A Region Based Convolutional Network for Tumor Detection and Classification in Breast Mammography

Ayelet Akselrod-Ballin(✉), Leonid Karlinsky, Sharon Alpert, Sharbell Hasoul, Rami Ben-Ari, and Ella Barkan

IBM Research - Haifa, Haifa, Israel
ayeletb@il.ibm.com

Abstract. This paper addresses the problem of detection and classification of tumors in breast mammograms. We introduce a novel system that integrates several modules including a breast segmentation module and a fibroglandular tissue segmentation module into a modified cascaded region-based convolutional network. The method is evaluated on a large multi-center clinical dataset and compared to ground truth annotated by expert radiologists. Preliminary experimental results show the high accuracy and efficiency obtained by the suggested network structure. As the volume and complexity of data in healthcare continues to accelerate generalizing such an approach may have a profound impact on patient care in many applications.

1 Introduction

Breast cancer is the second leading cause of death for women [1]. Despite the advances in imaging technology, Mammography (MG), X-ray imaging of the breast, remains the primary modality for screening and diagnosis of breast cancer. In current practice, screening for early detection is performed for asymptomatic individuals, where an expert radiologist examines the images and performs detection and classification of potential abnormalities. Nevertheless, MG analysis is challenging, due to the subtle fine-grained (FG) visual categories and large variability of appearance in abnormalities (e.g. different sizes, shapes, boundaries, and intensities) [2], making abnormalities difficult to detect and classify, even by an expert radiologist. The problem is further complicated by the non-rigidity of the breast, and the varying viewing conditions, leading to significant intra-expert and inter-expert variability. Still, due to the accelerated advances in technology, the information overload, the limited amount of expert time, and diagnosis errors, it is essential to augment the radiologist with decision supporting computational tools utilizing image processing and machine learning (ML) technology.

The objective of this paper is to introduce a novel algorithm for detection and classification of abnormalities based on a powerful region-based convolutional networks approach. Classification is performed according to the Breast Imaging-Reporting

A. Akselrod-Ballin and L. Karlinsky—contributed equally to this work.

G. Carneiro et al. (Eds.): LABELS 2016/DLMIA 2016, LNCS 10008, pp. 197–205, 2016.
DOI: 10.1007/978-3-319-46976-8_21

and Data System (BI-RADS) score [3]. The BI-RADS score ranges from 0 to 6 and is defined as (0 more information is needed,1 negative, 2 benign finding, 3 probably benign <2 % likelihood of cancer, 4 suspicious abnormality, 5 highly suggestive of malignancy, and 6 proven malignancy). This study dealt only with scores 1 to 5 based on radiological features and without correlation with clinical information. We demonstrate our results on tumor detection and classification where due to the small amount of training data in some of the score classes, we include the three major clinical classes of normal {1}, benign {2} and malignant {3, 4, 5}. Figure 1, demonstrates examples of tumors with different BI-RADS score.

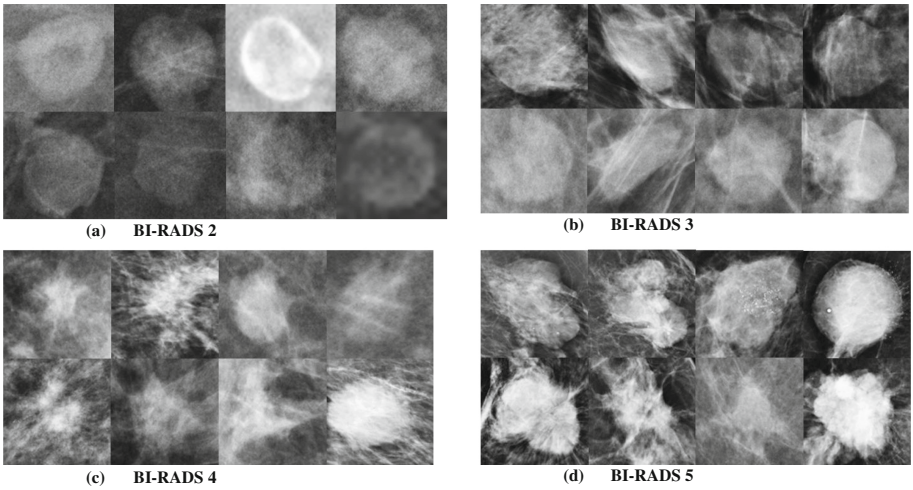

Fig. 1. Examples of tumors with different BI-RADS score.

Deep Neural Networks (DNN) have shown outstanding performance in image recognition tasks such as classification and detection and have been applied successfully in many fields [4]. Recently deep learning methods have also been utilized in medical imaging tasks where an additional difficulty stems from the limited number large labeled datasets, which are required for efficient training and avoiding parameter over fitting. Nevertheless, several deep learning algorithms for breast mammography lesion and calcification classification have been reported, for detailed report see [5, 6].

Prominent work in the field, include [7], where a multi-view convolutional neural network (CNN) based approach was developed. First a separate CNN model was trained for each unregistered view and each segmentation map fine tuning from an Imagenet pre-trained model. Then, using the features obtained the authors trained another CNN classifier that estimated the BI-RADS score. Arevalo et al. [8] used convolutional architectures within a supervised learning framework taking advantage of expert knowledge represented by previously manually segmented lesions by radiologists in both mammographic views. Most of the previous work has focused on binary classification of micro-calcification (MC) or tumors which is different than performing detection localization and classification of abnormalities. The study of [9] is

closely related to our Region-Based CNN (R-CNN) work. The authors present a system consisting of four modules, the first is a multi-scale deep belief nets (DBN) with a Gaussian mixture model (GMM) classifier for region proposal generation, the second is a cascade of R-CNNs, the thirds is a cascade of two random forests and finally, a post processing step merges regions with a high overlap ratio.

A variety of methods have been reported in the computer vision literature for region proposals. Classically, R-CNN [10], solves the detection problem by two modules, the first generates generic region proposals (image regions likely to contain objects) using low level cues such as color and texture, while the second computes CNN features and classifies them into object categories. Fast R-CNN [11] modified some of the drawbacks of R-CNN, by replacing the extraction of region proposal images by direct pooling of region features within a single CNN operating on the entire image and also introducing a regression stage for region box refinement. Recently, Faster R-CNN [12], introduced end-to-end joint training of both the region proposal module and the region classification module, both represented by CNNs with significant weight sharing. This also led to much higher detection speeds and higher detection quality than the original R-CNN [10]. The advantage of joint end-to-end training is that it allows avoiding sequentially-trained tasks and also that it may improve results since the shared representation allows the tasks to influence each other.

Our contribution is twofold: First the best of our knowledge this is the first work to utilizing state-of-the-art Faster-R-CNN in a full efficient pipeline for tumor detection and classification tasks in mammography images. Second we modify the original CNN architecture used in Faster-R-CNN in order to adapt the deep learning process to the specific problem domain, training and testing on a large multi-center dataset for both detection and classification tasks.

2 Methods

The system takes as input a Mammography (MG) image of $\sim$4k $\times$ 3k pixel and begins by segmenting the breast tissue automatically, removing the background and pectoral muscle and cropping image accordingly. In the second step a priori anatomical image is generated detecting the Fibro-glandular tissue based on a fuzzy logic approach [13]. The third step divides the images based on a grid representation to multiple overlapping sub images (parts) which are then used to train and test a modified Faster-RCNN [12]. The fourth step integrates the results obtained from the parts onto the entire image. Finally the output produces the detection and classification results represented by bounding-boxes with a confidence/probability score where pixels that were not detected are defined as normal breast tissue class. Figure 2 shows the outline of the system.

2.1 Breast Anatomical Segmentation and Candidate Extraction

To obtain accurate detection of tumors it is essential to perform segmentation of the breast and the pectoral muscle. Additionally, we generate an anatomical prior identifying the fibroglandular tissue to be used for reduction of FP. The fibroglandular

segmentation is based on the approach described in [13]. Due to the lack of space we refer the reader to the paper for details.

As a result of GPU memory limitations a grid was defined on the image, dividing the image to overlapping parts of fixed size 800 × 800. The use of overlapping parts can also be seen as a data augmentation technique as each lesion appears in several overlapping parts. This allows us to create several samples from each lesion appearing in the original data.

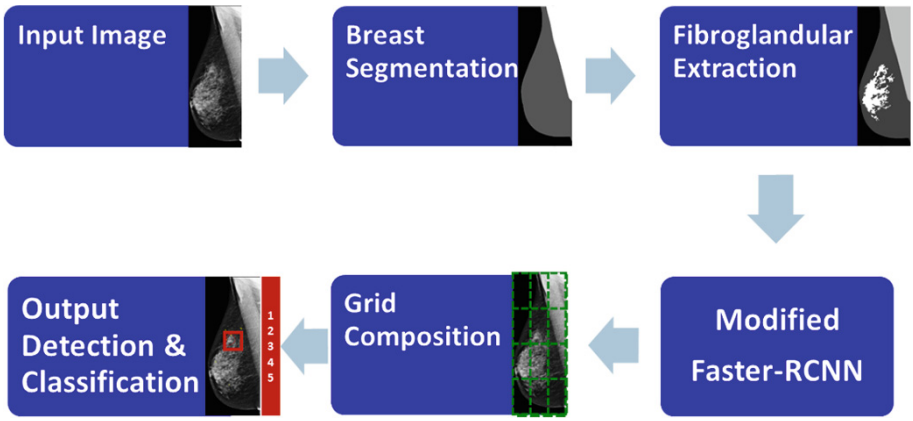

Fig. 2. (a) System outline. Given an MG Input (1) Breast tissue segmentation removal of background and pectoral muscle (2) Fibro-glandular tissue extraction. (3) The image is divided to multiple overlapping subparts utilized in the Faster-RCNN training and testing. (4) Composition: the bounding boxes detected are projected on image grid. The process ends with tumor detection and the corresponding BI-RADS classification.

2.2 Modified Faster-R-CNN

The system is composed of two main components: (1) a region proposal network (RPN), a deep fully convolutional network that is trained to detect windows on the input image that are likely to contain objects of interest regardless of their class. The RPN simultaneously predicts objects bounds and objectness scores at each position on a wide range of scales and aspect ratios, following which top scoring 500 predictions are kept. (2) Fast R-CNN detection network that is trained to classify candidate object windows returned by the RPN, each window is classified into one of the classes of interest or rejected as a false alarm. Both RPN and Fast-RCNN share a large common set of bottom layers allowing computation re-use and hence a significant speed-up at test time.

Faster-R-CNN training is a four-step alternating algorithm consisting of the following steps: (1) Train RPN, initialized with an ImageNet pre-trained model. (2) Train a separate classification network using the proposals generated in step 1. As proposed in [11], the network operates on the entire image and uses pooling from the proposal regions in order to generate deep features for classification. (3) Use the network from stage 2 to initialize the RPN but freeze the parameters of the shared bottom layers and

only fine-tune the top RPN layers (non-shared). (4) Keep the shared layers frozen and repeat stage 2 using the proposals generated in step 3. We have modified the original Faster-RCNN net of [12] to include information from the finer bottom levels during classification (Fast-RCNN) stage. This allows to consider low level (higher resolution) information such as color and texture during the classification stage and improves the classification results. When considering visually similar objects such as malignant lesions and patches of suspicious normal tissue, intricate visual details need to be taken into account in order to make the distinction. In other words, the problem addressed in this paper is inherently fine grained and significantly different than the one of classifying visually dis-similar objects as in many visual challenges such as PASCLA-VOC used for testing in [12]. Therefore, features from lower levels of the CNN need to be taken into the account when making the decision as they are the only ones looking on the considered region proposals in the high enough resolution. Figure 3, provides a schematic view of the modification in the network.

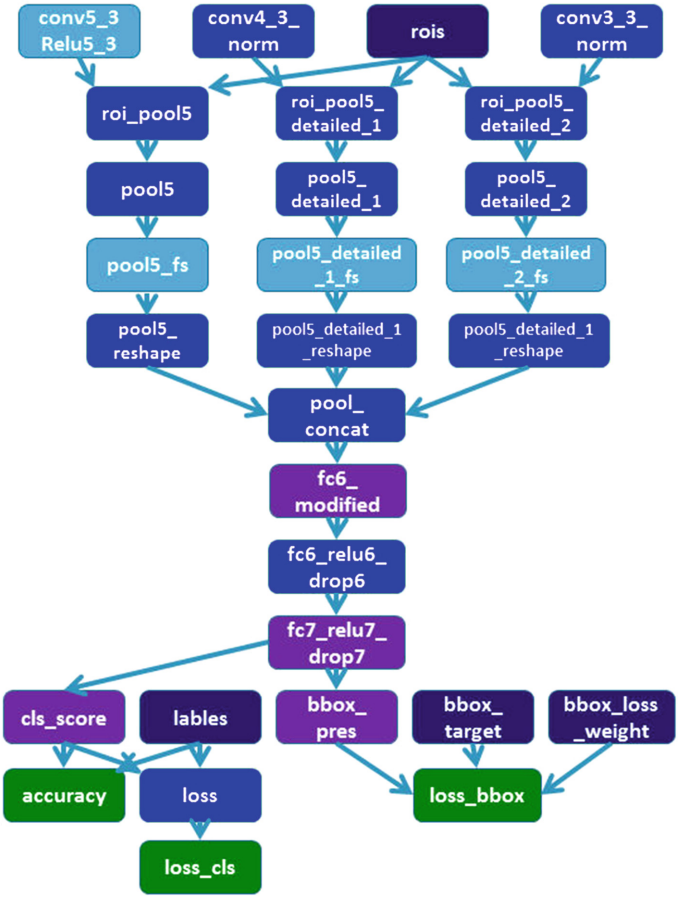

Fig. 3. Schematic zoomed view of the modified layers in the Faster R-CNN architecture. The boxes connected to conv4_3 and conv3_3 represent additional layers in the modified network.

Given a training set of images and RoIs $\{I_i, r_{1i}, c_{1i}, r_{2i}, c_{2i}\}$, where each RoI is defined by a four-tuple that specifies its top-left corner (r; c) and bottom right corner and the class and a test set of images $\{I_j\}$. The output includes all the objects in the test set and their bounding boxes represented in a six-tuple ($\{I_j, r_{1j}, c_{1j}, r_{2j}, c_{2j}\}$, class, and score). Namely, given a test input the system will automatically detect and classify into one of three classes including normal tissue (BI-RADS 1), benign (BI-RADS 2), and malignant tumors (BI-RADS 3, 4, 5).

Our implementation is based on the Faster R-CNN architecture [12] using VGG-16 architecture [14] and caffe implementation [15]. The system was trained on a single TitanX GPU with 12 GB on chip memory, and i7 Intel CPU with 64 GB RAM. Training times required $\sim$36 h, while testing takes 0.2 s per image. During training 2000 top-scoring boxes are sampled from the RPN, during testing top scoring 500 boxes are sampled using standard non-maximal suppression (NMS) based on box overlap. We used the SGD solver with learning rate of 0.001, batch size 2, momentum 0.99, and 60 epochs.

3 Experiments and Results

The experiments were conducted on a large multicenter hospital data set annotated and examined by expert radiologists (the fourth author). The data sets consists of approximately 850 images distributed in terms of BI-RADS to 400, 200, 150, 100 corresponding to classes 2, 3, 4, 5 respectively.

The dataset was split into training (80 %), and testing (20 %) sets following a stratified sampling per patient. Such that a particular patient belongs to only one of the subsets. The total number of patients in experiment (1) was $\sim$300, where training and testing of the model were performed with 4000 and 750 image parts respectively. In experiment (2), we excluded from both sets images corresponding to BI-RADS {3}, to demonstrate the recognition difficulties which the intermediate class brings. The number of patients in experiment (2) was $\sim$220, where training and testing were performed with 3000 and 700 image parts respectively.

Results in the field are commonly evaluated on the DDSM-BCRP [16] and INBreast [17] datasets. The INBreast comprises a set of 115 cases containing 410 images, where only 116 images contain benign or malignant masses. The DDSM-BCRP dataset contains 79 images of malignant masses, where commonly 39 cases are used for training and 40 cases for testing. State of the art results in this domain were reported in [7] which produced an Area under curve (AUC) for the receiver operating characteristics (ROC) of 0.91 $\pm$ .05 on INBreast and 0.97 $\pm$ .0.03 on DDSM. [8] obtained an AUC score of 0.826. Finally [9] present a table comparing previous results showing that their approach produces the best results to date in both datasets: with TPR 0.96 $\pm$ .0.03 @ 1.2 FPI (false positive per image) and TPR = 0.87 $\pm$ .0.14 @ 0.8 FPI for INBreast; and TPR = 0.75 @ 4.8 FPI and TPR = 0.70 @ 4 FPI for DDSM-BCRP. They also obtained the best results reported in respect to running time of 20 s. Table 1 and Figs. 4 and 5. It is difficult to compare to previously published results in the field as most of the previous work focused on binary classification of masses. Also previous work mainly report on DDSM

and INBreast which are limited in size. Our method is evaluated on a large multi-center data set, while most methods report results on smaller data sets of tumors. Also we evaluate both the detection, localization and classification of tumors and report the AUC of precision recall, commonly used in computer vision literature [4, 10]. Nevertheless, examining only classification results, the methodology can be considered as promising in terms of accuracy and extremely efficient in terms of time (Table 1) and Figs. 4 and 5.

Table 1. Results of two experiments for Detection and Classification of tumors.

Experiment	Detection AUC (average precision)	Classification accuracy	#images	Running time
(1) benign BI-RADS- 2 & malignant {345}	0.6	0.78	850	0.2 s
(2) benign BI-RADS- 2 & malignant {45}	0.72	0.77	850	0.2 s

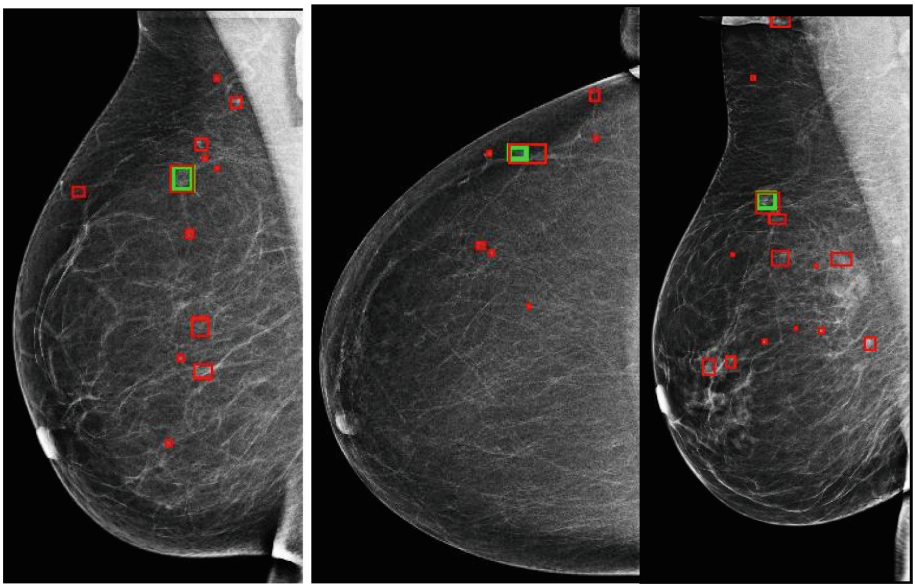

Fig. 4. Illustration of bounding box composition on full breast images. Ground truth annotation is highlighted in green while automatic detection can be viewed in red. (Color figure online)

4 Summary

In this paper we introduced a novel approach for detection and classification of breast tumors based on the state-of-the-art approach of Faster-RCNN. Our preliminary results show the promise of this approach to efficiently and accurately detect and classify

(a) Benign - correctly classified

(b)Malignant - wrongfully classified as benign

(c) Malignant - correctly classified

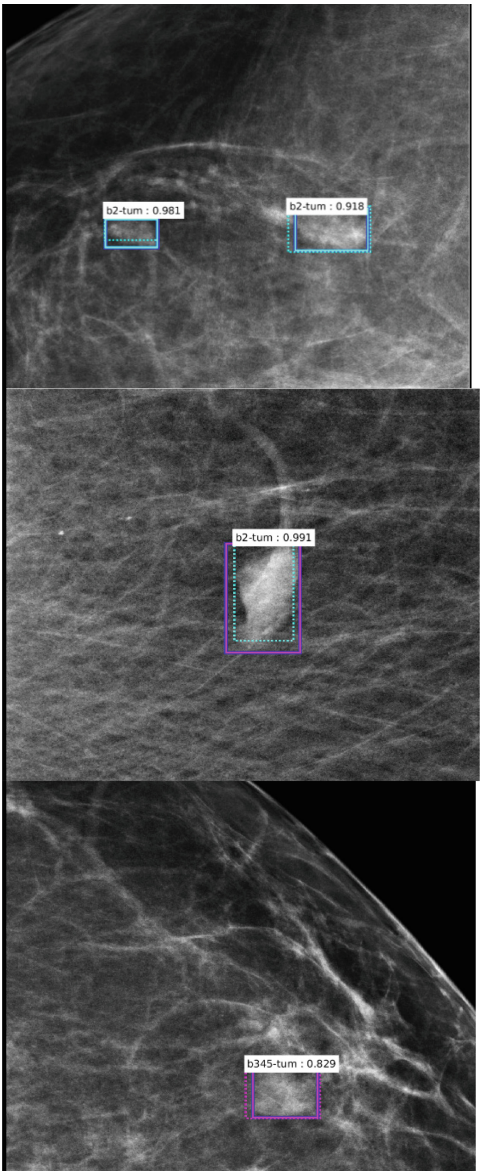

Fig. 5. Detection and Classification results on parts. We show typical detection and classification results, where benign and malignant masses are in cyan and magenta respectively. Ground truth annotation are marked by a solid line and automatic detection are marked in dashed line. (Color figure online)

breast abnormalities on real up-to-date data. Future work will extend this work to a multi-view approach, take into account associated features, such as breast density, perform segmentation in boxes, and evaluate on publically available dataset.

References

1. Jemal, A., Siegel, R., et al.: Cancer statistics, 2008. CA Cancer J. Clin. **58**(2), 71–96 (2008)
2. American Cancer Society: Cancer facts and figures, Atlanta, GA (2015)
3. Narváez, F., Díaz, G., Romero, E.: Automatic BI-RADS description of mammographic masses. In: Martí, J., Oliver, A., Freixenet, J., Martí, R. (eds.) IWDM 2010. LNCS, vol. 6136, pp. 673–681. Springer, Heidelberg (2010)
4. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the (NIPS 2012), pp. 1097–1105 (2012)
5. Giger, M.L., Karssemeijer, N., Schnabel, J.A.: Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. Annu. Rev. Biomed. Eng. **15**, 327–357 (2013)
6. Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., et al.: A review of automatic mass detection and segmentation in mammographic images. Med. Image Anal. **14**, 87–110 (2010)
7. Carneiro, G., Nascimento, J., Bradley, A.P.: Unregistered multiview mammogram anaysis with pre-trained deep learning models. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A. F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 652–660. Springer, Heidelberg (2015)
8. Arevalo, J., Gonzales, F.A., Ramos-Pollán, R., Oliveira, J.L., Lopez, M.A.G.: Representation learning for mammography mass lesion classification with convolutional neural networks. Comput. Methods Programs Biomed. **127**, 248–257 (2016)
9. Dhungel, N., Carneiro, G., Bradley, A.: Automated mass detection from mammograms using deep learning and random forest. In: DICTA (2015)
10. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Computer Vision and Pattern Recognition (CVPR), pp. 580–587 (2014)
11. Girshickv, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448 (2015)
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS (2015)
13. Ben-Ari, R., Zlotnick, A., Hashoul, S.: A weakly labeled approach for breast tissue segmentation and breast density estimation in digital mammography. In: ISBI (2016)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
15. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding (2014). arXiv:1408.5093
16. Heath, M., Bowyer, K., et al.: The digital database for screening mammography. In: Proceedings of the 5th International Workshop on Digital Mammography, pp. 212–218 (2000)
17. Moreira, I.C., Amaral, I., et al.: Inbreast: toward a full-field digital mammographic database. Acad. Radiol. **19**(2), 236–248 (2012)