Electronic Research Archive

ERA, 32(10): 5846–5867.
DOI: 10.3934/era.2024270
Received: 30 July 2024
Revised: 05 October 2024
Accepted: 17 October 2024
Published: 29 October 2024

http://www.aimspress.com/journal/ERA

Research article

Breast mass lesion area detection method based on an improved YOLOv8 model

Yihua Lan[1,2,*], Yingjie Lv[1,2], Jiashu Xu[1,2], Yingqi Zhang[1,2] and Yanhong Zhang[1,2]

[1] School of Artificial Intelligence and Software Engineering, Nanyang Normal University, Nanyang 473061, China
[2] Henan Engineering Research Center of Intelligent Processing for Big Data of Digital Image, Nanyang 473061, China

* **Correspondence:** Email: yihualan@nynu.edu.cn; Tel: +8618568782511.

Abstract: Breast cancer has a very high incidence rate worldwide, and effective screening and early diagnosis are particularly important. In this paper, two improved You Only Look Once version 8 (YOLOv8) models, the YOLOv8-GHOST and YOLOv8-P2 models, are proposed to address the difficulty of distinguishing lesions from normal tissues in mammography images. The YOLOv8-GHOST model incorporates GHOSTConv and C3GHOST modules into the original YOLOv8 model to capture richer feature information while using only 57% of the number of parameters required by the original model. The YOLOv8-P2 algorithm significantly reduces the number of necessary parameters by streamlining the number of channels in the feature map. This paper proposes the YOLOv8-GHOST-P2 model by combining the above two improvements. Experiments conducted on the MIAS and DDSM datasets show that the new models achieved significantly improved computational efficiency while maintaining high detection accuracy. Compared with the traditional YOLOv8 method, the three new models improved and achieved F1 scores of 98.38%, 98.8%, and 98.57%, while the number of parameters reduced by 42.9%, 46.64%, and 2.8%. These improvements provide a more efficient and accurate tool for clinical breast cancer screening and lay the foundation for subsequent studies. Future work will explore the potential applications of the developed models to other medical image analysis tasks.

Keywords: YOLOv8; deep learning; breast cancer; target detection; convolutional neural network

1. Introduction

Breast cancer is one of the most common tumors in women, and its incidence rate is increasing annually. According to the latest global malignant tumor burden data released by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) for 2020, the number of new breast cancer cases in China is 420,000, accounting for 19.9% of all malignant tumors in women [1]. Early diagnosis is essential for reducing breast cancer mortality, and mammography has been shown to be an effective method for detecting early-stage tumors.

Mammography images have the following characteristics:

1) Due to the intertwining of diseased and normal tissues, the gradual transition of the diseased area into normal tissue, the blurring of the edges of the diseased area, and the low degree of contrast between the two types of areas, the boundary between diseased and normal tissues in the breast structure is unclear. Owing to the employed shooting angle, the three-dimensional tissue structure is projected onto a two-dimensional plane, which makes the lesions appear superimposed on each other. In this scenario, it is difficult to distinguish boundaries with the naked eye.

2) When taking mammogram images, photon noise introduces a certain amount of image noise, which hinders the accuracy of the subsequent image recognition task.

3) One of the major problems faced by mammography images is their lack of clarity, mainly due to the small percentage of diseased tissue contained in the images, resulting in overlapping glandular cells. In addition, lesion areas are of various sizes, are close in color to the surrounding lines, and have very low contrast levels, further reducing the overall clarity of mammogram images. When performing deep model training, features unrelated to target identification may be captured, posing several challenges to the deep learning training process, such as reducing the diagnostic accuracy achieved in real-world applications.

Accurately interpreting mammograms is still a challenging task for radiologists, and avoiding missing lesions while not misclassifying normal tissue as abnormal lesions is difficult and time-consuming. Therefore, reducing false positives and improving sensitivity have become important issues to be addressed in mammography.

In recent years, the rapid development of computer-aided diagnosis (CAD) technology has provided a new way to solve this problem. However, the existing CAD systems still face many challenges in terms of processing mammography images, and the problems of low detection sensitivity and high false-positive rates make radiologists skeptical of the accuracy and reliability of computer-aided diagnostic systems when they are used for mass identification.

2. Previous work

Mahoro and her team [2] used two models, YOLOv7 and YOLOv8, on the VinDr-Mammo dataset to test different data augmentation techniques: contrast-limited adaptive histogram equalization, median filtering, and bilateral filtering. The results showed that the average accuracy (mAP) of YOLOv8 is 0.65, better than the 0.53 of YOLOv7. Intasam and colleagues [3] compared three different optimizers [the stochastic gradient descent (SGD) optimizer, the adaptive moment estimation (Adam) optimizer, and the AdamW optimizer with weight decay regularization] on the YOLOv5s model and tested them using a dataset containing six categories (benign mass, malignant mass, benign calcification, malignant calcification, benign associated features, malignant associated

features). The results showed that when using the SGD optimizer, the mAP of the model reaches 0.91, the precision is 0.92, and the recall is 0.85. Gao and his team [4] used mask-RCNN and YOLOv3 to detect squamous cell carcinoma from esophageal endoscopy videos. The average accuracies of the YOLOv3 classification and detection processes were 0.85 and 0.74, respectively, and for mask-RCNN, the accuracy of segmentation was 0.63. However, the dataset used in this study was small, resulting in uncertainty in the general applicability of the results. Yang and his team [5] proposed a new method for breast cancer pathology image classification based on deep learning and wavelet transform. Breast cancer pathology images were classified via the YOLOv8 network model, and the classification accuracy of the proposed method was compared with that of YOLOv8 on the original BreaKHis dataset. It was found that the developed algorithm was able to achieve improved classification accuracy for images with different magnification levels.

Al-Antari and his team [6] proposed an integrated CAD system for the detection and classification of lesion areas. First, the YOLO model was used to detect the lesion area in the entire mammography. Then, three classification models were used: regular feedforward CNN, ResNet50, and InceptionResNet-V2 for classification. The CAD system achieved significant improvements in the improved classification module. Al-masni and his team [7] also proposed a CAD system for detection and classification based on the YOLO model. The results showed that the overall accuracy of detection was 96.33%, and the overall accuracy of classification was 85.52%. Kassahun and his team [8] designed a system for the detection, segmentation, and classification of breast lumps, in which the detection phase used the YOLO model for the initial detection of breast lumps. In the publicly available RadImageNet dataset, using DenseNet-121 combined with the YOLOv5m model, the IoU threshold was 0.5, and the mAP was 0.718. Touazi and his team [9] designed a row detection and segmentation system using the CBIS-DDSM dataset, in which YOLOv5, V7, and V8 models were used for object detection and the ViT Nest-based SegNest architecture was used for breast cancer mass segmentation. The detection accuracy of the YOLOv8 m model is 59%, and the Dice loss of the SegNest model is 90.15%. This showed enhanced diagnosis of breast lesions, improving detection efficiency and accurate early detection methods.

Despite the progress made in these studies, several limitations remain. For example, some studies have failed to significantly achieve improved detection accuracy despite optimizing the YOLO model; other studies have embedded YOLO into more complex systems, such as those that combine semantic segmentation, instance segmentation, and even image classification tasks, which has broadened the application scope but indirectly exposed the accuracy limitations of YOLO itself.

Based on an in-depth analysis of existing studies, this study proposes three core innovative strategies aimed at improving the performance of the YOLO algorithm in breast cancer detection tasks.

1) YOLOv8-GHOST model: In the previously used CNN models for recognizing lumps in mammogram images, it is often necessary to utilize many convolutional layers for performing superposition operations to achieve higher recognition accuracy. Although increasing the number of convolutional layers can enhance the recognition ability of the employed model, which is beneficial for improving the recognition effect of the model [10], adding more convolutional layers makes the model demand more training data, the computational power requirement increases, and the complexity of the model training process increases significantly [11]. Through reparameterization, the YOLOv8-GHOST model enables the entire architecture to achieve results similar to those of more complex models, with nearly half the number of parameters required by the original model.

2) YOLOv8-P2 model: In mammogram radiography images, the percentage of breast lesion

areas is very small. General neural network models have low accuracy in terms of recognizing such small and complex targets. The YOLOv8-P2 model is able to generate a series of feature maps in addition to performing the traditional feature extraction process, which can be considered a high-dimensional extension or diverse representation of the original features. It is easier to capture subtle changes and detailed features that are difficult to recognize directly in tumor images, and the proposed approach improves the ability to capture small and complex targets.

3) YOLOv8-GHOST-P2 model: The depth of a network should be proportional to the size of the dataset used for training [12]. Convolutional operations with multiple convolutional kernels can be implemented on the input image data; although it is possible to separately extract the low-level and high-level features of images, many of the obtained features will show redundancy, which will undoubtedly cause problems such as underfitting or overfitting during training. The YOLOv8-GHOST-P2 model combines the first two models to greatly reduce its computational and memory footprints while maintaining and improving accuracy. The response rate is particularly critical in medical image processing, as medical image data are often voluminous, and real-time diagnosis requires fast responses.

The main goal of this study is to significantly improve the accuracy and efficiency of breast cancer detection through these innovative models, providing a more reliable tool for aiding clinical diagnosis tasks. We expect that these improvements will promote the application of CAD technology in the early diagnosis of breast cancer and ultimately improve the survival rate and quality of life of patients.

3. Related work

3.1. Target detection

Object detection techniques are designed to automatically detect and localize all objects of interest from image data and simultaneously determine the categories to which they belong, as well as their precise spatial locations.

Target detection methods can be categorized into two types according to their processes: one-stage and two-stage approaches. *One-stage approach* refers to the simultaneous prediction of the location and category of an object directly from the input image without pre-generating candidate regions (region proposals). A unified neural network model is used to perform bounding box regression and category classification, thus significantly accelerating the detection process. Representative algorithms include YOLO and the single-shot detector (SSD). In contrast, the *two-stage approach* is more complex, as it first generates a set of candidate regions that may contain objects and then performs a detailed analysis of these regions, including fine-grained categorization and bounding box adjustment. This process is usually accomplished by two separate modules; the first, called a region proposal network (RPN), is responsible for generating high-quality candidate regions, and the second module performs classification and bounding box fine-tuning for each candidate region. Representative two-stage algorithms include the region-based convolutional neural network (R-CNN) family (e.g., fast R-CNN and faster R-CNN). Each of these two approaches has its own characteristics, with single-stage detection being slightly less accurate than two-stage detection, although much faster.

An R-CNN is a CNN model that was originally created specifically for target detection. In the

R-CNN family of models, such as the vanilla R-CNN and fast R-CNN, a selective search (SS) algorithm is used to perform candidate box extraction. SS is a method based on traditional target detection algorithms that involves extracting candidate regions in an image and then classifying and regressing these candidate regions to finally obtain target detection results. The advantage of this method is that it is fast, but its accuracy is relatively low. Zoph and his team [13] have proposed the architecture search concept, which yields improved target detection accuracy by finding the optimal or near-optimal network architecture through a search space, a search strategy, and a performance evaluation. The advantage of this approach is that it reduces the burden of manually designing network architectures for RPNs and automatically discovers superior models to attain improved performance.

The CNN family performs well in medical image classification tasks but requires considerable data preparation and image processing work. Cai and his team [14] proposed an improved R-CNN algorithm for detecting mitosis in breast cancer histology images. Compared with previous studies, better results were obtained on the MICCAI TUPAC 2016 dataset. The relevant data in an image is called the region of interest (ROI). Extracting the ROI from an image is the most time-consuming step in medical image processing, and even today, it is mainly a manual task. The faster R-CNN, which was proposed in 2016 [15], improves upon the fast R-CNN by introducing an RPN instead of SS specifically for extracting candidate frames. The RPN significantly improves the efficiency of the candidate region extraction process in the target detection framework and estimates the likelihood of the presence of a target in each proposed frame and the location of the frame in parallel. Mahmood and his team [16] proposed a faster R-CNN-based multistage mitotic cell detection method, and good results were produced on the ICPR and ICPR 2014 datasets. In addition, the concept of anchors (anchor boxes) was introduced to generate multiple bounding boxes with different scaling ratios and aspect ratios centered on each pixel; these bounding boxes were called anchor boxes, which have been used in later models.

3.2. YOLO

First proposed by Redmon and others [17] in 2016, YOLO introduced the concept of one-stage object detection. A one-stage target detection method extracts features directly in a network to predict the class and location of the target object; this type of approach can recognize and classify objects at once, making it faster than other networks used for object detection. Unlike other object detection networks, YOLO looks at the entire input image and learns its context. The method treats the object detection task as a regression problem, where the input image is first divided into $N \times N$ grid cells, and each grid is responsible for detecting the object whose center is located in that area. Each grid cell predicts a fixed number of bounding boxes for the object by resizing a predetermined number of anchor boxes with different aspect ratios. For each grid, n bounding boxes and their corresponding confidence scores are predicted. A confidence score measures the probability of the presence of an object within the corresponding bounding box and the accuracy of the bounding box. Additionally, each grid predicts a probability distribution of m categories, indicating the likelihood that an object within that grid belongs to each category. For each object that may be predicted by multiple grids or bounding boxes, the best bounding box prediction is ultimately filtered via non-maximum suppression (NMS) to reduce the number of duplicate detections.

In addition to the use of YOLO for breast cancer detection in this study, YOLO can also be employed for the detection of other conditions. Sindhu and colleagues [18] used the YOLO-based

DetectNet architecture to detect nodules in lung computed tomography (CT) scans. Utilizing a single convolutional network to simultaneously predict multiple bounding boxes, experiments have shown that a single neural network for nodule detection can produce a fairly low false-positive rate while achieving high sensitivity and accuracy. Ünver and his team [19] proposed a skin lesion segmentation network that combines YOLO and the GrabCut algorithm. The network first removes hairs from lesions, and this step is followed by lesion location detection and segmentation. A 90% sensitivity was achieved on the ISBI 2017 dataset, outperforming other deep learning–based methods. MedYOLO [20] is a 3D medical image analysis framework based on YOLOv5, which specializes in the segmentation and localization of organs in complex medical scans. In addition, with its implementation of target tracking features, YOLO can identify and track the locations of anatomical structures in real-time during surgery, improving the accuracy and safety of surgery.

The YOLO algorithm surpasses faster R-CNN with its efficient real-time monitoring capabilities, end-to-end streamlined process, low resource requirements, lightweight nature for easy debugging and deployment, and accurate localization of the global field of view, which not only accelerates the disease diagnosis process but also broadens the application boundaries of medical technology and brings substantial benefits to healthcare. Thus, YOLO is well-suited for the real-time processing of large-scale medical image data.

4. The proposed method

4.1. Model structure design

The detection model proposed in this paper is obtained by improving YOLOv8. YOLOv8 is divided into two main parts: a backbone network and a head. These two parts also refer to the efficient latency approximation network (ELAN) in YOLOv7, which is a method used to accelerate the inference processes of neural networks by sacrificing a certain amount of accuracy. The ELAN can significantly reduce the imposed computational and memory requirements while keeping the utilized model as accurate as possible, thus improving the real-time inference speeds of neural networks. In the backbone network, YOLOv8 replaces the C3 structure of YOLOv5 with a C2f structure that is richer in gradients and adjusts the number of channels for models with different scales. For the head, YOLOv8 switches to a decoupled head structure, which separates the classification and detection heads and switches from the anchor-based mode to the anchor-free mode.

As mentioned earlier, breast cancer lesions often share similar morphological and density characteristics with normal breast tissue, leading to low contrast between the lesion areas and the surrounding healthy tissue. This challenge is particularly evident when the lesion gradually transitions into normal tissue, causing the boundaries to become blurred and indistinct. To address this issue, we introduced the GHOST module, which effectively enhances the contrast between the target areas (i.e., lesion regions) and the background (i.e., normal tissue), thereby improving the accuracy of lesion detection. The GHOST module was originally proposed by Han et al. [21]. It is a lightweight feature extraction module that is capable of capturing features that are more likely to be overlooked in breast tissues but are similar to normal tissues by introducing global and local information, thus improving the resulting detection accuracy. The model skillfully integrates the powerful learning capabilities of deep learning and an adaptive feature selection strategy, which enables it to not only effectively differentiate between masses but also maintain high degrees of

sensitivity and specificity when confronted with highly similar tissue features, thereby greatly improving the accuracy and efficiency of breast lesion detection.

Adequate or redundant feature layer information ensures a comprehensive understanding of the input data, and many similarities are observed among different feature layers. Considering that redundant mammogram image information may have an impact on the model, we do not choose to remove these redundancies when designing the lightweight model but rather opt to obtain redundant information with a lower computational cost. For a given feature layer, only some of the true features are generated via convolutional operations, and the remaining similar features are obtained by conducting linear operations on the true feature layer. Finally, the true and similar feature layers are spliced together to form a complete feature layer. In terms of computational complexity, this approach is much less costly than depthwise separable convolution, which uses depth convolution to process the spatial information contained in each feature channel and then uses point convolution for inter-channel feature fusion. Figure 1 shows the introduced GHOST module.

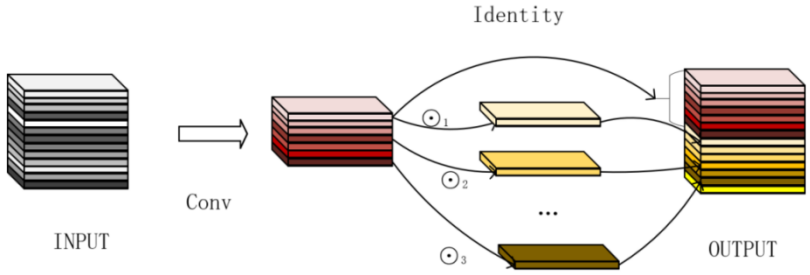


Figure 1. GHOST module.

As shown in Figure 1, the feature input map is decomposed into two parts: a base feature map, which is directly used as the output after implementing the standard convolution operation, and multiple generated feature maps, which are obtained by performing a series of element-level operations, such as point-by-point convolution, on the base feature map. Finally, the base feature maps and the generated GHOST feature maps are combined to form the final output feature maps. By generating a small number of base feature maps and many GHOST feature maps, the diversity of the observed features can be maintained. The model generates additional feature maps through simple linear combination and replication operations, which reduce the number of required standard convolutions, significantly reduce the computational cost and the number of parameters of the model, and make the network run faster and exhibit higher resource efficiency.

The parameter calculations of the GHOST module are given below. The input is set to h × w × c, the output is set to h' × w' × n, the size of the convolution kernel is set to k, and the linear operation in the GHOST module is a deep convolution. The number of ordinary convolution calculations is shown in Eq (1).

$$\text{cost} = h' \times w' \times n \times k \times k \times c \tag{1}$$

The GHOST module has the number of calculations shown in Eq (2).

$$\text{Cost} = h' \times w' \times \frac{n}{s} \times k \times k \times c + (s-1) \times h' \times w' \times \frac{n}{s} \times k \times k \tag{2}$$

The compression rate is shown in Eq (3).

$$r = \frac{n \times c}{\frac{n}{s} \times c + (s-1) \times \frac{n}{s}} = \frac{s \times c}{s + c - 1} \approx s \tag{3}$$

s denotes a parameter in the GHOST module that regulates the balance between its computational cost and model complexity. It is a positive integer for characterizing the number of repeated linear operations executed in the module. When s = 1, the module degenerates to a normal convolution; as s increases, the computational cost decreases accordingly. The three improved YOLOv8 models proposed in this paper are based on the GHOST module.

4.2. YOLOv8-GHOST

Figure 2 shows the overall structural framework of YOLOv8-GHOST. As shown in Figure 2, the overall YOLOv8-GHOST model is divided into three main parts, where the backbone part is mainly responsible for generating predictions, including bounding boxes, category probabilities, and confidence levels. Owing to the way GHOST convolution generates redundant features, the number of required convolution kernels is reduced, which decreases the number of network parameters and significantly improves its inference speed, making the whole network more efficient for quickly inferring large amounts of image data. Despite the reduction in the number of parameters, the model is able to more efficiently capture and preserve the key features of breast cancer lesion regions versus those of other regions while still accurately identifying and distinguishing lesion regions even in the simplified network.

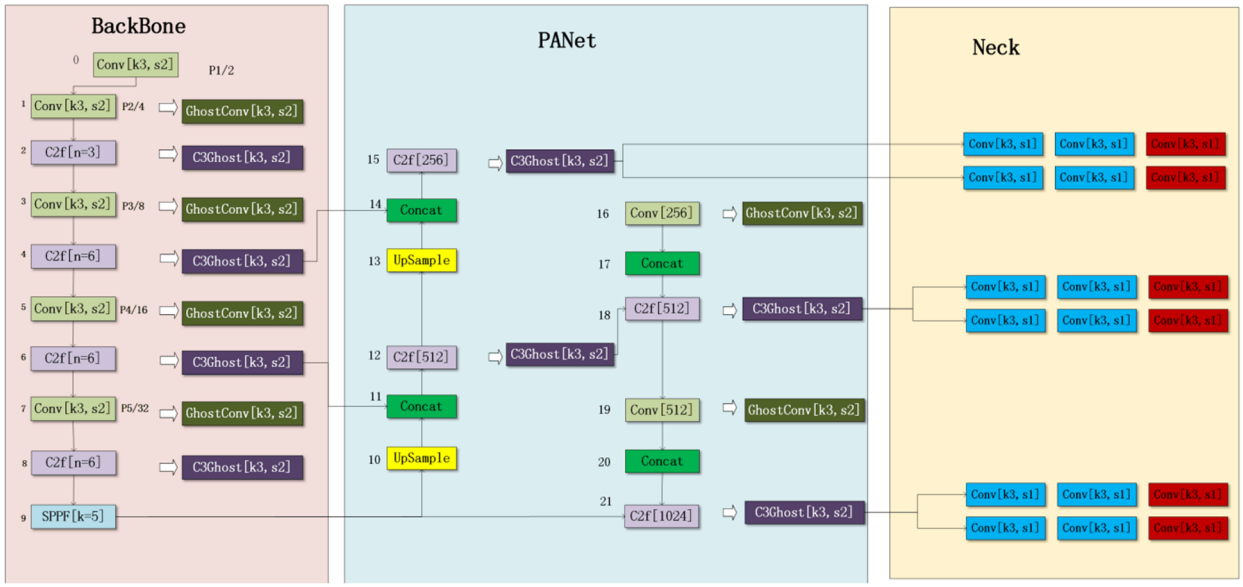


Figure 2. The YOLOv8-GHOST model.

The initial layer uses a Conv layer with 64 output channels and a convolution kernel size of 3 × 3 with a step size of 2 for the initial downsampling process. The subsequent convolutional layers are replaced with GHOST convolutions for downsampling, and the number of channels is gradually increased according to different feature pyramid levels. After each downsampling layer, the C2f layer

of YOLOv8 is replaced with a C3GHOST module to enhance the feature representations, and a GHOST-bottleneck connection is used, as shown in Figure 3.

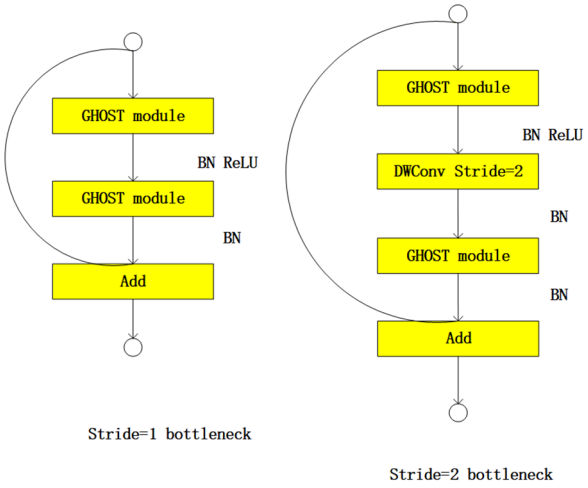


Figure 3. GHOST-bottleneck connection.

As shown in Figure 3, the GHOST-bottleneck connection consists of two main GHOST modules; the first module is used as an expansion layer to increase the number of channels, whereas the second module is used to decrease the number of channels to match the number of short channels. After four downsampling steps, the SPPF layer of the original YOLOv8 model is used for multiscale feature fusion to enhance the ability of the model to detect objects with different sizes.

The C3GHOST module is an enhanced version of the original C3 module that incorporates a GHOST convolution module. The C3 module was originally designed to build lightweight semantic segmentation networks that provide good feature representations with low computational complexity by integrating basic components such as convolutional layers, batch normalization, and activation functions. GHOST convolution, on the other hand, increases the number of output channels possessed by the convolutional layer without increasing the incurred computational cost. It generates a small number of real filters from a standard convolutional layer and then generates more "virtual" filters via linear mapping. These virtual filters are linear combinations of real filters, so despite the increase in the number of output channels, the total number of parameters is greatly reduced compared with that of traditional convolution, making the entire network lighter.

The head network is responsible for performing upsampling, fusion, and finally object detection on the features extracted from the backbone network. The feature maps acquired at different scales are connected through upsampling via nearest-neighbor interpolation, and then another feature extraction process is performed using multiple C3GHOST modules. The feature maps are spliced between the different layers by means of a concatenation layer, and after finishing feature fusion, additional downsampling and feature enhancement processes are performed via GHOSTConv, which receives the feature maps from the three different layers as inputs in the final detection header section. The number of network parameters is greatly reduced to speed up the detection process and achieve improved accuracy.

A comparison among the different parameter quantities in the backbone components of the models is shown in Figure 4.

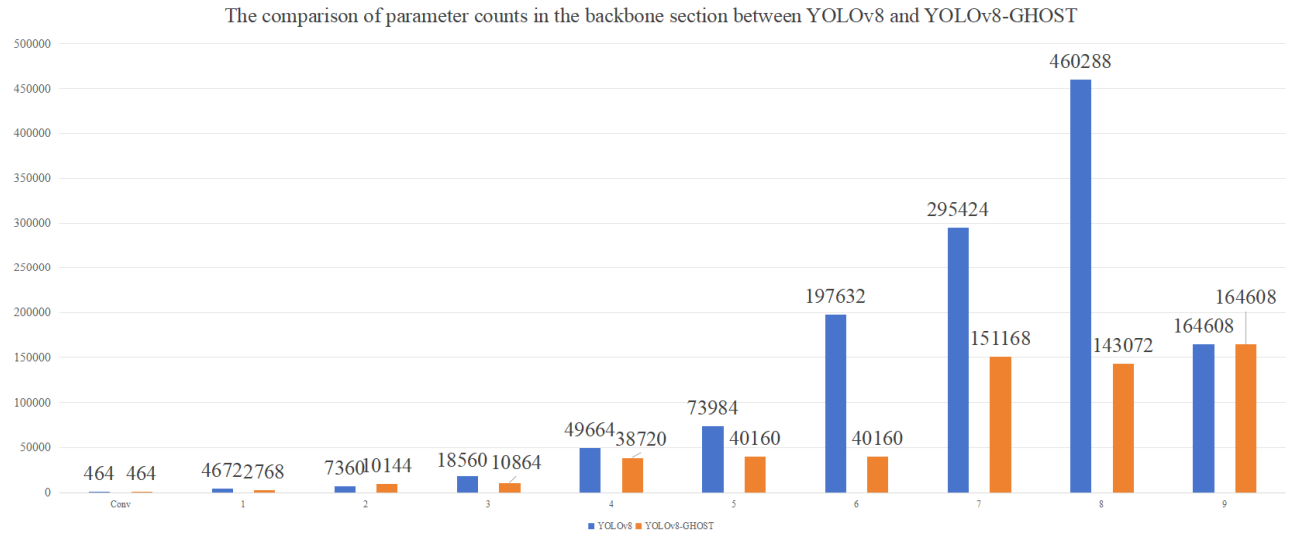


Figure 4. Comparison among the parametric quantities in the backbone sections of the YOLOv8 and YOLOv8-GHOST models.

Excluding the Conv in layer 0 and the SPPF layer in the last layer, YOLOv8 exhibits an exponential increase in the number of parameters from the first layer to the eighth layer. The odd-numbered layers are the introduced GHOSTConv module, and the even-numbered layers are the introduced C3GHOST module. Figure 4 clearly shows that the differences between the numbers of parameters in layers 1 and 2, layers 3 and 4, and layers 5 and 6 are extremely small. In contrast, the number of parameters in the C2f layer in YOLOv8 is several times the number of parameters in the C3GHOST layer.

4.3. YOLOv8-P2

This model is an improved YOLOv8 model with the addition of feature-level processing in the head module. The P2 architecture means that the output layer incorporates the information from the P2, P3, P4, and P5 layers and, compared with the conventional model, has an additional input from the P2 layer (4-fold downsampling). The P3 layer represents 23, i.e., 8-fold downsampling; the P4 layer represents 24, i.e., 16-fold downsampling; and the P5 layer represents 25, i.e., 32-fold downsampling. In mammograms, the percentages of diseased breast tissue are extremely small, and the model addresses this challenge by adding additional feature layers. P2 corresponds to an earlier feature extraction stage and has a relatively large size, which enables it to capture high-resolution image features. Particular attention is given to small targets, enabling rich multiscale detection and ensuring the accurate identification of lesion regions in complex breast tissues. This design not only improves the meticulousness of detection but also greatly enhances the accuracy of early breast cancer screening.

In YOLOv8 and YOLOv8-P2, most of the structures are the same; only the last few modules are different. The difference lies in the number of parameters in the detection module: in YOLOv8-P2, the number of parameters contained in the detection module is 617,496, and in YOLOv8, the number of parameters in the detection module is 751,702. This is because the YOLOv8 model considers fewer feature mapping scales (scale 32 is missing) but has a relatively large number of feature

mapping channels at each scale, which leads to an increase in the overall number of parameters. In the YOLO series or other similar multiscale detection models, the detection module is responsible for generating the final detection results, and its parameter count is closely related to the number of input feature maps and the number of channels per feature map. Therefore, despite the reduced input scales of the YOLOv8 model, more feature channels are retained per scale, leading to an increase in the overall number of parameters.

4.4. YOLOv8-GHOST-P2

This model combines the improved YOLOv8-GHOST and YOLOv8-P2 models from the previous section. In Figure 5, the upper structure is the head network of the YOLOv8-GHOST module, and the lower structure is the head network of the improved YOLOv8-GHOST-P2 module. Compared to the YOLOv8 model, the number of parameters is higher, about 53% of the original model, and the inference process is faster.

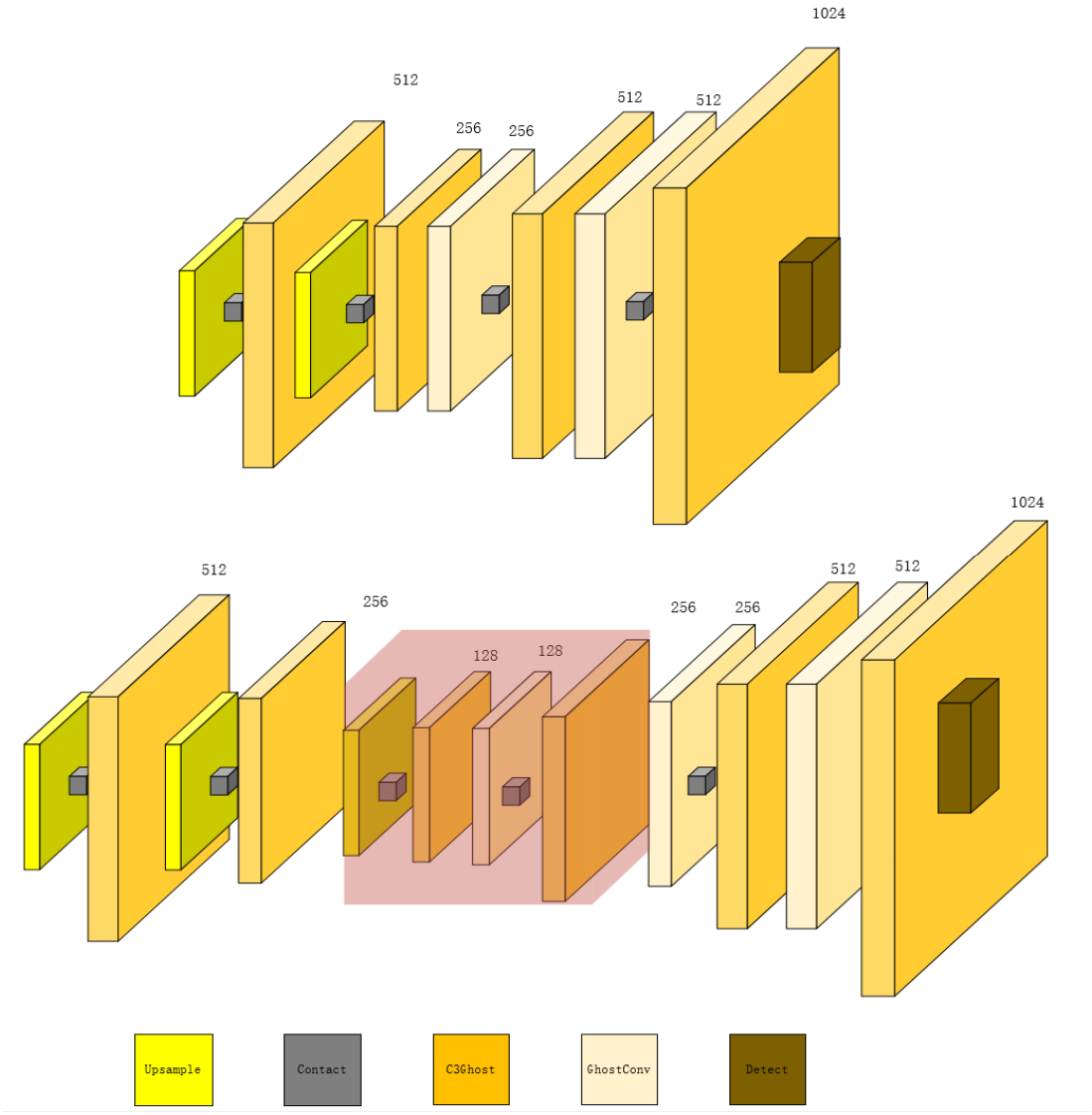


Figure 5. Head section of the improved YOLOv8-GHOST-P2 module.

5. Experimental results and analysis

This section describes the dataset and its processing, evaluation metrics, experimental setup, and its results.

5.1. Introduction to the dataset

5.1.1. MIAS

The MIAS dataset provides a data annotation file, which lists very detailed information about the data, including data indices and organizational features such as the background, abnormality categories, the severity levels of the abnormalities, the central coordinates of the abnormality locations, and the central radii of the abnormality sites. The MIAS dataset contains 322 images, each with a resolution of approximately 4000 × 2000 after decompression, among which 207 images are normal and lesion-free images and 114 images are abnormal; among these 115 abnormal images, 64 are benign lesions, and 51 are malignant lesions [22].

In YOLO, the labeled data corresponding to images are usually stored in TXT files, and each image corresponds to a TXT file, which records the bounding box information and category labels of all the target objects in that image and mainly contains the following columns of information, with each column of data separated by a space:

1) Class Index: This index indicates which category the given object belongs to within the predefined list of categories. Usually, categories are counted from 0, so if N categories are contained in a dataset, the index range is 0 to N–1.

2) Bounding Box Center X: This is the X coordinate of the center point of the bounding box of the target object, which is scaled relative to the width of the image. The value range is usually between 0 and 1, representing a percentage of the image width.

3) Bounding Box Center Y: Similarly, this is the y-coordinate of the center point of the bounding box, and it is also proportional to the height of the image.

4) Bounding Box Width Ratio: This is the ratio of the width of the bounding box to the width of the image, which is again between 0 and 1.

5) Boundary Box Height Ratio: This is the ratio of the height of the bounding box relative to the height of the image, and its value range is also 0 to 1.

For the MIAS dataset, we first transform the information contained in the raw data annotation files into a format that can be directly parsed by the YOLO algorithm. We then perform a median filtering operation on the data to remove the noise from the images. The brightness and contrast levels of the image are subsequently enhanced via histogram equalization. We flip the processed image vertically and horizontally, and to increase the number of input images, we rotate the image at random angles with 10-degree increments. Finally, we chose a total of 10,000 images as the training set, among which 5584 are benign and 4416 are malignant; 2321 images constitute the validation set, among which 1298 are benign and 1023 are malignant. The distribution of the dataset is shown in Figure 6 below. The image presentation process is represented in Figure 7.

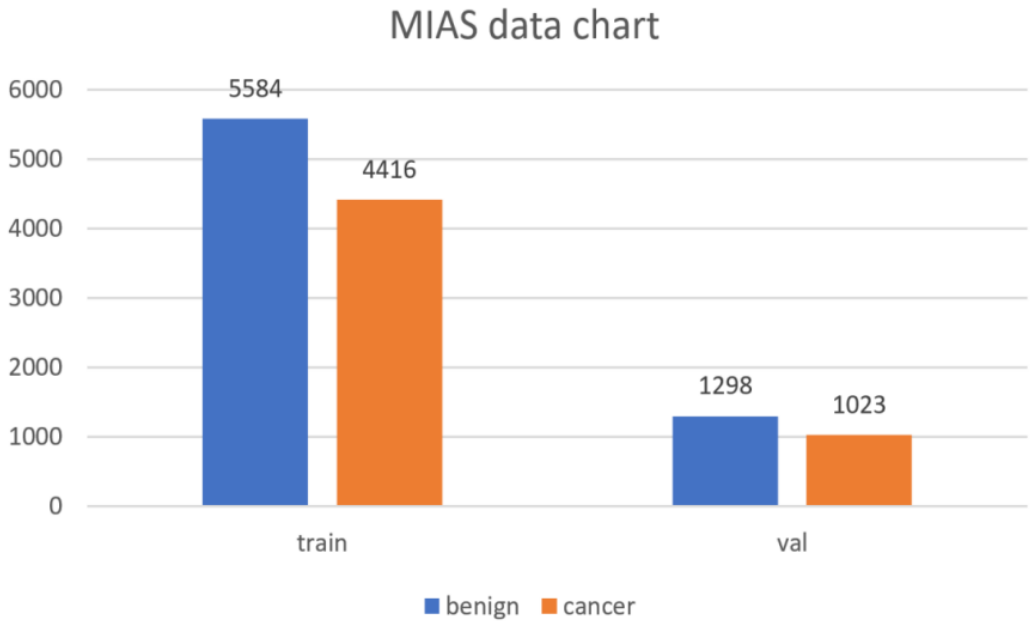


Figure 6. Distribution of the MIAS dataset.

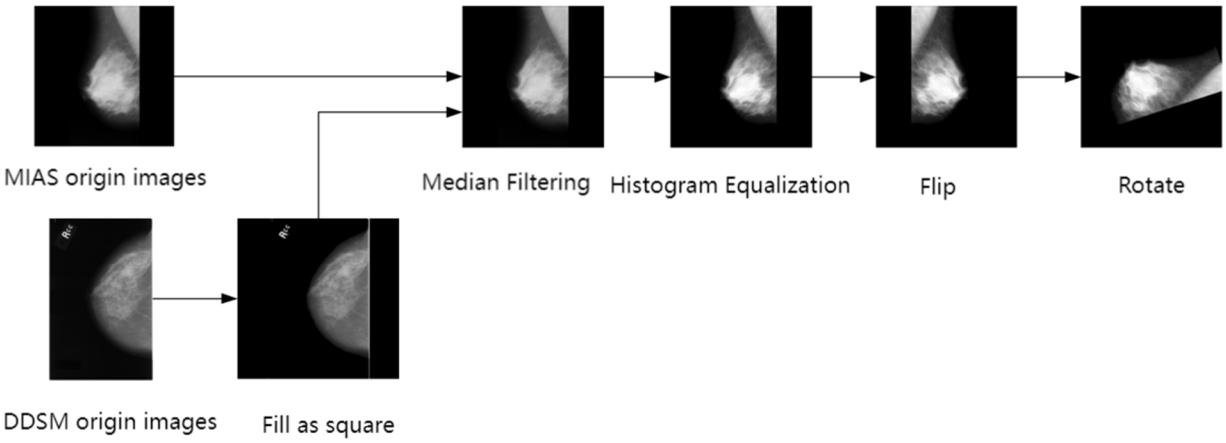


Figure 7. Flowchart of the image preprocessing strategy.

5.1.2. DDSM

Another dataset used in this thesis is the Digital Database for Screening Mammography (DDSM [23]). It includes four categories of data: cancer, normal, benign, and benign_without_callback. The DDSM contains 2620 examples, each of which has a view of the patient's left and right breasts at two different orientations: the lateral-canceled (CC) position, which shows the structure of the breast from the outside to the inside, and the non-lateral-canceled position, which is used to visualize the posterior part of the breast and the axillary region. The actual cases also detail the specifics of the

different breast lesions that were precisely labeled by a surgeon, such as the nature of each lesion, which may range from a microcalcified spot or a benign mass to a malignant mass, as well as the distinctive presenting features of these lesions and the lesion boundary information as outlined by the surgeon.

In this work, we mainly use the CC view because it is a projection of the breast in the vertical direction, which provides more consistent imaging angles and conditions across individuals and helps reduce noise due to the use of a differential image acquisition method. In addition, the CC view is clearer in terms of abnormalities located in the center region of a real breast. The DDSM contains 2391 CC views (935 malignant examples, 872 benign examples, and 584 normal examples), which are decompressed to a resolution of approximately 2400 × 4000. After screening, we obtain 917 malignant examples and 855 benign examples.

In this dataset, since each image has a different size, if the images are directly input into the training process, it will not only lead to image distortion but also greatly reduce the detection accuracy of the model. Therefore, we first individually fill the original images and transform them into squares. The subsequent steps are the same as those of the MIAS operation, and finally, we set the rotation angle to 72° for random flipping. Finally, we chose a total of 12,927 images as the training set, among which 5584 are benign and 4416 are malignant; 3151 images constitute the validation set, among which 1298 are benign and 1023 are malignant. The distribution of the dataset is shown in Figure 8 below. The overall process is shown in Figure 9.

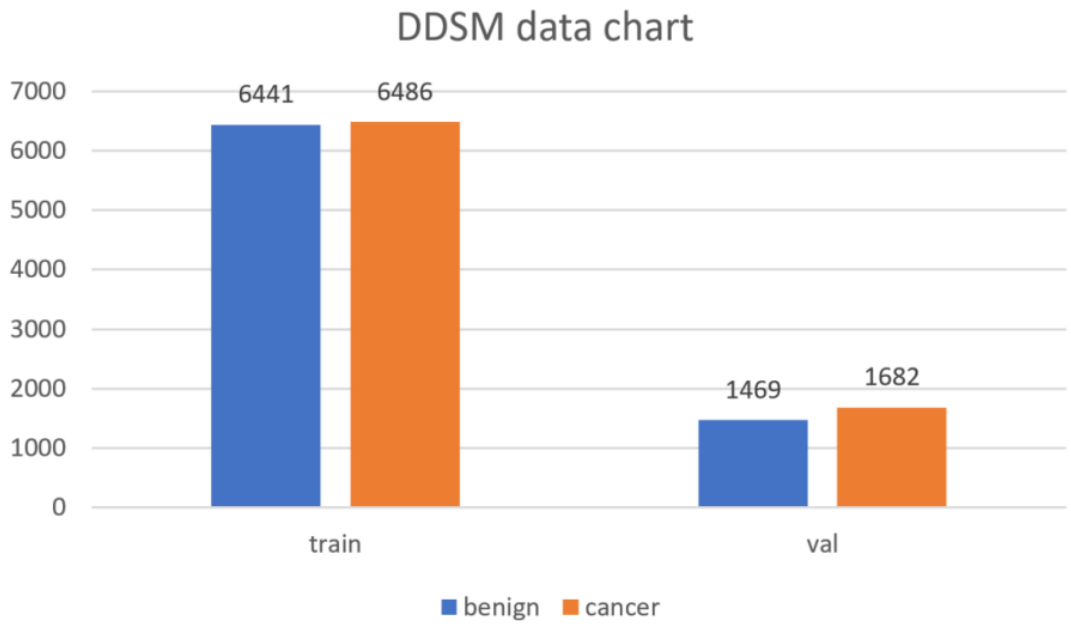


Figure 8. Distribution of the DDSM dataset.

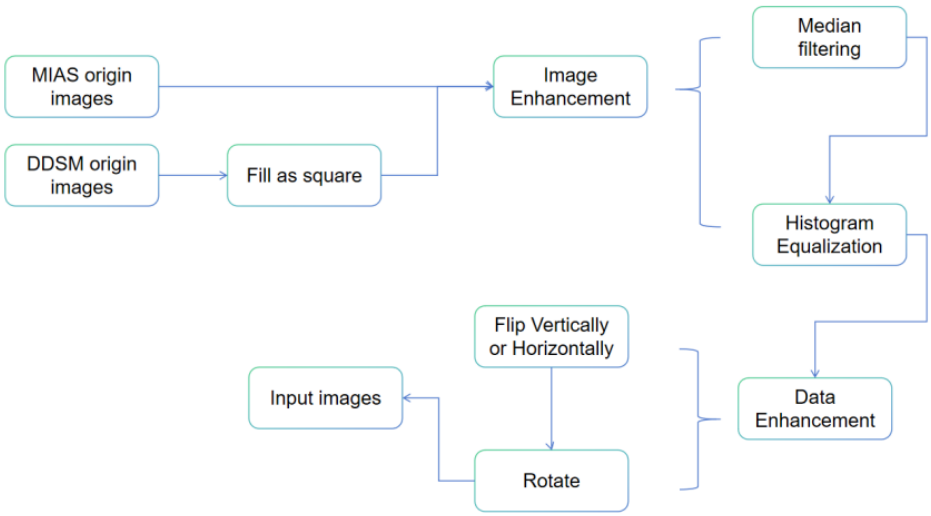


Figure 9. Flowchart of the image preprocessing strategy.

5.2. Evaluation methodology

In the experiment involving the identification of breast cancer using the YOLO model, we use several model evaluation metrics to assess the performance of the tested models. These metrics include average precision (AP), which is usually more convincing when the mAP is 50–95, recall, precision, the F1 score, and a loss function.

The AP refers to calculating the precision and recall achieved for a category under different confidence thresholds; when plotting the PR curve on the basis of the precision and recall values, the area under the PR curve is the AP. The mAP is obtained by averaging the APs achieved over all the different categories.

The formulas for F1 score, precision (PPV, the positive predictive value), and recall (TPR, the true-positive rate) are as follows:

$$F1 = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

A true positive (TP) denotes a positive sample that the model predicts as positive; a true negative (TN) denotes a negative sample that the model predicts to be negative. In target detection, due to the addition of background detection, a false positive (FP) denotes a negative sample that the model predicts as positive, i.e., an example that is predicted to be a background sample. A false negative denotes a positive sample that the model predicts as negative, i.e., a background sample that is predicted to be an example.

The values of the above evaluation indices are between 0 and 1. Higher index values indicate better detection performance for the tested model.

The loss function includes a classification loss (cls_loss), a localization loss (box_loss), and an object existence loss (obj_loss), and the loss function can directly reflect the classification accuracy, bounding box localization accuracy, and object existence judgment performance of each model. The smaller the loss value is, the better the corresponding model effect.

The categorization loss (cls_loss) is used to measure the difference between the category labels predicted by a model and the true labels, where different cells are responsible for predicting the categories of the objects in their regions. The formula for this function is shown in (7):

$$\text{cls\_loss} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(y_{ic}^{*}) \tag{7}$$

N denotes the number of samples, C denotes the number of categories, y_{ic} denotes the actual category label, and y_{ic}^{*} is the predicted probability.

The positioning loss (box_loss) is used to measure the difference between the predicted bounding box and the real bounding box in terms of the position and size dimensions. The associated formula is shown in (8):

$$\text{box\_loss} = \sum_{i=1}^{N}\text{SmoothL1}(b_i^{*} - b_i) \tag{8}$$

N denotes the number of predicted bounding boxes, b_i^{*}denotes the predicted bounding box coordinates, and bi denotes the real bounding box parameters.

The object presence loss (obj_loss) evaluates the accuracy of a model prediction regarding whether an object is present in a grid cell. Each grid cell has a confidence score that indicates the probability of the presence of an object in that cell. The corresponding formula is shown in (9):

$$\text{obj\_loss} = -\sum_{i=1}^{N}\left(t_i\log(t_i^{*}) + (1-t_i)\log(1-t_i^{*})\right) \tag{9}$$

N denotes the number of grid cells, t_i denotes the true target presence label (0 or 1, indicating the presence or lack of an object, respectively), and t_i^{*} denotes the predicted target presence probability.

5.3. Experimental setting

In this work, we experiment with the above network model to detect and recognize benign and malignant masses in breast cancer images. During the experiment, the initial learning rate lr0 is set to 0.01, the final learning rate lrf is set to 0.05, the momentum parameter is set to 0.95, the number of epochs is set to 150, and the optimizer is set to SGD. Since the resolution of each image in MIAS is much smaller than that in the DDSM, the batch size of MIAS is set to 90, and the batch size of the DDSM is 24. The experimental environment of this study includes an RTX 4090 GPU, the Python version is 3.10.12, and the PyTorch version is 2.0.1.

5.4. Experimental results and analysis

Since YOLO target detection requires not only identifying the location of the lesion but also classifying the type of lesion, Table 1 presents comprehensive classification results for the entire MIAS dataset, including malignant and benign cases. Table 2 focuses specifically on the performance of cancer detection, highlighting the models' ability to identify malignant lesions. Table 3 showcases the results for benign lesion detection, demonstrating the models' capability to distinguish benign

abnormalities. All three tables evaluate the performance of four advanced YOLO-based model configurations: YOLOv8, YOLOv8-GHOST, YOLOv8-GHOST-P2, and YOLOv8-P2. For each model, the tables display three key performance indicators: precision, recall, and F1 score. The experimental results clearly demonstrate that the F1 scores of the YOLOv8-GHOST-P2 model and the YOLOv8-P2 model are significantly higher than those of the original model. This improvement is due to the P2 layer, which possesses a higher resolution and more effectively captures the fine details of breast lesion areas. During the multiscale feature fusion process, the P2 layer not only inherits the semantic information from the upper layers but also retains its own rich spatial details through interaction with deeper feature maps. Compared to other layers in the pyramid network, the P2 layer exhibits greater sensitivity and accuracy, ensuring that even the smallest lesion regions in the entire mammogram are detected, thus significantly enhancing the overall detection performance.

Table 1. Overall classification results for the MIAS dataset.

	Precision	Recall	F1
YOLOv8 model	99.82%	97.02%	98.4%
YOLOv8-GHOST model	99.69%	97.11%	98.38%
YOLOv8-GHOST-P2 model	99.34%	98.27%	98.8%
YOLOv8-P2 model	99%	98.14%	98.57%

Table 2. Classification results for cancer detection in the MIAS dataset.

	Precision	Recall	F1
YOLOv8 model	99.90%	100%	99.95%
YOLOv8-GHOST model	99.90%	100%	99.95%
YOLOv8-GHOST-P2 model	100%	100%	100%
YOLOv8-P2 model	100%	100%	100%

Table 3. Classification results for benign lesion detection in the MIAS dataset.

	Precision	Recall	F1
YOLOv8 model	99.76%	94.68%	97.15%
YOLOv8-GHOST model	99.51%	94.83%	97.12%
YOLOv8-GHOST-P2 model	98.82%	96.91%	97.86%
YOLOv8-P2 model	98.2%	96.68%	97.43%

In the above experimental table, it is easy to see that the recall rates of the three models are significantly higher than that of the original model, and the overall "false negatives" of the models are low. This is because the traditional convolutional operation may lead to the loss of some key information when generating feature maps, especially when addressing small features, and the GHOST module can more comprehensively capture the details of the input image when generating features through linear operations, avoiding the omission of the target and thus increasing the recall rate and reducing the FN rate.

Table 4 presents the overall detection classification results for both malignant and benign lesions. Table 5 focuses on the detection results for malignant lesions, while Table 6 highlights the results for benign lesions. Table 7 lists the number of parameters of each model, from which it can be

seen that the YOLOv8-GHOST model achieves significant parameter scale compression compared with the basic version of the YOLOv8 model, and its total parameter size is only about 57.10% of the original model, reflecting efficient parameter utilization. Furthermore, the YOLOv8-GHOST-P2 model continues to be optimized on this basis, and the number of parameters is reduced to about 53.35% of the original model, which significantly improves the lightweight degree of the model. In contrast, although the YOLOv8-P2 model is also committed to reducing the number of parameters, its parameter scale is still close to the level of the original model, which is about 97.2% of the original model.

Table 4. Overall classification results for the DDSM dataset.

	Precision	Recall	F1
YOLOv8 model	94.51%	97.43%	95.95%
YOLOv8-GHOST model	91.64%	96.72%	94.11%
YOLOv8-GHOST-P2 model	74.31%	97.41%	84.31%
YOLOv8-P2 model	87.10%	96.46%	91.53%

Table 5. Classification results for cancer detection in the DDSM dataset.

	Precision	Recall	F1
YOLOv8 model	94.53%	97.55%	96.02%
YOLOv8-GHOST model	91.65%	96.64%	94.34%
YOLOv8-GHOST-P2 model	74.35%	97.56%	84.4%
YOLOv8-P2 model	87.4%	96.75%	91.84%

Table 6. Classification results for benign lesion detection in the DDSM dataset.

	Precision	Recall	F1
YOLOv8 model	94.49%	97.26%	95.86%
YOLOv8-GHOST model	91.62%	96.83%	94.15%
YOLOv8-GHOST-P2 model	74.27%	97.21%	84.21%
YOLOv8-P2 model	88.29%	96.08%	91.13%

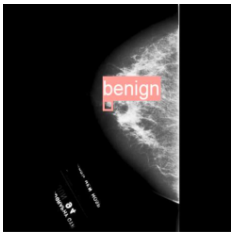
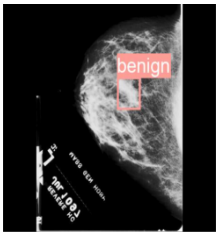
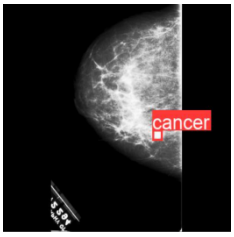
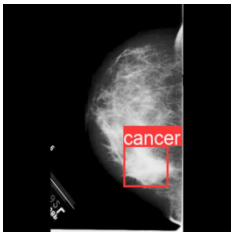
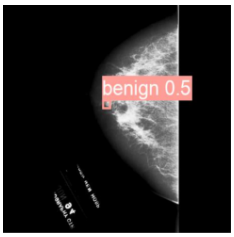
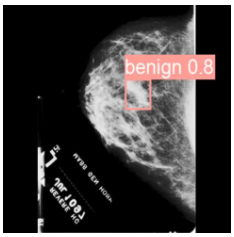
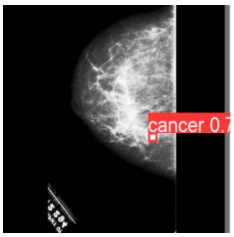
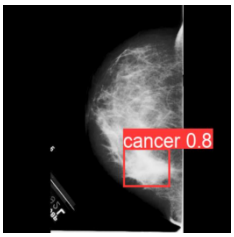
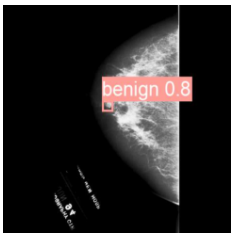
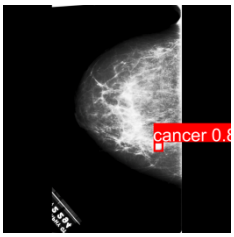
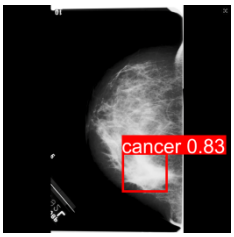
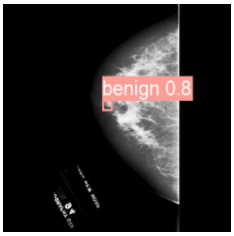
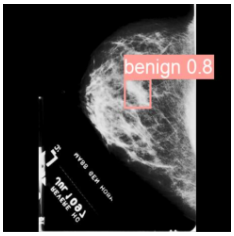
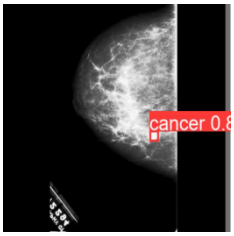
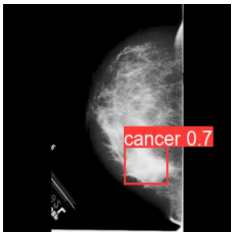
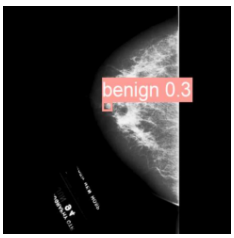
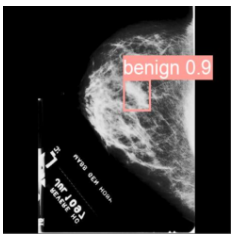
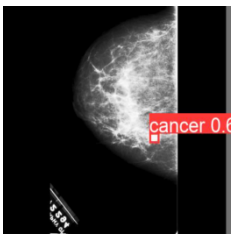
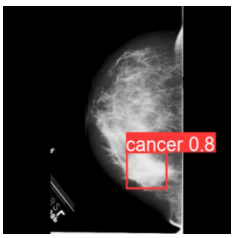
Table 7. Comparison among the numbers of parameters in different models.

	Number of participants	Percentage of the original model	Comparison with the original model
YOLOv8 model	3,011,238	-	-
YOLOv8-GHOST model	1,719,354	57.10%	-42.90%
YOLOv8-GHOST-P2 model	1,606,624	53.35%	-46.64%
YOLOv8-P2 model	2,926,824	97.2%	-2.8%

Data in these tables demonstrate that the significant reduction in the number of parameters is primarily due to the integration of the GHOST module. Moreover, the combined use of the GHOST module and the P2 layer not only enhances the model's operational efficiency but also improves detection accuracy, all while substantially reducing the number of required parameters.

5.5. Presentation of the forecasting results

Table 8. Projections yielded by the tested models.

Val labels				
YOLOv8 model				
YOLOv8-GHOST model				
YOLOv8-P2 model				
YOLOv8-GHOST-P2 model				

In Table 8, the first row contains the validation labels, which show the locations and sizes of the lesion areas. The last four columns indicate the detection results produced by different models for the same figure. The numbers in each figure are the model confidence scores, which are composed of two parts: an object existence probability (objectness score) and a class probability. The objectness score is the probability that an object is contained in the prediction box of the utilized model and is usually output as a value between 0 and 1 by the sigmoid function. The class probability is the probability that an object in the prediction box belongs to a particular class, and for each class, the probability value is again obtained via the softmax function, which ensures that the probabilities of all classes sum to 1. The final confidence score is the product of the object presence probability and the maximum class probability, which is mainly used to determine the reliability of the detection

bounding box. The maximum value is 1. A higher confidence level indicates a more likely lesion region. The above table shows that all four models have high accuracy in terms of detecting larger lesion regions, and the YOLOv8-GHOST-P2 model even reaches a confidence level of 0.9. The YOLOv8 model achieves a confidence level of 0.8 with respect to detecting lesion regions, indicating high stability when detecting lesion regions possessing different sizes. These models are able to accurately recognize and detect both small lesions and larger lesion areas.

6. Conclusions and ideas for future work

The YOLO series has continuously evolved, and this study explores the application of YOLOv8 for breast cancer detection. Early diagnosis is critical for reducing breast cancer mortality, but traditional methods often struggle with the complexity of mammography images. This study is the first to incorporate the GHOST module into YOLOv8 for breast cancer detection, which enhances the model's ability to distinguish between normal tissue and lesion areas, capturing and preserving key features more effectively. This leads to improved accuracy and efficiency in detecting breast cancer lesions. Our results show that the improved YOLOv8 model achieves comparable performance to the original model while requiring fewer parameters. Our YOLOv8-GHOST modification reduces parameters while maintaining performance, and YOLOv8-P2 improves small lesion detection, showing promising results on the MIAS dataset. However, performance on the DDSM dataset was limited due to its size, resolution, and insufficient training cycles. To improve, we plan to increase training cycles, refine the architecture, and explore pre-training on larger datasets. We also aim to extend our model's application to lung nodule detection, brain tumor segmentation, retinal disease identification, and bone fracture detection, addressing current limitations and broadening its impact in medical imaging.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The work is partially supported by the Research and Practice Project of Higher Education Teaching Reform in Henan Province in 2023 (Postgraduate Education) (Grant No. 2023SJGLX082Y), the Specialized Program for Basic and Frontier Technology Research of Nanyang City, 2023 (Grant No. 23JCQY2029) and the Young Core Faculty Support Program of Higher Education Institutions in Henan Province (Grant No. 2019GGJS184).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. X. Pei, R. Zhou, Current status and future of Chinese medicine diagnosis and treatment of breast cancer (in Chinese), *Beijing Tradit. Chin. Med.*, **42** (2023), 704–707. https://doi.org/10.16025/j.1674-1307.2023.07.001
2. E. Mahoro, M. A. Akhloufi, Breast masses detection on mammograms using recent one-shot deep object detectors, in *2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)*, IEEE, (2023), 1–4. https://doi.org/10.1109/BioSMART58455.2023.10162036
3. A. Intasam, Y. Promworn, A. Juhong, S. Thanasitthichai, S. Khwayotha, T. Jiranantanakorn, et al., Optimizing the hyperparameter tuning of yolov5 for breast cancer detection, in *2023 9th International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, IEEE, (2023), 184–187. https://doi.org/10.1109/ICEAST58324.2023.10157611
4. X. Gao, B. Braden, S. Taylor, W. Pang, Towards real-time detection of squamous pre-cancers from oesophageal endoscopic videos, in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, IEEE, (2019), 1606–1612. https://doi.org/10.1109/ICMLA.2019.00264
5. Y. Yang, J. Wang, Research on breast cancer pathological image classification method based on wavelet transform and YOLOv8, *J. X-Ray Sci. Technol.*, **32** (2024), 677–687. https://doi.org/10.3233/XST-230296
6. M. A. Al-Antari, S. Han, T. Kim, Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms, *Comput. Methods Programs Biomed.*, **196** (2020), 105584. https://doi.org/10.1016/j.cmpb.2020.105584
7. M. A. Al-masni, M. A. Al-antari, J. M. Park, G. Gi, T. Y. Kim, P. Rivera, et al., Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network, in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, (2017), 1230–1233. https://doi.org/10.1109/EMBC.2017.8037053
8. R. K. Kassahun, M. Molinara, A. Bria, C. Marrocco, F Tortorella, Breast mass detection and classification using transfer learning on OPTIMAM dataset through RadImageNet weights, in *International Conference on Image Analysis and Processing*, Springer, (2024), 71–82. https://doi.org/10.1007/978-3-031-51026-7_7
9. F. Touazi, D. Gaceb, M. Chirane, S. Hrzallah, Two-stage approach for semantic image segmentation of breast cancer: Deep learning and mass detection in mammographic images, in *6th International Workshop on Informatics & Data-Driven Medicine (IDDM)*, CEUR Workshop Proceedings, (2023), 62–76. https://doi.org/10.1109/iciprob54042.2022.9798724
10. W. Ouyang, P. Luo, X. Zeng, S. Qiu, Y. Tian, H. Li, et al., Deepid-net: Multi-stage and deformable deep convolutional neural networks for object detection, preprint, arXiv:1409.3505.
11. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2015), 1–9. https://doi.org/10.1109/CVPR.2015.7298594
12. S. Cheng, G. Shang, L. Zhang, Handwritten digit recognition based on improved VGG16 network, in *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*, (2019), 110693B. https://doi.org/10.1117/12.2524281

13. B. Zoph, Q. V. Le, Neural architecture search with reinforcement learning, preprint, arXiv:1611.01578.
14. D. Cai, X. Sun, N. Zhou, X. Han, J. Yao, Efficient mitosis detection in breast cancer histology images by RCNN, in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, (2019), 919–922. https://doi.org/10.1109/ISBI.2019.8759461
15. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2016), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
16. T. Mahmood, M. Arsalan, M. Owais, M. B. Lee, K. R. Park, Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs, *J. Clin. Med.*, **9** (2020), 749. https://doi.org/10.3390/jcm9030749
17. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, preprint, arXiv:1506.02640.
18. S. R. Sindhu, J. George, S. Skaria, V. V. Varun, Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans, in *Medical Imaging 2018: Computer-Aided Diagnosis*, SPIE, **10575** (2018), 347–355. https://doi.org/10.1117/12.2293699
19. H. M. Ünver, E. Ayan, Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm, *Diagnostics*, **9** (2019), 72. https://doi.org/10.3390/diagnostics9030072
20. J. Sobek, J. R. M. Inojosa, B. J. M. Inojosa, S. M. Rassoulinejad-Mousavi, G, M. Conte, F. Lopez-Jimenez, et al., MedYOLO: A medical image object detection framework, *J. Digit. Imaging. Inform. Med.*, (2024), 1–9. https://doi.org/10.1007/s10278-024-01138-2
21. K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: More features from cheap operations, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2020), 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165
22. J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, et al., Mammographic Image Analysis Society (MIAS) database v1.21, Apollo-University of Cambridge Repository, 2015. https://doi.org/10.17863/CAM.105113
23. M. Heath, K. Bowyer, D. Kopans, P. K. Jr, R. Moore, K. Chang, et al., Current status of the digital database for screening mammography, *Digital Mammography*, **13** (1998), 457–460. https://doi.org/10.1007/978-94-011-5318-8_75

AIMS Press