



Article

A Comparative Study of Deep Learning in Breast Ultrasound Lesion Detection: From Two-Stage to One-Stage, from Anchor-Based to Anchor-Free

Yu Wang¹, Qi Zhao¹, Baihua Zhang², Dingcheng Tian¹, Ruyi Zhang¹ and Wan Zhong^{3,*}

¹ College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110024, China

² Research Institute for Medical and Biological Engineering, Ningbo University, Ningbo 315000, China

³ General Hospital of Northern Theater Command, Shenyang 110024, China

* Correspondence: wzhong_88@163.com

How To Cite: Wang, Y.; Zhao, Q.; Zhang, B.; Tian, D.; Zhang, R.; Zhong, W. A Comparative Study of Deep Learning in Breast Ultrasound Lesion Detection: From Two-Stage to One-Stage, from Anchor-Based to Anchor-Free. *AI Medicine* **2024**, *1*(1), 5. <https://doi.org/10.53941/aim.2024.100005>.

Received: 16 July 2024

Revised: 26 August 2024

Accepted: 27 August 2024

Published: 4 September 2024

Abstract: Breast cancer is one of the most common tumors among women in the world, and its early screening is crucial to improve the survival rate of patients. Breast ultrasound, with the characteristics of non radiation, real-time imaging and easy operation, has become a common method for breast cancer detection. However, this method has some problems, such as low imaging quality and strong subjectivity of diagnosis results, which affect the accurate diagnosis of breast cancer. With the ongoing advancement of deep learning technology, intelligent breast cancer detection systems have effectively overcome these challenges, enhancing diagnostic accuracy and efficiency. This study uses nine popular deep learning object detection networks (including two-stage, one-stage, anchor-based, and anchor-free networks) for the detection of breast lesions and compares the results of these methods. The experiments show that the anchor-based Single Shot MultiBox Detector (SSD) network excels in overall performance, while the anchor-free Fully Convolutional One-stage Object Detector (FCOS) exhibits the best generalization ability. Moreover, the results also indicate that, in the context of breast lesion detection, anchor-based networks generally outperform anchor-free networks.

Keywords: deep learning; breast ultrasound image; breast cancer; breast lesion detection; object detection

1. Introduction

According to the global cancer statistics report of 2018, 11.6% of cancer patients worldwide are diagnosed with breast cancer, making it the second most prevalent cancer globally [1]. Each year, approximately 2.89 million women are diagnosed with breast cancer, accounting for 24.2% of all female cancer cases [1]. Clinical studies show that the survival rate of breast cancer is closely related to the early detection and staging of the disease; the earlier it is detected, the higher the possibility of survival [2]. Therefore, early screening for breast cancer is crucial.

In clinic, the detection of breast cancer typically relies on three types of medical imaging technologies, Mammography, Digital Breast Tomosynthesis (DBT) and medical ultrasound imaging. Each of these technologies has its own advantages and unique limitations in breast cancer detection. While Mammography can reach a detection sensitivity of up to 85% in general female populations, its sensitivity decreases to 47.8–64.4% in women with dense breast tissue [3]. This is due to the lower distinction between breast tissue and tumors in dense breast, leading to potential missed-detection. Moreover, Mammography carries radiation risks and is relatively costly. DBT also faces similar issues of high costs and radiation exposure. In contrast, breast ultrasound imaging is a comparatively lower-cost, non-ionizing radiation method that provides real-time imaging. It performs well in detecting hidden



breast cancers in dense breast tissues [4], thus becoming an important tool in breast cancer detection. However, the diagnostic results of breast ultrasound largely depends on the doctor's skill and experience level. Variations in training backgrounds and clinical experiences can lead to different diagnostic results for the same ultrasound images [5]. Additionally, ultrasound images often suffer from issues like noise interference, strong artifacts, and low contrast between tissue structures [6].

To address the above issues, many researchers have conducted research on the automated diagnosis of breast ultrasound images. Breast cancer automatic diagnosis typically includes two steps: lesion detection and lesion classification. In earlier studies, researchers generally used traditional digital image processing methods for lesion recognition and classification. For instance, Drucker et al. [7] identified breast lesion areas using radial gradient index filtering in a study on breast cancer classification, and then input the identified areas into a Bayesian classifier for benign-malignant lesion classification. In another study, Liu et al. [8] on lesion area identification, they initially conducted a preliminary analysis of breast ultrasound images using texture features, followed by refining the coarse identification results with active contour method, achieving precise segmentation of breast lesions to assist subsequent lesion classification. With the advancement of artificial intelligence algorithms, machine learning algorithms have been increasingly applied in the automated diagnosis of breast cancer. For example, Shan et al. [9] first determined the approximate location of breast lesions using traditional image processing methods in a lesion segmentation study, then extracted frequency and spatial domain features of the lesion area, and fed these features into a shallow artificial neural network for feature analysis, obtaining precise segmentation results of breast lesions. However, shallow artificial neural networks based on traditional machine learning algorithms still have limited feature extraction capabilities and cannot meet the requirements for high-precision breast lesion detection and classification.

With the development of computer hardware and advancements in deep learning algorithms, coupled with the powerful feature extraction and analysis capabilities of deep neural networks, deep learning has achieved remarkable successes in various fields. Consequently, researchers have shifted from using traditional machine learning algorithms to deep learning algorithms for automated breast cancer diagnosis. As mentioned earlier, automated breast cancer diagnosis mainly includes lesion detection and classification, which aligns well with the task of object detection in deep learning. Therefore, many researchers have applied deep learning object detection methods to the automated detection of breast cancer. Yap et al. [10] used Faster R-CNN [11] for the identification of breast lesions in ultrasound images and achieved good breast cancer detection performance through transfer learning and multi-feature image fusion methods. In a study on breast lesion detection, Wang et al. [12] used segmentation-based image enhancement techniques to enhance the contrast of breast ultrasound images, then input them into a Fully Convolutional One-stage Object Detector (FCOS) [13], achieving a mean average precision (mAP) of 90.2%. Cao et al. [14] compared the performance of five anchor-based object detection methods in detecting lesions in breast ultrasound images, with the Single Shot MultiBox Detector (SSD) [15] network achieving the best accuracy and Recall. Mo et al. [16] improved the preset anchor size of You Only Look Once (YOLO) V3 [17] using clustering methods and applied it to breast ultrasound lesion detection, achieving an mAP of 89.34%. Yu et al. [18] presented GFNet, a novel framework for breast mass detection, which integrates patch extraction, feature extraction, and mass detection modules. GFNet demonstrates high robustness and adaptability across different imaging devices.

As previously mentioned, researchers have used various categories of object detection networks for the automated detection of breast cancer, including Two-Stage (Faster R-CNN), One-Stage (YOLO V3), Anchor-based (SSD), and Anchor-free (FCOS) networks. However, in past work, there has been a scarcity of comparative studies on the performance of these different categories of object detection networks in detecting breast lesions. In this paper, we select nine popular object detection algorithms, encompassing Two-Stage, One-Stage, Anchor-based, and Anchor-free categories, and conduct a comprehensive comparison of their performance in breast lesion detection. The nine object detection networks are Faster R-CNN (Two-Stage, Anchor-based), SSD (One-Stage, Anchor-based), YOLO V3 (One-Stage, Anchor-based), RetinaNet [19] (One-Stage, Anchor-based), YOLOF [20] (One-Stage, Anchor-based), CornerNet [21] (One-Stage, Anchor-free), FCOS, TTFNet [22] (One-Stage, Anchor-free), and YOLOX [23] (One-Stage, Anchor-free).

2. Materials and Methods

In this section, we will introduce the datasets and object detection networks used in this study.

2.1. Datasets

This study uses data from two public datasets, BUS dataset [24] and BUSI dataset [25], with the images from these datasets as shown in Figure 1. As shown Figure 1, we can observe that compared to BUS dataset, the

ultrasound images in BUSI dataset have lower grayscale values and also contain more noise.

The BUS dataset from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell (Spain), where images were collected using the Siemens ACUSON Sequoia C512 17L5 HD linear array sensor (8.5 MHz). BUS dataset contains 163 breast ultrasound images with varying original size, averaging 760×570 , and each image includes one or more lesion areas. Of these 163 lesion images, 53 are malignant and 110 are benign. The malignant breast images include 40 cases of invasive ductal carcinoma, 4 cases of ductal carcinoma in situ, 2 cases of invasive lobular carcinoma, and 7 cases of other unspecified malignancies. In terms of benign breast images, there are 65 cases of unspecified cysts, 39 fibroadenomas, and 6 other types of benign lesions. All images were manually segmented and classified by radiologists, marking the lesion areas. Both the original breast images and the annotated images are saved in png format, and an xlsx file provides lesion type information for each image.

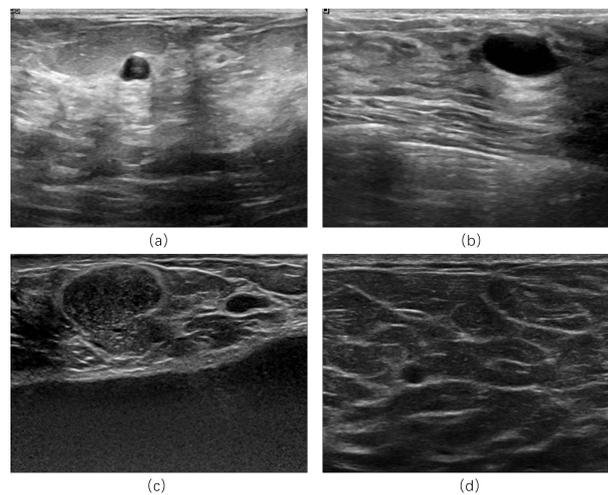


Figure 1. BUS and BUSI dataset images. (a,b) from BUS dataset, (c,d) from BUSI dataset.

BUSI dataset is from Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt, collected using the LOGIQ E9 and LOGIQ E9 Agile ultrasound systems. The breast ultrasound images were gathered from 600 female subjects aged between 25 and 75 years. Initially, this dataset contained a total of 1100 images. Each image's lesion area was manually segmented using Matlab software and classified as normal, benign, or malignant. However, after radiologists at Baheya Hospital removed duplicate and incorrectly annotated images, a total of 780 images remained, comprising 437 benign images, 210 malignant images, and 133 normal breast images (without lesions). Notably, the original size of BUSI images was 1280×1024 , but due to the presence of large amounts of irrelevant areas in the original images, they were cropped to a size of 500×500 and saved in png format.

Both BUS dataset and BUSI dataset contain accurate labels for breast lesion edge segmentation and benign-malignant classification. However, these labels are not suitable for the labeling requirements of object detection task. Therefore, we reprocess the labels of both BUS dataset and BUSI dataset to make them appropriate for breast lesion detection task, as shown in Figure 2. We traverse the points of the breast lesion contours in Figure 2b to locate the top, bottom, left, and right endpoints and then determine the top-left and bottom-right points of the lesion area and to obtain height and width of the lesion, as depicted in Figure 2c.

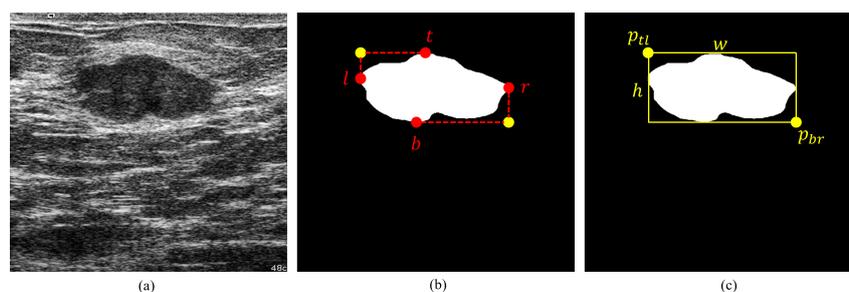


Figure 2. The process of creating labels for breast lesion detection. (a) Original ultrasound images; (b) ground truth in binary mask, yellow points represent the top-left and bottom-right corners of the ground truth; (c) represents a bounding box made according to the yellow points.

2.2. Deep Learning Neural Networks

Since the development of R-CNN [26], various highly accurate object detection networks based on deep learning have emerged. Generally, object detection networks can be categorized by the number of stages into two-stage and one-stage methods, or by the use of preset anchors into anchor-based and anchor-free methods. In this study, we select nine currently popular object detection networks and compare their performance in breast lesion detection tasks. The chosen networks include two-stage, one-stage, anchor-based, and anchor-free object detection methods, with specific descriptions of these networks provided in Table 1.

Table 1. Description of object detection networks.

Model	Number of Stage	Anchor Setting	Network Description
Faster R-CNN	Two-Stage	Anchor-based	Faster R-CNN introduced a Region Proposal Network to achieve real-time detection. Efficiency is improved through the sharing of convolutional features, and preset anchors are used to regress the position of the object, significantly enhancing detection speed and accuracy.
RetinaNet	One-Stage	Anchor-based	RetinaNet addresses the issue of class imbalance in object detection by introducing Focal Loss, which focuses on samples that are difficult to classify.
SSD	One-Stage	Anchor-based	SSD detects objects of various sizes effectively by predicting categories and bounding boxes on feature maps at multiple scales.
YOLO V3	One-Stage	Anchor-based	YOLO V3 can classify and locate in a single forward pass. It introduces multi-scale detection, using feature maps at three different scales to improve the detection of small objects.
YOLOF	One-Stage	Anchor-based	YOLOF simplifies the network structure by reducing the number of feature pyramid layers, maintaining high detection performance. This design lowers computational costs while increasing speed.
CornerNet	One-Stage	Anchor-free	CornerNet uses a corner detection method, locating objects by detecting their top-left and bottom-right corners.
FCOS	One-Stage	Anchor-free	FCOS predicts the size and center point of the object's bounding box directly on the feature map, offering a straightforward method to handle objects of various shapes and sizes.
TTFNet	One-Stage	Anchor-free	TTFNet uses a dense detection head and an efficient feature fusion strategy. It maintains high detection accuracy while significantly enhancing detection speed.
YOLOX	One-Stage	Anchor-free	YOLOX introduces an anchor-free design and decoupled head, and optimizes the label assignment strategy.

During the experimentation, we substitute the backbone of some networks to further compare the performance of different networks in breast lesion detection. We select ResNet [27], VGG [28], and DarkNet [29] as the backbones for most of the networks.

3. Results

In this section, we will introduce the performance metrics used in our experiments, the details of the experiments, and the performance results of each network. We chose the output results of the SSD network for demonstration, as shown in Figure 3. The network draws bounding boxes in different colors based on the predicted nature of the lesion, red for lesions predicted to be malignant and green for those predicted to be benign. The confidence level of the prediction is displayed above the bounding box.

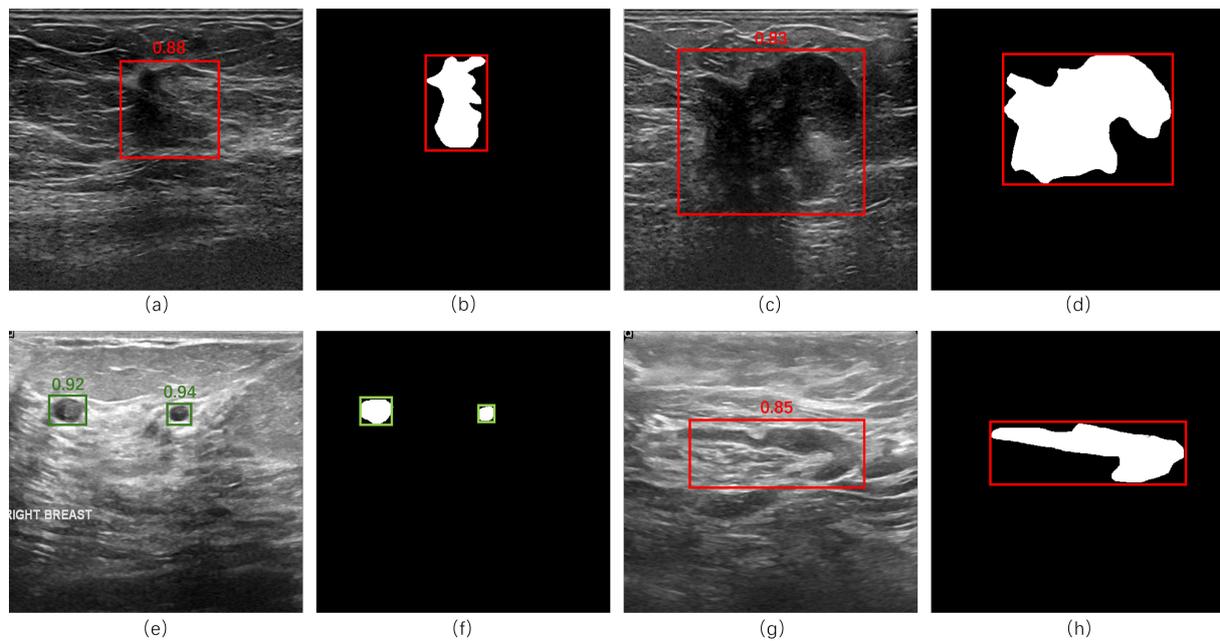


Figure 3. Breast lesion detection results of SSD network. (a,c,e,g) are prediction results. (b,d,f,h) are ground truth.

3.1. Performance Metrics

In this study, we use commonly used metrics in object detection, Average Precision (AP), Average Recall (AR), and Frames Per Second (FPS) as the performance metrics for our study.

AP represents the area under the Precision-Recall (PR) curve in object detection and is calculated based on the following values. First, it is necessary to compute the Intersection over Union (IoU) threshold (T) between the predicted and actual bounding boxes, as well as the confidence scores for the classification prediction of the bounding boxes. We have,

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

Then we have,

True Positives (TP): The prediction BBox with $\text{IoU} > T$ and meeting the category Confidence threshold.

False Positives (FP): The prediction BBox with $\text{IoU} < T$ and meeting the category Confidence threshold.

False Negatives (FN): The prediction BBox with $\text{IoU} = 0$.

Based on the TP, FP, and FN, we have,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

Based on different confidence thresholds for each category, we can plot the Precision-Recall (PR) curve, thereby determining the AP value. By adjusting various IoU thresholds, we can calculate AP50 ($T > 0.5$) and AP75 ($T > 0.75$). AR10 refers to the average recall rate when the IoU threshold is set to $T > 0.1$.

3.2. Experiment Implementation

In this study, we implement all comparative networks using PaddleDetection [30]. Each network trains for 300 epochs, evaluating performance on the validation set after each epoch. The model parameters that with the best performance on the validation set during these 300 epochs are retained as the final parameters. During the training process, the first five epochs use model warm-up, and for the remainder of training, a cosine learning rate decay strategy [31] reduces the learning rate to one percent of the initial rate. We apply random rotation as a preprocessing method. The image size is 320×320 .

We conduct model training on both the combined BUS+BUSI mixed-dataset and the single BUSI dataset. Both data groups are divided into training, validation, and test sets in an 8:1:1 ratio. All breast ultrasound images are resized to 320×320 with the learning rate set to 0.01 and the batch size set to 8, and during the training process, we use image augmentation methods such as random rotation, random flipping, and Mosaic [32].

3.3. Results of BUS+BUSI Mixed-Dataset

First, we evaluate the performance of object detection networks using BUS+BUSI mixed-dataset, with results shown in Table 2. From Table 2, we observe that within the anchor-based networks, YOLOV3-res34 performs best in terms of AP, reaching 0.637, and also leads in AP75 and AR10, indicating its advantages in accuracy. In terms of processing speed, SSD-vgg16 and SSD-res34, with nearly 30 FPS, outperform other networks. Additionally, SSD achieves the best result in AP50, indicating its excellent overall capabilities. Among the anchor-free networks, YOLOX-m leads with an AP of 0.563 and shows good performance in AP50, AP75, and AR10, exhibiting a balanced performance advantage. FCOS achieves slightly lower performance than YOLOX-m. On the other hand, although TTFNet reaches the highest FPS (38.37), it significantly behind in terms of accuracy.

Table 2. Performance comparison of different object detection networks on mixed-dataset.

	Model	AP	AP50	AP75	AR10	FPS
Anchor-based networks	Faster R-CNN-res50	0.573	0.882	0.672	0.677	14.01
	RetinaNet-res50	0.564	0.869	0.619	0.655	14.87
	SSD-res34	0.582	0.863	0.596	0.631	29.65
	SSD-vgg16	0.608	0.931	0.666	0.691	29.85
	YOLOF-res50	0.533	0.897	0.519	0.617	22.7
	YOLOV3-darknet53	0.632	0.925	0.678	0.683	19.5
	YOLOV3-res34	0.637	0.899	0.77	0.686	26.88
Anchor-free networks	CornerNet-res50	0.518	0.791	0.612	0.627	11.03
	FCOS-res50	0.541	0.821	0.62	0.629	17.18
	TTFNet	0.368	0.624	0.443	0.476	38.37
	YOLOX-m	0.563	0.887	0.651	0.69	26.53

Note: Bold font indicates the best performance results.

Figure 4 presents the performance results and AP-FPS plot of each network. In Figure 4b, the closer a network's performance is to the top-right corner, the stronger its overall performance. Overall, the two anchor-based object detection networks, YOLOV3 and SSD, show excellent performance, while the anchor-free networks are slightly behind the anchor-based networks in terms of performance.

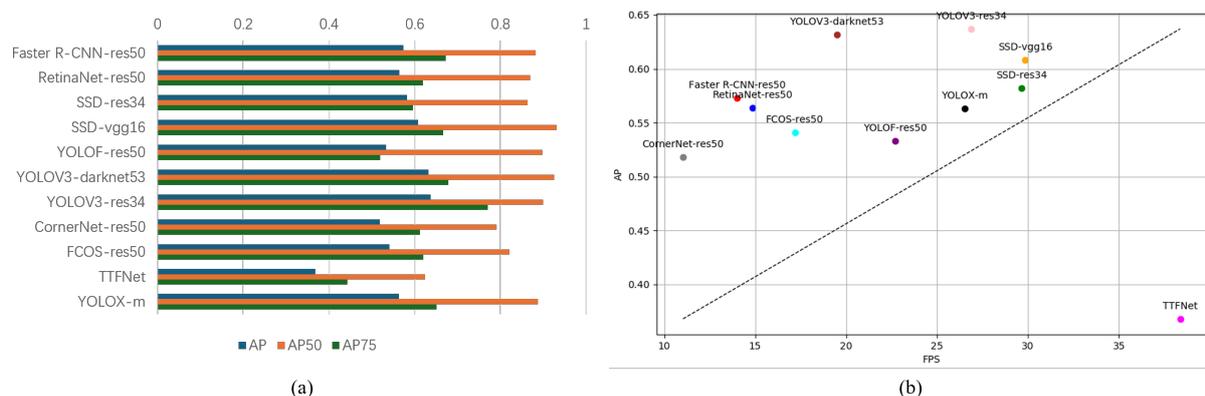


Figure 4. Networks performance results and AP-FPS plot on BUS+BUSI mixed-dataset. (a) is the performance results of different networks, (b) is the FPS and AP scatter plot of the networks.

3.4. Results of BUSI Dataset

Next, we compare the performance of the nine networks on BUSI dataset, with results shown in Table 3. Among the anchor-based networks, RetinaNet performs the best on BUSI dataset, achieving the highest AP and AR, as well as the second-highest AP75, but it shows some disadvantages in network speed. YOLOV3 and SSD, which perform well on BUS+BUSI mixed-dataset, still show excellent performance on BUSI dataset, achieving balanced results in both accuracy and network speed. For anchor-free networks, FCOS achieves an AP of 0.841, close to the best-performing anchor-based model RetinaNet-res50, and it achieves the best results among anchor-free networks in AP50, AP75, and AR. However, YOLOX-m, which performs relatively well on the BUS+BUSI mixed-dataset, has a significant decrease in performance on BUSI dataset.

Table 3. Performance comparison of different object detection networks on BUSI.

	Model	AP	AP50	AP75	AR10	FPS
Anchor-based networks	Faster R-CNN-res50	0.584	0.88	0.703	0.725	13.62
	RetinaNet-res50	0.849	0.962	0.927	0.885	17.42
	SSD-res34	0.813	0.939	0.924	0.845	31.4
	SSD-vgg16	0.791	0.965	0.947	0.845	29.52
	YOLOF-res50	0.823	0.962	0.877	0.856	23.3
	YOLOV3-darknet53	0.769	0.979	0.919	0.81	19.55
	YOLOV3-res34	0.791	0.966	0.95	0.825	25.52
Anchor-free networks	CornerNet-res50	0.535	0.823	0.694	0.702	11.42
	FCOS-res50	0.841	0.928	0.888	0.881	18.4
	TTFNet	0.394	0.658	0.485	0.526	38.11
	YOLOX-m	0.578	0.853	0.695	0.714	26.14

Note: Bold font indicates the best performance results.

Figure 5 presents the performance results and AP-FPS plot of each network on the BUSI dataset. Overall, YOLOV3 and SSD still demonstrate the most overall performance, similar to the results with BUS+BUSI mixed-dataset. Although RetinaNet and FCOS show impressive performance in AP, their lower FPS affects their overall performance.

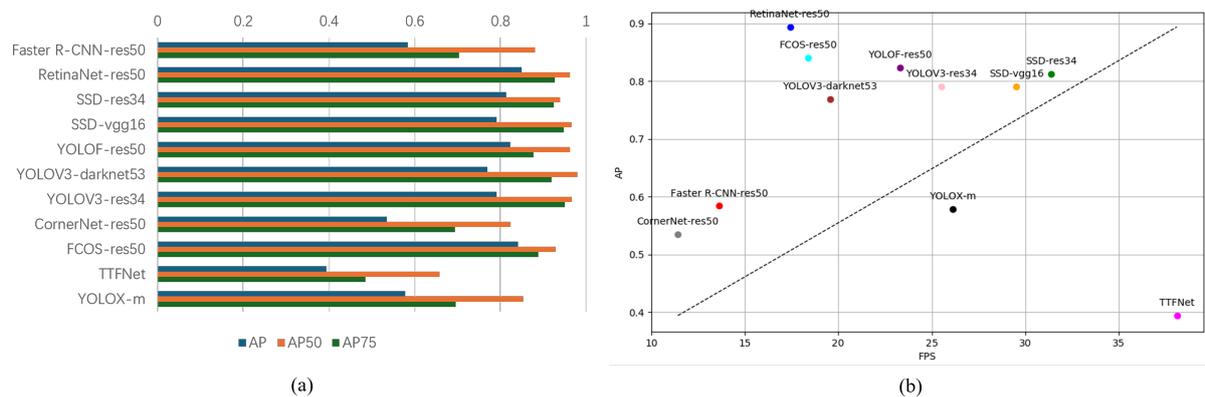


Figure 5. Networks performance results and AP-FPS plot on BUSI dataset. (a) is the performance results of different networks, (b) is the FPS and AP scatter plot of the networks.

3.5. Results of Generalization Performance

In medical image analysis, the generalization ability of a model is particularly important, as it directly relates to the model's practicality and reliability. A model with good generalization ability can adapt to a diverse range of cases, reducing the risk of misdiagnosis and missed diagnosis, thereby enhancing the accuracy and reliability of diagnoses. It ensures that the model accurately identifies and classifies data that differ in lesion shape, size or appearance from the training data. Strong generalization also means that the model can adapt to images from different devices and protocols, enhancing its application value in real clinical environments. Therefore, in this study, we compare the generalization performance of the nine networks. We train the models using BUSI dataset and validate them on BUS dataset, with validation results shown in Table 4.

From Table 4, we observe that FCOS achieves excellent performance in generalization, achieving the best results in AP, AP50, and AR, and the second-best in AP50, demonstrating its strong generalization ability. RetinaNet, which performs well on the BUSI dataset, also achieves good results, with the second-best AP. As shown in Figure 6b, SSD approaches the top right corner, indicating excellent overall performance, achieving a balance between speed and accuracy.

Table 4. Performance comparison of different object detection networks training on BUSI and testing on BUS.

	Model	AP	AP50	AP75	AR10	FPS
Anchor-based networks	Faster R-CNN-res50	0.603	0.925	0.676	0.625	7.14
	RetinaNet-res50	0.835	1	0.911	0.849	10.45
	SSD-res34	0.83	0.96	0.889	0.863	14.72
	SSD-vgg16	0.771	1	0.94	0.817	13.63
	YOLOF-res50	0.804	0.943	0.804	0.844	11.79
	YOLOV3-darknet53	0.762	1	0.952	0.787	10.45
	YOLOV3-res34	0.773	0.995	0.924	0.8	12.82
Anchor-free networks	CornerNet-res50	0.514	0.883	0.574	0.587	6.35
	FCOS-res50	0.871	1	0.946	0.894	11.15
	TTFNet	0.349	0.615	0.418	0.427	20.24
	YOLOX-m	0.573	0.888	0.611	0.619	12.25

Note: Bold font indicates the best performance results.

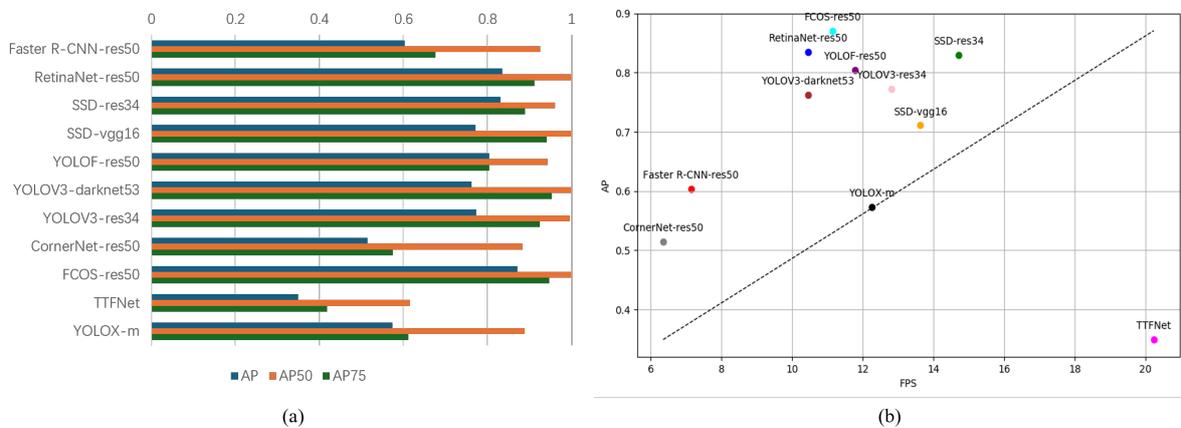


Figure 6. Networks performance results and AP-FPS plot of generalization experiments. (a) is the performance results of different networks, (b) is the FPS and AP scatter plot of the networks.

4. Conclusions

This study comprehensively compares the performance of nine object detection networks in breast lesion detection, encompassing four types: two-stage, one-stage, anchor-based, and anchor-free. This range covers all current types of object detection networks, ensuring a comprehensive and representative evaluation. We validate model performance on two datasets and compare their generalization abilities. The results demonstrate the strengths and limitations of different types of networks in breast lesion detection tasks. In terms of performance on a single dataset, anchor-based networks generally outperform anchor-free networks. Notably, the SSD model, while maintaining a high AP, also exhibits rapid detection speed, proving its practicality and effectiveness in breast cancer detection. This also indicates that anchor-based methods have strong detection capabilities for common lesion types in breast ultrasound images. The superior performance of anchor-based networks can be attributed to their predefined anchor boxes, which provide better assistance in detecting objects of varying sizes and aspect ratios. These anchor boxes serve as priors, helping the network focus on regions of interest, thereby enabling more accurate localization and classification of lesions.

However, in the comparison of generalization performance, the anchor-free network FCOS shows superior performance. This finding highlights the advantage of anchor-free networks in handling lesions with varying shapes and sizes. Since the FCOS network does not rely on preset anchors, it can adapt more flexibly to targets of different sizes, thereby performing better on new or unknown datasets. This is particularly important for breast cancer detection, as lesion shapes and sizes can vary among patients.

Early detection of breast cancer is crucial for improving patient survival rates, and developing accurate and rapid breast cancer auxiliary diagnostic systems is essential. In summary, this research provides valuable insights for the early detection and diagnosis of breast cancer, offering important guidance for the development of efficient and accurate breast cancer auxiliary diagnostic systems in the future.

Author Contributions

Y.W.: methodology, software and writing; Q.Z.: data preprocess; B.Z.: investigation; D.T. and R.Z.: data post-process; W.Z.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Research Data Policies at <https://doi.org/10.1109/JBHI.2017.2731873> and <https://doi.org/10.1016/j.dib.2019.104863>.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J. Clin.* **2018**, *68*, 394–424.
2. Sun, Y.S.; Zhao, Z.; Yang, Z.N.; Xu, F.; Lu, H.J.; Zhu, Z.Y.; Shi, W.; Jiang, J.; Yao, P.P.; Zhu, H.P. Risk factors and preventions of breast cancer. *Int. J. Biol. Sci.* **2017**, *13*, 1387.
3. Elmore, J.G.; Armstrong, K.; Lehman, C.D.; Fletcher, S.W. Screening for breast cancer. *JAMA* **2005**, *293*, 1245–1256.
4. Geisel, J.; Raghu, M.; Hooley, R. *The Role of Ultrasound in Breast Cancer Screening: The Case for and against Ultrasound*; Seminars in Ultrasound, CT and MRI. Elsevier: Amsterdam, The Netherlands, 2018; Volume 39, pp. 25–34.
5. Qian, X.; Pei, J.; Zheng, H.; Xie, X.; Yan, L.; Zhang, H.; Han, C.; Gao, X.; Zhang, H.; Zheng, W.; et al. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. *Nat. Biomed. Eng.* **2021**, *5*, 522–532.
6. Zhu, Z.; Wang, S.H.; Zhang, Y.D. A Survey of Convolutional Neural Network in Breast Cancer. *Comput. Model. Eng. Sci.* **2023**, *136*, 2127–2172.
7. Drukker, K.; Giger, M.L.; Horsch, K.; Kupinski, M.A.; Vyborny, C.J.; Mendelson, E.B. Computerized lesion detection on breast ultrasound. *Med. Phys.* **2002**, *29*, 1438–1446.
8. Liu, B.; Cheng, H.; Huang, J.; Tian, J.; Liu, J.; Tang, X. Automated segmentation of ultrasonic breast lesions using statistical texture classification and active contour based on probability distance. *Ultrasound Med. Biol.* **2009**, *35*, 1309–1324.
9. Shan, J.; Cheng, H.; Wang, Y. Completely automated segmentation approach for breast ultrasound images using multiple-domain features. *Ultrasound Med. Biol.* **2012**, *38*, 262–275.
10. Yap, M.H.; Goyal, M.; Osman, F.; Martí, R.; Denton, E.; Juette, A.; Zwiggelaar, R. Breast ultrasound region of interest detection and lesion localisation. *Artif. Intell. Med.* **2020**, *107*, 101880.
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149.
12. Wang, Y.; Yao, Y. Breast lesion detection using an anchor-free network from ultrasound images with segmentation-based enhancement. *Sci. Rep.* **2022**, *12*, 14720.
13. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9627–9636.
14. Cao, Z.; Duan, L.; Yang, G.; Yue, T.; Chen, Q. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Med. Imaging* **2019**, *19*, 1–9.
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Part I 14, pp. 21–37.
16. Mo, W.; Zhu, Y.; Wang, C. A method for localization and classification of breast ultrasound tumors. In Proceedings of the Advances in Swarm Intelligence: 11th International Conference, ICSI 2020, Belgrade, Serbia, 14–20 July 2020; pp. 564–574.
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Yu, X.; Zhu, Z.; Alon, Y.; Guttery, D.S.; Zhang, Y. GFNet: A Deep Learning Framework for Breast Mass Detection. *Electronics* **2023**, *12*, 1583.

19. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
20. Chen, Q.; Wang, Y.; Yang, T.; Zhang, X.; Cheng, J.; Sun, J. You only look one-level feature. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13039–13048.
21. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
22. Liu, Z.; Zheng, T.; Xu, G.; Yang, Z.; Liu, H.; Cai, D. Training-time-friendly network for real-time object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, New York, USA, 7–12 February 2020; Volume 34, pp. 11685–11692.
23. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
24. Yap, M.H.; Pons, G.; Marti, J.; Ganau, S.; Sentis, M.; Zwigelaar, R.; Davison, A.K.; Marti, R. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1218–1226.
25. Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **2020**, *28*, 104863.
26. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
29. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
30. PaddlePaddle. *PaddleDetection, Object Detection and Instance Segmentation Toolkit Based on PaddlePaddle*; PaddlePaddle: Haidian, Beijing, 2019.
31. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
32. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.