model quickly rises to about 0.75, while the values for Our Model-RFAConv-LASK, Our Model-RFAConv-EUCB, Our Model-LSKA-EUCB, Our Model-RFAConv, Our Model-LSKA, Our Model-

EUCB, and YOLO11 are 0.50, 0.40, 0.45, 0.55, 0.65, 0.60, and 0.70, respectively. At the final stage (Epoch 100), the proposed model's Map50 value stabilizes at about 0.88, while the values for Our Model-RFAConv-LASK, Our Model-RFAConv-EUCB, Our Model-LSKA-EUCB, Our Model-RFAConv, Our Model-LSKA, Our Model-EUCB, and YOLO11 are 0.68, 0.62, 0.65, 0.72, 0.78, 0.75, and 0.82 respectively.

The inference results comparison of the proposed model with ablation experimental models in breast tumor detection tasks is demonstrated in Fig. 10. The proposed model outperforms other ablation experimental models in terms of detection accuracy and confidence.

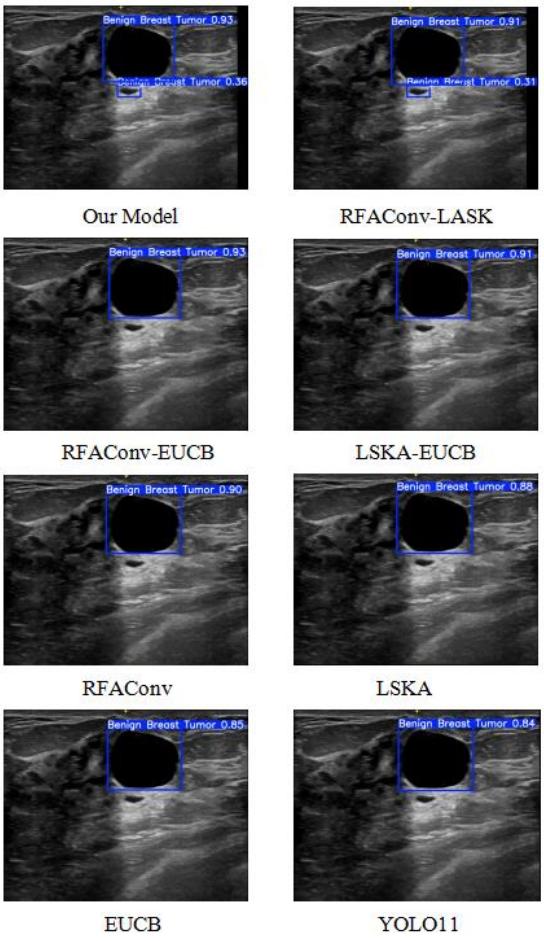

Fig. 10 .Ablation Study Inference Effect Diagram

The confidence of the proposed model for benign tumor detection reaches 0.93, while the confidence levels for Our Model-RFAConv-LASK, Our Model-RFAConv-EUCB, Our Model-LSKA-EUCB, Our Model-RFAConv, Our Model-LSKA, Our Model-

EUCB, and YOLO11 are 0.91, 0.93, 0.91, 0.90, 0.88, 0.85, and 0.84, respectively. This indicates that the proposed model captures tumor features more precisely and can identify tumor types with greater confidence.

Table 2. Results of Ablation Study

Models	Map50	Map50-90	FPS
-EUCB	0.827	0.525	28.5
-LSKA	0.829	0.542	29.3
-RFAConv	0.848	0.558	31.5
-LSKA-EUCB	0.831	0.546	30.4
-RFACon-	0.828	0.545	32.1
vLASK	0.83	0.538	32.3
-RFACon-	**0.876**	**0.584**	**28.4**
vEUCB			
Our Model			

The results shown in Table 2 indicate that, while using each technique alone improves model performance to some extent, integrating them yields significantly better data on Map50, Map50-90 and FPS. The experimental results fully demonstrate that the RFAConv, EUCB, LSKA technologies have significant advantages in breast tumor object detection. These methods significantly boost the model's feature extraction ability, address the limitations of convolution kernel parameter sharing, and enhance computational efficiency. In clinical practice, they enable more accurate and efficient breast tumor detection, offering robust support for early diagnosis and treatment.

Conclusion

This study, by integrating feature enhancement and dynamic loss techniques, proposes a model that incorporates three technological improvements in the feature enhancement, including RFAConv, EUCB, and LSKA technologies. After enhancing model performance and reducing computational costs, it integrates the UIoU loss function to improve model detection accuracy through dynamic box weighting strategies. Experimental results show that the proposed method yields a maximum 7.8% improvement on mean average precision (MAP) over the existing approach. Ablation experiments of each component have been performed, demonstrating optimal performance of the integrated model. Future research will explore the combination of the techniques in this study with the latest deep learning technologies to further enhance model performance; additionally, we will further investigate the application in the diagnosis of other diseases, such as breast cysts, to assess its universality and effectiveness in disease detection.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

This study was supported by Guangdong Basic and Applied Basic Research Foundation (2023A1515012783) of China.

Competing interests

The authors declare no competing interests.

References

[1] Umamaheswari, T., Babu, Y.M.M.: CNN-FS-IFuzzy: A New Enhanced Learning Model Enabled by Adaptive Tumor Segmentation for Breast Cancer Diagnosis using 3D Mammogram Images. Knowledge-Based Systems 290, 111443 (2024)
[2] Huang, H., Pedrycz, W.: Review of medical image processing using quantum-enabled algorithms. Artificial Intelligence Review (2024)
[3] Alirezazadeh, P., Dornaika, F., Charafeddine, J.: Mises-Fisher similarity-based boosted additive angular margin loss for breast cancer classification. Artificial Intelligence Review (2024)
[4] Singh, G., Kamalja, A., Patil, R., Karwa, A., Tripathi, A., Chavan, P.: A comprehensive assessment of artificial intelligence applications for cancer diagnosis. Artificial Intelligence Review (2024).
[5] Abd El-Mawla, N., Berbar, M.A., El-Fishawy, N.A., El-Rashidy, M.A.: A novel deep learning approach (Bi-xBcNet-96) considering green AI to discover breast cancer using mammography images. Neural Computing and Applications (2024).
[6] Yi, S., Chen, Z., She, F., Wang, T., Yang, X., Chen, D., Luo, X.: IDC-Net: Breast cancer classification network based on BI-RADS 4. Pattern Recognition 150, 110323 (2024)
[7] Addo, D., Zhou, S., Sarpong, K., Nartey, O.T., Abdullah, M.A., Ukwuoma, C.C., Al-antari, M.A.: A hybrid lightweight breast cancer classification framework using the histopathological images. Biocybernetics and Biomedical Engineering 43(4), 1257–1270 (2023)
[8] Munshi, R.M., Cascone, L., Alturki, N., Saidani, O., Alshardan, A., Umer, M.: A novel approach for breast cancer detection using optimized ensemble learning framework and XAI. Image and Vision Computing (2024), 104910.
[9] He, Q., Yang, Q., Xie, M.: HCTNet: A hybrid CNN-transformer network for breast ultrasound image segmentation. Computers in Biology and Medicine 153, 106629 (2023)

[10] Yu, C., Wang, Y., Tang, C., Feng, W., Lv, J.: EU-Net: Automatic U-Net neural architecture search with differential evolutionary algorithm for medical image segmentation. Computers in Biology and Medicine 107579 (2023)
[11] Thakur, N., Kumar, P., Kumar, A.: Reinforcement learning (RL)-based semantic segmentation and attention based backpropagation convolutional neural network (ABB-CNN) for breast cancer identification and classification using mammogram images. Neural Computing and Applications (2024), 1-18 (2024)
[12] Afrifa, S., Varadarajan, V., Zhang, T., Appiahene, P., Gyamfi, D., Mensah Gyening, R.O., Mensah, J., Berchie, S.O.: Deep learning based capsule networks for breast cancer classification using ultrasound images. Current Cancer Reports (2024)
[13] Atrey, K., Singh, B. K., Bodhey, N. K.: Integration of ultrasound and mammogram for multimodal classification of breast cancer using hybrid residual neural network and machine learning. Image and Vision Computing 43(5), 104987 (2024)
[14] Lu, S.-Y., Wang, S.-H., Zhang, Y.-D.: BCDNet: An Optimized Deep Network for Ultrasound Breast Cancer Detection. IRBM 44(4), 1–10 (2023)
[15] Thakur, N., Kumar, P., Kumar, A.: A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities. Multimedia Tools and Applications (2023)
[16] Harshvardhan, G.M., Mori, K., Verma, S., Athanasiou, L.: Machine learning applications in breast cancer prediction using mammography. Image and Vision Computing (2024).
[17] Rama, P.: FA-WSI-CNN Model for Predicting Breast Cancer using Deep Learning. International Journal on Recent and Innovation Trends in Computing and Communication 11(9), 1-10 (2023)
[18] Yang, L., Zhang, B., Ren, F., Gu, J., Gao, J., Wu, J., Li, D., Jia, H., Li, G., Zong, J., Zhang, J., Yang, X., Zhang, X., Du, B., Wang, X., Li, N.: Rapid Segmentation and Diagnosis of Breast Tumor Ultrasound Images at the Sonographer Level Using Deep Learning. Bioengineering 10(10), 1220 (2023)
[19] Al-Dhabyani W, Gomaa M, Khaled H, et al. Dataset of breast ultrasound images[J]. Data in Brief, 2020, 28: 104863.dib.2019.104863.