Improving Breast Cancer Detection using Symmetry Information with Deep Learning

Yeman Brhane Hagos[1,3], Albert Gubern Mérida[1], and Jonas Teuwen[1,2]

[1] Radboud University Medical Center, Department of Radiology and Nuclear Medicine, Nijmegen, the Netherlands
[2] Delft University of Technology, the Netherlands
[3] University of Burgundy (France), University of Cassino and Southern Lazio (Italy) and University of Girona (Spain)

Abstract. Convolutional Neural Networks (CNN) have had a huge success in many areas of computer vision and medical image analysis. However, there is still an immense potential for performance improvement in mammogram breast cancer detection Computer-Aided Detection (CAD) systems by integrating all the information that radiologist utilizes, such as symmetry and temporal data. In this work, we proposed a patch based multi-input CNN that learns symmetrical difference to detect breast masses. The network was trained on a large-scale dataset of 28294 mammogram images. The performance was compared to a baseline architecture without symmetry context using Area Under the ROC Curve (AUC) and Competition Performance Metric (CPM). At candidate level, AUC value of 0.933 with 95% confidence interval of [0.920 , 0.954] was obtained when symmetry information is incorporated in comparison with baseline architecture which yielded AUC value of 0.929 with [0.919 , 0.947] confidence interval. By incorporating symmetrical information, although there was no a significant candidate level performance again ($p = 0.111$), we have found a compelling result at exam level with CPM value of 0.733 ($p = 0.001$). We believe that including temporal data, and adding benign class to the dataset could improve the detection performance.

Keywords: Breast cancer · Digital mammography · Convolutional neural networks · Symmetry · Deep learning · Mass detection

1 Introduction

Breast cancer is the second most common cause of cancer death in women after lung cancer in the United States, which covers around 30% of cancers diagnosed and the chance of women dying from breast cancer is 2.6% [1]. Mammography is the main imaging modality used to detect breast abnormalities at an early stage. Breast masses are most dense and appear in grey to white pixel intensity with oval or irregular shape [4]. Normally, irregular shaped masses are suspicious [4], [5]. Breast cancer screening has shown a reduction in mortality rate of between 40% and 45% for women who were undergoing mammogram screening regularly [6]. However, mammogram screening has drawbacks due to False Positive (FP)

recalls, such as FP biopsy and cost associated with the unnecessary follow up [8]. Therefore, it is necessary to increase sensitivity for early stage detection and increase specificity to reduce FP detection.

Nowadays, with a massive amount of data and computational power, Deep Learning (DL) has shown a remarkable success in natural language processing [9] and object detection and recognition [10]. This has opened an interest in applying DL in medical image processing and analysis. However, care should be taken as the way we as humans interpret natural images and medical images are different in some cases. Eventually, the performance of DL method will be compared with the radiologist and thus, the CNN should preferably be given all the information that radiologist utilize. For instance, during the reading of screening mammograms, radiologists use priors, multiple views and look for asymmetries between the two breasts.

DL has been explored for Digital Mammogram (DM) image analysis. Some of them work directly on the whole image [8], [5], and others focused on patch based [11]. [8] proposed a multi-view single stage CNN breast mammogram classification that works at original resolution. To address memory issue, aggressive convolution and pooling layers with stride greater than one were proposed. It is stated also that this approach suffers from loss of spatial information. In the work by [11], incorporating symmetry and temporal context improves detection of malignant soft tissue lesion, in which random forest classifier was used for mass detection and CNN for classification.

In this study, we conducted an investigation to analyze the performance gain of integrating symmetry information into a CNN to detect malignant lesions on a large scale mammography database. First, a database of 7196 exams which contains 28294 images was collected from different sites in the Netherlands. Previous work by [12] was employed to detect suspicious candidates locations. Then, patches centered on the points were extracted to train a two input CNN to reduce FP candidates. Left and right breast images were considered as contra-later images to each other, and a patch in a primary image and an exact reflection or mirror on the contra-lateral were considered as a pair of inputs to the network.

2 Materials and Methods

2.1 Dataset

The mammogram images used were collected from General Electric, Siemens, and Hologic from women attending for diagnostic purpose between 2000 and 2016. The images are anonymized and approved by the regional ethics board after summary review, with a waiver of a full review and informed consent [13]. The database contains 7196 exams. For most of the exams, Medio Lateral Oblique (MLO) and Cranio-Caudal (CC) views of both breasts are provided, resulting in 28294 DM images in totals. All images with malignant lesions were histopathologically confirmed, while normal exams were selected if they had at least two years of negative follow-up. From 7196 DM exams, 2883 exams (42%) contained

a total of 3023 biopsy-verified malignant lesions. The exact distribution of the dataset is shown in Table 1. In the whole dataset, 1315 exams does not have either left or right breast images of MLO and/or CC views.

Training, validation and test data split was done at patient level to evaluate the generalization of the the model developed. Data was randomly split into training (50%), validation (10%) and testing (40%) while making sure exams from each vendors present in each partition proportionally.

Table 1. Distribution of DM dataset used including their vendor.

	General Electric	Siemens	Hologic
number of studies	2248	1518	3430
normal images	7771	5842	12288
images with malignant lesions	1292	255	1476

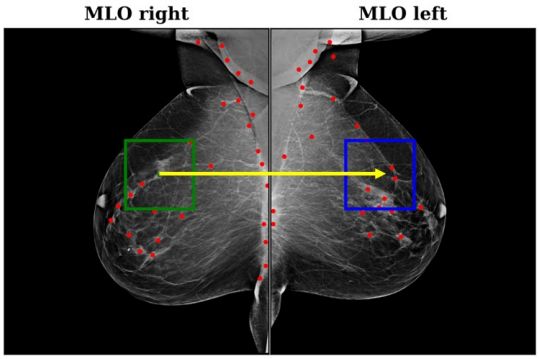


Fig. 1. An illustrative example showing center location of suspected masses in a sample MLO views of right and left breast mammogram images, and patches used to train symmetry CNN model. The green box represents a patch centered on a positive candidate on MLO view of right breast DM image, and its corresponding symmetry patch at the same location on the contra-lateral image is displayed in blue.

2.2 Candidate Selection

Previous work by [12] was employed to detect suspicious mass candidates. Fig. 1) shows sample MLO view DM images of left and right breast. Likelihood of a pixel to be part of a mass was computed using local lines and distribution of gradient orientation features. Then, a global threshold was applied to the likelihood image to generate regions that are considered as suspicious. The red and green points correspond to suspected candidate center locations of mass.

The green point is a true mass and others are false positive candidates. Table 2 details the number of suspicious candidates from training, validation and test data.

Table 2. Number of suspicious candidates. The numbers after + indicate candidates from exams without left or right breast images. Positive refers to candidates inside malignant mass and negative candidates are outside a mass.

Candidates	Training	Validation	Test
negative	337366+2359	61833 +1093	250293+6154
positive	2217+58	927+30	727+67

3 Patch Extraction and Augmentation

To extract patches, the contra-lateral images were flipped horizontally to place both images in the same space. Maximum size of the mass in our dataset was about 5cm, and a patch size of 300×300 pixels (6cm $\times$ 6cm) was considered to provide enough context to discriminate soft tissue lesions. We reduced the number of training negative candidates by ensuring a sufficient distance (at least 2cm) from a lesion and an inter-negative candidate distance of 1.4cm. This resulted in 253476 (74.6%) negatives patches.

As an augmentation scheme, initially positive patches were flipped and Gaussian blurred with standard deviation between $[0.2, 3]$. Then, with probability of 0.5 one of the three augmentations were applied to both negative and positive patches: scaling, translation around the center and rotation. The parameters for these augmentations were uniformly selected from $[0.88, 1.25]$, $[-25, 25]$ and $[-30°, 30°]$, respectively.

4 Network Architecture and Training

In addition to incorporating symmetry information, a single input baseline architecture was trained. The baseline architecture is a variant of VGG architecture [15] as shown in Fig. 2a and it consists of feature extraction and classification parts. The feature extraction section has a series of seven convolutional layers with $\{16, 32, 32, 64, 64, 128, 128\}$ filters each followed by a max pooling layer. Convolution was performed with a stride of $(1, 1)$ and valid padding. The classification part is composed of three dense layers with $\{300, 300, 2\}$ neurons and with dropout (rate= 0.5) regularization after the first dense layer. ReLU activation was chosen for all layers but softmax for the last. Global Average Pooling (GAP) [16] was applied after the last convolutional layer while the other layers are followed by 3×3 max pooling. The advantage of GAP over flattening is it minimizes overfitting by reducing the number of parameters. The symmetry model has two inputs, the primary patch and a contra-lateral patch as shown in

Fig. 2b. The parallel streams were transfer learned from the baseline architecture in Fig. 2a without weight sharing. The features from the parallel stream were concatenated and fed to the classifier. The classification part is similar to the baseline model. For exams without a contra-lateral image, zero matrices were used as a symmetry image.

Weights of both networks were initialized using Glorot weight initialization and optimized using Stochastic Gradient Decent (SGD) with time-based learning rate scheduler with an initial learning rate (ILR) of 10^{-2} for baseline architecture and 10^{-3} for the symmetry model, decay rate(ILR/200), and momentum (0.9). Mini-batch size of 64 was used and for each epoch, all positives samples were presented twice and an equal number of randomly sampled negatives, ensuring balance in each batch. Model with highest AUC on validation was selected as the best model. Furthermore, we monitored AUC for early stopping with patience of 20 epochs.

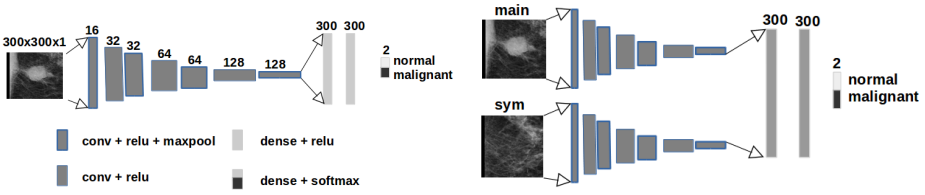


Fig. 2. CNN architectures: a) Baseline architecture b) Symmetry model architecture

5 Result

All the experiments were conducted in Keras[14], and results presented are on a separately held 40% of the data. Candidate level quantitative evaluation was done using AUC and Free Receiver Operating Characteristic (FROC) along with CPM [17] for image and exam level performance analysis. Moreover, a 95% confidence interval and p-values of AUC and CPM were computed using bootstrapping [18], using 1000 bootstraps.

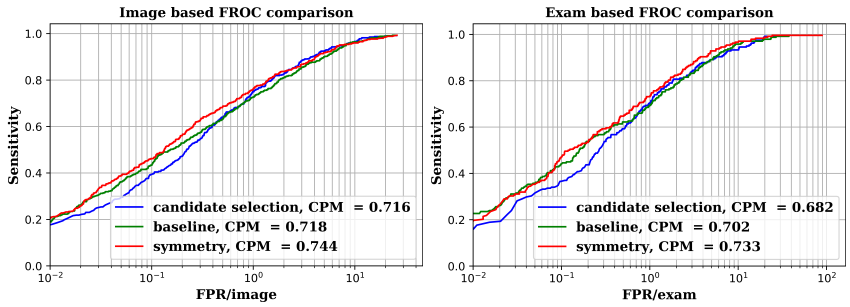
Table 3 reveals AUC values of the candidate selection, baseline and symmetry models. AUC value of 0.896 with 95% confidence interval of [0.879 , 0.913] was obtained by the model used for candidate selection. The baseline architecture that processes a single Region of Interest (ROI) image yielded an AUC value of 0.929 with 95% confidence interval [0.916 , 0.942], which is significantly better than the candidate selection stage performance ($p = 0.004$). Incorporating symmetry information improved the AUC to 0.933 with [0.919 , 0.947] 95% confidence interval, although it was not significant ($p = 0.111$) in comparison with baseline architecture. For symmetry model, a zero matrix was used as a substitute when contra-lateral image is missing and a separate evaluation is presented

in Table 3. The separate evaluation resulted in an AUC value of 0.866 with 95% significance interval of [0.788 , 0.930]. A symmetry model was trained without augmentation and the best model resulted in AUC value of 0.91. This shows the proposed augmentation has significantly improved detection AUC. Fig. 3 present

Table 3. AUC comparison of candidate selection, baseline and symmetry network. symmetry* represents evaluation of symmetry model on candidates with missing contra-lateral patch.

	candidate selection	baseline	symmetry	symmetry*
AUC	0.896	0.929	0.933	0.866

image and exam based FROC comparison of the three models. In our test set, the symmetry model showed a better performance ($p = 0.001$) compared to the baseline architecture at both image and exam level. At an image level, CPM value of 0.716, 0.718, and 0.744 with 95% confidence interval of [0.682 , 0.750], [0.679 , 0.756] , and [0.723 , 0.794] was obtained for candidate selection, baseline and symmetry model, respectively. Moreover, during exam level evaluation sensitivity of the model that incorporates symmetry context was found to be better than the other model throughout the whole False Positive Rate (FPR) range, resulting in CPM value and confidence interval of 0.733 [0.721 , 0.823] compared to 0.682 [0.671 , 0.746] and 0.702 [0.687 , 0.772] for candidate selection and baseline model, respectively.

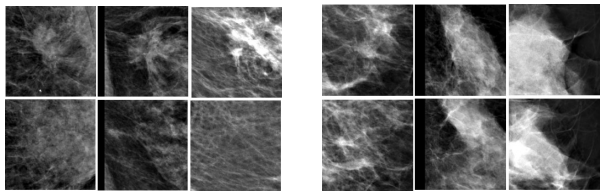


a) b)

Fig. 3. FROC comparison of candidate selection, baseline and symmetry models; a) Image based FROC. b) Exam based FROC.

6 Discussion

The proposed patch augmentation method showed an improvement in the generalization of the CNN model and thus, the performance of the classifier. Symmetry



a) b)

Fig. 4. Sample patches with an improved prediction using symmetry model: a) Positive patches that were misclassified by baseline architecture and correctly classified by symmetry model. b) Negative patches that were misclassified by baseline architecture and correctly classified by symmetry model. The top and bottom row images are primary and contra-lateral pairs, respectively.

model trained without patch augmentation yielded AUC value of 0.91 on a test set, in comparison to 0.933 when augmentation was applied. Moreover, incorporating symmetry information helps in learning distinctive features when there is a low-intensity contrast between mass and the background as shown in Fig. 4a. For the malignant candidates in Fig. 4a, without symmetry context malignancy probability was found to be below 0.2, however, integrating symmetrical information increased the malignancy prediction to a value greater than 0.7. Moreover, the negative patches in Fig. 4b were predicted as malignant masses by the baseline model (probability greater than 0.9), however, after including symmetrical context, their malignancy probability was found less than 0.1.

One of the main limitations of this work is that only soft tissue lesions were studied and detecting calcification will be of added value. Secondly, some benign abnormalities were found to be difficult for the network to differentiate from malignant candidates. We expect that separating the benign candidates from the normal and training with three classes could improve the detection performance. As studied in [11], integrating temporal context could also improve the performance of the model.

7 Conclusions

In this work, we proposed a deep learning approach that integrates symmetrical information to improve breast mass detection from mammogram images. Previous work by Karssemeijer et al. [12] was used to detect suspicious candidates. The FP candidates were reduced by learning symmetrical differences between primary and contra-lateral patches. AUC was employed as a performance measure at candidate level, whilst CPM was computed for image and exam level evaluation. We have found that our proposed approach reduces FP predictions compared to baseline architecture. An AUC value 0.933 ($p = 0.111$) with 95% confidence interval of [0.919 , 0.947] was obtained at candidate level and 0.733 ($p = 0.001$) CPM with 95% confidence interval of [0.721 , 0.823] was achieved with our symmetry model.

Training with a dataset which includes more time points could possibly improve reliability and detection accuracy[11], and will be part of our future work.

References

1. Rakhlin, Alexander, et al. "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis." arXiv preprint arXiv:1802.00752 (2018).
2. Bick, Ulrich. "Mammography: How to interpret microcalcifications." Diseases of the Abdomen and Pelvis 20142017. Springer, Milano, 2014. 313-318.
3. Giger, Maryellen L., Nico Karssemeijer, and Julia A. Schnabel. "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer." Annual review of biomedical engineering 15 (2013): 327-357.
4. Oliver, Arnau, et al. "A review of automatic mass detection and segmentation in mammographic images." Medical image analysis 14.2 (2010): 87-110.
5. Dhungel, Neeraj, Gustavo Carneiro, and Andrew P. Bradley. "Fully automated classification of mammograms using deep residual neural networks." Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017.
6. Feig, Stephen A. "Effect of service screening mammography on population mortality from breast carcinoma." Cancer 95.3 (2002): 451-457.
7. Oeffinger, Kevin C., et al. "Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society." Jama 314.15 (2015): 1599-1614.
8. Geras, Krzysztof J., et al. "High-resolution breast cancer screening with multi-view deep convolutional neural networks." arXiv preprint arXiv:1703.07047 (2017).
9. Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).
10. Wang, Xiaogang. "Deep learning in object recognition, detection, and segmentation." Foundations and Trends in Signal Processing 8.4 (2016): 217-382.
11. Kooi, Thijs, et al. "Large-scale deep learning for computer-aided detection of mammographic lesions." Medical image analysis 35 (2017): 303-312.
12. Karssemeijer, Nico. "Local orientation distribution as a function of spatial scale for detection of masses in mammograms." Biennial International Conference on Information Processing in Medical Imaging. Springer, Berlin, Heidelberg, 1999.
13. de Moor, Timothy, et al. "Automated soft tissue lesion detection and segmentation in digital mammography using a u-net deep learning network." arXiv preprint arXiv:1802.06865 (2018).
14. Chollet, Franois. "Keras." (2015): 128.
15. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
16. Lin, Min, Qiang Chen, and Shuicheng Yan. "Network in network." arXiv preprint arXiv:1312.4400 (2013).
17. Setio, Arnaud Arindra Adiyoso, et al. "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge." Medical image analysis 42 (2017): 1-13.
18. Efron, Bradley, and Robert J. Tibshirani. An introduction to the bootstrap. CRC press, 1994.

View publication stats