

Detection of Masses in Mammogram Images Based on the Enhanced RetinaNet Network With INbreast Dataset

Mingzhao Wang¹, Ran Liu¹, Joseph Luttrell IV², Chaoyang Zhang², Juanying Xie¹

¹School of Computer Science, Shaanxi Normal University, Xian, People's Republic of China; ²School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS, USA

Correspondence: Juanying Xie, School of Computer Science, Shaanxi Normal University, No. 620, West Chang'an Avenue, Chang'an District, Xi'an, 710119, Shaanxi, People's Republic of China, Tel +86 13088965815, Email xiejuanying@snnu.edu.cn; Chaoyang Zhang, School of Computing Sciences and Computer Engineering, University of Southern Mississippi, 118 College Drive, Hattiesburg, MS, 39406-0001, USA, Email chaoyang.zhang@usm.edu

Purpose: Breast cancer is the most common major public health problems of women in the world. Until now, analyzing mammogram images is still the main method used by doctors to diagnose and detect breast cancers. However, this process usually depends on the experience of radiologists and is always very time consuming.

Patients and Methods: We propose to introduce deep learning technology into the process for the facilitation of computer-aided diagnosis (CAD), and address the challenges of class imbalance, enhance the detection of small masses and multiple targets, and reduce false positives and negatives in mammogram analysis. Therefore, we adopted and enhanced RetinaNet to detect masses in mammogram images. Specifically, we introduced a novel modification to the network structure, where the feature map M5 is processed by the ReLU function prior to the original convolution kernel. This strategic adjustment was designed to prevent the loss of resolution for small mass features. Additionally, we introduced transfer learning techniques into training process through leveraging pre-trained weights from other RetinaNet applications, and fine-tuned our improved model using the INbreast dataset.

Results: The aforementioned innovations facilitate superior performance of the enhanced RetinaNet model on the public dataset INbreast, as evidenced by a mAP (mean average precision) of 1.0000 and TPR (true positive rate) of 1.00 at 0.00 FPPI (false positive per image) on the INbreast dataset.

Conclusion: The experimental results demonstrate that our enhanced RetinaNet model defeats the existing models by having more generalization performance than other published studies, and it can also be applied to other types of patients to assist doctors in making a proper diagnosis.

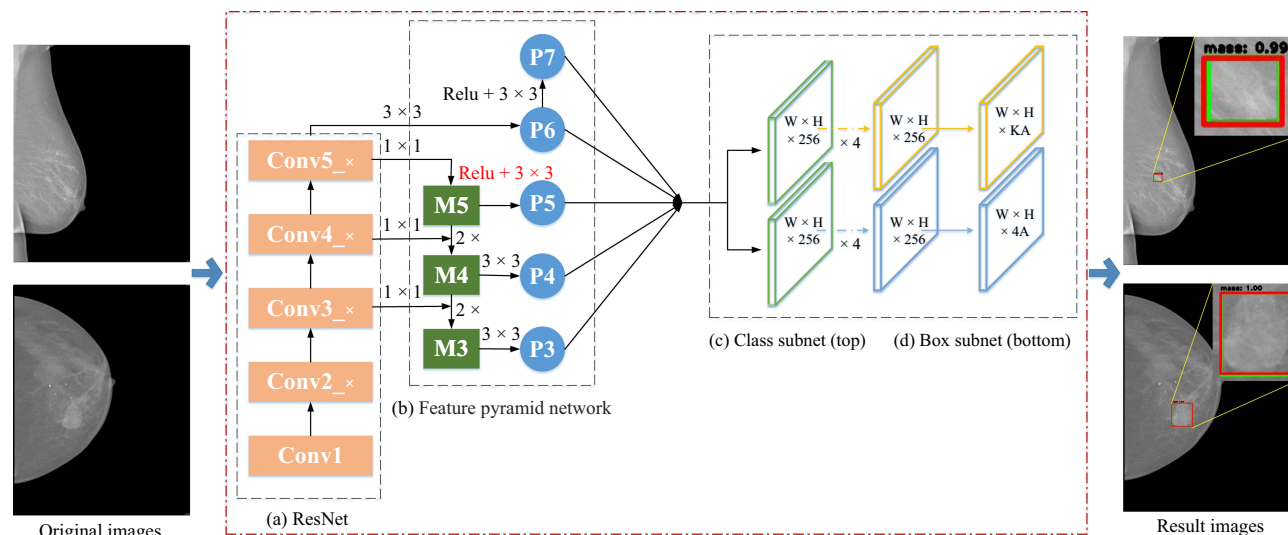
Keywords: computer-aided diagnosis, deep learning, object detection, RetinaNet, transfer learning

Introduction

Breast cancer is one of the most common and deadly cancers in women and accounts for about 15% of all cancer deaths in women.^{1,2} The World Health Organization reported that 67,0000 women died of breast cancer in 2022.² The death rate of breast cancer continues to increase globally. In order to improve the prognosis and survival rate of breast cancer, early diagnosis of patients has become crucial. Although they have been used for more than 100 years, mammograms are still the primary screening tool for the clinical diagnosis of breast cancer.³ Doctors usually make a diagnosis by examining a patient's mammogram and combining it with the patient's clinical information. However, making a correct diagnosis is a challenging task because it is dependent on the doctor's experience. Therefore, computer-aided diagnosis (CAD) systems for medical images are often used to assist medical doctors in detecting and diagnosing lesions in breast imaging examinations by providing them with auxiliary advice and thus reducing their burden.⁴⁻⁶

However, traditional CAD systems have several limitations that can impact the accuracy and efficiency of breast cancer detection. Firstly, the traditional CAD systems are usually dependent on manually extracted features. This requires

Graphical Abstract



people to select those useful features based on prior knowledge of the data, which is inefficient and computationally intensive. Secondly, the features can be affected by external factors such as patient breast density and imaging equipment.^{7,8} As a result, poor diagnosis results are often produced by these types of approaches. In addition, traditional CAD systems may be less effective in detecting breast cancer in women with dense breast tissue, as the dense tissue can obscure the visibility of tumors.^{9,10} This can lead to a higher risk of missed diagnoses in these patients.

Deep learning methods have recently advanced to an unprecedented stage and have attracted much research attention because of the rapidly emerging big data era and vast improvements in computer hardware. Introducing deep learning technology in CAD systems can potentially address these limitations and offer several benefits. By using a standardized deep learning, the variability in interpretation among radiologists can be minimized, leading to more consistent and reliable results. Furthermore, they can avoid the limitations of traditional CAD systems due to their automatic feature extraction capabilities.^{11,12} In this way, deep learning methods have attained breakthrough achievements in image recognition,¹³ natural language processing,¹⁴ bioinformatics,¹⁵ object detection,¹⁶ and many other fields^{17,18} without relying on manually extracted features. For example, we have used deep learning to classify and analyze breast cancer histopathological images, and we obtained the best results so far.¹⁹ Although we can successfully classify breast cancer histopathological images into benign and malignant classes and further into eight classes, including Adenosis, Fibroadenoma, Phyllodes Tumor, Tubular Adenoma, Ductal Carcinoma, Lobular Carcinoma, Mucinous Carcinoma, and Papillary Carcinoma, our study was limited to these breast cancer histopathological images. Therefore, it cannot significantly assist medical doctors in early breast cancer screening using mammograms. Furthermore, most studies have focused on the classification of various types of breast cancer,^{13,20–23} but there have been few studies about lesion region detection in mammogram images.

Our research in this paper is dedicated to deep learning-based detection of lesion regions in mammogram images with the aim of enhancing the efficiency and accuracy of breast cancer screening and treatment, as well as providing assistance for medical doctors to administer clinical treatment to breast cancer patients in a timely manner. Specifically, we aim to address the class imbalance problem, reduce the high false positive and false negative rates in existing models, and focus on detecting small masses and multiple targets in mammogram analysis. To achieve these goals, we initially adapt the end-to-end one-stage object detection model RetinaNet²⁴ to the mammogram masses detection problem and conduct experiments on the INbreast database.²⁵ Subsequently, we propose improvements to the RetinaNet network and combine transfer learning and fine-tuning techniques to further enhance RetinaNet model and our experimental results. Ultimately,

our approach achieves superior performance on the INbreast database compared to previous studies in the literature, thus contributing to more effective breast cancer screening and clinical treatment.

The enhanced detection capabilities of our proposed method hold significant implications for patient care. By improving the accuracy of mass detection in mammogram images, we can reduce the number of false negatives, which is of utmost importance as missed detections can lead to delayed diagnosis and treatment, negatively affecting patient prognosis. At the same time, minimizing false positives is equally crucial as it reduces unnecessary patient anxiety and further diagnostic procedures. Our approach aims to strike a balance between these two aspects, ultimately leading to more reliable and effective breast cancer screening. This improvement not only benefits individual patients by providing more accurate diagnoses but also has the potential to enhance the overall efficiency and effectiveness of breast cancer screening programs.

Related Works

There have been many classification studies on the INbreast database. For example, in a 2015 study by Carneiro et al,²⁶ researchers divided the images in the INbreast database into two categories (benign and malignant) by using a convolutional neural network pre-trained on the ImageNet dataset (a very large and popular dataset in the field of computer vision). They attained good experimental results in terms of an AUC (Area Under the ROC Curve) of 0.91.

However, compared with the classification task, it is much more important to detect the location of breast lesions in the mammogram images so that medical doctors can make a correct diagnosis and administer further appropriate clinical treatment to patients. As a result, there have been some studies published that used the INbreast dataset and focused on detecting breast lesions. In 2015, Dhungel et al²⁷ proposed a new method for detecting masses in mammography images in the INbreast database. They adopted R-CNN (Region-based Convolutional Neural Network), a two-stage object detection method and deep learning technique cascaded with a random forest classifier, and obtained experimental results of 0.96@1.2 (TPR@FPPI). TPR refers to True Positive Rate and FPPI denotes False Positive Per Image. In 2017, they²⁸ combined the deep learning technique with Bayesian optimization to perform further research with the INbreast database and achieved results of 0.90@1.3 in mammography mass detection. Also in 2017, Akselrod-Ballin et al²⁹ proposed to use Faster R-CNN (Faster Region-based Convolutional Neural Network) to detect masses in the INbreast database and achieved experimental results of 0.93@0.56. In 2018, Ribli et al³⁰ used Faster R-CNN to do more challenging research on the INbreast database. They tried to detect all types of lesions in the mammography images, such as masses and calcification, and obtained experimental results of 0.90@0.3 (TPR@FPPI). They also found that masses are the most common type of lesion in mammography images. Furthermore, a 2018 study by Jung et al³¹ proposed to use the one-stage object detection model RetinaNet to detect lesions in images in the INbreast database due to the advantages offered by its speed and high precision. They obtained results of 0.97@3.0 on INbreast and also used the parameters obtained by training RetinaNet on their inhouse database GURO. These parameters will be helpful for related research in the future. The study on INbreast was done by Agarwal et al³² in 2019, in which the InceptionV3³³ network was combined with transfer learning to obtain results of 0.98@1.67 for mass detection in mammography images. They further proposed a fully automated framework to detect masses in full-field digital mammograms based on the Faster R-CNN network in 2020, and they obtained 0.99@1.17 for malignant masses and 0.85@1.0 for benign masses on the INbreast dataset.³⁴ At the same year, Gao et al³⁵ proposed a multi-task deep learning model enabled by feature transfer and obtained 0.91@3.67 and 0.91@5 for malignant and benign masses on the INbreast dataset, respectively. They also tested Mask R-CNN on the INbreast dataset by obtaining 0.85@3.67 and 0.85@5 for malignant and benign masses, respectively.³⁵ Li et al³⁶ proposed a convolutional neural network-based bilateral image analysis method for mass detection in mammograms and obtained 0.88@1.12 on the INbreast dataset. Min et al³⁷ put forward a scheme based on the sifting architecture for detecting mammographic lesions, obtaining the 0.90@0.9 on INbreast dataset. Yan et al³⁸ proposed the combined matching and classification network (CMCNet) which was used on the INbreast dataset, obtaining results of 0.96@0.23 for detecting masses in mammograms. Zhang et al³⁹ proposed to use the YOLOv3 network for mass detection in mammogram images and obtained TPR@FPPI results of 0.9@1.7 for the INbreast dataset. Elmoufidi⁴⁰ developed a framework for automatizing the analysis of mammograms' suspicious areas. It utilizes BEMD (Bidimensional empirical mode decomposition) to decompose ROIs (Regions of interest) into BIMFs (Bidimensional

intrinsic mode function) and SVM (Support vector machine) for classification, significantly improving diagnostic accuracy compared to recent literature. In the past two years, Aslan and colleagues⁴¹ introduced two types of end-to-end deep learning models tailored for the analysis of mammography images, comprising a CNN (Convolutional neural network) and a more complex model integrating CNN with BiLSTM (Bidirectional long short-term memory) to leverage temporal feature extraction. Their findings revealed that these models achieved commendable classification accuracies for benign samples, with 0.95 and 0.971 on the INbreast dataset, respectively, underscoring the efficacy of these deep learning architectures in breast imaging diagnostics. A deep learning-based model was introduced for the Severity Classification of Breast Cancer in Chakravarthy's work.⁴² It employs transfer learning for feature extraction and enhances the wKNN (Weighted k-nearest neighbor) algorithm using optimization algorithm for transforming non-linear input features into linearly separable feature vectors. The proposed model achieved the best performance in INbreast dataset, with TPR of 0.807 and accuracy of 0.832, when CSO algorithm transformed features are using with wKNN classify. Chakravarthy et al⁴³ proposed a CAD framework leveraging deep learning for breast cancer diagnosis. Four distinct experiments are conducted to identify the most effective classification strategy, including pre-trained Deep CNNs, feature extraction with SVM using various kernels, deep feature fusion for improvement, and PCA to reduce computational burden. The INbreast dataset achieved an impressive classification accuracy of 0.966 through the fusing of deep features. Prinzi et al⁴⁴ developed an automated data-driven model using Yolo-based architecture for breast cancer detection in mammograms. The comparison of various Yolo models highlighted the efficacy of the YoloV5s model with transfer learning, achieving a mean TPR of 0.742 ± 0.146 in INbreast dataset.

Although masses are very common lesions in mammography, the aforementioned studies show that the best TPR@FPPI for mass detection in the mammography images of the INbreast database is no more than $0.98@1.67$, which was achieved using the InceptionV3 network. In addition, although it was reported that the best TPR@FPPI of $0.99@1.17$ was achieved using Faster R-CNN for malignant masses on the INbreast dataset, the TPR@FPPI for benign masses was only $0.85@1.0$.³⁴ Therefore, it is necessary for us to advance the study of mass detection in mammography. However, masses in mammography images are typically small, which makes detecting them a very challenging task. It is reported that, in addition to having high speed and accuracy, the one-stage object detection network RetinaNet has advantages that enhance its ability to detect small targets. Although Jung et al³¹ proposed to use RetinaNet to detect lesions in images from the INbreast database and obtained a TPR @ FPPI of $0.97@3.0$, their results are not better than the $0.98@1.67$ achieved using the InceptionV3 network reported in Agarwal's work.³²

With the advantages of RetinaNet and the challenges of mass detection in mammography in mind, we prefer to adopt the RetinaNet network to perform mass detection on the mammography images in the INbreast database. The choice of the RetinaNet network was driven by its unique advantages. Its one-stage architecture provides a balance between speed and accuracy, making it suitable for clinical applications. The focal loss function it proposed effectively tackles the class imbalance issue prevalent in mammogram images, where normal regions are abundant compared to mass regions. Additionally, the hierarchical feature maps in RetinaNet enable better detection of small and multiple sizes of masses, which is essential given the diverse size range of masses in mammograms. These characteristics distinguish RetinaNet from other commonly used models, such as Faster R-CNN and YOLO, and make it a particularly apt choice for our mammogram analysis task. Here, we try to improve the TPR while also reducing the FPPI for mass detection using the techniques introduced in the next section.

Materials and Methods

This section introduces the data and the networks used in this paper. For constructing the datasets in this study, we used the publicly available INbreast image database. The network architectures used to train our models include the existing RetinaNet and our improved RetinaNet.

Data

The INbreast dataset used in this paper is a mammographic dataset, which can be acquired at a breast center located in a university hospital (Centro Hospitalar de S. João [CHSJ], Breast Centre, Porto) with the permission of the Portuguese National Committee of Data Protection and Hospital's Ethics Committee.²⁵ It consists of 410 X-ray breast images from

115 patients, depicting the craniocaudal (CC) and mediolateral oblique (MLO) views of both breasts for each patient. There are 90 patients who have not had a mastectomy (they have two breasts) and thus can provide four breast X-ray images. The remaining 25 patients have had a mastectomy (they have only one breast) and can only provide two breast X-ray images. Figure 1a–d show four breast images from a patient that did not have a mastectomy. The size of the INbreast images is 3328×4048 or 2560×3328 pixels, which, based on the size of the breast being imaged, depends on the size of the compression plate the researchers used to collect the images.

There are six categories in the INbreast database, including asymmetry, calcification, distortion, mass, multi-cases, and normal status. Since masses are the most common lesions, we focus on detecting masses in the mammogram images in this paper. The experts provided accurate contour annotation in XML file format for the lesion sites in 343 mammogram images in the INbreast dataset. However, there were 67 mammogram images that we could not use in our experiments because they were left without any labels. A total of 116 masses can be found among 107 images, with an average of about 1.1 masses per image.²⁵ According to this description of the INbreast dataset in the reference²⁵, it can be guessed that all images labeled as masses are included within the 343 labeled images. In other words, none of the 67 mammogram images without labels contain masses. The detail information of the images in the INbreast dataset is displayed in Figure 1e.

As a result of the differences between the INbreast dataset and the PASCAL VOC (Visual Object Classes) dataset that RetinaNet was originally tested with, we first needed to process these 343 images before they could be used. Specifically, we needed to convert the INbreast dataset's annotations for the masses in the images, which are in the form of an array of points representing the contour line centered around regions of interest, to the rectangular bounding box target format used by RetinaNet and the PASCAL VOC dataset. In order to accomplish this, we used the point with the minimum horizontal and vertical coordinates in each mass contour line in the mammogram images as the upper-left point of each rectangular bounding box and the point with the maximum coordinates as the bottom-right point. By doing this, the

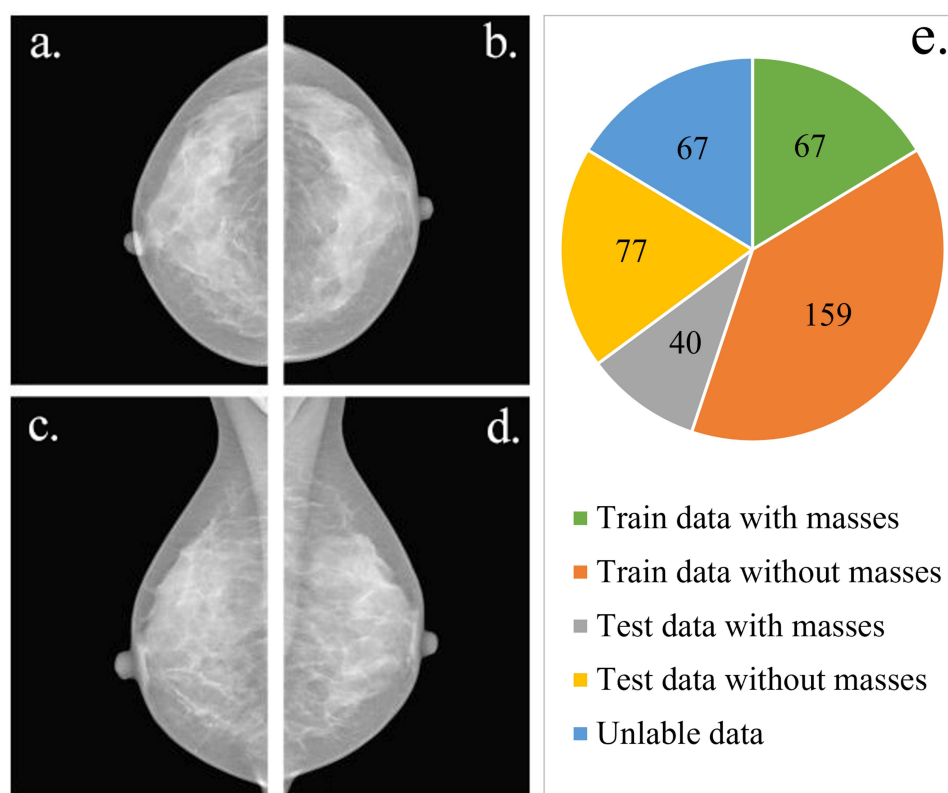


Figure 1 Detail information of the INbreast dataset used in this study. Subfigures (a)–(d) show the four mammograms of a patient from two views,²⁵ (a) The CC view of the patient's right breast, (b) The CC view of the patient's left breast, (c) the MLO view of the patient's right breast, (d) The MLO view of the patient's left breast; Subfigure. (e) Presents the number of images with and without masses included in the training and test subsets, and the number of unlabeled images, respectively.

rectangular bounding box target format abides by the rules of the PASCAL VOC dataset. It should be noted that the directions of the horizontal and vertical coordinates we used here are toward right and down, respectively. In addition, we normalize images following the normalization protocol utilized in the original RetinaNet.

RetinaNet Network

Deep learning-based object detection algorithms can be divided into two families.⁴⁵ The first family is composed of two-stage detection algorithms, such as R-CNN⁴⁶ and its variants. This kind of algorithm divides the object detection task into two stages that involve generating candidate regions before performing classification and regression. As a result, they typically can achieve high detection accuracy, but they often have very slow detection speed. The second family contains one-stage deep learning-based object detection algorithms, such as YOLO⁴⁷ and its variants. These algorithms simplify object detection to an end-to-end classification and regression task without generating candidate regions. As a result, they are typically fast but are often less accurate, compared with two-stage networks. How to achieve both high speed and high accuracy has been a challenging task in the advancement of deep learning-based object detection studies, and many researchers have devoted themselves to making improvements in this area. In 2018, Lin et al²⁴ pointed out that this challenge mainly comes from the extreme imbalance typically found in sample category distributions. They proposed a novel loss function named Focal Loss (FL) as an alternative to the traditional cross-entropy loss, and they developed the RetinaNet network to test this FL function. Focal Loss is defined in equation (1)-(3).

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (1)$$

$$\alpha_t = \begin{cases} \alpha, & y = 1 \\ 1 - \alpha, & \text{otherwise} \end{cases} \quad (2)$$

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3)$$

Here, y is the label, $y=1$ denotes that the sample belongs to the positive class. $p \in [0, 1]$ for binary classification represents the probability that a sample will be predicted as a positive sample. p_t represents the model's estimated probability p for positive class, otherwise is the $1-p$ for negative class. $\alpha \in [0, 1]$ is a weight factor used to adjust the contributions to the FL from positive and negative samples. Similar to the definition of p_t , α_t is the α for positive class, otherwise is the $1 - \alpha$ for negative class. $\gamma > 0$ is the modulation factor used to suppress the weights of samples that are easily classified and to lead the model to focus on samples that are more difficult to classify. We set $\alpha = 0.25$ and $\gamma = 2$ in our experiments. Thanks to improvements from the FL function, it was finally proved that the RetinaNet network was capable of real-time object detection with both high accuracy and speed.

Figure 2 shows a diagram of RetinaNet and comprises the four parts: (a) is the residual network used to extract features from input images. It uses layer jumping connections to avoid the problem of gradient disappearance caused by increasing network depth. (b) is the FPN (Feature Pyramid Network)⁴⁸ used to generate multi-scale feature maps. (c) is the full convolution subnetwork for classification. It is used to classify the feature maps generated by (b). (d) is the fully convolutional regression subnetwork. This section is parallel to the subnetwork in (c) and shares the same structure, but its function is to perform regression on the feature maps generated by (b). To understand the details of the RetinaNet network, so as to improve it, we present its detailed structure in Figure 3.

The residual network shown in Figure 3a is composed of five convolutional residual blocks, each of which can transfer the output information of the first few layers of the network to a deeper network layer by jumping with a layer connection. This helps the network avoid the gradient disappearance problem and ensures good performance while training deeper layers of the model. Figure 3b is the feature pyramid network, which can achieve accurate location detection via adding a connection between the reconstruction layer and the corresponding feature map. Finally, there are five different scale feature maps (P3, P4, P5, P6 and P7) both for the classification subnet in Figure 3c and the regression subnet in Figure 3d. This allows for target classification and regression to be performed simultaneously. The whole network is an end-to-end and one-stage object detection model. Once an image of an object is used for input, the network

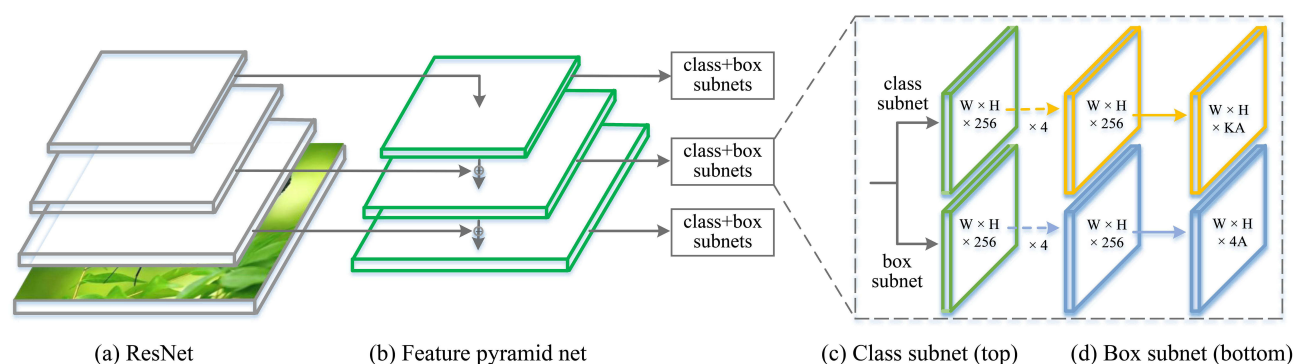


Figure 2 The diagram of the RetinaNet network structure.²⁴

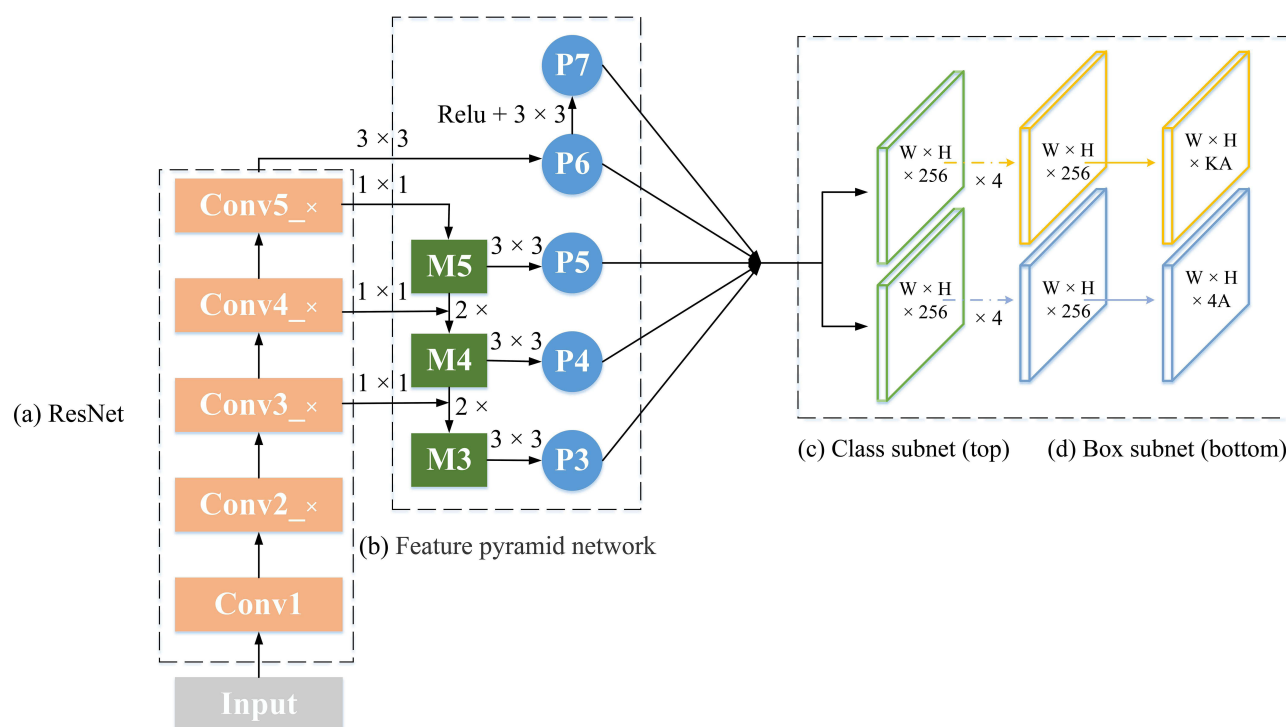


Figure 3 The diagram of the refined network structure of RetinaNet.

will output both classification and regression results after a series of forward propagations. Small objects can be detected by the RetinaNet network using the multi-scale feature maps produced by the embedded feature pyramid network. Therefore, it is a suitable network for detecting masses in breast X-ray images. Additional details concerning the RetinaNet network can be found in Lin's work.²⁴

The Improved RetinaNet Network

To improve the ability of RetinaNet to detect very small masses in breast X-ray images, we propose the improved RetinaNet network shown in Figure 4. The improved part of the network is depicted in red font. Specifically, the feature map M5 will first be processed by the nonlinear function ReLU (Rectified Linear Unit) before being processed by the original 3×3 convolution kernel. This improvement was made to prevent the resolution of features for very small masses in the X-ray images of breast from becoming too low, resulting they cannot be detected out, though they are still informative after being processed by ResNet50. The ReLU function can strengthen some features of feature map M5, so as to enhance their resolution and guarantee that very small masses will be detected in the X-ray images. Therefore, we

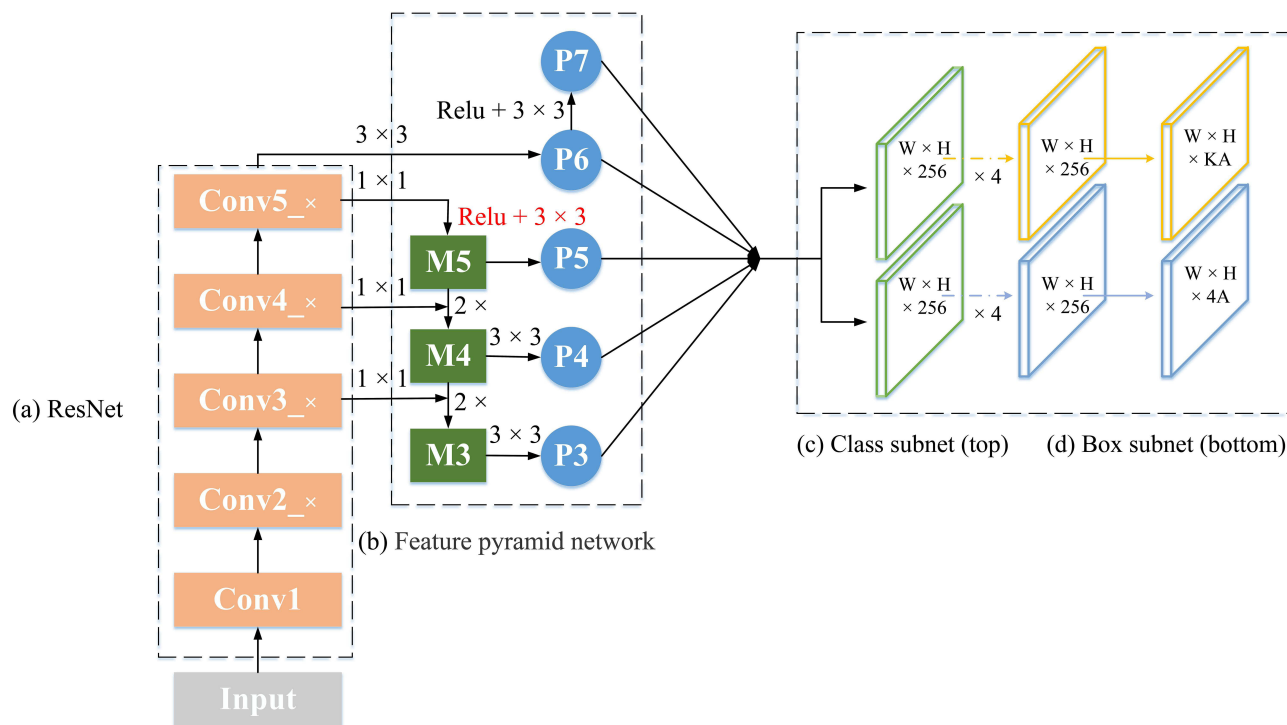


Figure 4 The diagram of the improved RetinaNet network structure.

use the ReLU function to enhance the feature information of the high-level feature map M5 of RetinaNet network, so as to achieve the effect of enhancing resolution. Then a convolution operation is performed to improve the detection accuracy of the very small masses in breast X-ray images.

Evaluation Metrics

Although sensitivity (recall), specificity, precision and F1-score are popular metrics to evaluate the performance of a model, this study focus on the detection of masses in mammogram images, so the metrics which are popular and more comprehensive in object detection will be utilized to value the performance of the model. The evaluation metrics used in this paper are mAP (mean Average Precision)⁴⁹ and FROC (Free-Response Receiver Operating Characteristic Curves).⁵⁰ mAP is the most popular metric used in the object detection field. It is defined as the mean of the AP (Average Precision), which is the area under the precision-recall curve (PR curve), and is within the interval of [0, 1]. Each point in the curve corresponds to a specific threshold, the models with better overall performance typically exhibit higher mAP values. FROC is a common metric for cancer diagnosis that uses the x-axis to denote false positives per normal case and the y-axis for sensitivity. It is used to evaluate the performance of a model for each image by generating a series of points derived by changing the threshold that determines the True Positive Rate (TPR) and the False Positives per Image (FPPI). Finally, the performance of the model is evaluated by the minimum FPPI value (TPR@FPPI) when the TPR has reached the maximum value. Our experiments used the implementations of these two metrics that are available as library functions included with Python Sklearn and Matlab2017b evaluateDetectionMissRate, respectively.

Results

It is well known that RetinaNet was very successful on the PASCAL VOC database, and its parameters have been used by many researchers across a wide variety of different experiments thanks to the proliferation of transfer learning techniques. We mostly used the default parameters of RetinaNet, but we chose to tune a few of them, such as batch_size and num_epoch (set to 1 and 50, respectively, in our experiments. Here, the batch_size parameter was set to be 1 because of the memory limits of our computers). In addition, Adam (Adaptive Moment Estimation) optimizer is adopted to train

our model. The parameters for Adam are $\beta_1=0.9$ and $\beta_2=0.999$. The learning rate is $1e-6$. We performed experiments on the original INbreast database and a modified version, which we obtained by using image augmentation techniques. All of our experiments were carried out using the Tensorflow and Keras frameworks.

Experiments on the Raw Dataset

We randomly partitioned the 343 mammogram images in the INbreast dataset into training and test subsets in a 7:3 ratio under the condition that the distributions of the test and training subsets are similar. Finally, the training and test subsets include 226 and 117 mammogram images respectively. The number of images with and without masses included in the training and test subsets is detailed in Figure 1e. Figure 3 shows that the residual network is the first part of RetinaNet, which comprises several stacked convolutional layers. ResNet50,⁵¹ ResNet101,⁵¹ and ResNet152⁵¹ are three of the most popular residual networks that we found in our literature review. Therefore, we conducted our experiments by using these three residual networks as the backbone of the RetinaNet network for training and testing. First, we used the 226 mammogram images in the training subset to train the RetinaNet network with ResNet50, ResNet101, and ResNet152 backbones. Then, we tested the network with the 117 mammogram images in the test subset. The test results are recorded here as train_raw_50, train_raw_101, and train_raw_152, respectively. Figure 5 displays the curves depicting the change in training loss over each iteration (epoch) observed when using ResNet50, ResNet101 and ResNet152 as the backbone networks. Figure 6a displays the PR curves with the values of mAP for the three resulting models. Figure 6b displays the FROC curves of these models and their TPR@FPPI values.

The results in Figure 5 show that all the three training loss curves descend as the number of iterations increases. The loss decreases rapidly in the early stages of training, while it gradually stabilizes in the later stages of training. These results also show that the ResNet50 backbone exhibits the most rapidly decreasing training loss curve, while the curve belonging to ResNet101 decreases at the slowest rate. This is one of the reasons why we choose RetinaNet with ResNet50 backbone for detecting masses in the INbreast mammogram images.

The results in Figure 6 show that the mAP and TPR@FPPI values of each model are good regardless of the backbone used for RetinaNet (ResNet50, ResNet101, or ResNet152). Specifically, the results in Figure 6a show that the mAP

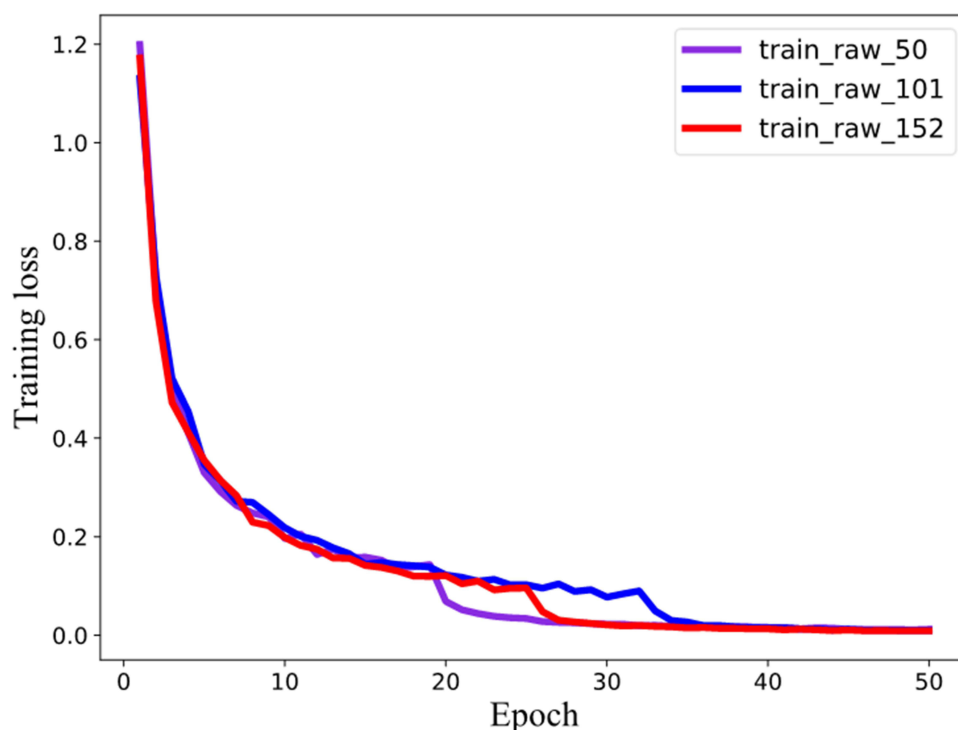


Figure 5 The training loss curves of the three RetinaNet models.

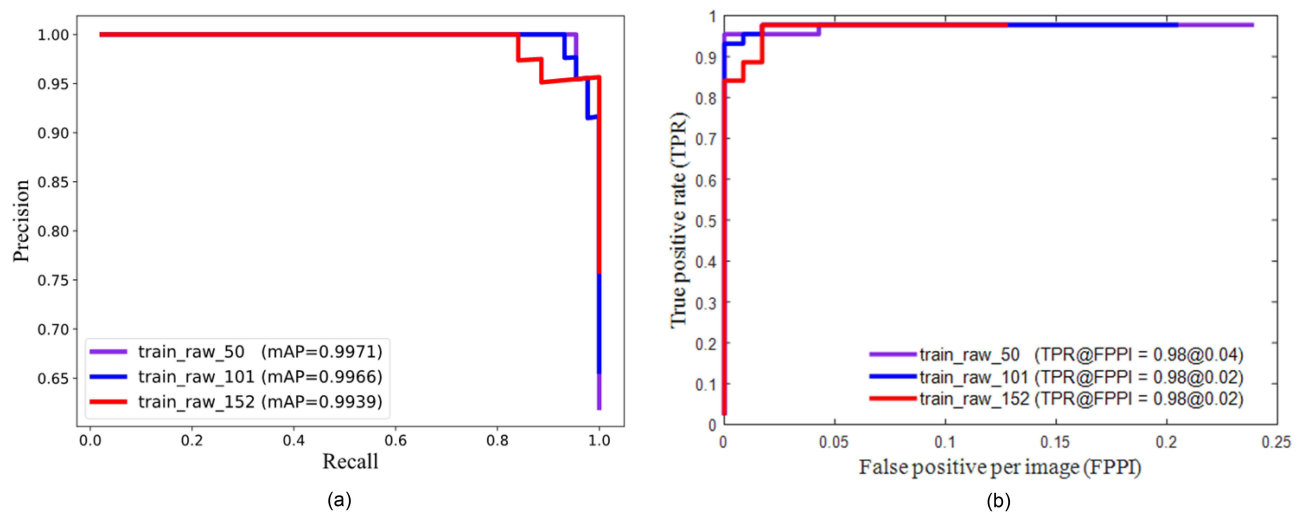


Figure 6 Performance comparison of three RetinaNet models in terms of PR and FROC curves, respectively. (a) PR curves with mAP values, (b) FROC curves with TPR@FPPI values.

values of the three RetinaNet models are 0.9971, 0.9966 and 0.9939, respectively, which are inversely proportional to the depth of the RetinaNet network. These results demonstrate that the rule of thumb, ie the deeper the network, the better the performance of a deep learning network, is not always correct. Therefore, we need take the data properties, domain-specific target metrics, and task-specific empirical optimization into consideration to design the proper architecture of a network for a specific task.

The results in Figure 6b show that the minimum FPPI values of three RetinaNet models are not same when they reached the same maximum TPR values. Although the minimum FPPI value of the RetinaNet using ResNet50 backbone is higher 2% than that of the other two RetinaNet models with ResNet101 and ResNet152, respectively, as the backbones, its minimum FPPI value is acceptable, considering its maximum TPR value and its efficient. Therefore, our models based on RetinaNet are well suited for detecting masses in the INbreast mammogram images.

To further compare the detection capability of our three models, Figure 7 shows the specific results for two mammogram images in the test subset. Here, the green box represents the ground truth location of the mass, and the red box depicts the location of the mass predicted by our three models. To display the results clearly, the ground truth and the detected results are enlarged and displayed in the up-right corner of each image.

The results in Figure 7 show an example of how the RetinaNet model with the ResNet50 backbone outperformed the other two models in the INbreast mass detection test. This was due to the ground-truth box and the predicted box for the masses in the test images nearly completely overlapping. In addition, the results in Figures 5 and 6 show that, for the INbreast database, our RetinaNet model using ResNet50 as the backbone was faster and more accurate than our RetinaNet models that used ResNet101 and ResNet152 for their backbones. This is likely caused by the images in the INbreast database having a relatively simple black background. Therefore, this well-defined sample space does not need a very deep network to learn the features for images, and our RetinaNet model with ResNet50 as the backbone is capable of learning the features used to detect masses in the INbreast database. However, our RetinaNet models with the ResNet101 and ResNet152 backbones may be too deep to learn these features and most likely succumbed to overfitting. Furthermore, these two deeper networks are more complex and will require too much training time. Therefore, we will conduct further experiments on RetinaNet with ResNet50 as the backbone network.

Experiments on the Augmented Dataset

To avoid the probable overfitting caused by the limited 226 X-ray breast images in the training subset, this subsection focuses on experiments that we performed using our augmented training subset. The mammogram images in the training subset were augmented by randomly applying the following operations: flipping horizontally, flipping vertically, rotating

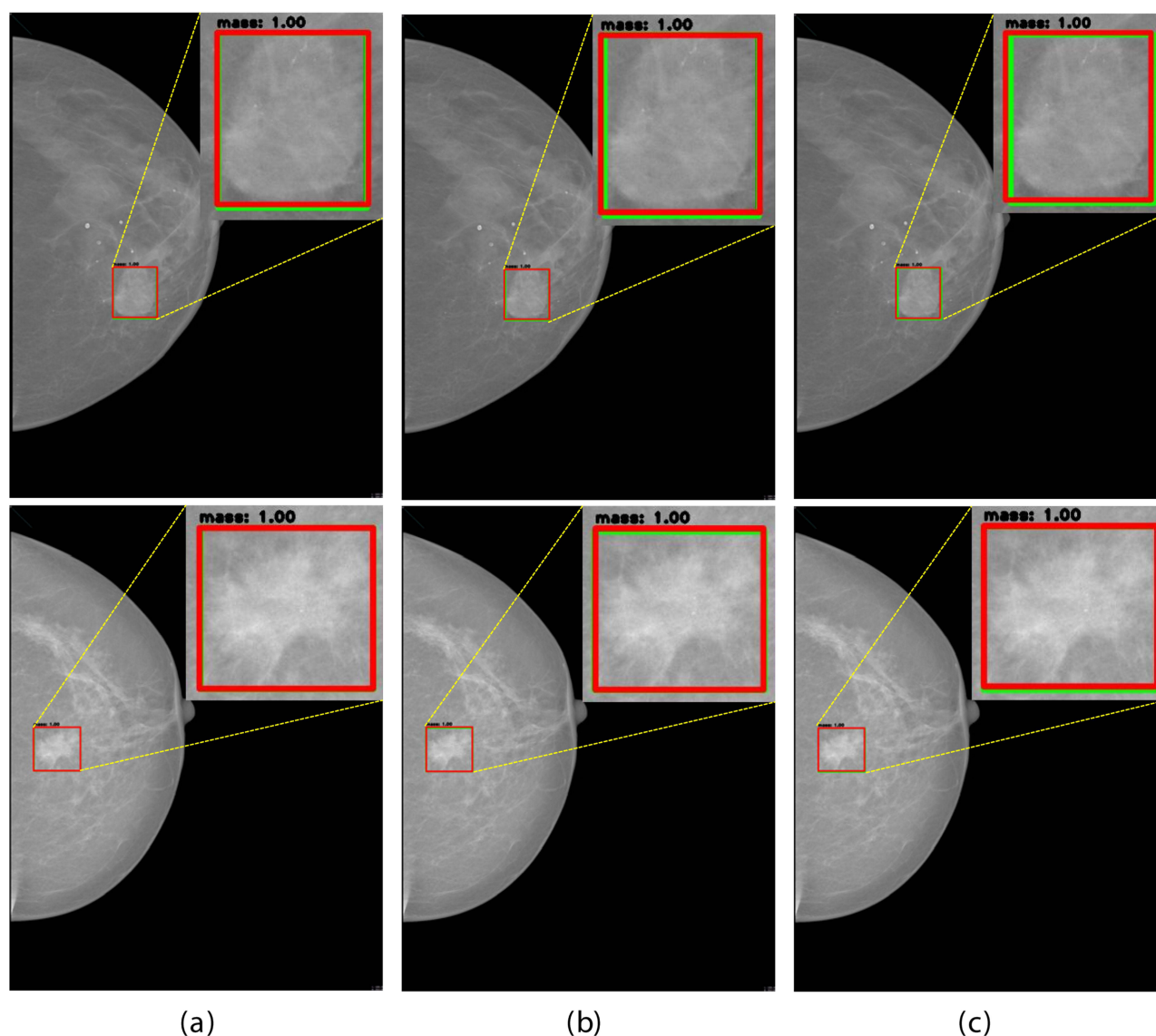


Figure 7 The ground truth (green boxes) and predicted (red boxes) locations of the masses in two mammogram images from the test subset generated using different backbones for RetinaNet. (a) ResNet 50, (b) ResNet101, (c) ResNet152.

up to 90 degrees counterclockwise, and rotating up to 180 degrees counterclockwise. Finally, the training subset is augmented to comprise 1,130 mammogram images including the original ones in the training subset and their augmented counterparts. These 1,130 images were used to train the RetinaNet network with ResNet50 as the backbone. The 117 images in the original test subset were used to test this trained RetinaNet model, and the test results were recorded as train_aug_50.

Figure 8 compares the training loss curves of the original training subset and the augmented training subset. Figure 9a shows the PR curves and the mAP values of the two RetinaNet models trained by the original training subset and the augmented training subset. Figure 9b depicts the FROC curves and the TPR@FPPI values of these two RetinaNet models. Figure 10 compares the performance of the RetinaNet models trained using the original and augmented training subsets on one image having multiple masses, from the test subset, where the green boxes represent the ground-truth and the red boxes represent the predicted results. The ground truth and the detected results are enlarged and displayed in the up-left corner of each image, so as to display the predicted results clearly.

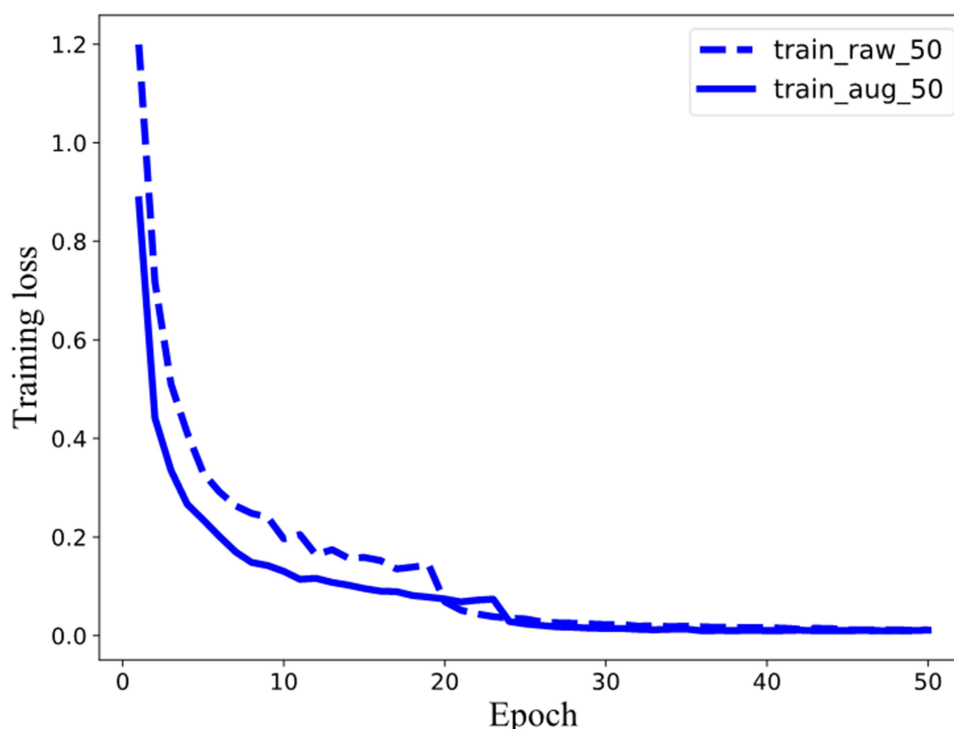


Figure 8 The training loss curves of the RetinaNet models for the original and augmented training subsets.

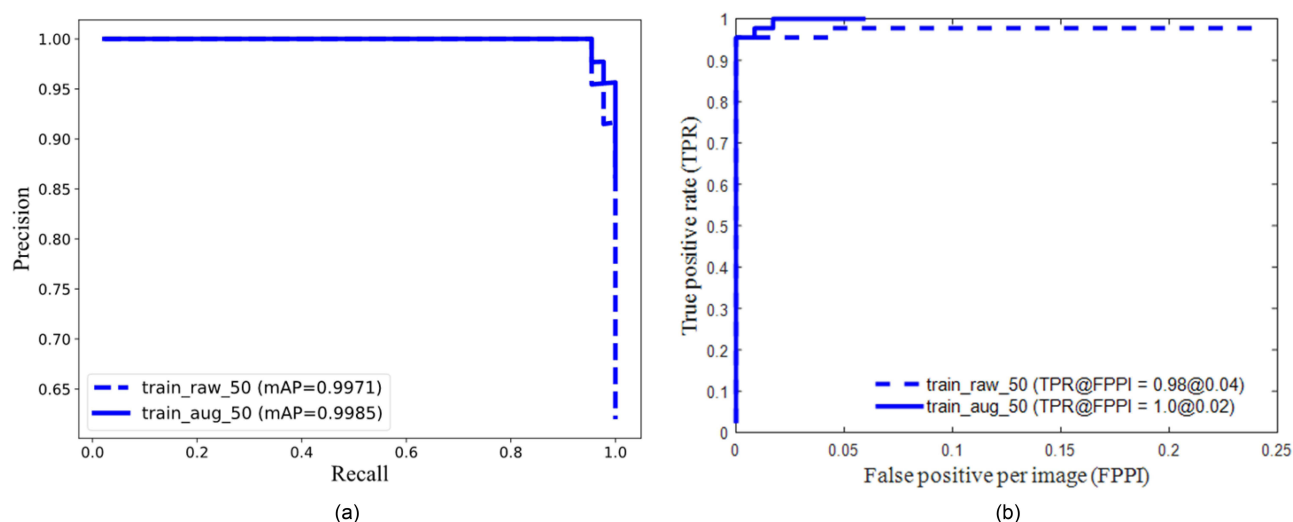


Figure 9 The PR and FROC curves of the RetinaNet models trained using the original and augmented training subsets, respectively. (a) PR curves, (b) FROC curves.

The results in Figure 8 show that the RetinaNet model's training loss on the augmented training subset converges faster than that of the original training subset. Furthermore, this model converges to a smaller training loss value, which indicates this model is more accurate than the RetinaNet model trained on the original training subset.

The results in Figure 9a show that the RetinaNet model trained on the augmented training subset obtains a higher mAP on test subset than the model from the original training subset. Furthermore, the results in Figure 9b show that the RetinaNet model trained on the augmented training subset obtains a better TPR@FPPI on test subset than the model trained with the original training subset. The maximum TPR value of the model from the augmented training subset is

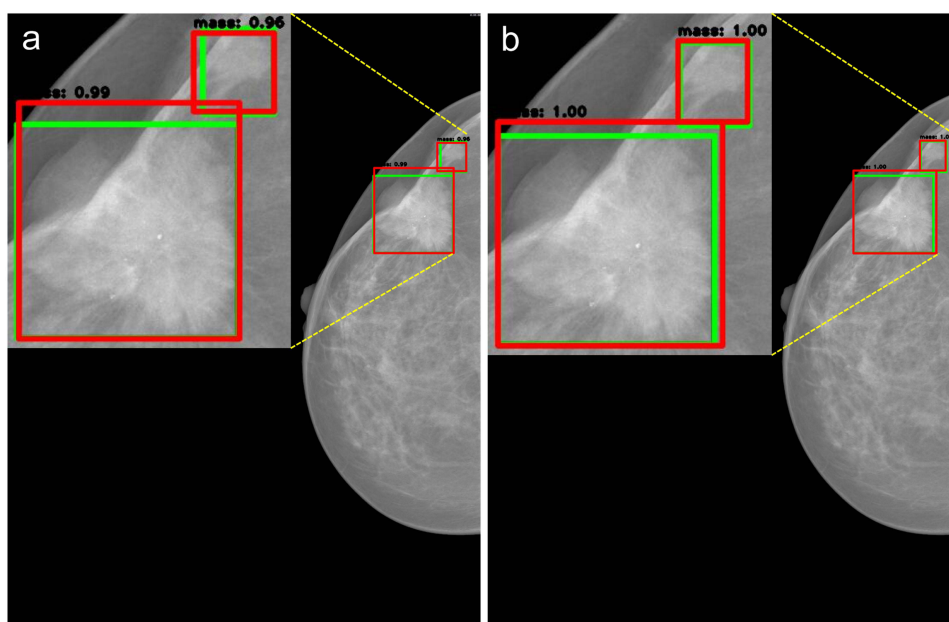


Figure 10 The test comparison of two RetinaNet models on one image in the test subset. (a) RetinaNet model trained on the original training subset, (b) RetinaNet model trained on the augmented training subset.

higher than that of the RetinaNet model trained on the original training subset, while its minimum FPPI value is lower than that of the model from the original training subset when they reach the maximum TPR, respectively.

The results in Figure 10 show that the RetinaNet model trained on the augmented training subset has more accurate detection capability than the model trained on the original 226 breast X-ray images in the INbreast database, especially for very small masses, which demonstrates important role of image augmentation in training the RetinaNet model.

The rule of thumb in here is that the image augmentation guarantees the number of images containing small masses which make the RetinaNet have more chances to extract the features of small masses, which leads to the capability of the model to detect out small masses correctly. But the RetinaNet trained on raw training subset may not extract the features of small masses extensively, which subsequently affects the capability of the model in detecting the small masses. Therefore, we will train our proposed RetinaNet using the augmented training subset.

Test Experiments With Our Improved RetinaNet Trained on the Augmented Dataset

This subsection will test the power of our improved RetinaNet network shown in Figure 4. We trained this improved RetinaNet on our augmented training subset because we have proved that the original RetinaNet network can obtain better performance when trained on this subset.

In addition, it is well known that pre-training datasets can influence the model's performance, such as the size and diversity, quality, imbalance, and domain of the pre-training datasets will lead different influence to the model. A larger and more diverse pre-training dataset will provide the model with a broader understanding of the real world, which can be beneficial when transferring knowledge to a new task. However, larger datasets also require more computational resources for pre-training. The quality of the data, including the absence of noise and the presence of high-quality labels, is crucial because high-quality data can lead to better-learned representations, which can transfer more effectively to new tasks. Class imbalance of pre-training data can lead to biased models that perform poorly on minority classes, therefore, balancing the pre-training datasets or using techniques to mitigate imbalance can improve transferability. The domain of a pre-training dataset can either help or hinder transfer learning. If the domains of the pre-training dataset and target tasks are similar, the transfer is likely to be more effective. However, if the domains are different, the model may need to "unlearn" some irrelevant features.

Considering these aforementioned aspects of pre-training datasets on the performance of the model, we choose to utilize the specific pre-training dataset to pre-train our improved RetinaNet model, so as to enhance its capability in detecting the masses, particularly the very small masses in the X-ray images of breast. Here, we introduce the specific transfer learning techniques that we used to advance the training process of RetinaNet. First, we applied the weights of RetinaNet trained by Jung et al³¹ using their inhouse database called GURO to our improved RetinaNet in Figure 4 as its initial weights. Then, we performed fine-tuning on our RetinaNet model using the 1,130 mammogram images in our augmented training subset. We recorded these test results as improved_train_aug_50.

Figure 11 displays the training loss curves of the original model and our improved RetinaNet model. Figure 12a shows the PR curves and the mAP values of the obtained two RetinaNet models on the test dataset. Figure 12b shows the FROC curves and the TPR@FPPI values of the two RetinaNet models. Figure 13 compares the performance of the original RetinaNet and our improved RetinaNet on some images in the test subset, where the green boxes represent the ground-truth and the red boxes show the predicted results. The ground truth and the detected results by models are enlarged and displayed in the up-right or up-left corner of each image, so as to display the detected results clearly.

The results in Figure 11 show that our improved RetinaNet converges faster than the original RetinaNet. The training loss of our improved RetinaNet is always lower than that of the original RetinaNet during training process. They converge to the same training loss value after training 40 Epochs.

The results in Figure 12 demonstrate that the improved RetinaNet outperformed the original RetinaNet with a mAP of 1 and a TPR@FPPI of 1.0@0.00, which means that the improved RetinaNet can detect all the masses present in the 117 mammograms in the test subset. Although the performance of the original RetinaNet is good with the mAP of 0.9985, its minimum FPPI is 0.02 when it achieves its maximum TPR of 1, while our improved RetinaNet obtains the minimum FPPI of 0 when achieving its maximum TPR of 1. We therefore say that our improved RetinaNet is perfect.

Therefore, we tested the unlabeled 67 mammogram images in the INbreast database with our trained and improved RetinaNet model. We found they were all predicted to be without masses. We believe that this information should be useful for future studies using this dataset. For example, we can add the unlabeled 67 mammogram images in the INbreast database to the training subset to improve the robustness of the model.

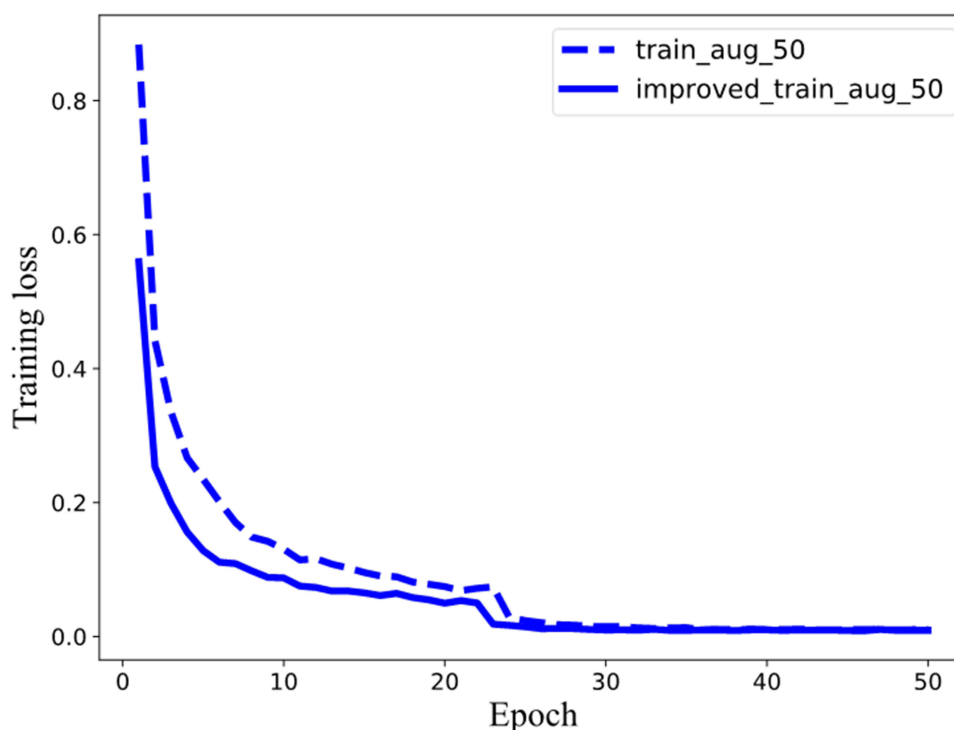


Figure 11 The training loss curves of the original RetinaNet and our improved RetinaNet.

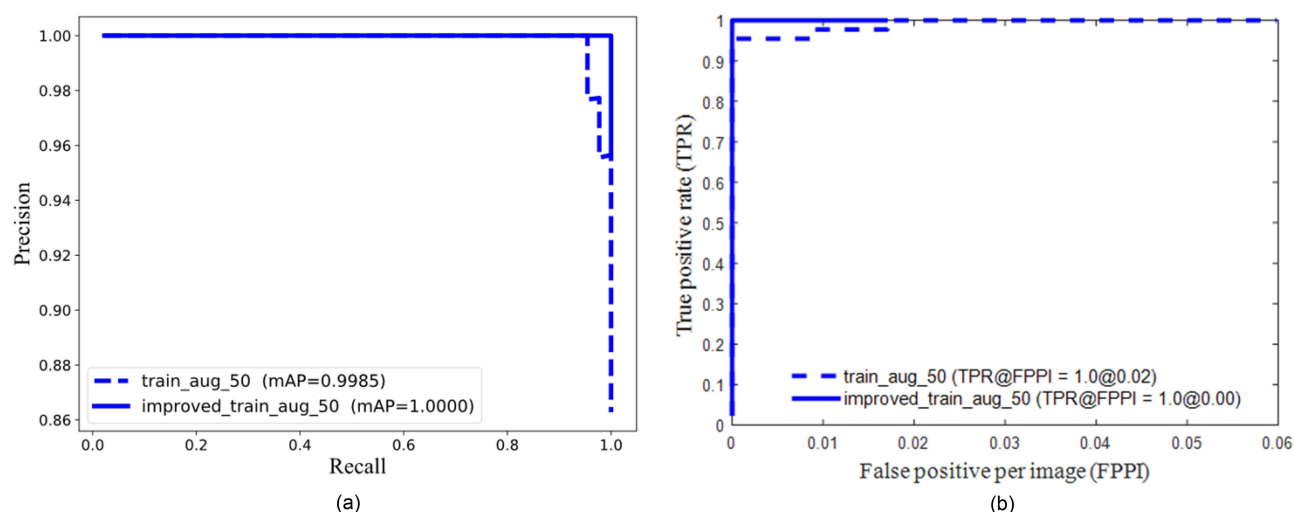


Figure 12 The PR and FROC curves of the original model and our improved RetinaNet. (a) ResNet 50, (b) FROC curves.

After performing a comprehensive analysis, we believe that the following reasons lead to our improved RetinaNet model achieving such high performance. First, the features of feature map M5 were enhanced by our extra ReLU activation function, which enhanced the resolution of the extracted features and lead to the model being able to more accurately detect very small masses. Second, we utilized transfer learning techniques by adopting the weights trained using the inhouse GURO dataset from one of the South Korean hospitals as the initial weights of our RetinaNet model before using the INbreast database to fine-tune these weights. We believe that the weights obtained from training with two completely different mammogram databases lead to the good generalization performance of our model and help it to successfully avoid overfitting.

As a result, our new RetinaNet model can be successfully applied even for patients with widely varying types of breast masses. This is especially true for the patients with multiple breast masses, which can be seen in Figure 13.

The results in Figure 13 show that our improved RetinaNet model successfully detects both the instances of very small masses and the cases of multiple masses in the mammography images of the INbreast database. It solved the challenge of making RetinaNet be able to detect multiple masses in these mammography images. Furthermore, in addition to improving the poor multi-mass detection of the original RetinaNet, it also improved the detection accuracy for specific masses in the INbreast database, which can be seen from the experimental results in the third mammography image in Figure 13.

The results in Figure 13 also demonstrate the significance of our improvement to RetinaNet. It has obtained the capability to detect small and multiple masses by adding of Relu function to enhance features in feature map M5 before its being processed by the original convolution kernel.

Experimental Comparisons

This subsection compares our experimental results with those from other researchers on the same public database (INbreast). Since the mAP metric was not used by other researchers, we only compare the results in terms of TPR@FPPI or TPR. Table 1 displays the comparison of the existing experimental results with our own results. The value in bold font in Table 1 refer to the best experimental result. Three studies^{34,35,52} presented the two types of TPR@FPPI for detecting malignant and benign masses on the INbreast dataset, respectively.

The results in Table 1 show that our experimental results obtained on the INbreast dataset in this paper are all superior to that of the available studies that we have found. Even our RetinaNet model trained on the original 226 mammogram images in the training subset can outperform the majority of studies in terms of TPR@FPPI with its value of 0.98@0.04. However, the performance of the models built by Agarwal et al,³⁴ Zebari et al,⁵⁵ Muduli et al,⁵⁷ Ayana et al⁶⁰, Adedigba et al⁶¹ was better on the INbreast dataset and could correctly identify more patients. Although this TPR value is the same as the results of Agarwal et al,³²

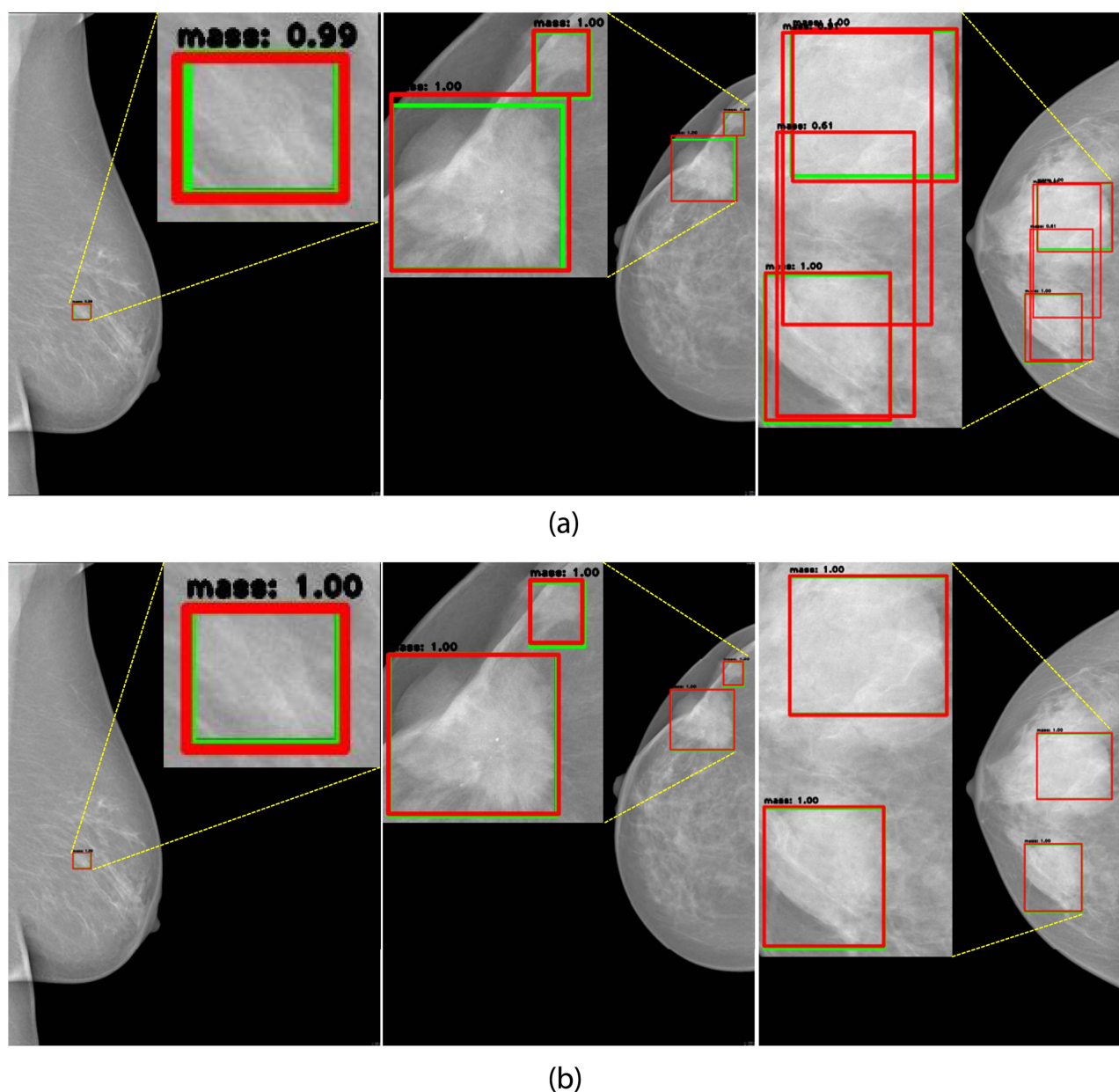


Figure 13 The performance comparison of the original and our improved RetinaNet models on three mammogram images in the test subset. (a) the original RetinaNet, (b) our improved RetinaNet.

it was obtained with a considerably lower FPPI value. In addition, although the study by Agarwal et al³⁴ achieved the TPR@FPPI of 0.99@1.17 for detecting malignant masses, its TPR@FPPI for benign masses is only 0.85@1.

The results in Table 1 also show that the experimental results can be further improved in terms of the metric of TPR@FPPI when the original training subset is augmented by our augmentation techniques. In addition, the results show that our improved RetinaNet model can achieve the best performance in terms of TPR@FPPI (1.00@0.00) among all the studies we have found, which demonstrates that this model can successfully avoid over-fitting and easily handle several of the challenging tasks of performing mass detection on the INbreast database, such as very small mass detection and multi-mass detection. Furthermore, thanks to improvements in the model's generalization capabilities granted by our use of transfer learning techniques, it can be applied to different kinds of mass detections.

Table 1 The Comparison of the Existing Experimental Results on the INbreast Database With Our Own

Previous works	Year	Classifiers	TPR@FPPI or TPR
Kozegar et al ⁵³	2013	AdaBoost, Bagging and Rotation Forest	0.87@3.67
Dhungel et al ²⁷	2015	M-DBN+R-CNN+RF	0.87@3.67
Dhungel et al ²⁸	2017	GMM+m-DBM+R-CNN+RF	0.90@1.3
Akselrod-Ballin et al ²⁹	2017	Faster-RCNN	0.93@0.56
Ribli et al ³⁰	2018	Faster R-CNN	0.90@0.3
Jung et al ³¹	2018	RetinaNet	0.97@3.0
Agarwal et al ³²	2019	InceptionV3	0.98@1.67
Min et al ³⁷	2019	Ensemble learning	0.90@0.9
Agarwal et al ³⁴	2020	Faster R-CNN	0.99@1.17 and 0.85@1
Gao et al ³⁵	2020	FT-MTL-Net	0.91@3.67 and 0.91@5
He et al ⁵⁴ used in Ref 33	2020	Mask-RCNN	0.85@3.67 and 0.85@5
Li et al ³⁶	2020	Siamese-Faster-RCNN	0.88@1.12
Yan et al ³⁸	2021	Faster R-CNN+YOLOv3	0.96@0.23
Zebari et al ⁵⁵	2021	M-FD+ANN	0.984
Zhang et al ³⁹	2022	YOLOv3	0.95@1.7
Xu et al ⁵⁶	2022	ARF-Net	0.858
Muduli et al ⁵⁷	2022	CNN	0.994
Chakravarthy et al ⁵²	2022	ResNet18+ELM	0.980 and 0.960
Singh et al ⁵⁸	2022	KNN	0.920
Elmoufidi et al ⁴⁰	2022	SVM	0.976
Lou et al ⁵⁹	2022	ECA-Net50	0.928
Ayana et al ⁶⁰	2022	EfficientNetB2	0.999
Adedigba et al ⁶¹	2022	CNN	0.997
Chakravarthy et al ⁴²	2023	wKNN+Transfer learning	0.807
Aslan et al ⁴¹	2023	CNN	0.950
Aslan et al ⁴¹	2023	CNN+BiLSTM	0.971
Prinzi et al ⁴⁴	2024	YoloV5s	0.742
Our_train_raw_50	\	RetinaNet	0.98@0.04
Our_train_aug_50	\	RetinaNet	1.00@0.02
Our_improved_train_aug_50	\	RetinaNet	1.00@0.00

Notes: The value in bold font in Table 1 refer to the best experimental result.

Abbreviations: m-DBN, multi-scale deep belief network; RF, Random forest; GMM, Gaussian mixture model; FT-MTL-Net, MTL (Multi-task deep learning) model enabled by feature transfer; M-FD, Multi fractal dimension; ANN, Artificial neural network; ARF-Net, Adaptive receptive field network; ELM, Extreme learning machine; KNN, k-nearest neighbor; ECA-Net50, Efficient channel attention-Net50; wKNN, weighted k-nearest neighbor; BiLSTM, Bidirectional long short term memories.

Discussion

This paper performed a deep learning-based approach by introducing RetinaNet and transfer learning to detect masses in mammograms. A series of experiments were conducted to compare the detection performance of the RetinaNet network and our improved RetinaNet on the INbreast dataset and its augmented dataset, respectively.

Our study's innovations lie in the comprehensive optimization and improvement of the RetinaNet model for mammogram image analysis. By carefully adapting the model to the specific characteristics of the INbreast dataset, we have achieved a level of performance that surpasses previous studies. The selection of the ResNet50 backbone, based on the analysis of training loss curves and the nature of the dataset, was a crucial step in enhancing the model's efficiency and accuracy. This approach challenges the traditional notion of simply relying on deeper network architectures and emphasizes the importance of customizing models to fit the data at hand.

The implementation of data augmentation techniques further distinguished our work. By expanding the training subset through horizontal and vertical flipping and various rotations, we provided the model with a more diverse set of samples. This not only improved its ability to detect small masses but also enhanced its overall generalization capabilities. The augmented

dataset enabled the model to learn more robust and discriminative features, leading to better performance in terms of mAP and TPR@FPPI. Moreover, our proposed improvements to the RetinaNet model, such as the addition of the ReLU activation function to enhance the feature map M5 and the utilization of transfer learning with pre-trained weights from the GURO dataset, were innovative strategies. These modifications allowed the model to converge faster and achieve a perfect mAP of 1 and a TPR@FPPI of 1.0@0.00, successfully detecting all masses in the test subset. This level of performance is a significant milestone in the field, as it addresses the challenges of detecting small and multiple masses, which are often difficult to identify with traditional methods. In comparison with existing research on the INbreast dataset, our results stand out. We have demonstrated that our model outperforms previous studies in terms of both accuracy and detection capabilities. This not only validates the effectiveness of our proposed techniques but also sets a new benchmark for future research in this area.

Our research has the potential to impact the field of breast cancer screening in several ways. The improved detection accuracy can lead to earlier diagnosis, which is crucial for improving patient outcomes. By providing more reliable and detailed information to medical doctors, our model can assist in more informed clinical decision-making, potentially reducing the need for unnecessary biopsies and improving the overall efficiency of the screening process.

It is noteworthy that when these 67 unlabeled images were utilized as an independent test dataset, our proposed enhanced RetinaNet model demonstrated that there were not masses in the unlabeled 67 mammogram images in INbreast database. This result demonstrates that our guess is correct about whether these 67 images contain masses or not, which also ascertained the efficiency of our enhanced RetinaNet. We believe that this information should be useful for future studies using this dataset. For example, we can add these unlabeled 67 mammogram images in the INbreast database to the training subset to improve the robustness of the model.

In conclusion, our results show that RetinaNet is the most suitable network to detect masses in the images of the INbreast database, based on several models we investigated. However, our initial focus and the characteristics of our training dataset have led to the model's inability to simultaneously detect asymmetry, calcifications, and other pathological changes when detecting masses in patient mammograms.

To address these limitations, we will pay attention to the subtleties and complexities of these specific pathological features in the mammogram images in the INbreast database, such as calcifications, to obtain the fine-grained label information of the images for training an enhanced RetinaNet with strong capability. In addition, we intend to incorporate attention mechanism and meta-learning and vision transformer techniques to enhance the capability of our model in our future work. Attention mechanism will learn the subtle features of specific category to enhance the capability of the model in identifying various breast pathologies. Meta-learning will enable our model to learn more generalized patterns from a diverse range of mammogram images, enhancing its ability to detect asymmetry and calcifications, especially in few-shot learning scenarios. Vision transformers, with their enhanced attention mechanisms, will capture fine-grained details and spatial relationships, assisting in identifying asymmetry and calcification patterns that traditional CNNs may miss. The integration of these advanced techniques is expected to significantly expand our model's detection capabilities, making it more comprehensive and accurate in identifying various breast pathologies. We believe that through these innovative approaches, we can overcome the existing limitations and contribute to a more robust and effective breast cancer screening tool, attracting more researchers to further advance this field.

In clinical practice, it's crucial to improve our enhanced RetinaNet model's performance for early breast cancer detection, especially in dense breast tissue and across diverse patient demographics. For the future, we will first collaborate with radiology to gather more data on relevant patients to optimize the model's sensitivity and specificity for dense breast tissue. Second, we will evaluate the model's performance in different clinical settings and populations to ensure its adaptability and make necessary adjustments. Finally, we will partner with industry to integrate the model into existing CAD systems, enhancing screening efficiency and providing more accurate diagnostic information. These efforts will enhance our research's practical value and promote breast cancer detection technology.

Conclusion

In this study, we delved into the potential of the RetinaNet model for mass detection in the INbreast database. We found the RetinaNet with ResNet50 outperformed those with ResNet101 and ResNet152, and the augmentation of training subset further improved the performance.

For addressing prevalent issues such as class imbalance and improving detection accuracy of small and multiple masses, we have innovatively enhanced the RetinaNet. By adding an additional ReLU activation function to the feature map M5, we sharpened its sensitivity to small and multiple masses. Complemented with transfer learning, our enhanced RetinaNet outperformed existing models in the literature, achieving a mAP of 1 and a TPR@FPPI of 1.0@0.00.

These innovative strategies and enhancements we have put forward possess the potential to revolutionize the modalities of breast cancer detection and diagnosis in clinical practice, thereby forging novel pathways for the furtherance of research and development within this domain. We are doing further studies building on these findings to address remaining limitations and advance the capabilities of the enhanced deep learning models for breast cancer screening and treatment.

Abbreviations

Adam, Adaptive Moment Estimation; ANN, Artificial neural network; ARF-Net, Adaptive receptive field network; BiLSTM, Bidirectional long short-term memory; BIMFs, Bidimensional intrinsic mode function; CAD, computer-aided diagnosis; CNN, Convolutional neural network; ELM, Extreme learning machine; ECA-Net50, Efficient channel attention-Net50; FPN, Feature Pyramid Network; FL, Focal Loss; FT-MTL-Net, MTL (Multi-task deep learning) model enabled by feature transfer; FPPI, false positive per image; FROC, free-response receiver operating characteristic curves; GMM, gaussian mixture model; KNN, k-nearest neighbor; mAP, mean average precision; M-FD, Multi fractal dimension; m-DBN, multi-scale deep belief network; PR curve, precision-recall curve; RF, Random forest; R-CNN, region-based convolutional neural network; ROIs, regions of interest; SVM, Support vector machine; TPR, true positive rate; VOC, Visual Object Classes; wKNN, Weighted k-nearest neighbor.

Ethical Approval

This study has obtained approval from the Academic Committee of Shaanxi Normal University [IRB code: 202414061] and has been conducted in compliance with the Declaration of Helsinki.

Acknowledgments

The authors would like to thank Moreira et al researchers for their publishing this INbreast database for us to use, and acknowledge other related researchers as well for their work to stimulate this research.

Funding

This study is supported in part by the National Natural Science Foundation of China under Grant No. 62076159, 12031010 and 61673251, the Fundamental Research Funds for the Central Universities under Grant No. GK202207017, and is also supported by the China Postdoctoral Science Foundation under Grant No. 2024M761906, and the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. of 2024JC-YBQN-0662.

Disclosure

The author(s) report no conflicts of interest in this work.

References

- Boyle P, Levin B. 2008. 412–417.
- World health organization. Breast cancer; 2024. Available from: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Accessed December 18, 2024.
- Hussein H, Abbas E, Keshavarzi S, et al. Supplemental breast cancer screening in women with dense breasts and negative mammography: a systematic review and meta-analysis. *Radiology*. 2023;306:e221785. doi:10.1148/radiol.221785
- Sahu A, Das PK, Meher S. Recent advancements in machine learning and deep learning-based breast cancer detection using mammograms. *Phys. Medica*. 2023;114:103138. doi:10.1016/j.ejmp.2023.103138
- Oza P, Sharma P, Patel S, et al. Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey. *Neural Comput Appl*. 2022;34:1815–1836. doi:10.1007/s00521-021-06804-y
- Kontos D. The promise of AI in personalized breast cancer screening: are we there yet? *Nat Rev Clin Oncol*. 2024;21:403–404. doi:10.1038/s41571-024-00877-z

7. Jabeen K, A KM, Hameed MA, et al. A novel fusion framework of deep bottleneck residual convolutional neural network for breast cancer classification from mammogram images. *Front Oncol.* **2024**;14:1347856. doi:10.3389/fonc.2024.1347856
8. Wang J, Khan MA, Wang S. SqueezeNet-guided svm for breast cancer diagnosis. *Comput. Mater. & Continua.* **2023**;76:2201–2216. doi:10.32604/cmc.2023.041191
9. Loizidou K, Elia R, Pitris C. Computer-aided breast cancer detection and classification in mammography: a comprehensive review. *Comput Biol Medicine.* **2023**;153:106554. doi:10.1016/j.combiomed.2023.106554
10. Abhisheka B, Biswas SK, Purkayastha B. A comprehensive review on breast cancer detection, classification and segmentation using deep learning. *Arch. Comput. Methods Eng.* **2023**;30:5023–5052. doi:10.1007/s11831-023-09968-z
11. Rehman S, A KM, Masood A, et al. BRMI-Net: deep learning features and flower pollination-controlled regula falsi-based feature selection framework for breast cancer recognition in mammography images. *Diagnostics.* **2023**;13:1618. doi:10.3390/diagnostics13091618
12. Fatima M, Khan MA, Shaheen S, et al. B2c3netf2: breast cancer classification using an end-to-end deep learning feature fusion and satin bowerbird optimization controlled newton raphson feature selection. *CAAI T Intell Techno.* **2023**;8:1374–1390. doi:10.1049/cit2.12219
13. Zhang Z, Wu H, Zhao H, et al. A novel deep learning model for medical image segmentation with convolutional neural network and transformer. *Interdiscip. Sci. Comput. Life Sci.* **2023**;15:663–677. doi:10.1007/s12539-023-00585-9
14. Wang Y, Wang W, Chen Q, et al. Fusing external knowledge resources for natural language understanding techniques: a survey. *Inform Fusion.* **2023**;92:190–204. doi:10.1016/j.inffus.2022.11.025
15. Jossa-Bastidas O, Sanchez AO, Bravo-Lamas L, et al. Iot system for gluten prediction in flour samples using NIRS technology, deep and machine learning techniques. *Electronics.* **2023**;12:1916. doi:10.3390/electronics12081916
16. Xie J, Hou Q, Shi Y, et al. The automatic identification of butterfly species. *J. Comput. Res. Dev.* **2018**;55:1609–1618.
17. Peng Y, Xu Y, Wang M, et al. The nnU-Net based method for automatic segmenting fetal brain tissues. *Health Inf Sci. Syst.* **2023**;11:17. doi:10.1007/s13755-023-00220-3
18. Xie J, Peng Y, Wang M. The squeeze & excitation normalization based nnu-net for segmenting head & neck tumors. *Chinese J. Electron.* **2024**;33:766–775. doi:10.23919/cje.2022.00.306
19. Xie J, Liu R, Luttrell IV, et al. Deep learning based analysis of histopathological images of breast cancer. *Front. Genetics.* **2019**;10:80. doi:10.3389/fgene.2019.00080
20. Yu X, Kang C, Guttery DS, et al. Resnet-scds-50 for breast abnormality classification. *IEEE ACM T Comput Bi.* **2021**;18:94–102.
21. Xie L, Zhang L, Hu T, et al. Neural networks model based on an automated multi-scale method for mammogram classification. *Knowledge-Based Syst.* **2020**;208:106465. doi:10.1016/j.knosys.2020.106465
22. El houbi EM, Yassin NI. Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. *Biomed. Signal Process Control.* **2021**;70:102954. doi:10.1016/j.bspc.2021.102954
23. Das R, Kaur K, Walia E. Feature generalization for breast cancer detection in histopathological images. *Interdiscip Sci Comput Life Sci.* **2022**;14:566–581. doi:10.1007/s12539-022-00515-1
24. Lin TY, Goyal P, Girshick R, He K, Doll'ar P. P. Doll ar, "Focal loss for dense object detection Proc. IEEE Int. Conf. Comput. Vis **2017**.
25. Moreira IC, Amaral I, Domingues I, et al. INBreast: toward a full-field digital mammographic database. *Acad Radiol.* **2012**;19:236–248. doi:10.1016/j.acra.2011.09.014
26. Carneiro G, Nascimento J, Bradley AP. **2015. Unregistered Multiview Mammogram Analysis With Pre-Trained Deep Learning Models. Medical Image Computing and Computer-Assisted Intervention–MICCAI; 2015.** Munich, Germany. Berlin: Springer; Oct 5-9.
27. Dhungel N, Carneiro G, Bradley AP. Automated mass detection in mammograms using cascaded deep learning and random forests. In: *Proceedings of the Digital Image Computing: Techniques and Applications (DICTA); 2015.* IEEE; **2015**:Nov23–25.
28. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Anal.* **2017**;37:114–128. doi:10.1016/j.media.2017.01.009
29. Akselrod-Ballin A, Karlinsky L, Hazan A, et al. Deep learning for automatic detection of abnormal findings in breast mammography. *deep learning in medical image analysis and multimodal learning for clinical decision support.* In: *Held in Conjunction With MICCAI 2017; 2017.* QC, Canada; Berlin: Québec City; **2017**.
30. Ribli D, Horv'ath A, Unger Z, et al. Detecting and classifying lesions in mammograms with deep learning. *Sci. Reports.* **2018**;8:4165.
31. Jung H, Kim B, Lee I, et al. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. *PLoS One.* **2018**;13:e0203355. doi:10.1371/journal.pone.0203355
32. Agarwal R, Diaz O, Llad'o X, et al. Automatic mass detection in mammograms using deep convolutional neural networks. *J. Med. Imaging.* **2019**;6:031409. doi:10.1117/1.JMI.6.3.031409
33. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* **2016**.
34. Agarwal R, Diaz O, Yap MH, Llad'o X, Marti R. Deep learning for mass detection in full field digital mammograms. *Comput Biol Med.* **2020**;121:103774. doi:10.1016/j.combiomed.2020.103774
35. Gao F, Yoon H, Wu T, et al. A feature transfer enabled multi-task deep learning model on medical imaging. *Expert Systems With Applications.* **2020**;143:112957. doi:10.1016/j.eswa.2019.112957
36. Li Y, Zhang L, Chen H, et al. Mass detection in mammograms by bilateral analysis using convolution neural network. *Comput Meth Prog Bio.* **2020**;195:105518. doi:10.1016/j.cmpb.2020.105518
37. Min H, Chandr SS, Crozier S, et al. Multi-scale sifting for mammographic mass detection and segmentation. *Biomed. Phys & Eng. Express.* **2019**;5:025022. doi:10.1088/2057-1976/aafc07
38. Yan Y, Conze PH, Lamard M, et al. Towards improved breast mass detection using dual-view mammogram matching. *Med Image Analysis.* **2021**;71:102083. doi:10.1016/j.media.2021.102083
39. Zhang L, Li Y, Chen H, et al. W Wu, Anchor-free yolov3 for mass detection in mammogram. *Expert Syst Appl.* **2022**;191:116273. doi:10.1016/j.eswa.2021.116273
40. Elmoufidi A. Deep multiple instance learning for automatic breast cancer assessment using digital mammography. *IEEE T Instrum Meas.* **2022**;71:4503813. doi:10.1109/TIM.2022.3177141

41. Aslan MF. A hybrid end-to-end learning approach for breast cancer diagnosis: convolutional recurrent network. *Comput Electr Eng.* **2023**;105:108562. doi:10.1016/j.compeleceng.2022.108562
42. Chakravarthy SS, Bharanidharan N, Rajaguru H. Deep learning-based metaheuristic weighted k-nearest neighbor algorithm for the severity classification of breast cancer. *IRBM.* **2023**;44:100749. doi:10.1016/j.irbm.2022.100749
43. Sannasi Chakravarthy S, Bharanidharan N, Rajaguru H. Multi-deep cnn based experimentations for early diagnosis of breast cancer. *IETE J. Res.* **2023**;69:7326–7341. doi:10.1080/03772063.2022.2028584
44. Prinzi F, Insalaco M, Orlando A, et al. A yolo-based model for breast cancer detection in mammograms. *Cogn Comput.* **2024**;16:107–120. doi:10.1007/s12559-023-10189-6
45. Xie J, Liu R. The study progress of object detection algorithms based on deep learning. *J Shaanxi Norm Univ.* **2019**;47:1–9.
46. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* **2014**.
47. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* **2016**.
48. Lin T, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017.* **2017**.
49. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. *Int. Journal Computer Vision.* **2010**;88:303–338. doi:10.1007/s11263-009-0275-4
50. Bunch PC, Hamilton JF, Sanderson GK, Simmons AH. A free response approach to the measurement and characterization of radiographic observer performance. *Application of Optical Instrumentation in Medicine*; **1977**. Vol. 1977.
51. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* **2016**.
52. Chakravarthy SS, Rajaguru H. Automatic detection and classification of mammograms using improved extreme learning machine with deep learning. *Irbm.* **2022**;43:49–61. doi:10.1016/j.irbm.2020.12.004
53. Kozegar E, Soryani M, Minaei B, et al. Assessment of a novel mass detection algorithm in mammograms. *J Cancer Research Therapeutics.* **2013**;9:592–600. doi:10.4103/0973-1482.126453
54. He K, Gkioxari G, Doll'ar P, et al. Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision.* **2017**.
55. Zebari DA, Ibrahim DA, Zeebaree DQ. Breast cancer detection using mammogram images with improved multi-fractal dimension approach and feature fusion. *Appl Sci.* **2021**;11:12122. doi:10.3390/app112412122
56. Xu C, Qi Y, Wang Y, et al. Arf-net: an adaptive receptive field network for breast mass segmentation in whole mammograms and ultrasound images. *Biomed Signal Process Control.* **2022**;71:103178. doi:10.1016/j.bspc.2021.103178
57. Muduli D, Dash R, Majhi B. Automated diagnosis of breast cancer using multi-modal datasets: a deep convolution neural network based approach. *Biomed Signal Process Control.* **2022**;71:102825.
58. Singh H, Sharma V, Singh D. Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor. *Vis Comput for Ind Biomed Art.* **2022**;5:3. doi:10.1186/s42492-021-00100-1
59. Lou Q, Li Y, Qian Y, et al. Mammogram classification based on a novel convolutional neural network with efficient channel attention. *Comput. Biol. Medicine.* **2022**;150:106082. doi:10.1016/j.compbiomed.2022.106082
60. Ayana G, Park J, Choe S. Patchless multi-stage transfer learning for improved mammographic breast mass classification. *Cancers.* **2022**;14:1280. doi:10.3390/cancers14051280
61. Adedigba AP, Adeshina SA, Aibinu AM. Performance evaluation of deep learning models on mammogram classification using small dataset. *Bioengineering.* **2022**;9:161. doi:10.3390/bioengineering9040161