Deep Learning Based Lesion Detection For Mammograms

Zhenjie Cao[1], Zhicheng Yang[1,2†], Xinya Liu[1,3†], Yanbo Zhang[1*], Shibin Wu[4],
Ruei-Sung Lin[1], Lingyun Huang[4], Mei Han[1], Jie Ma[5*]
[1]Ping An Institute of Intelligence, Ping An Technology, Palo Alto, CA, USA
[2]Department of Computer Science, University of California, Davis, CA, USA
[3]Department of Biostatistics, University of California, Los Angeles, CA, USA
[4]Ping An Technology, Shenzhen, Guangdong, China
[5]Shenzhen People's Hospital, Shenzhen, Guangdong, China

***Abstract*—Mammogram plays an essential role in reducing breast cancer deaths by detecting cancer early. In this study, Convolutional Neural Networks (CNN) based Faster R-CNN model was applied to detect mass and calcification in breast cancer effectively. Mammography dataset contains public dataset including DDSM, INbreast, and BCD, as well as private dataset, which was obtained from Shenzhen People's Hospital, China. Final detection result for public dataset was 0.804 in Average Precision (AP), and 0.975 in recall for mass detection, and 0.686 AP and 0.925 recall in calcification detection. Result for private dataset was higher with 0.902 AP and 0.978 recall for mass detection, and 0.605 AP and 0.834 recall for calcification detection.**

***Index Terms*—Mammogram; Deep learning; Object detection**

I. Introduction

Medical imaging has been a revolutionary way for medical professionals to diagnose and treat medical diseases over the past decades. However, interpreting medical imaging requires the well-trained professional domain knowledge accumulated during the long-time clinical experience of a doctor. With the rapid development of Artificial Intelligence with the emphasis of *deep learning* methods, numerous Convolutional Neural Networks (CNN) based approaches are proposed. It effectively assists doctors (or even automatically) to locate lesions and determine if they are benign, or malignant (commonly known as *cancer*) in medical images.

Among these applications on medical images, detecting lesions in mammograms is gaining the increasing attention during recent years. Mammography is an effective and efficient way to detect breast lesions. Breast lesions in mammograms are typically categorized into four types: mass, calcifications, asymmetry, and architectural distortion. In particular, mass and calcification are two most commonly concerned detection targets among four types of breast lesions in mammograms. Mass is the hard but painless breast areas with irregular edges; Calcification refers to small calcium deposits in breast tissues.

Past studies proposed traditional machine learning methods in K-means and support vector machine (SVM) classification [1]. Abstract patterns were extracted from malignant and benign tumors, and used in prediction model training. A novel modified Faster R-CNN model [2] was introduced in [3] for automatic detection of abnormalities in medical imaging.

In this paper, we present our preliminary results of the lesion detection in mammograms using the improved Faster R-CNN. We investigate the detection performance on various public datasets and our private dataset. The discussion and our ongoing work are presented at last.

II. Methods and Materials

A. Methods

Faster R-CNN is a two-stage deep learning-based object detection model. In the Faster R-CNN, a backbone network is adopted to generate feature maps. Based on feature maps, region candidates are firstly by a region proposal networks (RPN), then a CNN-based network is used to classify the object class and detect the bounding box. In our work, we improved the Faster R-CNN to get better results for lesion detection in mammograms.On the one side, to adapt the various sizes of lesions, we applied the feature pyramid network (FPN) upon the backbone to produce multi-scale features. On the other side, the original Faster R-CNN usually obtains many false positives for mass detection. Hence, we replaced the original loss function in Faster R-CNN with the state-of-the-art focal loss, which applies a weighting strategy and focuses on hard examples [4].

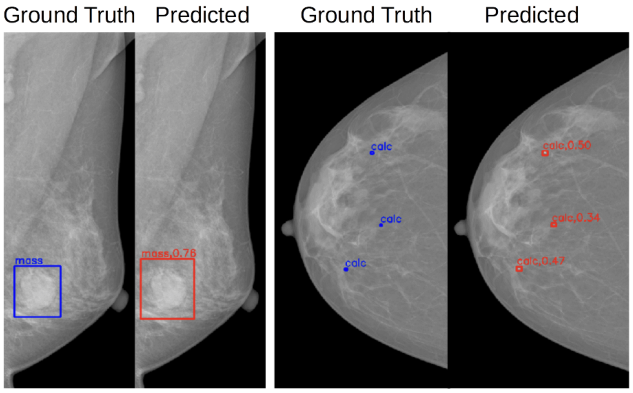
We trained models for mass and calcification detection, respectively. The backbone of our Faster R-CNN was ResNet50, where the hyper parameters were loaded from the pre-trained model on ImageNet. Each original training image was down-sampled to a small size to ensure that the short edge was 1200 pixels. The batch size was 4, and the Adam optimization method was used to train the model. There were 500 steps in each epoch, the training stopped after 200 epoch.

B. Datasets

Our datasets comprise two parts, the public datasets and our private dataset. The public datasets consists of the Digital Database for Screening Mammography (DDSM) [5], INbreast [6], and Breast Cancer Digital repository (BCD) [7]. DDSM

* Corresponding authors: Jie Ma (email: majie688@hotmail.com) and Yanbo Zhang (email: yanbozhang007@gmail.com).

† This work was done during Zhicheng Yang and Xinya Liu's internships at Ping An Institute of Intelligence, Palo Alto, CA, USA



(a) Mass (b) Calcifications

Fig. 1: Examples of detected (a) mass and (b) calcification in mammograms. Bounding boxes in ground truth are labeled in blue, and the predicted lesions are highlighted in red bounding boxes, respectively.

has ∼1,600 images for mass and ∼1,900 images for calcification. BCD dataset consists of 535 mass images. INbreast has a total of 115 patient cases, of which 90 cases are from women with both breasts and 25 cases are from mastectomy cases. These datasets were combined as one public dataset.

Our private dataset collected from Shenzhen People's Hospital, China, contains a total of 1,432 images for 567 patients, out of which 627 images have mass and 1,001 images have calcification. The private dataset provides the most comprehensive images with fine quality, and images are also carefully and precisely labeled and checked by multiple doctors.

In our experiment, for public and private datasets, data were randomly split as training and testing data, respectively. About 80% data were used for training, and the rest were used as test data.

III. Preliminary Results

Results from public dataset, including DDSM, INbreast, and BCD was summarized in Table I. Average precision (AP) is commonly used evaluation metric in object detection models. It is calculated as mean precision over several recall levels. When the short edge of an image is resized to 1200 pixels, result shows that mass detection performs the best as 0.804 in AP and 0.975 in recall. However in calcification detection, the best AP is 0.686 and recall is 0.925. We assume that due to the fact that calcification tends to cluster together as small and densely distributed dots, as well as multiple pooling layers in Faster R-CNN would dilute image features for neural network to learn, calcification detection results were much lower than those from mass detection.

Regarding results for private dataset (Table II), the precision and recall for resized images were higher than those tested on public dataset (0.902 AP and 0.978 recall for mass). This attributes to the better image quality of the private dataset. Currently, we are still accepting new images from China, and the performance has the potential to be improved. Some

TABLE I: Prediction results using Faster R-CNN to train and test on public datasets (DDSM + INbreast + BCD).

Type	AP	Recall
Mass	0.804	0.975
Calcification	0.686	0.925

TABLE II: Prediction results using Faster R-CNN to train and test on our datasets (Shenzhen People's Hospital, China).

Type	AP	Recall
Mass	0.902	0.978
Calcification	0.605	0.834

representative examples of detected mass and calcifications are shown in Fig. 1.

IV. Discussion and Ongoing Work

Studies have shown that breast density has been a risk of breast cancer in woman is in relation to the higher breast density. Relationship of race to mammographic density may explain different breast cancer risks among race groups [8]. The public dataset and private datasets were collected from people of different races. This is one major difference between them.

Data annotation consistency is another part to consider in the future. For example, inconsistency was found in our private dataset. Since multiple doctors participated in the process to label images with calcifications, they might treat those microcalcifications gather together differently. Some of them tended to label them together as a flat sheet, the others might label them as small dots respectively. However, label inconsistency may influence the accuracy and sensitivity of our detection model.

Asymmetry and architectural distortion are two other major lesion types in breast cancer, and will be focused later when we obtained more labeled data from Shenzhen People's Hospital in the future.

References

[1] S. S. Bichen Zheng, Sang Won Yoon, "Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithms," in *Expert Systems with Applications*, 2014, pp. 1476–1482.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[3] H. B. A. B. H. Y. S. E. B. Ayelet Akselrod-Ballin, Leonid. KarlinskyAlon, "Deep learning for automatic detection of abnormal findings in breast mammography," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision*, 2017, pp. 321–329.

[4] R. G. K. H. P. D. Tsung-Yi Lin, Priya Goyal, "Focal loss for dense object detection," in *The IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[5] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, p. 170177, 2017.

[6] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[7] M. G. Lopez, N. G. Posada, D. C. Moura, R. R. Pollán, J. M. F. Valiente, C. S. Ortega, M. Solar, G. Diaz-Herrero, I. Ramos, J. Loureiro *et al.*, "Bcdr: a breast cancer digital repository," in *15th International conference on experimental mechanics*, 2012.

[8] W. J. G. L. B. H. M. M. C. P. G. U. Zhengjia Chen, Anna H. Wu, "Does mammographic density reflect ethnic differences in breast cancer incidence rates?" in *American Journal of Epidemiology*, 2004, p. 140–147.