2017 IEEE 30th International Symposium on Computer-Based Medical Systems

The Effect of Mammogram Preprocessing on Microcalcification Detection with Convolutional Neural Networks

Agnese Marchesi*, Alessandro Bria†, Claudio Marrocco*, Mario Molinara*, Jan-Jurre Mordang‡, Francesco Tortorella*, Nico Karssemeijer‡
*Department of Electrical and Information Engineering, University of Cassino and Southern Latium, Cassino, Italy
Email: {a.marchesi, c.marrocco, m.molinara, tortorella}@unicas.it
†Department of Molecular Medicine, University La Sapienza of Rome, Italy
Email: a.bria@{uniroma1.it,unicas.it}
‡Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
Email: {jan-jurre.mordang, nico.karssemeijer}@radboudumc.nl

***Abstract*—Microcalcifications are an early mammographic indicator of breast cancer. To assist screening radiologists in reading mammograms, machine learning techniques have been developed for the automated detection of microcalcifications. In the last few years, Convolutional Neural Networks (CNNs) have achieved state-of-the-art performance in many computer vision and medical image analysis applications. A key step in CNN-based detection is image preprocessing, including brightness and contrast variations. In this work, we investigate the influence of preprocessing of digital mammograms on the microcalcification detection performance of two CNNs inspired by the popular AlexNet and VGGnet. We tested two preprocessing methods commonly applied to unprocessed raw digital mammograms: (i) the logarithmic transformation adopted by different manufacturers for the presentation of the image to the radiologists; and (ii) the square-root of image intensity that stabilizes the intensity-dependent noise present in the mammogram. Experiments were performed on $1,066$ mammograms acquired with GE Senographe systems. Both preprocessing methods yielded statistically significantly better microcalcification detection performance. Results of the square-root transform were superior to those obtained with the log transform.**

***Keywords*-preprocessing; Convolutional Neural Network; microcalcification detection; digital mammograms.**

I. Introduction

Breast cancer is the first leading cause of cancer death among women worldwide [1]. Early detection increments the survival rate, increases treatment options and helps to improve the life quality of the patient [2]. Digital mammography is the most widely used tool in screening programs for early breast cancer diagnosis. In the last few decades, Computer-Aided Detection (CADe) systems have been developed to assist radiologists in reading the large number of acquired screening mammograms [3]. Several studies have shown that radiologists performance in breast cancer detection can be significantly improved when assisted by CADe systems [4]. Among the lesions that can appear in a mammogram, microcalcifications (MCs) are of particular interest because they represent an early sign of breast cancer [5]. They are calcium deposits of size between 0.1 mm and 1 mm and can be difficult to detect even for an expert radiologist [6]. CADe of MC often builds on supervised learning techniques [7]–[10] that yield powerful binary classifiers able to determine whether a MC is present at a pixel location.

In recent years, Deep Learning techniques grew in popularity because of their outstanding performance in computer vision [11]. Among various methodological variants, Convolutional Neural Networks (CNNs) have received the largest share of research in medical imaging [12], [13] and were also effective for MC detection [14], [15]. A typical CNN architecture is a sequence of feed-forward layers where convolutional filters are interlaced with nonlinear activation functions and pooling. The convolutional layers determine a set of high-level abstract features, whereas the last fully connected layers perform the classification.

An essential step when training the CNNs is image preprocessing. Among preprocessing techniques, those including brightness and contrast variations have shown to be particularly useful to improve the CNN performance. In [16] a preprocessing contrast-extracting layer was used to improve the recognition rate. Jarrett et al. [17] proposed a local contrast normalization layer to normalize the responses across all the features in output from each convolutional layer. This approach has been partially modified in [18] where local response normalization layers are introduced to implement a brightness normalization. Successively, local contrast normalization has been used in [19] as preprocessing step together with global contrast normalization, and also in [20] to normalize brightness and color variations of RGB images.

Preprocessing techniques are commonly applied in digital mammograms [21]–[23], and recently, the effect of contrast enhancement techniques on a CNN has been studied for glaucoma detection [24] and vessels segmentation [25] in retinal fundus images. In this work, we analyze how contrast enhancement influences the performance of two CNNs for

2372-9198/17 $31.00 © 2017 IEEE
DOI 10.1109/CBMS.2017.29

the detection of MCs in digital mammograms. To this end, we tested two transformations commonly applied to unprocessed raw digital mammograms: (i) the logarithmic transformation adopted by different manufacturers for the presentation of the image to the radiologists; and (ii) the square-root of image intensity that stabilizes the intensity-dependent noise present in the mammogram.

II. Dataset

For this study, we collected a database consisting of $1,066$ mammograms acquired with GE Senographe systems (GE, Fairfield, Connecticut, United States) in Radboud University Medical Center (Nijmegen, The Netherlands). All mammograms were acquired with standard clinical settings, and unprocessed raw DICOM 'FOR PROCESSING' images were used in this work. A total of $7,579$ individual MCs were annotated by an experienced reader who marked the center of each MC based on the diagnostic reports. To feed the CNNs, we extracted a dataset of patches of size 12×12 pixels from the mammograms. The patches containing MCs (*positive* samples) were taken by centering the detector window at the groundtruth MC centers, yielding the same number of samples as the individually labeled MCs. The background patches (*negative* samples) were randomly extracted from the remaining regions of the images, totalizing $27,017,503$ samples.

III. Preprocessing

A. Logarithmic transformation

Digital mammography systems frequently apply a transformation to the raw image to make it more suitable for viewing on a computer monitor. These transformations often include a logarithm-like operation [26] that brightens the intensities of the images and enhances the contrast of the lesions with respect to the background tissue. The logarithmic transform of image intensity y is given by:

$$T_{\mathrm{LOG}}(y) = c \log(y+1) \tag{1}$$

where c is a constant used to scale the range of the log function to match the input domain. Denoting with L the number of gray levels of the input image, c must be:

$$c = \frac{L-1}{\log L} \tag{2}$$

so that the transformed image will have intensities in the range $[0, L-1]$ as the input image. The final form of the log transform is then:

$$T_{\mathrm{LOG}}(y) = \frac{L-1}{\log L} \log(y+1) \tag{3}$$

which is plotted in Fig. 1a for $L = 2^{14}$. The histogram and a close-up of a mammogram image after the application of the log transform are shown in Fig. 1b and Fig. 1e, respectively.

B. Square-root transformation

On digital mammograms, the detectability of MCs is hampered by the presence of quantum noise, which affects both radiologists [27] and CADe performance [28]. Quantum noise is an intensity-dependent noise caused by fluctuations in the number of photons hitting the detector. In [29] we proposed a nonparametric square-root transform that equalizes the noise to a constant level:

$$T_{\mathrm{SQR}}(y) = \sqrt{y(L-1)} \tag{4}$$

In addition to noise equalization, this transform also enhances the contrast of MCs. Specifically, the square-root of image intensity belongs to the family of gamma correction transformations in which the corrected pixel intensity is a power function of the original intensity:

$$T(y) = cy^{\gamma} \tag{5}$$

that in our case yields $\gamma = \frac{1}{2}$ and $c = \sqrt{L-1}$. The plot of this function is shown in Fig. 1a for $L = 2^{14}$. Its effect on the histogram of image intensities is to expand the low gray levels (see Fig. 1b), enhancing the contrast of MCs with respect to the surrounding tissue (see Fig. 1d).

IV. Convolutional Neural Networks

A CNN is an ensemble of neurons each featuring several weighted inputs and one output, performing convolution of inputs with weights and transforming the outcome according to a nonlinear activation function. Neurons are arranged in layers and usually share the same weights so as to produce a feature map and reduce the number of parameters. In a typical CNN architecture, convolutional layers are equipped with the Rectified Linear Units (ReLUs) and are intertwined with max-pooling layers. ReLUs apply a nonsaturating activation function $f(x) = \max(0, x)$ which allows the network to easily obtain sparse representations. Max-pooling layers aggregate the outputs of multiple neurons and return the maximum, which results in less training time and lower network complexity. The final decision is made through one or more fully connected layers where each neuron is fed with the outputs of all the neurons of the previous layer. Dropout layers usually follow a fully connected layer to reduce overfitting. The term dropout indicates that, at each training stage, a fixed percentage of outputs coming from the previous layer is ignored in the training of the successive layer.

In this study, we implemented two CNNs inspired by the AlexNet [18] and the VGGnet [30]. The first model is composed by five convolutional and three fully connected layers. Local Response Normalization (LRN) layers follow the first and second convolutional layers, whereas max-pooling layers follow both LRN layers and the last convolutional layer. The ReLU nonlinearity is applied to the output of every convolutional and fully connected layer.

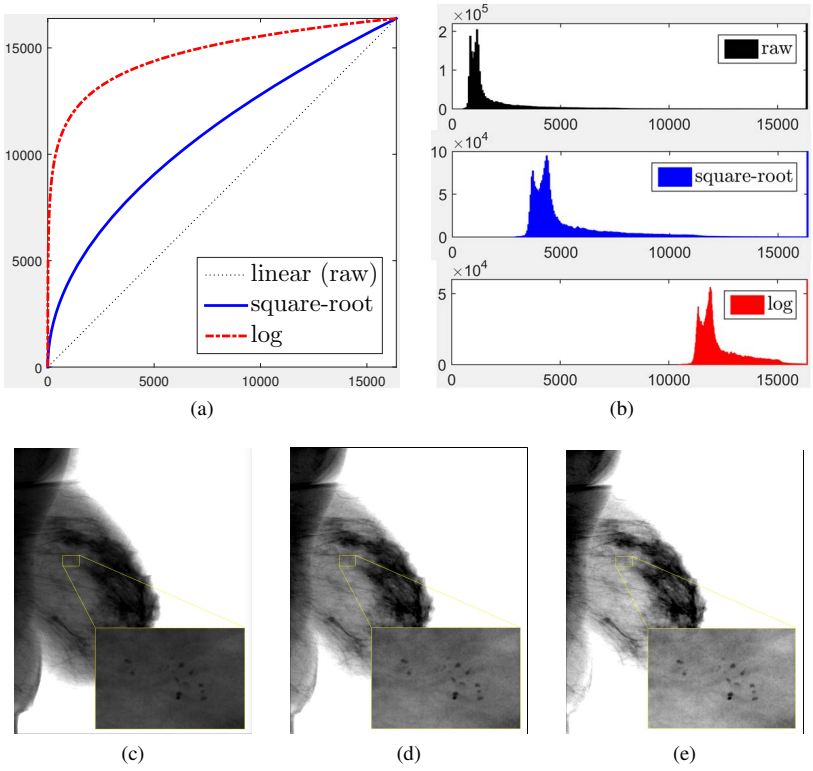

Figure 1. (a) Plots of the gray level transformations used as mammogram preprocessing methods and (b) the corresponding histograms obtained from a mediolateral oblique mammogram originally coded with $L = 16,384$ gray levels (14 bits). The same mammogram is shown in (c) without preprocessing, in (d) after square-root transformation, and in (e) after log transformation. In the close-ups of (c-e) it is shown a cluster of malignant microcalcifications.

The parameters of each layer are reported in Table I. The second model consists of two stacks of two convolutional layers followed by one max-pooling layer. ReLU is used as activation function for each convolutional layer. The final layers are three fully connected layers. The parameters of each layer are shown in Table II.

V. Experiments

We applied the two CNNs to the unprocessed raw mammograms and to the mammograms processed with log and square-root transforms. We used 2-fold cross validation to train and test the networks. In each cross validation step, the CNN was trained on the 50% of the samples and tested on the other 50%. Before training, positive and negative samples were balanced by means of data augmentation using flipping, rotation, and replication. Each network was trained to minimize the Softmax loss function by means of backpropagation and Mini-Batch Stochastic Gradient Descent, with mini-batches of 32 samples. Standardization was applied to the inputs by mean subtraction and normalization to unit variance [31]. Weights of each learning layer were initialized using the algorithm of Glorot and Bengio [32]. The learning rate was set to the initial value of 10^{-3} and decreased during training by a factor of 10 every 6 epochs. Momentum and weight decay were set respectively to 0.9 and $5 \cdot 10^{-4}$. The dropout was performed with a probability of 0.5. For the LRN layers of the AlexNet we set the following parameters: $k = 1$, $n = 5$, $\alpha = 10^{-4}$, and $\beta = 0.75$. The learning was stopped after 30 epochs (1 epoch $= 844,297$ iterations), i.e. when the loss function did not decrease significantly. We used the Caffe framework [33] for the implementation of both networks, and all the experiments were performed on a computer with 2 Intel Xeon e5-2609 processors, 256 GB of RAM and 2 GPU NVIDIA TitanX Pascal.

VI. Results

The CNN-based MC detectors without and with the two preprocessing methods have been evaluated in terms of Receiver Operating Characteristics (ROC) curve by plotting True Positive Rate (TPR) against False Positive Rate (FPR) for a series of thresholds on the CNN output associated to each sample. Furthermore, the mean sensitivity of the ROC curve in the specificity range on a logarithmic scale was calculated and compared. The mean sensitivity is defined as [14]:

$$\overline{S}(a,b) = \frac{1}{ln(b) - ln(a)} \int_a^b \frac{s(f)}{f} df \qquad (6)$$

Table I
ARCHITECTURE OF THE ALEXNET-BASED CNN

Layer	Type	Output size	Kernel Size	Stride	Padding
0	Input	$1 \times 12 \times 12$			
1	Convolutional	$96 \times 12 \times 12$	3×3	1	1
2	ReLU	$96 \times 12 \times 12$			
3	LRN	$96 \times 12 \times 12$			
4	Max pooling	$96 \times 6 \times 6$	2×2	2	1
5	Convolutional	$256 \times 6 \times 6$	3×3	1	1
6	ReLU	$256 \times 6 \times 6$			
7	LRN	$256 \times 6 \times 6$			
8	Max pooling	$256 \times 3 \times 3$	2×2	2	1
9	Convolutional	$384 \times 3 \times 3$	3×3	1	1
10	ReLU	$384 \times 3 \times 3$			
11	Convolutional	$384 \times 3 \times 3$	3×3	1	1
12	ReLU	$384 \times 3 \times 3$			
13	Convolutional	$256 \times 5 \times 5$	3×3	1	1
14	ReLU	$256 \times 5 \times 5$			
15	Max pooling	$256 \times 3 \times 3$	3×3	1	1
16	FC	$1,024$	1×1		
17	ReLU	$1,024$			
18	Dropout	$1,024$			
19	FC	512	1×1		
20	ReLU	512			
21	Dropout	512			
22	FC	2	1×1		

Table II
ARCHITECTURE OF THE VGGNET-BASED CNN

Layer	Type	Output size	Kernel Size	Stride	Padding
0	Input	$1 \times 12 \times 12$			
1	Convolutional	$32 \times 12 \times 12$	3×3	1	1
2	ReLU	$32 \times 12 \times 12$			
3	Convolutional	$32 \times 12 \times 12$	3×3	1	1
4	ReLU	$32 \times 32 \times 12$			
5	Max pooling	$32 \times 6 \times 6$	2×2	2	1
6	Convolutional	$32 \times 6 \times 6$	3×3	1	1
7	ReLU	$32 \times 6 \times 6$			
8	Convolutional	$32 \times 6 \times 6$	3×3	1	1
9	ReLU	$32 \times 6 \times 6$			
10	Max pooling	$32 \times 3 \times 3$	2×2	2	1
11	FC	256	1×1		
12	Dropout	256			
13	FC	256	1×1		
14	Dropout	256			
15	FC	2	1×1		

where a and b are the lower and upper bound of the false positive fraction and were set, respectively, to 10^{-6} and 10^{-1} and $s(f)$ is the sensitivity at the false positive fraction f. Statistical comparisons were performed by means of bootstrapping [34] as in [35]. On the test set, average ROC curves were calculated over 1,000 bootstraps, and are reported in Fig. 2. Additionally, the mean sensitivity was calculated for each bootstrap and p-values were computed for testing significance [36]. The statistical significance level was chosen as $\alpha = 0.05$ but, due to the number of comparisons $m = 3$, we applied the Bonferroni correction [37], so that performance differences were considered statistically significant if $p < 0.01\overline{6}$. Comparative results are reported in Table III. The mean sensitivity obtained on unprocessed raw images was 62.01 and 63.53 for the AlexNet- and the VGGnet-based CNNs, respectively. Results with preprocessing were statistically significantly better than without preprocessing, on average by 15.41 and 16.70 for the AlexNet- and VGGnet-based CNNs, respectively. Results of the square-root transform were on average superior by 1.50 to those of logarithmic transform.

VII. DISCUSSION AND CONCLUSIONS

It has been shown that the performance of MC detection with CNNs can be greatly improved if mammograms are preprocessed in order to enhance the global contrast of the image. Indeed, it is known that the first layers of a CNN automatically learn and extract low-level features. We can suppose that improving the contrast of MCs is beneficial for these layers that capture contrast and spatial information in the salient regions. Consequently, this may positively influence the learning task of the subsequent layers that are aimed at capturing more complex features.

We investigated and experimentally evaluated two different preprocessing approaches. The first is the logarithmic transform, commonly adopted by mammography manufacturers to present the mammogram to the radiologists. The second is the square-root transform which, in addition to contrast enhancement, also acts as noise stabilization across the image. Results of the square-root transform were statistically significantly better than those of the log transform. This suggests that eliminating the noise dependence on the image intensity may have a positive impact on the learning process of the first layers of the CNN.

Future work will be focused on experimenting other contrast enhancement techniques such as CLAHE [38], both as standalone preprocessing and in combination with the noise-equalizing square-root transform. In addition, we will also analyze the impact of mammogram preprocessing on a wider range of CNN architectures.

Table III
COMPARATIVE RESULTS OF MEAN MC DETECTION SENSITIVITY $\overline{S}$ IN THE FPR RANGE $[10^{-6}, 10^{-1}]$ FOR DIFFERENT PREPROCESSING METHODS (RAW = NONE, LOG = LOGARITHMIC, SQR = SQUARE-ROOT). STATISTICALLY SIGNIFICANTLY DIFFERENCES ARE LISTED IN BOLD.

CNN architecture	$\overline{S}_{\mathrm{RAW}}$	$\overline{S}_{\mathrm{LOG}}$	$\overline{S}_{\mathrm{SQR}}$	$\overline{S}_{\mathrm{LOG}} - \overline{S}_{\mathrm{RAW}}$	$\overline{S}_{\mathrm{SQR}} - \overline{S}_{\mathrm{RAW}}$	$\overline{S}_{\mathrm{SQR}} - \overline{S}_{\mathrm{LOG}}$
AlexNet	62.01	72.86	74.65	$+\mathbf{10.85}$ ($p < 0.001$)	$+\mathbf{12.64}$ ($p < 0.001$)	$+\mathbf{1.79}$ ($p < 0.001$)
VGGnet	63.53	77.35	78.55	$+\mathbf{13.82}$ ($p < 0.001$)	$+\mathbf{15.02}$ ($p < 0.001$)	$+\mathbf{1.20}$ ($p < 0.001$)

ACKNOWLEDGMENT

The authors affiliated to the University of Cassino and Southern Latium gratefully acknowledge the support of NVIDIA Corporation for the donation of the Titan X Pascal GPUs used for this research.

REFERENCES

[1] S. McGuire, "World cancer report 2014," 2015.

[2] American Cancer Society, "Cancer facts & figures 2016," 2016.

[3] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.

[4] L. H. Eadie, P. Taylor, and A. P. Gibson, "A systematic review of computer-assisted diagnosis in diagnostic cancer imaging," *European Journal of Radiology*, vol. 81, no. 1, pp. e70 – e76, 2012.

[5] D. B. Kopans, *Breast imaging*. Lippincott Williams & Wilkins, 2007.

[6] H. J. Yoon, B. Zheng, B. Sahiner, and D. P. Chakraborty, "Evaluating computer-aided detection algorithms," *Medical physics*, vol. 34, no. 6, pp. 2024–2038, 2007.

[7] H. Jing, Y. Yang, and R. M. Nishikawa, "Detection of clustered microcalcifications using spatial point process modeling," *Phys. Med. Biol.*, vol. 56, no. 1, pp. 1–17, 2011.

[8] C. Marrocco, M. Molinara, F. Tortorella, P. Rinaldi, L. Bonomo, A. Ferrarotti, C. Aragno, and S. S. lo Moriello, "Detection of cluster of microcalcifications based on watershed segmentation algorithm," in *Computer-Based Medical Systems (CBMS), 2012 25th Int. Symp. on*, pp. 1–5, IEEE, 2012.

[9] A. Bria, C. Marrocco, M. Molinara, and F. Tortorella, "A ranking-based cascade approach for unbalanced data," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 3439–3442, IEEE, 2012.

[10] A. Bria, C. Marrocco, N. Karssemeijer, M. Molinara, and F. Tortorella, "Deep cascade classifiers to detect clusters of microcalcifications," in *International Workshop on Digital Mammography*, pp. 415–422, Springer, 2016.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[13] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[14] J.-J. Mordang, T. Janssen, A. Bria, T. Kooi, A. Gubern-Mérida, and N. Karssemeijer, "Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks," in *International Workshop on Digital Mammography*, pp. 35–42, Springer, 2016.

[15] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, "Discrimination of breast cancer with microcalcifications on mammography by deep learning," *Scientific reports*, vol. 6, 2016.

[16] D. C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 1237, Barcelona, Spain, 2011.

[17] K. Jarrett, K. Kavukcuoglu, Y. LeCun, *et al.*, "What is the best multi-stage architecture for object recognition?," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 2146–2153, IEEE, 2009.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[19] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 3288–3291, IEEE, 2012.

[20] M. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *Proc. of the Int. Conf. on Learning Representation (ICLR)*, 2013.

[21] H. Mirzaalian, M. R. Ahmadzadeh, S. Sadri, and M. Jafari, "Pre-processing algorithms on digital mammograms.," in *MVA*, pp. 118–121, 2007.

[22] I. K. Maitra, S. Nag, and S. K. Bandyopadhyay, "Technique for preprocessing of digital mammogram," *Computer methods and programs in biomedicine*, vol. 107, no. 2, pp. 175–188, 2012.

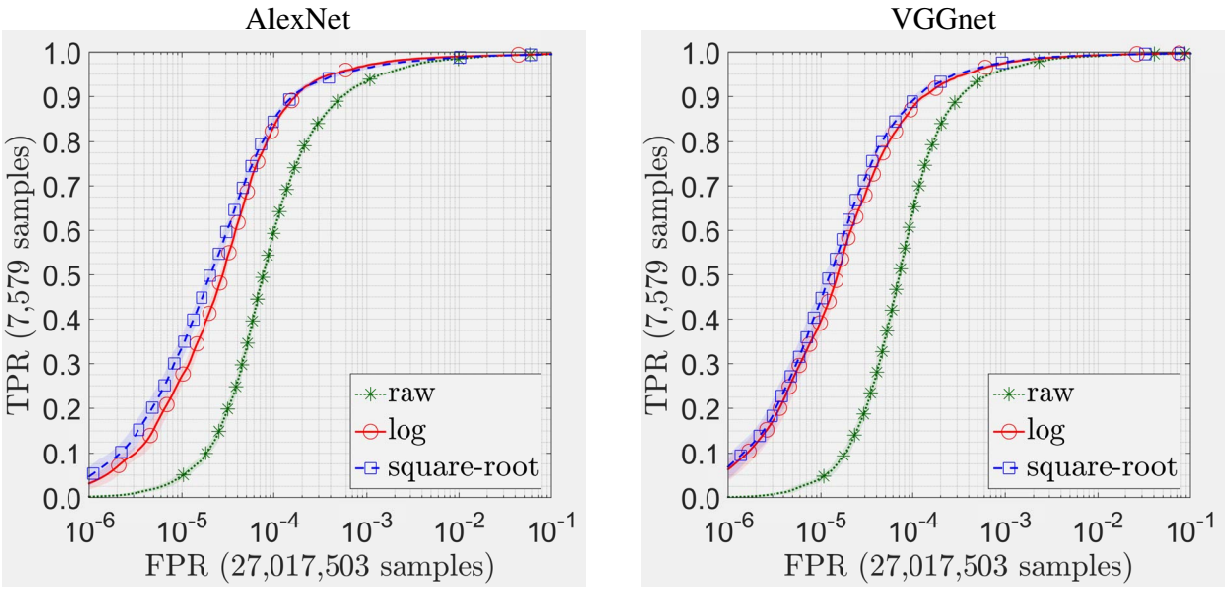

Figure 2. Average ROC curves of the CNN-based detectors obtained from $1,000$ bootstrap iterations. Confidence bands indicate 95% confidence intervals along the TPR axis.

[23] M. Molinara, C. Marrocco, and F. Tortorella, "Automatic segmentation of the pectoral muscle in mediolateral oblique mammograms," in *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th Int. Symp. on*, pp. 506–509, IEEE, 2013.

[24] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *12th International Symposium on Medical Information Processing and Analysis*, vol. 10160, pp. 101600U–101600U–10, 2017.

[25] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2369–2380, 2016.

[26] I. Bankman, *Handbook of medical image processing and analysis*. academic press, 2008.

[27] R. S. Saunders, J. A. Baker, D. M. Delong, J. P. Johnson, and E. Samei, "Does image quality matter? impact of resolution and noise on mammographic task performance," *Medical physics*, vol. 34, no. 10, pp. 3971–3981, 2007.

[28] L. C. S. Romualdo, M. A. C. Vieira, H. Schiabel, N. D. A. Mascarenhas, and L. R. Borges, "Mammographic Image Denoising and Enhancement Using the Anscombe Transformation, Adaptive Wiener Filtering, and the Modulation Transfer Function," *Journal of Digital Imaging*, vol. 26, no. 2, pp. 183–197, 2013.

[29] A. Bria, C. Marrocco, J.-J. Mordang, N. Karssemeijer, M. Molinara, and F. Tortorella, "Lut-qne: Look-up-table quantum noise equalization in digital mammograms," in *International Workshop on Digital Mammography*, pp. 27–34, Springer, 2016.

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

[31] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*, pp. 9–48, Springer, 2012.

[32] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks.," in *Aistats*, vol. 9, pp. 249–256, 2010.

[33] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[34] F. W. Samuelson and N. Petrick, "Comparing image detection algorithms using resampling," in *IEEE Int. Symp. Biomed. Imag.*, pp. 1312–1315, 2006.

[35] A. Bria, C. Marrocco, M. Molinara, and F. Tortorella, "An effective learning strategy for cascaded object detection," *Information Sciences*, vol. 340, pp. 17–26, 2016.

[36] H. Bornefalk and A. B. Hermansson, "On the comparison of FROC curves in mammography CAD systems," *Medical Physics*, vol. 32, no. 2, pp. 412–417, 2005.

[37] O. J. Dunn, "Multiple Comparisons Among Means," *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.

[38] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, pp. 474–485, Academic Press Professional, Inc., 1994.