On Improving Breast Density Segmentation using Conditional Generative Adversarial Networks

Nasibeh SAFFARI[a,1], Hatem A. RASHWAN[a], Blas HERRERA[a], Santiago ROMANI[a], Meritxell ARENAS[b], Domènec PUIG[a]

[a] *DEIM, Universitat Rovira i Virgili, Tarragona, Spain*

[b] *Hospital Sant Joan de Reus, Reus, Spain*

Abstract. Breast density is a crucial factor to follow-up the relapse of breast cancer in mammograms and the risk of local recurrence after conservative surgery and/or radiotherapy. Accurate breast density estimation with visual assessment is still a challenge due to faint contrast and significant variations in background fatty tissues in mammograms. The important key of breast density estimation is to properly detect the dense tissues in a mammographic image. Thus, this paper presents an automatic deep breast density segmentation using conditional Generative Adversarial Networks (cGAN) that consist of two successive deep networks: generator and discriminator. The generator network learns the mapping from the input mammogram to the output binary mask detection the area of the dense tissues. In turn, the discriminator learns a loss function to train this mapping by comparing the ground-truth and the predicted mask under observing the input mammogram as a condition. The performance of the proposed model was evaluated on the public INbreast mammographic datasets. The proposed model can segment the dense regions with overall recall, precision and F-score about 95%, 92%, and 93%, respectively, outperforming state-of-the-art of breast density segmentation. The proposed model can segment more than 40 images with a size of 512x512 per second on a recent GPU.

Keywords. Breast cancer, Breast density estimation, Deep learning, Generative Adversarial Networks, Mammograms

1. Introduction

Breast cancer is the most common cause of death among women, which is prevalent in both developed and developing countries. However early detection of breast cancer improves the treatment of the patients and increases their chance of survival. Several risk factors for breast cancer have been recognized, such as age, family profile, genetics, and breast density. Since breast masses have very similar appearance to dense tissues, many studies have indicated that breast density is a strong risk factor for developing breast cancer [1] [2] [3] [4] [5]. Mammographic density has been associated with the risk of local recurrence after conservative surgery and/or radiotherapy, as it is a predictor of local recurrence, especially in patients who have not received it [6]. Therefore, accurate breast density estimation is an important step during the screening procedure, because women with dense breasts tissues have three-fold to six-fold higher probability to face the risk of breast cancer [7] [8] [9]. Thus, breast density estimation is used to predict the

[1]Corresponding Author, Nasibeh Saffari, DEIM, Universitat Rovira i Virgili, Av. Paisos Catalans 26, 43007 Tarragona, Spain; E-mail: Nasibeh.saffari@urv.cat

presence of tumors at the early stage to help doctors for planning an appropriate treatment, either chemotherapy or radiotherapy.

The amount of fibro glandular tissue content in the breast is commonly referred to as breast percent density (PD %) [10]. According to the research presented by Norman et al. [11], one of the factors that hinder the detection of masses by the specialists is the type of the breast density, which can be dense (fibrous) or non-dense (fat). Furthermore, Breast Imaging Reporting And Data System standard (BI_RADS) [12], presented by the American College of Cancer, provides four categories: BI_RADS I, BI_RADA II, BI_RADS III and BI_RADS IV, as shown in Figure 1.

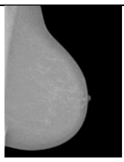	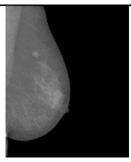	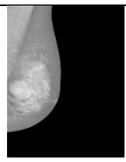	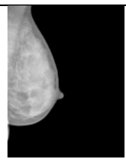
BI_RADS I Fatty breast	**BI_RADS II** Average density	**BI_RADS** III Heterogeneously dense	**BI_RADS IV** Extremely dense
(0-25%)	(26-50%)	(51-75%)	(75-100%)

Figure 1. Breast density tissues classification in INbreast mammograms.

X-Ray mammography is considered as the most popular method utilized by radiologists for screening and early detection. The common screening mammographic views are craniocaudally (CC) and mediolateral oblique (MLO). The CC mammographic view is captured from the superior view of a horizontally compressed breast. In turn, the MLO view is captured from the side of a diagonally compressed. Analysis of mammograms is not easily feasible for every case, for this reason, computer techniques to diagnose and classify the breast density have attracted the researchers' attention in recent decade [1] [2] [10] [11] [13] [14].

In the literature, several approaches for breast density estimation in mammogram analysis have been developed and presented. These methods can be classified into *hand-crafted* based methods and *deep learning* based methods. *Hand-crafted* methods are based on extracting features from mammograms to represent the dense tissues. Traditional machine learning algorithms used to segment or classify the features. For instance, a fully automated quantification of breast density percent based on adaptive multiclass fuzzy c-means (FCM) clustering and support vector machine (SVM) classification has been proposed by B. M. Keller et al. [9]. In addition, the LBP descriptor [15], Gabor filters [14] and GLCM features [16] have been used to extract features from the input mammograms for segmenting and estimating breast density.

In turn, many researches have recently been proposed based on *deep learning* systems. Deep learning networks, such as Convolutional Neural Networks (CNN), automatically learn features from raw images directly and they can properly represent the object at different scales and orientations. An unsupervised deep learning method based on CNN to learn the characteristics of dense and fatty tissues have been presented by M. Kallenberg [17]. The input mammogram is divided into a set of sub-images that are classified into dense or fatty regions. Their proposed model yielded a strong correlation to the manual breast density segmented by expert radiologists. In addition, a deep learning system consisting of four convolutional layers with a maximum pooling have been developed by C. K. Ahn [18] to segment the breast density regions in mammograms. The convolutional layers are trained as auto-encoders and their weights and bias terms are fine-tuned using softmax regression.

It should be considered that increasing the quality of dense tissues segmentation, increases the accuracy of breast density estimation. The novelty of the current paper is that a deep model has been proposed using conditional Generative Adversarial Network (cGAN) [18] to improve the segmentation of the breast dense regions in mammograms. cGAN is a deep learning model that can learn the statistical invariant features (texture, color, etc.) of an input image and then generate nearly synthetic images, which look like the input image. cGAN networks consist of two successive networks: generator and discriminator. The generator network learns to map the input image to the target output, in turn the discriminator learns a loss function to train this mapping by comparing the ground-truth and the predicted output. Finally, the cGAN network optimizes a loss function that combines a conventional binary cross-entropy loss with an adversarial term (i.e., discriminator), which encourages the generator to produce outputs that cannot be distinguished from ground-truth ones. The main objective of the present paper is to enhance the accuracy of breast density segmentation using the cGAN network. To the best of authors' knowledge, this is the first application of the cGAN network for segmentation of dense tissues in mammographic images.

The remaining part of this paper is structured as follows. Section 2 is devoted to the proposed cGAN model. In order to demonstrate the effectiveness of the proposed work, some experimental results and a comparative study are shown in Section 3. In addition, conclusion and future work are explained in Section 4.

2. Proposed Methods

An overview of the proposed methodology is presented in Figure 3, followed by a summary of each step in the process. The cGAN is a conditional variation of the GAN, where the generator is instructed to generate a real sample having specific characteristics rather than a generic sample from full distribution [19]. We assumed that the cGAN structure is well- suited to accurately outline the breast density area, especially when the training data is limited and our experimental results support our hypothesis.

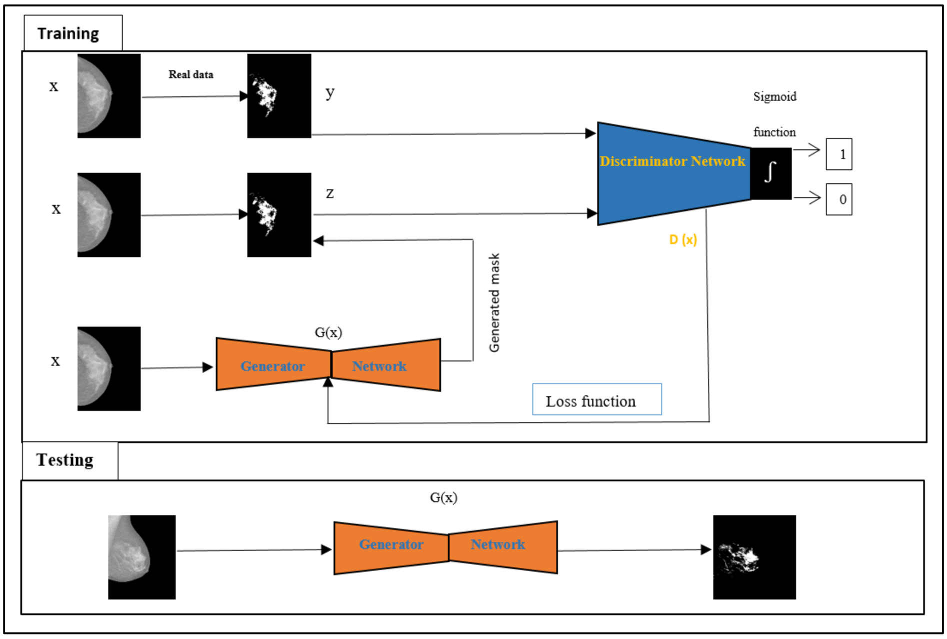

Figure 2. Conditional Generative Adversarial Network framework.

As demonstrated in Figure 2, the cGAN network is comprised of two main networks, generator and discriminator. In our model, the generator Network G takes a mammographic image and tries to generate a mask image of the areas related to dense tissues (i.e., 0 for non-dense pixels and 1 for dense pixels). The generator network then generates a data latterly fed into a discriminator network. The discriminator D learns a loss function to train this mapping by comparing the ground-truth and the predicted output, but with observing the input image as a condition in order to improve the network optimization as proposed in [19].

The generator networks follow an encoder-decoder architecture of U-net with skip connection [20]. The encoder includes down sampling 8 convolutional layers. The first layer uses 7x7 convolution to generate 64 feature maps and the final layer generate 512 feature maps with a 1x1 size. While, the six middle layers are from the pre-trained ResNet-101 [20]. To improve the segmentation performance, we also used a U-net architecture based on skip connections in which the input of each decoder is concatenated to of its corresponding convolutional output of the encoder. In turn, the decoder includes up sampling 8 convolutional layers with a reverse ordering layers that are similarly structured to the encoder network. On the other hand, the discriminator network consists of 5 convolutional layers. The first layer of the discriminator used 64 filters of 3x3 and a stride of 2x2. The final layer of the discriminator produces 512 feature maps with a size of 30x30 followed by Sigmoid as an activation function.

The proposed cGAN model has been trained over a loss function resulting from combining a content and adversarial losses. The content loss follows a classical approach in which the predicted dense mask is a pixel-wise compared to the corresponding one from ground-truth. In turn, the adversarial loss depends of the real/fake prediction of the discriminator over the ground-truth and the predicted foreground mask with observing the input image. The training process of this cGAN can be expressed as an optimization of the following function presented in Eq. (1), which mathematically describes the training of cGAN.

$$G *= argmin_G \, max_D L_{GD}(D,G) + \lambda \, L_{MSE} \tag{1}$$

Where $\lambda = 10$, and $L_{GD}(D,G)$, the binary cross entropy (BCE) of the adversarial, can be computed as:

$$L_{cGAN} \, (D,G) = E_{x,y}\big[\log\big(D(x,y)\big)\big] + E_{x,y}\big[\log\big(1 - D(x,z)\big)\big] \tag{2}$$

In Eq. (2), the first term is entropy of the discriminator D with real data (i.e., the input image is x and the ground-truth is y, both images are concatenated). The second term is entropy with a fake input data (i.e., the input image is x and the generated image is z, both images are concatenated) passes through the generator, which is then passed through the discriminator to identify the fakeness (i.e. the log probability that the data from generated is fake if it equals to 0), and the content loss function computed between z and its corresponding ground-truth y, can be defined as:

$$L_{MSE}(G) = \frac{1}{N}\sum_{j=1}^{N} \|y - z\|_2 \tag{3}$$

Where N is the number of the pixels per input image.

During training, the discriminator tries to maximize (1), while the task of the generator is exactly the opposite that tries to minimize (1). For our experiments, we have used an Adam [21] optimizer with learning rate 0.0002, and batch size equals 4, in addition to an optimal number of epochs equals 200.

3. Experimental results

3.1. Database

To evaluate our results, we have used the INbreast dataset [22], which consists of 410 mammographic images and presented from MLO and CC views for both side of breast (right and left). We manually segmented the 410 images to generate the ground-truth masks assessing by two radiologists of Hospital Sant Joan de Reus. This dataset has been divided into three sets, 250 images for training (~60%), 60 images for validation (~15%) and 100 images for testing (~25%).

3.2. Pre-processing

Re-sizing: The mammographic images of the INbreast dataset used in this work have large sizes of 2000x2000 or more. In order to reduce the computation time, all mammograms were re-sized to a size of 512 × 512 and normalized to scale (0...1).
Removing pectoral muscle: Since the intensity similarity and the overlap between the pectoral muscles and the glandular dense tissues can cause false positive detection of dense regions. Pectoral muscle area extraction can help to reduce the false positives [23],

hence, we manually removed pectoral muscles of all images in the INbreast dataset using a threshold method a proposed in [24].

3.3. Result and discussion

We compared the proposed cGAN to three common methods of semantic segmentation based on deep learning models FCN8 [25], FCN32 [25] and Vgg-Segnet [26]. The three aforementioned methods depend on the generator networks. We computed seven evaluation metrics; recall, specificity, false positive rate(FPR), false negative rate (FNR), PBC, precision and F-score, over the segmented images resulted with the four tested methods. For the best segmentation performance, recall, specificity, precision and F-score metrics values should be high, however False Negative Rate(FNR), False Positive Rate **(**FPR) and PBC should be low.

Quantitative measurement results with the testing set (i.e., 100 images) are summarized in Table1. As shown, the GAN-Unet yielded the best results among the four tested methods with F-score of 94% of. In addition, it outperformed all the three methods tested over the six evaluation metrics. Respectively, the FCN-32 achieved the worst results among the evaluated methods with F-score about 50%, since this network consists of 32 convolutional layers that need a lot of images to train for a proper optimization and that is not available for all cases of medical images datasets. In addition, the SegNet network provided acceptable results with F-score of 73%.

Qualitative results with the proposed model and the three tested methods are shown in Figure 3. Supporting the quantitative results shown in Table 1, the segmented images resulted from the cGAN model properly segment the dense regions including the small regions and preserving the small details and boundaries of the dense tissues. In turn, the other methods, FCN-8, FCN-32 and Vgg-Segnet, smoothed the boundaries of the dense tissues and have not detected the small regions.

Table 1. Recall, Specificity, FPR, FNR, PBC, Precision and F-score with the cGAN-Unet, FCN8, FCN32 and VGG-SegNet methods evaluated on the testing set of the INBreast dataset. The best results are marked in a bold text.

METHODS	RE CALL	SPECIFICITY	FPR	FNR	PBC	PRECISION	F-SCORE
cGAN-Unet[20]	**0.957**	**0.999**	**0.000**	**0.043**	**0.022**	**0.916**	**0.936**
F C N-8[25]	0.748	0.997	0.002	0.252	0.506	0.693	0.721
FCN-32[25]	0.572	0.997	0.004	0.428	0.795	0.595	0.584
Vgg-Segnet[26]	0.832	0.996	0.004	0.169	0.558	0.662	0.738

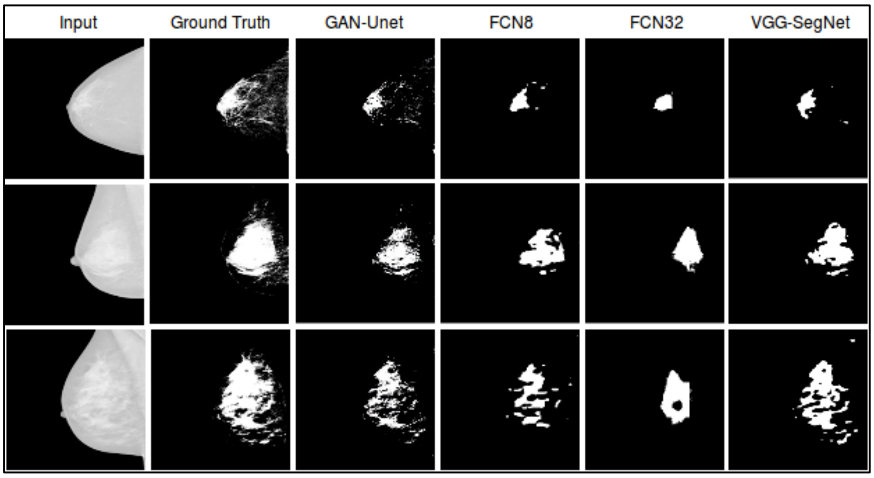

Figure 3. Examples of the breast density segmentation: (col 1) breast mammogram images, (col 2) ground truth mask, (col 3) generated masks with our proposed method cGAN-Unet, (col 4) FCN 8, (col 5) FCN 32 and (col 6) VGG-SegNet.

4. Conclusion

This paper proposed to use the conditional Generative Adversarial Network for promising breast density segmentation. The cGAN network consists of two networks. The first network is an encoder-decoder network, called generator, that maps the input mammogram to a segmented image including the dense areas, and The second network is called discriminator that learns a loss function to enforce the generator to produce an output, which is similar to the ground-truth. Using of the network help the proposed model to train by a small set of mammographic images. The proposed method can segment the dense regions in mammographic images with overall recall, precision and F-score around 95%, 92%, and 93%, respectively outperforming the common semantic segmentation methods (FCN and SegNet). Future work aims to use the density segmented in mammographic images to generate a more accurate breast density estimation.

Acknowledgements

This work was supported by the Spanish project: DPI2016-77415-R and the program of Beatriu de Pinós (BP 2016): 2016 BP 00063.

References

[1] A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. R. Denton, and R. Zwiggelaar, “A novel breast tissue density classification methodology,” IEEE Trans. Inf. Technol. Biomed. 12, 55–65, 2008.

[2] A. Oliver, M. Tortajada, X. Lladó, J. Freixenet , S. Ganau , L. Tortajada , M. Vilagran , M. Sentís , R. Martí, "Breast Density Analysis Using an Automatic Density Segmentation Algorithm," Digital Imaging, vol. 28, no. 5, p. 604–612, 2015.

[3] N. M. Parthaláin, R. Jensen, Q. Shen and R Zwiggelaar, "Fuzzy-rough approaches for mammographic risk analysis," Intelligent Data Analysis, vol. 14, no. 2, p. 225–244, 2010.

[4] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jongl, and M. J. Yaffe,, "Automated analysis of mammographic densities," Physics in Medicine and Biology, vol. 41, no. 5, p. 909–923, 1996.

[5] W. He, E. Denton, K. Stafford and R. Zwiggelaar, "Mammographic Image Segmentation and Risk Classification Based on Mammographic Parenchymal Patterns and Geometric Moments," Biomed. Signal Processing and Control, vol. 6, no. 3, p. 321–329, 2011.

[6] Cil, T., Fishell, E., Hanna, W., Sun, P., Rawlinson, E., Narod, S. A., & McCready, D. R. "Mammographic density and the risk of breast cancer recurrence after breast conserving surgery". Cancer: Interdisciplinary International Journal of the American Cancer Society, 115(24), 5780-5787. 2009.

[7] Cancer Research UK, "Breast cancer statistics Available online,"http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer, 2014.

[8] Breast Cancer U.S, "Breast Cancer Statistics," Available online: http://www.breastcancer.org/symptoms/understand_bc/statistics, 2016.

[9] McCormack VA, dos Santos Silva I , "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," Cancer Epidemiol Biomarkers Prev, vol. 15, no. 6, p. 1159–1169, 2006.

[10] B. M. Keller, D. L. Nathan, Y. Wang, Y. Zheng, J. C. Gee, E. F. Conant, and D. Kontos, "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation," Medical Physics, vol. 39, no. 8, pp. 4903-4917, 2012.

[11] Norman F. Boyd, D.Sc., Helen Guo, Lisa J. Martin, Limei Sun, Jennifer Stone, Eve Fishell, F.R.C.P.C., Roberta A. F.R.C.P.C., Greg Hislop, F.R.C.P.C., Anna Chiarelli, Salomon Minkin, and Martin J. Yaffe, "Mammographic Density and the Risk and Detection of Breast Cancer," New England Journal of Medicine, vol. 356, no. 3, p. 227–236, 2007.

[12] Susan G. Orel, Nicole Kay, Carol Reynolds, Daniel C. Sullivan, "BI-RADS categorization as a predictor of malignancy," Radiology, vol. 211, no. 3, pp. 845-50, 1999.

[13] A. Rampun , P. J Morrow , B. Scotney , J. Winder, "Breast Density Classification using Local Quinary Patterns with Various Neighbourhood Topologies," Multidisciplinary Digital Publishing Institute, 2017.

[14] Z. Gamdonkar, K. Tay,W. Ryder, P. C. Brennan, C. Mello-Thoms, " Idensity: An automatic Gabor filter-based algorithm for breast density assessment," In SPIE medical imaging.International Society for Optics and Photonics., p. 941607, 2015.

[15] Y. Zheng, B. M. Keller, S. Ray, Y. Wang, E. F. Conant, J. C. Gee, and D. Kontos, "Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment," Medical Physics, vol. 42, no. 7, p. 4149–4160, 2015。

[16] S. Petroudi, I. Constantinou, M. Pattichis, C. Tziakouri, K. Marias, C. Pattichis, "Evaluation of spatial dependence matrices on multiscale instantaneous amplitude formammogram classification," In 6th European conference of the international federation formedical and biological engineering-Springer, p. 156–159, 2015.

[17] M. Kallenberg, K. Petersen, M. Nielsen, A. Ng, P. Diao, C. Igel, C. Vachon, K. Holland, N. Karssemeijer and M. Lillholm, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," IEEE Trans. Med. Imaging, vol. 35, p. 1322–1331, 2016.

[18] C. K. Ahn, C. Heo, H. Jin., J. H. Kim, "A Novel Deep Learning-based Approach to High Accuracy Breast Density Estimation in Digital Mammography.," in Proceedings of the SPIE Medical Imaging, Orlando, FL, USA, 3 March 2017.

[19] P. Isola, J-Y. Zhu, T. Zhou, A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in cvpr, 2017.

[20] O. Ronneberger, P. Fischerand T. Brox, "U-net: Convolutional net- works for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

[21] D. P. Kingma ,J. Ba, "Adam: A method for stochastic optimization," in ICLR, 2015.

[22] I.C. Moreira, I. Domingues, M. J. Cardoso, Jaime S. Cardoso "INbreast: Toward a Full-field Digital Mammographic Database," Academic radiology, pp. 236-248, 2012.

[23] N. Saidin, H. AmyliaMat Sakim, U. Kalthum Ngah, I. Lutfi Shuaib, "Computer Aided Detection of Breast Density and Mass, and Visualization of Other Breast Anatomical Regions on Mammograms Using Graph Cuts," Computational and Mathematical Methods in Medicine, 2013.
[24] M. Abdel-Nasser, A. Moreno , H.A. Rashwan , D. Puig , Analyzing the evolution of breast tumors through flow fields and strain tensors, Pattern Recognit. Lett. 93 (Supplement C) (2017) 162–171. Pattern Recognition Techniques in Data Mining.
[25] J. Long, E. Shelhamer, T. Darrell, " Fully convolutional networks for semantic segmentation, in Proc. IEEE Conf " in Computer Vision and Pattern Recognition(CVPR), pp. 3431–3440. 2015.
[26] V. Badrinarayanan, A. Handa. R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 39. pp. 2481-2495. 2016.