Elastic Deformations for Data Augmentation in Breast Cancer Mass Detection

Eduardo Castro[1], Jaime S. Cardoso[1,2] and Jose Costa Pereira[1,3]
{eduardo.m.castro , jaime.cardoso , jose.c.pereira} @inesctec.pt

***Abstract*— Two limitations hamper performance of deep architectures for classification and/or detection in medical imaging: (i) the small amount of available data, and (ii) the class imbalance scenario. While millions of labeled images are available today to build classification tools for natural scenes, the amount of available annotated data for automatic breast cancer screening is limited to a few thousand images, at best. We address these limitations with a method for data augmentation, based on the introduction of random elastic deformations on images of mammograms. We validate this method on three publicly available datasets.**

Our proposed Convolutional Neural Network (CNN) architecture is trained for mass classification – in a conventional way –, and then used in the more interesting problem of mass detection in full mammograms by transforming the CNN into a Fully Convolutional Network (FCN).

I. INTRODUCTION

Worldwide, breast cancer is the most lethal form of cancer in women [1]. It is estimated that 1.7 million new cases and 520 thousand deaths happen due to it every year; making it one of the biggest health concerns in modern society. Similar to other forms of cancer, early detection is critical for successful treatment. The National Cancer Institute (NCI) states that when diagnosed in early stages, survival rates are nearly 100%, unfortunately dropping to 22% for later diagnosis [2]. One difficulty for early diagnosis arises from the fact that, initially, breast cancer is asymptomatic. As such, detection needs to be done through screening.

In mammograms, radiologists look for malignant lesions – often subtle – with an eye for detail. The screening of a large group of women substantially increases specialized workload. This is aggravated in double reading scenarios, which have been shown to yield better detection rates [3]. With this problem in mind, many algorithms have been proposed by the scientific community to assist radiologists during diagnosis, increasing early detection rates while also reducing workload [4]. These Computer Aided Detection systems traditionally work by first detecting, segmenting and characterizing lesions using heuristic algorithms and then classifying them using machine learning [5].

For detection and segmentation, approaches are often based on the bright appearence of masses [6] or by detecting object edges on the mammogram [7], [8]. Characterization is usually done with a wide set of features related to intensity, morphology and generic texture descriptors [9]. For the classification task, Support Vector Machines and Random Forests are still widely used, but other classification tools can be found [10].

The recent success of deep learning architectures in many visual recognition tasks [11] has motivated researchers to favor these methods instead of the traditional CAD pipeline for mammogram screening [9], [12]. These architectures include feature learning and classification in a single framework, and they can easily be adapted to perform detection and segmentation. Rather than using heuristics, these architectures work by minimizing a certain *loss function* that encodes the classification error from a set of labeled samples.

In breast cancer detection, as in other medical imaging applications, the number of healthy samples largely exceeds that of unhealthy ones. If care is not taken this will lead to biased models. Additionally, the cost of missing a malignant mass is much higher than a false detection (predicting malignant when in fact it's benign, or no mass exists). Due to this, CAD platforms tend to favor *high sensitivity* at the expense of more *false positives* per image. On the other hand, it is well established that deep learning models require a high number of labeled examples. Otherwise they tend to under-perform traditional classification methods. The relatively small size of mammogram datasets (particularly in unhealthy cases) and the general scarcity of labeled data in this area, makes it hard to train good deep learning models for mass detection.

To counter the effect of the small number of available samples, data augmentation techniques (*e.g.* rotations, scaling or translations) are often used. In this work, we explore a new technique of elastic distortions. The rationale of using this type of data augmentation is as follows. Because the breast is a non-rigid object which is stretched before the X-ray, the same breast can appear with slightly different appearance in different exams due to different stretching. As such, one can artificially introduce similar small elastic distortions to create new artificial samples that are physically plausible. We also propose a new *Convolutional Neural Network* (CNN) topology to show that elastic distortions on mammograms can significantly improve mass detection performance on CAD systems.

Our contributions in this manuscript are thus two-folded: (i) a new architecture for mass classification inspired in the *VGG model* of [13], and (ii) a data augmentation strategy using elastic deformations motivated by the nature of breast lesions, and the scarcity data for mass detection.

[1]INESCTEC, Porto 4200-465, Portugal
[2]Department of Electrical and Computer Engineering, School of Engineering, Porto University, Porto 4200-465, Portugal
[3]Department of Computer Science, School of Engineering, Porto University, Porto 4200-465, Portugal

II. METHODOLOGY

In this section, we describe our proposed framework for mass detection. We start by an overview of the preprocessing and patch extraction tasks. Then, we proceed to detail the proposed network topology and the adopted data augmentation strategy – based on elastic deformations of the mammograms. We conclude with an end-to-end overview of the proposed mass detection method.

Preprocessing: mammogram images are first processed to: (i) correct contrast by adaptive histogram equalization, (ii) separate the breast tissue from the background, and (iii) remove image artifacts (*e.g.* film boundary, watermarks). The first step is used to reduce variability between images; a contrast-limited adaptive histogram equalization (CLAHE) is applied. Also, image size is reduced by a factor of 12, using pixel area relation interpolation. In this method each pixel's intensity in the down-sampled image is set to the mean of the corresponding region in the original image. Resizing allows the overwhelming majority of lesions to fit an input of size of (76×76), dramatically reducing the required computational power. Finally pixel intensity range is centered around zero by remapping intensities linearly to the range $[-0.5, 0.5]$ which provides numerical stability during training [14]. Second, the image region containing tissue (ROI) is separated from the background using morphological filtering, image subtraction and thresholding; similar to [15]. Additionally, we only keep the binary object with the largest area and perform a dilation to guaranty the whole breast is inside the ROI. Pixels in the mammogram outside this region are set to zero, which eliminates most of the digitization artifacts.

Patch extraction: traditional data augmentation involves rotations, mirroring and translations. For rotation and mirroring, parameters are an angle in the range $[0, 2\pi[$ and a boolean value, respectively. The full image is rotated (mirrored) and patches of fixed-size – (76×76) – are collected from the rotated (mirrored) mammogram. Translations are added by collecting multiple nearby patches for each lesion. The full-scheme for data augmentation based on elastic deformations of the mammogram is detailed in Section II-B. For each lesion, a total of 9 positive patches centered in the red dots are collected, see Figure 1. Negative patches are collected by grid sampling the image with a fixed-step equal to 50% of the patch length. Their centers are represented by the blue dots of Figure 1.

A. Convolutional Neural Networks

The neocognitron [16], proposed in the 80's by Fukushima, triggered the development of Convolutional Neural Networks, a broader class of networks that have became a mainstream tool for visual recognition tasks. They model data using a composition of relatively simple functions called layers and parameterized by weights. The same model can solve many different tasks as long as a "good" set of weights can be found; usually through an optimization process called learning.

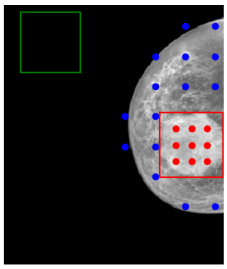


Fig. 1: Selected points for patch extraction. Blue and red dots indicate negative and positive patch centers, respectively. Red square indicates the lesion bounding box. Green square indicates patch size.

In this work a new architecture is proposed, inspired in the design principles described by Simonyan and Zisserman in [13]. All filters are 3×3 and the depth varies depending on the position of the layer within the network. *Rectifier Linear Units* (ReLU) are used after each convolutional layer, which have been shown to decrease the overall training time, while increasing the network's discriminative power [17]. Weight initialization is done as proposed by Gorot and Bengio in [18]. For training, categorical cross-entropy is minimized using *Adam* [19], a gradient descent algorithm which includes momentum and an adaptive learning rate for each variable. In practice these properties reduce the importance of weight initialization and hyper-parameter optimization in the final solution, while also allowing a faster convergence.

The ultimate goal is to obtain a probability map that emphasizes the regions of a mammogram that are more likely to have a lesion. We formulate this detection problem as a classification one. First, a CNN patch classifier is trained in a conventional way. This patch classifier is then transformed into a fully convolutional network , which classifies the whole image in one forward pass and outputs the same numerical result as screening with the patch classifier in a sliding window fashion, while taking much less time. The transformation is done by operating three changes in a layer-wise fashion. (1) The input layer is made bigger to fit a whole mammogram. (2) For max-pooling layers, the (2×2) filter is applied to four copies of the image starting at different pixel locations – $\{(0,0);(0,1);(1,0);(1,1)\}$ (as in [20]). (3) Dense layers are implemented as convolutional ones by reshaping weights and operations (as in [21]). A schematic with the proposed detection method is shown in Figure 2. The upper part of the diagram focuses on the patch-based training architecture, while the lower half emphasizes the changes made for inference on full mammograms.

B. Elastic Deformations

In a continuous body, a deformation results from a stress field induced by applied forces[1]. If the deformation recovers after the stress field has been removed the deformations are called elastic. This phenomena can realistically occur in

[1]changes in the temperature field can also cause deformations, but they are out of scope of the current application.

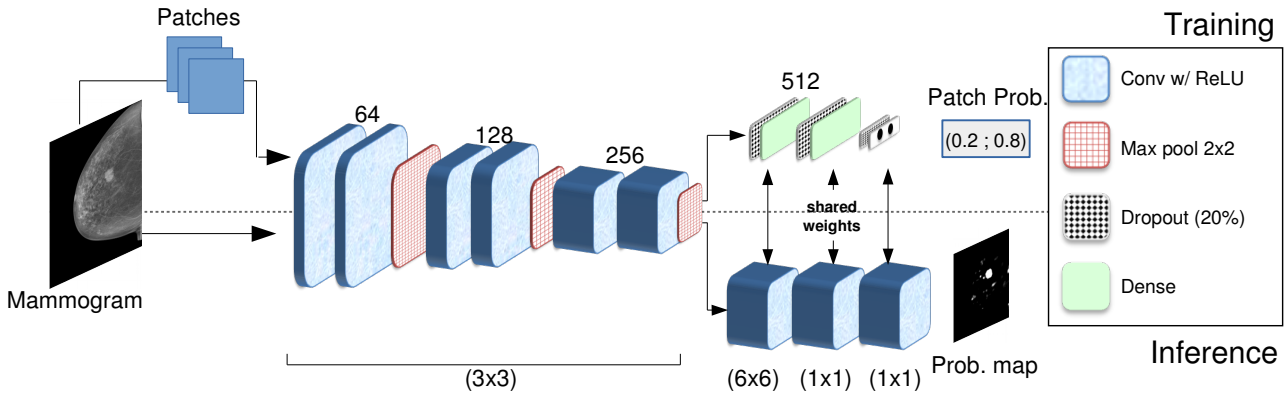
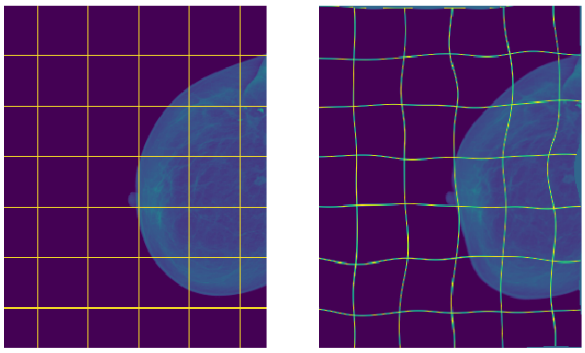


Fig. 2: Overview of the proposed framework. Numbers on top of layers correspond to either number of filters (convolutional layers) or neurons (dense layers). Numbers at the bottom correspond to the filter size of the convolutional layers.

breast cancer screening. For each exam, a slightly different stretch is applied due to the position of the breast and the strength used in the compression. As such, the same breast observed in two different screenings may have different appearances, which should not influence the decision of whether a lesion is present or not. Elastic deformations as a data augmentation technique, therefore, comes naturally as a method to model these variations.



(a) Original (b) Deformed

Fig. 3: Effects of performing elastic deformation on a mammogram.

In CNNs, these transformations have been used for data augmentation in handwritten digit recognition tasks [22]. To the best of our knowledge, this is the first work that uses elastic deformations to generate synthetic samples of cancer masses. Recognizing that breast tissue is not a rigid-body – and as such is subject to deformations – is a stepping stone that makes way for alternative methods of data augmentation. The synthetic samples are used to artificially increase the data available for training.

Obtaining a deformation on a mammogram image is done in two parts. First a random stress field is generated for the horizontal and vertical directions, Δ_x, and Δ_y respectively. For each pixel and direction, a random value in the range of $\alpha \times [-0.5, 0.5]$ is uniformly picked. To ensure close pixels have similar displacement the resulting horizontal and vertical images are applied a Gaussian filter separately (eq. (1) and (2)). These transformations have two parameters: the maximum value for the random initial displacement (α) and the strength of the smoothing operation, given by the standard deviation of the Gaussian filter (σ). We set these to be $\alpha = 300$ and $\sigma = 20$, based on the resulting patches appearance. After this, the stress field is applied to the image, the breast segmentation mask and mass annotations. This is done by moving each pixel to a new position (eq. (3)) and using spline interpolation of order one to obtain intensities at integer coordinates. An example of the effects of such deformation is show in Figure 3.

$$\Delta_x = G(\sigma) * (\alpha \times Rand(n, m)) \quad (1)$$

$$\Delta_y = G(\sigma) * (\alpha \times Rand(n, m)) \quad (2)$$

$$I_{trans}(j+\Delta_x(j,k), k+\Delta_y(j,k)) = I(j,k) \quad (3)$$

In the equations above, I and I_{trans} are the original and transformed images, respectively; and $n \times m$ are the mammogram dimensions.

C. End-to-end mass detection

In this section, we briefly sketch the proposed method end-to-end. From training (on patches) to inference (on full mammograms).

A network similar to the one shown in Figure 2 is trained using patches of masses and normal breast tissue. These are extracted as described in the beginning of Section II. Balanced batches of positive and negative images are fed to the classifier. Due to the existing imbalance, the model will repeat positive cases much more frequently than negative ones. We use data augmentation on the positive class to make its samples different from the originals.

During inference: (1) by transforming the trained network into a FCN, we obtain, for each test mammogram, a probability map that indicates a prediction of whether a lesion is present in each area of the exam image; (2) the resulting probability map is thresholded, with a value (≈ 0.5) depending on the dataset; (3) from the resulting binary image, small objects are removed via morphological opening; and finally (4) predicted lesion detections are the regions that survive the previous step. Bounding boxes are obtained for each region, and are evaluated against ground-truth annotations.

III. EXPERIMENTS

There are two sets of experiments. First, using the proposed CNN architecture we compare two strategies for data augmentation: traditional *vs.* elastic deformations. For this, 40 transformations of each image are computed before training. Second, we compare the performance of our proposed CNN architecture against state-of-the-art methods.

A. Datasets

We make use of three publicly available datasets summarized in Table I.

INbreast [23] is composed of high quality full-field digital (FFD) mammograms of both healthy and unhealthy patients as well as detailed lesion annotations. Because INbreast does not a have a standard train/test split, we use stratified five-fold cross validation with a proportion of $80/20\%$ (train/test) for a more robust measure of performance.

CBIS [24] is a Curated Breast Image Subset of DDSM [25], provided by *The Cancer Imaging Archive* [26] divided in a fixed train/test split. In particular: questionable cases were removed, images were standardized to the DICOM format, and lesion annotation was improved.

BCRP [25] is a fixed split (train/test) of the most challenging cases from DDSM [25].

Important to note that in the case of CBIS and BCRP, images were obtained by digitizing mammograms that were originally on film. This results in lower quality samples when compared to the FFD mammograms of INbreast. Both BCRP and CBIS are divided into masses and calcifications. We only use the former set of images.

TABLE I: Dataset summary used in the experiments.

	Cases		Images		Masses	
	train	test	train	test	train	test
INbreast	108		410		116	
CBIS	691	201	1231	361	1318	378
BCRP	39	40	156	160	84	87

B. Detection and Evaluation

At inference time, the full image (mammogram) is forward propagated through the FCN to obtain a probability map which is then thresholded and filtered. Isolated blobs are counted as detections. At this stage there is no data augmentation. Following common practice, the true positive (TP) criterion is met when the intersection over union (IoU) of the mass's bounding box and prediction bounding box is greater then 0.2. Region-based analysis is done by means of the *Free Response Operating Characteristic* (FROC) curve. This plots the true positive rate (TPR, or sensitivity) as a function of the number of false positives per image (FPI). The curves for the two data augmentation strategies for CBIS and BCRP are shown in Figure 4.

C. Results

FROC curves are summarized in Table II for TPR levels of 80% or 60% depending on the dataset. In the case of the BCRP, a more challenging dataset, a threshold of 0.6 was required to avoid multiple detections being connected in the thresholded image and thus considered as a single one. We also used a TPR of 60%, as a TPR of 80% is still not attainable by any CAD system that we are aware of. Comparison between our elastic deformation strategy and a traditional data augmentation method can also be seen in Figure 4 for different levels of FPI.

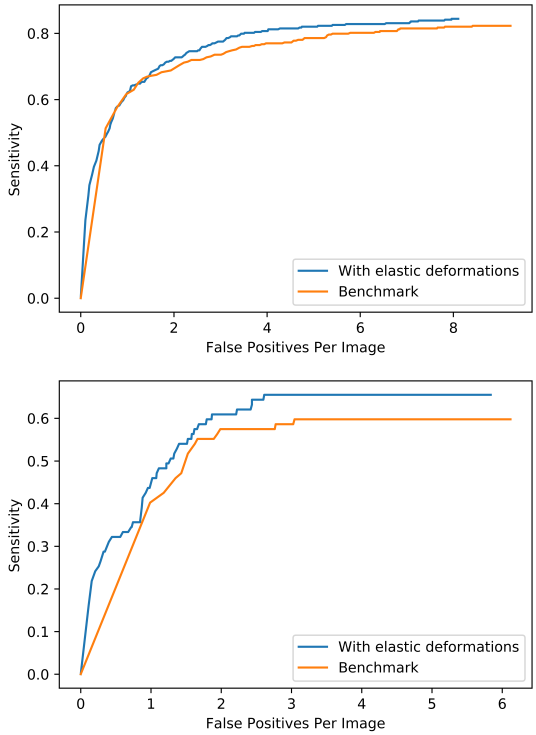


Fig. 4: FROC curve, showing sensitivity vs. number of false-positives per image. Top - CBIS; Bottom - BCRP

TABLE II: Number of false-positives per image (FPI) measured at 80% sensitivity (TPR) for "INbreast" and "CBIS", and at 60% for the more challenging "BCRP".

	INbreast [23]		CBIS [24]		BCRP [25]	
	TPR	FPI	TPR	FPI	TPR	FPI
elastic	0.8	1.171	0.8	**3.509**	0.6	**1.864**
benchmark	0.8	**0.912**	0.8	5.757	0.6	3.047
thres.	0.5		0.5		0.6	

Data augmentation: even-though the strategy based on elastic deformations is not a winner *across-the-board*, it is worth noting that in the cases where it does perform better – CBIS and BCRP datasets –, for the same level of sensitivity the number of FPI drops $\approx 40\%$ when compared to FPI achieved by traditional data augmentation (*i.e.* denoted "benchmark" in Table II).

CNN architecture: When used together with the elastic deformation strategy for data augmentation, the proposed CNN compares competitively with other known methods. Table III summarizes the best reported results for the public datasets used in these experiments. In particular, for the IN-

breast dataset, the recent work of Dhungel *et al.* [12] achieves impressive sensitivity (0.96 ± 0.03) values for a comparable number of FPI. We note however their framework is based on a much more complex pipeline that includes deep-belief networks, Gaussian mixture models, a cascade of R-CNN and a cascade of random forests. When compared to more traditional methods [27], [28] we see that our deep learning approach presents much less FP per image. However, a TPR of 0.7 was not achieved in the BCRP dataset.

TABLE III: Comparison to other methods, including known state-of-the-art for benchmark datasets.

	INbreast [23]		CBIS [24]		BCRP [25]	
	TPR	FPI	TPR	FPI	TPR	FPI
[12]	0.96 $\pm$0.03	1.2	-		0.75	4.8
	0.87 $\pm$0.14	0.8			0.70	4.0
[27]	0.8	2.5	-		-	
[28]	-		-		0.70	8.0
elastic	0.8	1.171	0.8	3.509	0.6	1.864

IV. CONCLUSIONS

During a real-life screening, different views of the same breast can be obtained from the application of different compression forces. This however should not change the outcome of screening. In this paper we aim at augmenting the training data by simulating this phenomena. The extensive experimental results using elastic deformations show that this data augmentation technique can be used to improve performance of CNN models in mass detection. Another strategy that we are currently studying, is to perform elastic deformations at the patch level. Perhaps less intuitive, this strategy does have the potential to generate masses with a finer-level of detail. Additionally, other types of stretching could be considered.

V. ACKNOWLEDGMENTS

This work was funded by project NanoSTIMA, Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics, NORTE-01-0145-FEDER-000016. Project funded by North Portugal Regional Operational Programme (NORTE 2020), under PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87–108, 2015.

[2] N. Howlader, A. Noone, M. Krapcho, D. Miller, K. Bishop, C. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D. Lewis, H. Chen, E. Feuer, and K. Cronin, "SEER Cancer Statistics Review, 1975-2014."

[3] I. Anttinen, M. Pamilo, M. Soiva, and M. Roiha, "Double reading of mammography screening films-one radiologist or two?" *Clinical Radiology*, vol. 48, no. 6, pp. 414–421, 1993.

[4] S. Bessa, I. Domingues, J. S. Cardoso, P. Passarinho, P. Cardoso, V. Rodrigues, and F. Lage, "Normal breast identification in screening mammography: a study on 18 000 images," in *IEEE Proc. Int. Conf. on Bioinformatics and Biomedicine*, 2014.

[5] A. Oliver, J. Freixenet, J. Marti, E. Pérez, J. Pont, E. R. Denton, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," *Medical image analysis*, vol. 14, no. 2, pp. 87–110, 2010.

[6] D. M. Catarious, A. H. Baydush, and C. E. Floyd, "Characterization of difference of gaussian filters in the detection of mammographic regions," *Medical Physics*, vol. 33, no. 11.

[7] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Medical Imaging*, vol. 18, no. 5, pp. 369–378, May 1999.

[8] J. S. Cardoso, I. Domingues, and H. P. Oliveira, "Closed shortest path in the original coordinates with an application to breast cancer," *Int. J. Pattern Recognition Artificial Intelligence*, vol. 29, 2015.

[9] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, and N. Karssemeijer, "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, 2017.

[10] J. Thongkam, G. Xu, and Y. Zhang, "Adaboost algorithm with random forests for predicting breast cancer survivability," in *Neural Networks, 2008*. IEEE, 2008, pp. 3062–3069.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, no. 7553, pp. 436–444, May.

[12] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests," in *Proc. of Int. Conf. of DICTA*. IEEE, 2015, pp. 1–8.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.

[14] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural network design*. Martin Hagan, 2014.

[15] D. C. Pereira, R. P. Ramos, and M. Z. do Nascimento, "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 1, pp. 88 – 101, 2014.

[16] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and cooperation in neural nets*. Springer, 1982, pp. 267–285.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[18] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feed-forward neural networks," in *Proc. Int. Conf.on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[19] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[20] A. Giusti, D. C. Ciresan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," *CoRR*, vol. abs/1302.1700, 2013.

[21] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015.

[22] X. Song, X. Gao, Y. Ding, and Z. Wang, "A handwritten chinese characters recognition method based on sample set expansion and cnn," *2016 3rd Int. Conf. on Systems and Informatics (ICSAI)*, 2016.

[23] 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.

[24] R. S. Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated breast imaging subset of DDSM," *The Cancer Imaging Archive (TCIA)*, 2016.

[25] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proc. 5th Int. Workshop on Digital Mammography*. Medical Physics Publishing, 2000, pp. 212–218.

[26] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[27] E. Kozegar, M. Soryani, B. Minaei, and I. Domingues, "Assessment of a novel mass detection algorithm in mammograms," *Journal of cancer research and therapeutics*, vol. 9, no. 4, p. 592, 2013.

[28] M. Beller, R. Stotzka, T. Müller, and H. Gemmeke, "An example-based system to support the segmentation of stellate lesions," *Bildverarbeitung für die Medizin 2005*, pp. 475–479, 2005.