The Influence of Image Normalization in Mammographic Classification with CNNs

Ana C. Perre
ana.perre@ipcb.pt

Luís A. Alexandre
lfbaa@ubi.pt

Luís C. Freire
luis.freire@estesl.ipl.pt

Faculdade Ciências da Saúde,
Universidade da Beira Interior,
Covilhã, Portugal

Dept. Informática,
Universidade da Beira Interior,
Covilhã, Portugal

Escola Superior de Tecnologia da Saúde de Lisboa,
Instituto Politécnico de Lisboa
Lisboa, Portugal

Abstract

In order to improve the performance of Convolutional Neural Networks (CNN) in the classification of mammographic images, many researchers choose to apply a normalization method during the pre-processing stage. In this work, we aim to assess the impact of six different normalization methods in the classification performance of two CNNs.

Results allow us to concluded that the effect of image normalization in the performance of the CNNs depends of which network is chosen to make the lesion classification; besides, the normalization method that seems to have the most positive impact is the one that subtracts the image mean and divide it by the corresponding standard deviation (best AUC mean with CNN-F = 0.786 and with Caffe = 0.790; best run AUC result was 0.793 with CNN-F and 0.791 with Caffe).

1 Introduction

Mammographic images are interpreted by highly trained radiologists. However, due to the frequent need of analyzing large amounts of images which are produced daily in medical institutions, they may misinterpret between normal and abnormal tissues [1]. Therefore, it is important to develop automatic or semi-automatic computer-assisted tools that can help radiologists in the detection and interpretation of suspicious regions on mammograms [2]. Convolutional Neural Networks (CNN) have been recently successfully used in the medical field for detection and classification of mammographic lesions [1, 2, 3].

To improve the performance of CNNs in this task , many researchers choose to apply a normalization pre-processing method to mammographic images in the pre-processing stage [1], which is justified by the fact that images are obtained with different exposure conditions and are affected by noise and some artifacts[2]. Furthermore, to perform an accurate analysis, it is necessary to achieve an optimal image contrast [2].

In the paper of [4], the authors found that the use, or not, of a pre-processing image normalization method could yield to different performances of the classification tool. Therefore, in this paper we intend to deepen understand such impact by using six different image normalization methods, being the first four methods variations of Global Contrast Normalization (GCN). Therefore, we have: (method 1) subtracting the image mean; (method 2) subtracting the image mean and dividing by the standard deviation; (method 3) Histogram equalization; (method 4) Histogram equalization in combination with method 2. (method 5) and (method 6) used the same GCN applied in method 1 and method 2, respectively, in combination with a local contrast normalization (LCN). Lastly, we tested the classification process on the same images without normalization, which we call "NoNORM" (see fig. 1 and 2).

2 Related Work

Rouhi *et al*. [2] used local area histogram equalization, that stretched the intensity of image pixels to extend the contrast, and then the median filtering, that is a nonlinear operation used to reduce noise ("salt and pepper" and speckle noise). The mammographic images has been normalized and whitened in Jiao *et al*. [1], in the first step the dataset was normalized to the range [0,1] by subtracting them by their mean, and in the second step, they used a method named PCA whitening by dividing the standard deviation of its elements. Arevalo *et al*. [3] applied two normalization types: (1) Global Contrast Normalization, by subtracting the mean of the intensities in the image to each pixel (the mean is calculated per image, not per pixel), and (2) Local Contrast Normalization, that mimics the behavior of the visual cortex and reduces statistical dependencies, which accentuates differences between input features and accelerates gradient-based learning.

3 Material and Methods

The dataset used was the BCDR-FM dataset (Film Mammography Dataset) from Breast Cancer Digital Repository [1]. The downloaded subset, named BCDR-F03 - "Film Mammography Dataset Number 3", comprises 736 grey-level digitized mammograms (426 benign and 310 malign mass lesions) from 344 patients. These are distributed into Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC) views with image size of 720×1168 (width×height) pixels and a bit depth of 8 bits per pixel in TIFF format [3].

In the pre-processing stage we cropped a ROI of 150×150 pixels (following the indications in [3]) using the information of the bounding box of the segmented region, preserving the aspect ratio, even when the lesion's dimensions are bigger than 150×150). When the lesion is next to the border of the image we translate the square crop, changing image coordinates and including the surrounding breast pattern, instead of zero-padding the outer portion of the crop. We have also performed data augmentation by using a combination of flipping and 90, 180 and 270 degrees' rotation transformations.

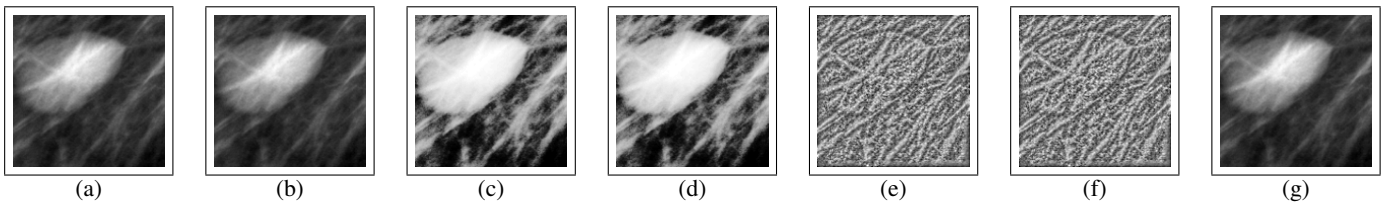
The networks used in this paper were previously used to perform classification in the ImageNet ILSVRC challenge data: the CNN-F (Fast, imagenet-vgg-f) model [5] and the Caffe reference model. The architecture of the CNN-F model consists in 8 learnable layers (5 convolutional layers and 3 fully-connected layers), and the fast processing is guaranteed by the 4 pixel stride in the first convolutional layer [5]. Caffe showed the best classification performance in our previous work and has a complete set of layers that are used for visual tasks such as classification and trains models by the fast and standard stochastic gradient descent algorithm [6]. In order to apply the pre-trained model to our problem, we have adapted the software MatConvNet [7] available for Matlab. Images were divided into 60% for training and 40% for testing, with an input size of 224×224 pixels (that is the size used for MatConvNet) and the parameters' exploration space comprised three fully connected layers, 50 epochs and five learning rate values (1e-2, 1e-3, 1e-4, 5e-2, 5e-3 and 5e-4).

4 Results and Discussion

Table 1 presents the results in terms of minimum and maximum of the images with the different methods of normalization. Note that in the Methods 1 and 3 the values range remains high, and in Methods 2, 4, 5 and 6 the range values are close to zero. Although, for example, the images *a*, *b* and *g* are visually similar, Table 1 shows that the minimum and maximum values are not the same.

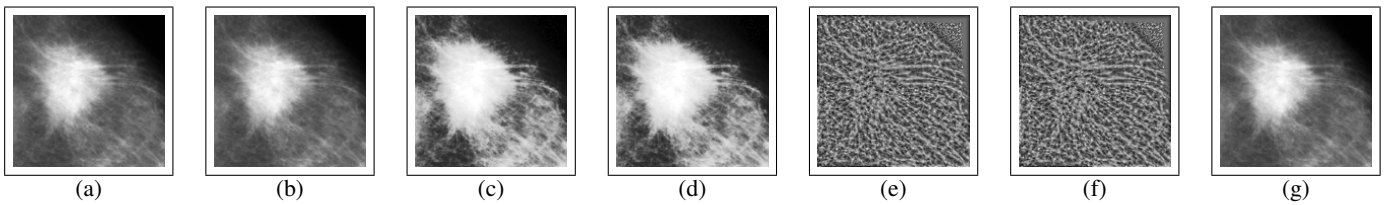
Table 2 shows the results of normalization tests, performed five times, with the CNN-F and Caffe reference model in terms of area under the curve (AUC) mean and standard deviation and the statistic values (*p* value) of comparison between the use or not of the different normalization methods. Note that only for Caffe and Method 2 the AUC value is statistically

[1] http://bcdr.inegi.up.pt



(a) (b) (c) (d) (e) (f) (g)

Figure 1: Examples of the same crop image (benign lesion) with the different normalization methods: (a-f) Methods 1 to 6 (g) NoNORM.



(a) (b) (c) (d) (e) (f) (g)

Figure 2: Examples of the same crop image (malign lesion) with the different normalization methods: (a-f) Methods 1 to 6 (g) NoNORM.

equal (p value = 0.119), for the others always exist a significance difference for better or worse performance with more or less significance value. The best performance achieved with CNN-F was 0.786 and with Caffe 0.790 both using Method 2. However, note that with Caffe, the difference in the results between Method 2 and the one without normalization is not significant (p value = 0.119), which leads to ask if whether important to make the image normalization. However, in the case of CNN-F, the AUC means with Method 2 reveal a significant improvement in the results, from 0.763 to 0.786 (p value = 2.28e-5), with a best run of 0.793 against 0.768.

The histogram normalization does not seems to have a great influence in the performance of the network with Caffe; however with CNN-F, one observes an increase in the classification performance, although the results are slightly lower than those obtained with Method 2.

While GCN seems to have some effect in the network performance, mostly in CNN-F, the LCN does not produce any improvement in the results; the best AUC mean with these methods is 0.742. The AUC values of Methods 5 and 6 are similar, which may be due to the fact both methods have the same minimum and maximum values.

Method	Benign lesion (Min/Max)	Malign lesion (Min/Max)
Met. 1	-49.68 / 111.32	-93.86 / 137.14
Met. 2	-1.52 / 3.41	-1.82 / 2.66
Met. 3	-127.35 / 127.65	-127.48 / 127.52
Met. 4	-1.70 / 1.70	-1.70 / 1.70
Met. 5	-3.07 / 2.69	-2.88 / 3.60
Met. 6	-3.07 / 2.69	-2.88 / 3.60
NoNORM	82 / 243	4 / 235

Table 1: Examples of minimum and maximum values of benign and malign images, presented in figure 1 and 2, after the different normalization methods.

Method	AUC Mean/Std CNN-F	p-value	AUC Mean/Std Caffe	p-value
Met. 1	0.767 / 0.003	0.0489	0.779 / 0.001	1.00e-6
Met. 2	0.786 / 0.005	1.13e-4	0.790 / 0.002	0.119
Met. 3	0.785 / 0.002	1.89e-6	0.781 / 0.003	1.89e-6
Met. 4	0.785 / 0.002	1.08e-6	0.782 / 0.003	0.004
Met. 5	0.730 / 0.003	1.17e-7	0.742 / 2.0e-4	1.03e-17
Met. 6	0.729 / 0.003	4.37e-8	0.741 / 4.0e-4	2.17e-14
NoNORM	0.763 / 0.003	-	0.789 / 2.0e-4	-

Table 2: Results of normalization tests with the CNN-F and Caffe reference model (AUC mean and standard deviation) and statistic values of comparison between the use or not of the different normalization methods (p value).

5 Conclusions

The effect of image normalization in the performance of the CNNs depends of which network is chosen to make the lesion classification. We have seen from the results that, for Caffe, the image normalization is not so important as much as for CNN-F. The method of image normalization that seems to have a bigger impact in the classification performance is the one that subtracts the image mean and divide by the standard deviation (Method 2). The use of LCN is associated with the worst results, which leads to believe that is not a good way to obtain a better CNN performance. The current study was made using scanned images; as future work, we intend to apply these methods to digital images, that are actually the most used in the medical field, with the aim of increasing the classification performance.

6 Acknowledgments

This work was supported by FCT - Fundação para a Ciência e a Tecnologia, through the UID/EEA/50008/2013 Project. The GTX Titan X used in this research was donated by the NVIDIA Corporation. The database used in this work was a courtesy of MA Guevara and coauthors, Breast Cancer Digital Repository Consortium.

References

[1] Zhicheng Jiao, Xinbo Gao, Ying Wang, and Jie Li. A deep feature based framework for breast masses classification. *Neurocomputing*, 197:221–231, jul 2016.

[2] Rahimeh Rouhi, Mehdi Jafari, Shohreh Kasaei, and Peiman Keshavarzian. Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Systems with Applications*, 42(3):990–1002, 2015.

[3] John Arevalo, Fabio A González, Raúl Ramos-Pollán, Jose L Oliveira, and Miguel Angel Guevara Lopez. Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 127:248–257, apr 2016.

[4] Ana Perre, Luis Alexandre, and Luis Freire. VI ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, VipIMAGE. *Lesion Classification in Mammograms Using Convolutional Neural Networks and Transfer Learning*, Oct 18-20 2017.

[5] Simonyan K. Vedaldi A. Zisserman A. Chatfield, K. Best Scientific Paper Award Return of the Devil in the Details: Delving Deep into Convolutional Nets. *British Machine Vision Conference*, 2014.

[6] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[7] Lenc K. Vedaldi, A. MatConvNet – Convolutional Neural Networks for MATLAB. *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.