Lightweight Deep Learning Pipeline for Detection, Segmentation and Classification of Breast Cancer Anomalies

Hugo S. Oliveira[1(✉)], João F. Teixeira[2,3], and Hélder P. Oliveira[1,3]

[1] Faculty of Sciences, University of Porto, Porto, Portugal
`hugo.soares@fe.up.pt`
[2] Faculty of Engineering, University of Porto, Porto, Portugal
[3] INESC TEC, Porto, Portugal

Abstract. The small amount of public available medical images hinders the use of deep learning techniques for mammogram automatic diagnosis. Deep learning methods require large annotated training sets to be effective, however medical datasets are costly to obtain and suffer from large variability. In this work, a lightweight deep learning pipeline to detect, segment and classify anomalies in mammogram images is presented. First, data augmentation using the ground-truth annotation is performed and used by a cascade segmentation and classification methods.

Results are obtained using the INbreast public database in the context of lesion detection and BI-RADS classification. Moreover, a pre-trained Convolutional Neural Network using ResNet50 is modified to generate the lesion regions proposals followed by a false positive reduction and contour refinement stages while a pre-trained VGG16 network is fine-tuned to classify mammograms.

The detection and segmentation stage results show that the cascade configuration achieves a DICE of 0.83 without massive training while the multi-class classification exhibits an MAE of 0.58 with data augmentation.

Keywords: Lesion detection · Segmentation · Classification · Deep learning

1 Introduction

Breast cancer is considered a massive health problem worldwide being accountable for 15% of cancer deaths among females between 40 and 55 years of age. Despite this fact, the most effective form to reduce the mortality rate its early diagnosis [7]. The majority of the early diagnoses are still manual, achieving a sensitivity of 84% and sensibility of 91% [6]. To improve the accuracy of this manual interpretation, a double reading by another clinical expert or Computer Aided Detection (CAD) system is put in place. CAD systems are useful in

the detection, segmentation, and classification of lesions. Mammograms lesions namely breast masses commonly exhibit low signal-to-noise ratio, inconsistent appearance, and irregular shape, hampering its correct segmentation and classification [11]. The major drawback of CAD systems are the large number of False Positives (FP), while missing large portions of True Positives (TP) [9]. Recently, Deep Learning (DL) based strategies increased segmentation and classification performance. A particular advantage of DL models is their ability to automatically learn a rich hierarchy of key representative features automatically, enabled to aid the expert interpretation of the breast mammogram images. Nevertheless, DL models are trained on datasets, and need to be adapted to work in the imaging domain where the number of annotated datasets is much smaller.

Mammogram diagnosis commonly encompasses lesion detection, segmentation and classification steps. Robust lesion segmentation plays a vital in mammogram diagnosis, due to the association between the lesion shape irregularities and the probability of cancer [6]. Ground Truth (GT) annotations tend to be limited among the different databases, making the design of a robust mass segmentation algorithm challenging. To address this problematic, a large number of methods have been proposed, ranging from level set approaches [10], up to ones based in Shortest Path (SP) [3] procedures. Concerning DL models, Dhungel in [4] makes use of Convolutional Neural Networks (CNN) and deep belief networks as potential functions in structured prediction models to segment and classify breast masses. The work is based on multi-scale Deep Belief Nets (m-DBN) and Gaussian Mixture Model (GMM) for candidate generation followed by a FP reduction step, based on the features provided by two CNN, and used by an SVM classifier finalized with a Random Forest (RF) for final candidate selection. Dhungel in [5] extends his previous work by adding a hypothesis refinement based on Bayesian Optimization and Level Set method for final contour refinement, while for mass classification, a CNN model trained in two stages is used to determine mass malignancy.

With the goal to obtain a lightweight deep learning pipeline to robustly detect, segment and classify mammogram image anomalies, we evaluate the potentialities of transfer learning techniques by reusing pre-trained DL models to facilitate training and circumvent the small annotated datasets problematic. CNN has the advantage of automatically learn representative features, contrary to the hand-crafted ones that may be less representative. For the task, an augmentation, segmentation, and classification techniques are proposed and evaluated on INbreast dataset [8]. The segmentation component consists in a cascade of methods for semantic segmentation, formed by an initial region proposal stage, a CNN classifier, for FP reduction, and a final graph-based segmentation method, for lesion contour refinement. Regarding multi-class classification, a pre-trained CNN is employed with the last layers reconfigured and fine-tuned to our training data to predict the Breast Imaging Reporting And Data System (BI-RADS) level. The accuracy of the segmentation and BI-RADS classification methods are compared against GT annotations using the following measures: True Positive rate (TPr), FP for detection, Dice Coefficient (DC) for segmenta-

tion and Mean Absolute Error (MAE) for classification. The results show that the system correlates well with the GT annotations and is able to detect 85% of the masses at three FP, with a DC of 83%, achieving an final MAE of 0.524 for classification without extensive training.

## 2   Proposed Framework and Experiments

The proposed work is divided into three main stages: first the dataset construction and corresponding data augmentation techniques, secondly the cascade segmentation procedure and third the mammogram malignancy prediction. Common data augmentation consist in images rotations and mirroring during training. In order to increase the robustness of the models, we encompass image transformation by the use of affine transformations, enabling a training set with n images be increased to $n \times (n-1)$ images by applying a single affine transformation. The dataset is constructed by cropping breast regions from original mammograms and images are zero padding until the $2^{11} \times 2^{11}$ size. Translations, rotations, shear and zoom transformations where employed to increase training set. Considering that BI-RADS 6 that corresponds to biopsied cases with fewer examples and BI-RADS 5 to highly suggestive of malignancy with a lower number of cases, we merged both classes into a single one (56). Dataset augmentation, encompasses only rotations, mirroring, and affine transformation with an maximum of 20% of deformation to maintain lesion contour appearance. Table 1 summarizes the training set with examples in Fig. 1.

To tune the ResNet50 for the segmentation task, the training set encompasses 40 patch samples from mass region box with a 0.9 overlap and 40 from breast region. The main objective is for models to learn the difference between masses and background. All initial images are subject to background removal and breast region is cropped and scaled until it reaches one of the minor axis length ($x$or$y$) of the original image. After this process images are then resized to 1/4 enabling to encompass the largest mass lesions inside a 224×224 box size to fit network input (Fig. 2), with the smaller mass lesion contour occupying a minimum 35×35 pixels box, crucial to maintaining relevant lesion features. Final dataset contains 44800 patch images from both classes.

Table 1. Dataset size for BI-RADS classification.

Data	1	2	3	4	56	Total
Original	67	220	23	43	57	410
Train 75%	50	165	17	32	43	307
Test 25%	17	55	6	10	15	101
Train Aug (A)	250	825	85	160	215	1535
Train Aug (A+T)	750	2475	255	480	645	4605
Train Aug (A+Af+R+T+M+Z)	19800	2040	3840	4440	2200	32320

A - Angle, Af - Affine, T - Translation, M - Mirror, R - Rotation, Z - Zoom

<div align="center">Original   Shear   Shear H</div>
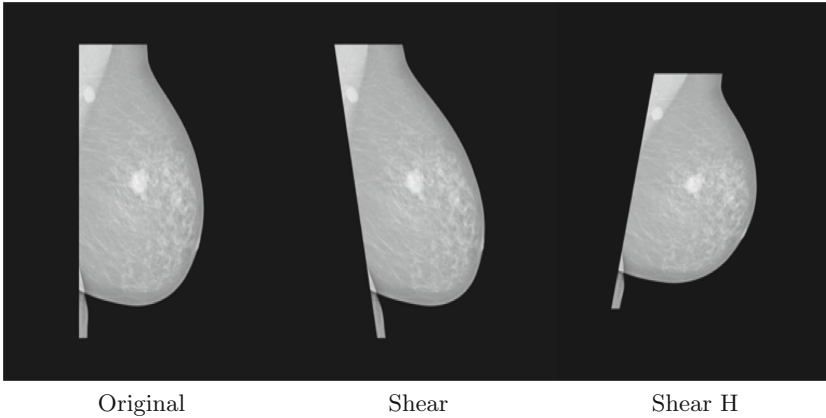
Fig. 1. Example of the constructed dataset (without mirroring).

For mass detection and segmentation, the first stage *(Resnet)* corresponds to the generation of the initial region candidates (Fig. 2), accomplished by the reuse a pre-existing CNN architecture trained in Imagenet[1], namely a ResNet50 with the final layer modified for to distinguish between mass/background images. The model is then re-trained on our sampled images patches. The choice of ResNet50 relies on the fact is composed of convolutional layers and a final global averaging pool layer, making this network suitable to compute Class Activation Maps (CAM)[2] directly without further training. The final model is then used to generate the region's proposals by sliding the image input model on larger images and attain the CAM. Regions similar to mass lesions exhibit higher activation's values, suggesting that the particular area may correspond to a Region of Interest (ROI). From CAM, square mass images candidates are taken from regions that present a CAM above the threshold T.

Since a higher number of regions may correspond to background areas, a second stage, the **FP** reduction consisting in a CNN classifier using a VGG architecture is trained using the same patch lesion/background dataset to classify the initial region's proposals as mass/background, enabling to discard FP detection's while attaining TP ones.

The third and final module of the segmentation component, the contour refinement **(Ref)**, operates only on positively identified regions. This stage consists of a SP operating in Cartesian Coordinates proposed by [3] to determine the outside boundary of convex objects. SP operating in the Cartesian Coordinates benefit from the fact that the graph is generated from the image on its original form, avoiding deformations associated with image transformations. An inverse cost function centered on the object is modulated to avoid small

---

[1] https://image-net.org/.

[2] https://jacobgil.github.io/deeplearning/class-activation-maps.

inner paths collapsing over the seed point being naturally favored when using Cartesian Coordinates.
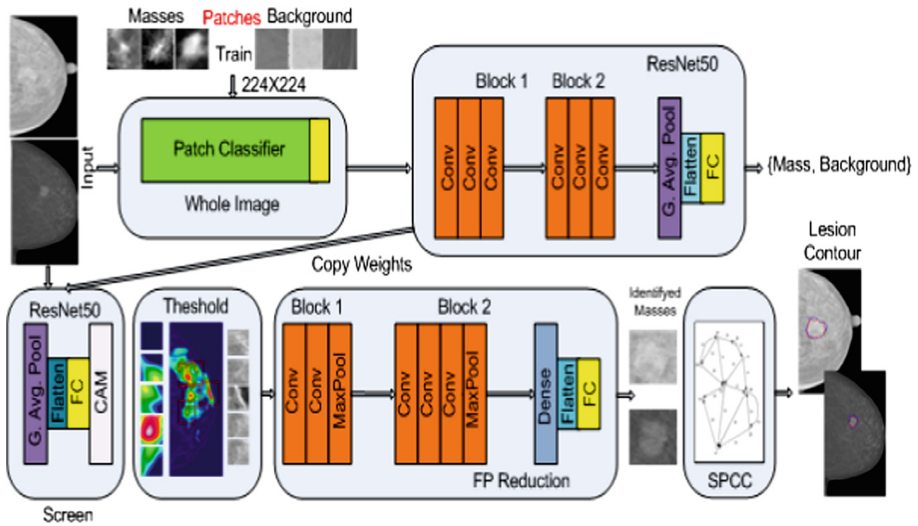
Fig. 2. Region proposal + FP reduction + contour refinement.

For BI-RADS determination, a pre-trained CNN is used, namely the VGG16 architecture trained on Imagenet. This choice is supported by the simplicity of VGG16 combined with good performance in medical context images. Since VGG16 has an input size of 224×224 with 3 channels being able to identify 1000 different classes, we resize our images dataset and replicate gray image channel among the 3 channels to fit network input and redefine to output layer to our 5 BI-RADS class problem. Table 1 summarizes the constructed dataset. Lower classes correspond to the normal cases that are the most common the population.

Both segmentation and classification performance is evaluated on INbreast [8] database. All the models are trained using two non-overlapping subsets with a 75% random split for training and testing. 5-fold cross-validation was used to determine the best parameters.

The initial region proposal (Resnet), the ResNet50 learning rate was set to $\alpha = 3 \times 10^{-3}$, $\lambda = 4 \times 10^{-4}$ and ADAM was the selected optimizer with ($\beta_1 = 0.9, \beta_2 = 0.995$ and $\epsilon = 10^{-6}$, trained for 30 epochs using the lesion/background images setting the batch size to 32. Only the new added layers are fine-tuned in the initial phase. Then, different parts of the network, deeper, middle and shallow layers where unfrozen individually and retrained during 10 epochs each, with learning rates set to 4×10^{-3} for deeper layers, 3×10^{-4} for

middle layers and 3×10^{-5} for shallow layers. This retrain strategy relies in the fact that low level features do not vary as much as high level features among different datasets.

After training, CAMs layer is included and due to memory constrains the model is slided over the whole image with a stride of $l = 5$ to generate image CAM. Regions that present CAM values above the threshold T are set to be candidates. Two distinct thresholds are evaluated for candidate generation, $T = 0.6$ and 0.8. Square image patches above the threshold are then evaluated by the FP reduction stage.

Concerning the FP reduction (FP), three different VGG architectures where trained and evaluated during 40 epochs, with the best model achieving a final accuracy in the patch test set of 0.915, with the parameters $\alpha = 2 \times 10^{-5}$, $\lambda = 3 \times 10^{-4}$ and ADAM optimizer with ($\beta_1 = 0.9, \beta_2 = 0.997$ and $\epsilon = 10^{-6}$).

For final contour refinement (Ref), a SP operating in Cartesian Coordinates is employed with the cost function corresponding the inverse of the radial distance combined with an exponential law for weight generation expressed as $\hat{f}(g) = f_l + (f_h - f_l)\frac{\exp\left((255-g)\cdot\beta\right)-1}{\exp\left(255\cdot\beta\right)-1}$, with $f_h, f_l, \beta \in \mathbb{R}$ set to be constant values ($f_h = 30, f_l = 2, \beta = 0.025$), with g being the minimum of the gradient on the two incident pixels. Results are evaluated using DC.

For BI-RADS class assessment, the VGG16 architecture pre-trained on Imagenet was used, with the new fully connected layers fine-tuned using our training data composed full breast images resized to fit network input. Initial training parameters where $\alpha = 2 \times 10^{-2}$, $\lambda = 1 \times 10^{-4}$ and ADAM as the optimizer with ($\beta_1 = 0.9, \beta_2 = 0.995$ and $\epsilon = 10^{-6}$). After training the final layer, we employ the same strategy used in the ResNet50 to retrain the deeper, middle and shallow layers of the network during 10 epochs also. The learning rates for deeper layers was set to 4×10^{-3}, 4×10^{-4} for middle layers and 4×10^{-5} for shallow layers. Results are evaluated using the MAE.

3   Results

Results are divided into two main components: segmentation and classification. Results on each stage of the segmentation cascade are compared with a State-of-the-art (SotA) method proposed by [5], that uses a Conditional Random Field (CRF) model with active contour refinement, and a manual approach proposed by Brake [1], listed in Table 2. The **method** column lists SotA works and the stages of the segmentation cascade, with a example of the segmentation stages exhibited on Fig. 3.

Several observations can be drawn from the segmentation stage:

- **Effect of the threshold T:** The region proposal stage presented an higher FP number and sensitivity of 10(1.8) and 0.85(0.1) respectively) when using a lower T.
- **Effect of the FP Reduction:** Some of the TP where rejected due to center shift initial detection, misleading the classifier.

Table 2. Performance evaluation of lesion region proposal + classifier + contour refinement. Results $mean(std)$.

Method	# FP	TP_r	DICE
Dhungel [5]	1.00(-)	0.90(-)	0.85(0.02)
ResNet ($T = 0.6$)	10(1.8)	0.86(0.09)	0.72(0.08)
ResNet ($T = 0.8$)	8(1.7)	0.83(0.10)	0.68(0.07)
Res.+FP+Ref ($T = 0.6$)	3(0.20)	0.85(0.07)	**0.83(0.10)**
Res.+FP+Ref ($T = 0.8$)	2(0.11)	0.76(0.07)	0.70(0.13)
Brake [1] (Man)			0.820(-)

Example   Heat   Contour
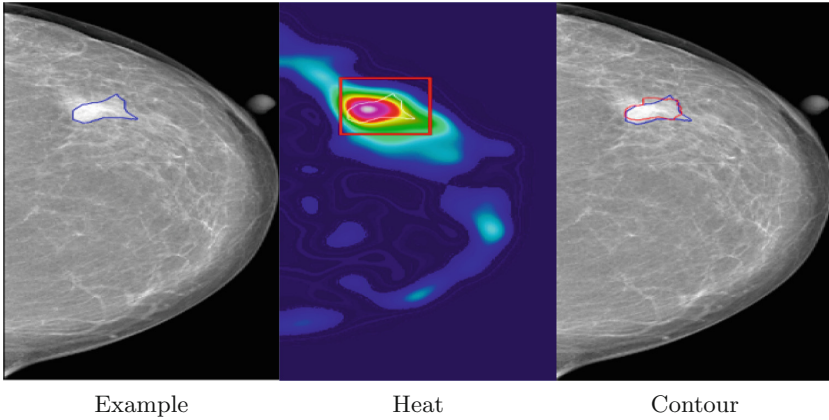
Fig. 3. Pairwise comparison between mammogram image and heatmaps (Blue - GT, Red - Detection). SP operate only on positive identified patches. (Color figure online)

– **Contour Refinement:** The SP exhibited similar accuracy when compared with the original work due to the similarities on the datasets Full Field Digital Mammography (FFDM).

Concerning the BI-RADS classifier, results are summarized in Table 3. The listed SotA method consist in Maximal-Coupled Learning using the GT annotation masks to extract features for BI-RADS classification [2].

Table 3. Attained accuracy in the test set, $mean(std)$.

Data	MAE
SotA - (Manual) Max. Coupled Learn. [2]	0.190(-)
Using Train Aug (A)	1.343(0.503)
Using Train Aug (A+Af+R+T+M+Z)	**0.584(0.011)**

Several observations can be drawn from the classification stage:

- **Effect of the data augmentation:** The affine data augmentation technique outperformed the simple rotation and mirroring of the images.
- **Effect of the image resizing:** Small calcifications that are associated with high malignancy level cannot be detected by the model and mislead the final BI-RADS level prediction.
- **Effect of pre-trained networks:** The use of pre-trained networks enabled to reuse the convolutional layers as robust feature extractor to generate a robust model without massive training data.

4   Conclusions and Future Work

The present work concerns the creation of a lightweight DL pipeline easily trained for detection, segmentation and classification of mammogram images.

Data augmentation without altering lesion shape appearance proved to be vital, enabling to generate a vast dataset improving model generalization. Only affine transformations such as zoom, shear with a maximum of 20%, translation, and rotation were considered. Shear with larger percentages and elastic deformation must be considered and asses their impact in classifier performance. Cropping and scaling enabled to create a dataset suitable to fit pre-trained network input without losing to much detail on smaller mass lesions.

Concerning the segmentation stage, the formulation of a cascade configuration enabled to train models separately and fine-tune individual stage parameters. The selection of segmentation threshold T proved to be the main bottleneck, with higher T values leading to a rejection of some of TP lesions that exhibited lower probability. Integrating both stages into a single one by using a Faster R-CNN architecture and fine-tune to our dataset can attenuate this problem. Contour refinement enabled to refine the lesion segmentation in great detail.

The BI-RADS level classification benefit from the use of a pre-trained network, enabling to obtain a robust classifier without extensive data and training time. However, BI-RADS report to the higher level must be carefully analyzed. While our approach does not beat the SotA, its prediction uses only images without using any GT contour annotation for feature extraction. Overhall, the reuse of pre-trained models enabled the creation of a well performing pipeline without extensive data and training.

References

1. te Brake, G.M., Karssemeijer, N., Hendriks, J.H.: An automatic method to discriminate malignant masses from normal tissue in digital mammograms. Phys. Med. Biol. **45**, 2843–2857 (2000)
2. Cardoso, J.S., Domingues, I.: Max-coupled learning: application to breast cancer. In: 2011 10th International Conference on Machine Learning and Applications and Workshops (2011)
3. Cardoso, J.S., Domingues, I., Oliveira, H.P.: Closed shortest path in the original coordinates with an application to breast cancer. Int. J. Pattern Recognit. Artif. Intell. (2015)
4. Dhungel, N., Carneiro, G., Bradley, A.P.: Automated mass detection in mammograms using cascaded deep learning and random forests. In: International Conference on Digital Image Computing: Techniques and Applications (2015)
5. Dhungel, N., Carneiro, G., Bradley, A.P.: A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. Image Anal. **37**, 114–128 (2017)
6. Giger, M.L.: Medical imaging and computers in the diagnosis of breast cancer. In: Photonic Innovations and Solutions for Complex Environments and Systems (PISCES) II. International Society for Optics and Photonics (2014)
7. Hela, B., Hela, M., Kamel, H., Sana, B., Najla, M.: Breast cancer detection: a review on mammograms analysis techniques. In: 10th International Multi-Conferences on Systems, Signals Devices (SSD 2013) (2013)
8. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. Acad. Radiol. **19**, 236–248 (2012)
9. Oliver, A., et al.: A review of automatic mass detection and segmentation in mammographic images. Med. Image Anal. **14**, 87–110 (2010)
10. Rahmati, P., Adler, A., Hamarneh, G.: Mammography segmentation with maximum likelihood active contours. Med. Image Anal. **16**, 1167–1186 (2012)
11. Tang, J., Rangayyan, R.M., Xu, J., El Naqa, I., Yang, Y.: Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE Trans. Inf. Technol. Biomed. **13**, 236–251 (2009)