Breast Tumor Detection in Ultrasound Images Using Deep Learning

Zhantao Cao[1](✉), Lixin Duan[1], Guowu Yang[1], Ting Yue[2], Qin Chen[3], Huazhu Fu[4], and Yanwu Xu[5]

[1] The Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, China
caozhantao@163.com

[2] School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

[3] Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China

[4] Agency for Science, Technology and Research, Singapore, Singapore

[5] Guangzhou Shiyuan Electronics Co., Ltd., Guangzhou, China

Abstract. Detecting tumor regions in breast ultrasound images has always been an interesting topic. Due to the complex structure of breasts and the existence of noise in the ultrasound images, traditional handcraft feature based methods usually cannot achieve satisfactory results. With the recent advance of deep learning, the performance of object detection has been boosted to a great extent, especially for general object detection. In this paper, we aim to systematically evaluate the performance of several existing state-of-the-art object detection methods for breast tumor detection. To achieve that, we have collected a new dataset consisting of 579 benign and 464 malignant lesion cases with the corresponding ultrasound images manually annotated by experienced clinicians. Comprehensive experimental results clearly show that the recently proposed convolutional neural network based method, Single Shot MultiBox Detector (SSD), outperforms other methods in terms of both precision and recall.

Keywords: Deep learning · Breast tumor detection

1 Introduction

Breast cancer is the second leading cause of female death. Early diagnosis is the key for breast cancer control, as it can reduce mortality dramatically (40% or more) [1]. Previously, mammography is the main modality for detecting of breast cancer. However, mammography not only causes health risks for patients, but also leads to unnecessary (65–85%) biopsy operation due to low specificity [1]. As a much better option, ultrasound imaging can increase the overall cancer detection by 17% and reduce unnecessary biopsies by 40% [1]. Currently, using

G. Wu et al. (Eds.): Patch-MI 2017, LNCS 10530, pp. 121–128, 2017.
DOI: 10.1007/978-3-319-67434-6_14

ultrasound techniques for tumor detection relies on doctor's experience, especially for the marks and measurements of tumors. Specifically, a doctor usually uses ultrasound instruments for tumor detection by first finding a good angle to wake the tumor clearly shown on the screen, and then keeping probe fixed for a long time using one hand, with another hand to mark and measure the tumor on the screen. It is a difficult task, because the slight shaking of hand holding the probe will cause big impact on the quality of breast ultrasound images; Based on this, computer aided automatic detection technology is highly demanded for locating regions of interest (ROIs), i.e., tumors, in breast ultrasound images.

Several previous methods discussed on how to automatically locate ROIs of breast tumors. In [2], A self-organizing map neural network was used for the detection of the breast tumor. The ROIs can be extracted automatically by employing local textures and a local gray level co-occurrence matrix which is a joint probability density function of two positions. Compared with the basic texture feature, the gray level co-occurrence matrix can reflect the comprehensive information about the direction, the interval and the amplitude of the image. In [3], Shan et al. developed an automatic ROI generation method which consisted of two parts: automatic seed point selection and region growing. However, the method depends on textural features, and these features are not effective for breast ultrasound images when there exists a fat region close to the tumor area or contrast is low [4]. In [5], a supervises learning method was proposed to categorize breast tissues into different classes by using a trained texture classifier, where background knowledge rules were used to select the final ROIs for the tissues. However, due to the inflexibility of the introduced constraints in the proposed method, its robustness was reduced. In [4], the authors improved the method in [5] by proposing a fully automatic and adaptive ROI generation method with flexible constraints. In their work, the ROI seed can be generated with high accuracy, and can also well distinguish the dataset tumor regions from normal regions. However, as shown in the experiments, the recall is still unsatisfactory, that average recall rate was low that benign was 27.69%, malignant was 30.91%, total was 29.29%.

Recently, deep learning techniques have attracted a lot of attention from researchers, because of the good data interpretability as well as the high discriminability power. Noticeably, deep convolutional neural networks (CNNs) have substantially improved the performance not only for image classification, but also for general object detection [6–9]. In order to take advantage of the recent developement of CNNs, in this work we employ the state-of-the-art CNN based detection methods to locate tumor regions in breast ultrasound images, and systematically evaluate them on our newly collected dataset consisting of both benign and malignant breast tumor images. So far in the literature, people have employed CNN based methods to handle detection tasks for other image modalities, such as mammograms [10]. To the best of our knowledge, there is little work that has comprehensively evaluated the performance of different CNN based detection methods for detecting tumors in breast ultrasound images. To this end, in this work we establish benchmarks for our newly collected dataset, and our study can potentially benefit other researchers working in the same area.

2 Related Work

2.1 Traditional Object Detection

The traditional object detection framework normally consists of three parts: (1) feature extraction; (2) proposal regions generation (including sliding window [11], Selective Search [12] and Objectness [13]); (3) proposal classification.

In the past, researchers usually studied hand-crafted features within the traditional detection framework. For example, Dalalet and Triggs [14] used SVM with histogram of oriented gradients (HOG) features for the pedestrian detection task. Felzenszwalb et al. [15] proposed a Deformable Part-based Model (DPM) using latent SVM, which achieved the best performance in the 2006 PASCAL person detection challenge. In [16], the authors used the K-SVD dictionary learning method to obtain a sparse expression of an image, which was called Histograms of Sparse Codes (HSC). HSC was used to replace HOG for classifier training and target detection. Although the performance has been considerably improved, the detection speed is quite slow. In [17], the author proposed an object detector based on co-occurrence features, which was three kinds of local co-occurrence features constructed by the traditional Haar, LBP, and HOG respectively. In addition, the author proposed a generalization and efficiency balanced framework for boosting training, where both high accuracy and good efficiency were achieved. Although the traditional detection method developed for many years, in recent years, it is generally acknowledged that progress has been slow.

2.2 CNN Based Object Detection

The remarkable progress of deep learning techniques, especially CNN, have largely promoted the research of visual object detection. In the following, we briefly review some state-of-the-art CNN based detection methods.

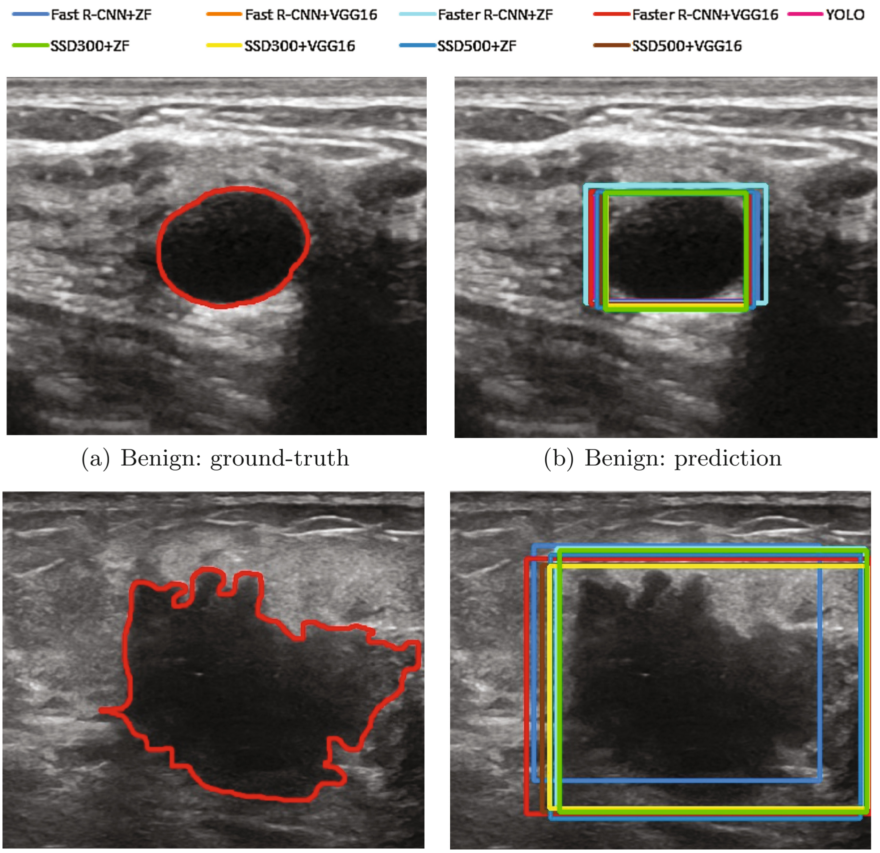
In 2014, Girshick et al. [18] proposed Region-based Convolutional Neural Networks (R-CNN), which combined the heuristic region proposal method and CNN. However, R-CNN has notable drawbacks: (1) the training phase is time-consuming; and (2) the detection phase is slow due to the repetitive feature extraction. In order to improve the speed of R-CNN, He et al. [19] introduced Spatial Pyramid Pooling Net (SPP-Net). Compared to R-CNN, SPP-Net does not require to resize proposed regions to a fixed size. Since the convolution process is done only once by caching the values, SPP-Net largely accelerates the detection time. However, two major issues still exist: (1) the training phase is quite complex; and (2) the fine-tuning stage could not update the convolutional layers, which somehow restricts SPP-Net to achieve better performance. To overcome those drawbacks, also inspired by SPP-net [19], Girshick [6] improved R-CNN by proposing Fast R-CNN which adds a ROI pooling layer to the last convolution layer as well as performs classification and bounding box regression simultaneously. However, as selective search is used for region proposals, the detection time is not very fast. To avoid the standalone step to generate regions, Ren et al. [7] proposed to integrate a so-called Region Proposal Network (RPN) into Fast R-CNN. Since the convolutional features of regions are shared, the region proposal step is almost cost free, making the detection phase of Faster R-CNN almost real-time. But the small scale objects cannot be well detected, due to the loss of detail information in the corresponding deep features.

Recently, researchers also investigated possible ways to avoid proposing regions at the very beginning for detection. For instance, You Only Look Once (YOLO) [8] employed a single convolutional neural network to predict the bounding boxes and class labels of detected regions. Since the YOLO limits the number of bounding boxes, it avoids repetitive detection of the same object and thus greatly improves the detection speed, making YOLO suitable for real-world applications. However, like Faster R-CNN, YOLO also has problems in detecting small scale objects. To deal with the issues as in YOLO, Liu et al. [9] proposed Single Shot MultiBox Detector (SSD) by generating bounding boxes of multiple sizes and aspect ratios from feature maps of different levels. However, these CNN-based methods only focus on general object detection. In this paper, we apply them to detecting tumors in our newly collected breast ultrasound dataset.

3 Dataset

Collecting a well defined dataset for breast ultrasound images is key to the research on breast tumor detection/classification. For that, we have been collaborating with Sichuan Provincial People's Hospital to have experienced clinicians annotate breast ultrasound images obtained from breast lesions patients. Specifically, the patients were told to get scanned by LOGIQ E9 (GE) and IU-Elite (PHILIPS) to generate those ultrasound images. Each ultrasound image was later reviewed and diagnosed by two or three clinicians. Based on the ratings obtained from the BI-RADS system [22], each diagnosed image was then grouped into 7 categories indexed from 0 to 6, where 0 means more information is needed, 1 negative, 2 benign finding, 3 probably benign (less than 2% likelihood of cancer), 4 suspicious abnormality, 5 highly suggestive of malignancy, and 6 proven malignancy. According to [22], some medical specialists proposed to further partition the fourth category (suspicious abnormality) into three sub-category, i.e., 4A (low suspicion for malignancy), 4B (intermediate suspicion of malignancy) and 4C (moderate concern, but not obvious for malignancy). For that, by following the professional instructions from our clinicians, we divide our dataset into two classes: benign and malignant. The benign class is constructed by the images grouped into categories 2, 3 and 4A, while the malignant class consists of the images from categories 4B, 4C, 5 and 6.

By working with the clinicians, we have collected 579 benign and 464 malignant cases from patients. Moreover, the tumor in each image has also been marked out by those experienced clinicians. Figure 1 showcases four ultrasound



(a) Benign: ground-truth

(b) Benign: prediction

(c) Malignant: ground-truth

(d) Malignant: prediction

Fig. 1. Ground-truth annotations and predicted bounding boxes of different methods, for four tumor cases from different patients.

images containing either benign or malignant tumors. To the best of our knowledge, there is no such a publicly available ultrasound image dataset as ours for breast tumors.

4 Experiments

4.1 Experimental Setup

In the experiments, we evaluate the performance of several state-of-the-art detection methods, i.e., Fast R-CNN [6], Faster R-CNN [7], YOLO [8], and SSD [9]. We also combine each CNN based detection method with different existing neural networks, e.g., VGG16 [20], ZFNet [21].

For evaluation metric, we employ average precision rate (APR) and average recall rate (ARR) over all test images [4] as well as the F_1 score for each method:

$$\mathrm{APR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|R_i^{gt} \cap R_i^{pred}\right|}{\left|R_i^{pred}\right|},\ \mathrm{ARR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|R_i^{gt} \cap R_i^{pred}\right|}{\left|R_i^{gt}\right|},\ \mathrm{F}_1 = \frac{2 \times \mathrm{APR} \times \mathrm{ARR}}{\mathrm{APR} + \mathrm{ARR}},$$

where N is the number of images, R_i^{gt} is the grount-truth tumor region, and R_i^{pred} is the predicted bounding box. A higher APR shows the higher overlapped rate between the ROI and the true tumor region, while a higher ARR indicates that ROI generated by the proposed method could be subject to the removal of additional non-tumor regions.

In the experiments, we prepare our data as follows. For the benign class, 285 cases are randomly selected as the training set, 191 cases as the validation set and 103 cases as the test set. For the malignant class, we sample 230 cases as training set, 154 cases as the validation set and 80 cases as test set. In total, we have 515 training cases, 345 validation cases and 183 test cases. It's worth noting that all experimental protocols were approved by Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital.

4.2 Results

In this paper, we compared the results of the different methods (the method in [4], Fast R-CNN, Faster R-CNN, YOLO, SSD) on the locating tumor ROIs in breast ultrasound images. For the deep architecture, we employ a medium-sized network VGG16 [20] and a small network ZFNet [21] for Fast R-CNN, Faster R-CNN and SSD. YOLO uses its original Darknet model [8].

The comparison of these baseline is listed in Table 1, where the APRs, ARRs and F_1 scores of different methods are compared on three settings, i.e., benign

Table 1. Average precision rates (APR), average recall rates (ARR) and F_1 scores of different methods under three settings.

Method	Benign			Malignant			Benign + Malignant		
	APR	ARR	F_1	APR	ARR	F_1	APR	ARR	F_1
Fully auto ROI [4]	66.95	14.16	23.38	78.22	19.23	30.87	71.86	16.36	26.65
Fast R-CNN+ZFNet	87.25	65.47	74.81	89.02	53.54	66.86	91.11	62.60	74.21
Fast R-CNN+VGG16	90.17	66.39	76.47	71.00	40.83	51.84	88.70	61.97	72.96
Faster R-CNN+ZFNet	93.14	66.25	77.43	86.37	46.83	60.73	92.42	62.23	74.38
Faster R-CNN+VGG16	93.01	67.08	77.95	90.36	52.05	66.05	92.37	62.54	74.58
YOLO	95.59	68.85	80.05	96.46	57.73	72.23	96.81	65.83	78.37
SSD300+ZFNet	**97.20**	**70.56**	**81.76**	96.44	54.91	69.97	**96.89**	**67.23**	**79.38**
SSD300+VGG16	96.03	69.76	80.82	**97.56**	**58.96**	**73.50**	96.42	66.70	78.85
SSD500+ZFNet	95.98	70.04	80.98	94.22	54.90	69.38	95.09	65.06	77.26
SSD500+VGG16	94.58	69.57	80.17	94.67	55.82	70.23	96.42	66.70	78.85

images only, malignant images only and both benign + malignant images. We can clearly observe that the CNN based methods perform much better than the method in [4]. Also, SSD300 in general achieves good results than other CNN based methods, which shows SSD300 is more suitable for the tumor detection task in this work. It is worth noting that SSD300 is better than SSD500 in all three settings by using either ZFNet or VGG16. The reason is as follows. SSD300 resizes images into 300×300, while SSD500 makes the size as 500×500. The region candidates in SSD300 cover a relatively larger area than those in SSD500. Since the tumor region takes a good portion in an image, SSD300 is able to better capture the region, which thus leads to better performance. Furthermore, SSD300+ZFNet is better than SSD300+VGG16 under the benign setting, but worse under the malignant setting. This interesting observation can be explained based on the model complexity of ZFNet and VGG16. Specifically, although ZFNet is a small neural network, it can well handle the easier case (i.e., benign), but is a bit underfitting for the harder case (i.e., malignant). In contrast, the larger VGG16 model is good at dealing with malignant tumors, while getting overfitting for the benign ones.

We also plot the resultant bounding boxes predicted by different methods for four tumor cases in Fig. 1.

5 Conclusion and Future Work

In this paper, we have mainly studied the existing state-of-the-art CNN based methods for tumor detection in breast ultrasound images. Due to the lack of publicly available dataset, we have collected a new one consisting of both benign and malignant cases, which are carefully annotated by experienced clinicians. Through comprehensive experiments, we find that SSD300 achieves the best performance in terms of APR, ARR and F_1 score.

Currently in our work, we only detected the tumor regions by using bounding boxes. In the future, we will conduct further investigation on the automatic segmentation of tumor areas.

Acknowledgement. This work is supported by grants from the National Natural Science Foundation of China (61572109) and the Fundamental Research Funds for the Central Universities (ZYGX2016J164).

References

1. Cheng, H.D., Shan, J., Ju, W., Guo, Y.H., Zhang, L.: Automated breast cancer detection and classification using ultrasound images: a survey. Pattern Recogn. **43**, 299–317 (2010)
2. Su, Y., Wang, Y.: Automatic detection of the region of interest from breast tumor ultrasound image. Chin. J. Biomed. Eng. **29**(2), 178–184 (2010)
3. Shan, J., Cheng, H.D., Wang, X.Y.: Completely automated segmentation approach for breast ultrasound images using multiple-domain features. Ultrasound Med. Biol. **38**(2), 262–275 (2012)

4. Xian, M., Zhang, Y.T., Cheng, H.D.: Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. Pattern Recogn. **48**(2), 485–497 (2015)
5. Liu, B., Cheng, H.D., Huang, J.H., Tian, J.W., Tang, X.L., Liu, J.F.: Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recogn. **43**(1), 280–298 (2010)
6. Girshick, R.: Fast R-CNN. In: ICCV, pp. 1440–1448 (2015)
7. Ren, S.Q., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS (2015)
8. Redmon, J., Divvala, S.K., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: CVPR, pp. 779–788 (2015)
9. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). doi:10.1007/978-3-319-46448-0_2
10. Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., Barkan, E.: A region based convolutional network for tumor detection and classification in breast mammography. In: Carneiro, G., et al. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 197–205. Springer, Cham (2016). doi:10.1007/978-3-319-46976-8_21
11. Viola, P., Jones, M.: Robust real-time face detection. In: IJCV (2004)
12. Sande, K., Uijlings, J., Gevers, T., Smeulders, A.: Segmentation as selective search for object recognition. In: ICCV (2011)
13. Alexe, B., Deselaers, T., Ferrari, V.: What is an object? In: CVPR (2010)
14. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
15. Felzenszwalb, P., McAllester, D., Ramaman, D.: A discriminatively trained and multiscale: deformable part model. In: CVPR, pp. 1–8 (2008)
16. Ren, X.F., Ramanan, D.: Histograms of sparse codes for object detection. In: CVPR, pp. 3246–3253 (2013)
17. Ren, H.Y., Li, Z.N.: Object detection using generalization and efficiency balanced co-occurrence features. In: ICCV, pp. 46–54 (2015)
18. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR, pp. 580–587 (2014)
19. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 346–361. Springer, Cham (2014). doi:10.1007/978-3-319-10578-9_23
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2014)
21. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). doi:10.1007/978-3-319-10590-1_53
22. BI-RADS. https://en.wikipedia.org/wiki/BI-RADS