Assessing the Potential of Multi-view approaches in Breast Cancer Mass Detection

Eduardo Castro[1,2]
https://eduardo-castro.github.io/

José Costa Pereira[1,2,3]
jose.c.pereira@inesctec.pt

Jaime S. Cardoso[1,2]
jaime.cardoso@inesctec.pt

[1] Faculty of Engineering of the University of Porto
University of Porto
Porto, Portugal

[2] Centre for Telecommunications and Multimedia
INESC TEC
Porto, Portugal

[3] Noah's Ark Lab
Huawei
London, UK

Abstract

In the mammography exam, two views with complementary information are obtained for each breast. The state-of-the-art algorithms used in this context do not fully leverage this complementary aspect of the exam. In this work we show the potential of multiview approaches to the problem of lesion detection. For this we compare two models, one which is trained to classify between lesion and non-lesion patches and the second which, given a patch of the lesion in one view ranks candidates of that same lesion on the other view. The second model outperforms the first, showing the potential of using information from one view to guide decision for the other.

1 Introduction

Worldwide, breast cancer is the most frequently diagnosed and most lethal form of cancer in women [3]. The cumulative risk of developing the disease before the age of 75 is a little over 5%. Early diagnosis is vital in addressing this disease as it frequently translates into a better prognosis and allows more treatment options such as breast-conserving surgery [2]. Due to this, different countries implemented screening programs to anticipate detection. Screening mammography is the most commonly used imaging exam in this context and has been shown to decrease mortality [1]. Recent works have focused on developing more accurate Computer-Aided Diagnosis (CAD) algorithms [6]. The developments in deep learning in recent years and, consequently, improved image recognition models have fueled and shaped these efforts [4].

The detection of breast cancer in the screening mammography exam consists of finding lesions, often subtle, indicative of the disease. For each breast, two images are obtained for different projections (views) of the breast, which complement each other information-wise. Often, radiologists observe the same lesion in both views before making a decision.

The state-of-the-art algorithms for breast cancer screening do not integrate the information at the lesion level. They fuse the knowledge of the two views either by averaging the "diagnosis" or aggregating image-level features. Lesion level integration has the potential to improve accuracy and the interpretability of the results returned by the algorithm. This work is a starting point in this direction. We show that a lesion's information in one view is useful to detect the lesion on the other view.

2 Methods

2.1 Baseline: Image Classification with CNNs

Convolutional Neural Networks (CNNs) are the most common type of neural network in vision applications. In recent years researchers have adopted these models in the context of Computed Aided Diagnosis tools for Breast Cancer screening. Image classification is commonly done by minimization of the cross-entropy loss function on a training set:

$$\mathcal{L}_H = \sum_{c=1}^{M} y_c \log(p_c) \tag{1}$$

where $y_c \in \{0,1\}$ is 1 if the label of the image is class c, and p_c is the probability assigned by the model that the image belongs to class c.

2.2 Multiview: An approach based on the Triplet Loss

In this work, we considered an alternative setting for image classification in which the patch corresponding to the lesion in the other view (anchor) is given. In this context, the model can obtain information on what the lesion might look like. The triplet loss [7] can thus be used to train a model that tells us if there is a correspondence between the anchor and the candidate. This loss function is given by:

$$\mathcal{L}_T = \max(d(a,p) - d(a,n) + margin, 0) \tag{2}$$

where a, p, n are feature representations for the *anchor*, the *positive* and the *negative* images and d is a measure of distance (euclidean norm in the case of this work). The minimization of this loss function leads to the desirable case where: $d(a,p) < d(a,n) + margin$. Thus, from all candidates it is expected that the patch that minimizes $d(a,x)$ is the correct one.

2.3 Hybrid: Aggregating the two decisions

The two proposed models are conceptually different. While the baseline learns to discriminate between positive and negative patches, the multiview approach relies on the anchor. As such, the information they base their decisions on may be complementary.

A third option is to use a hybrid strategy that relies on the decision of the two models. Here we propose a simple rule in which the final score, which ranks the candidates, is obtained with the **baseline** model plus a Δ if that candidate is the preferred one for the **multiview** model.

3 Experiments and Discussion

Due to its size and accessibility, CBIS-DDSM [5] is the leading publicly available dataset for developing breast cancer screening algorithms. This collection is an updated and standardized version of DDSM and contains approximately 10k images. Each finding in the dataset is associated with a segmentation mask and its pathology (malign or benign). Images were obtained from scanned film mammography. In this work, a subset of this data with around 1200 images was used.

The data was split at the patient level into three sets: train (70%), validation (10%), and test (20%). Positive patches were taken centered on each lesion's mask. A custom lesion detector was employed to obtain five false negatives per image in the dataset to serve as additional candidates. All patches were resized to 64×64.

A custom architecture was used for all experiments with eight convolutional and two fully-connected layers. The models were trained for around 80k iterations with a starting learning rate of 0.01, which was decreased one time by a factor of 10, using stochastic gradient descend with momentum, with a batch size of 32. Batch normalization and weight decay were used. Each experiment was repeated five times.

Each method's accuracy was computed by first ranking the candidates and then selecting their top choice for each image. This selection is considered correct if it corresponds to the lesion and incorrect otherwise. The top candidate for the **baseline** model was the one that maximized the probability of being positive while for the **multiview** model, the one that minimized the distance to the *anchor*. The Δ for the hybrid strategy was 0.25. Results are shown on table 1. In Figure 2, the sensitivity per average number of false positives is shown for the three models.

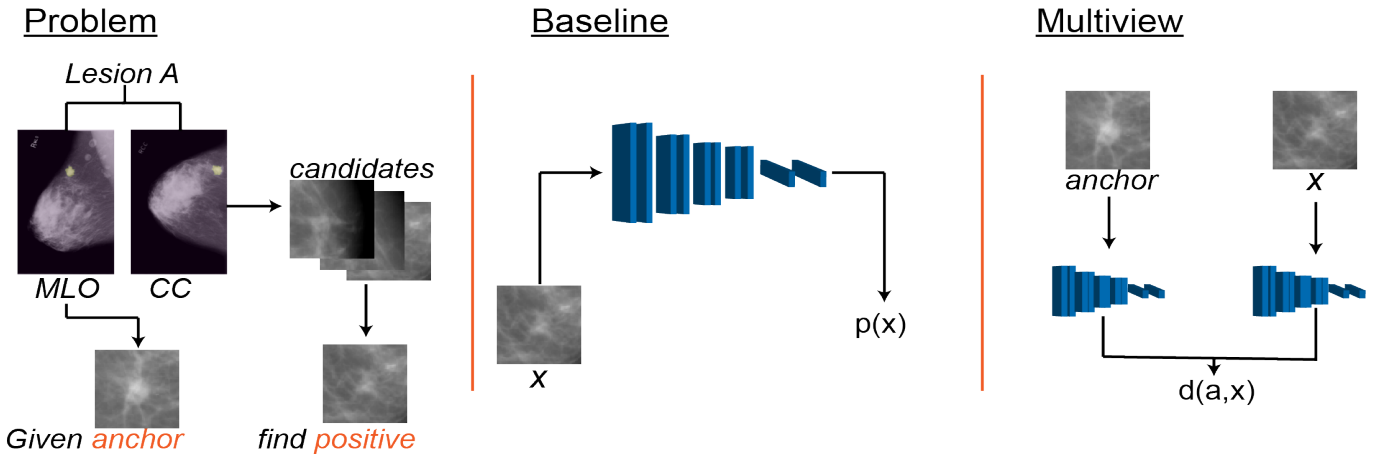

Figure 1: Diagram for the problem formulation and strategies followed in this work.

Table 1: Test set accuracy for each method.

Method	Baseline	Multiview	Hybrid
Accuracy (%)	76.13 ± 1.64	80.4 ± 0.6	82.02 ± 1.62

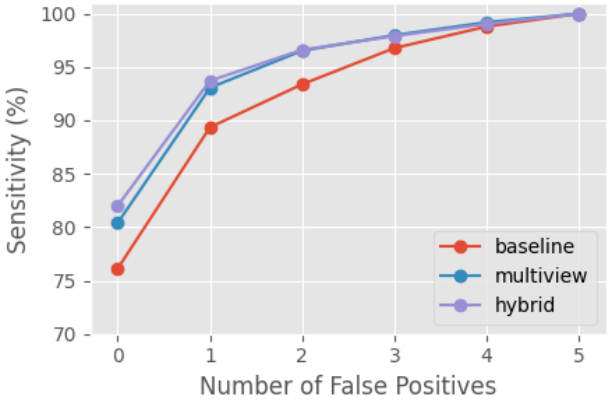

Figure 2: Sensitivity per false positive.

As shown, the **multiview** approach outperforms the baseline, showing that the appearance of the anchor image can be used to enhance the detection of the same lesion on the other view. When combined in the hybrid strategy, the two models can enhance one another, suggesting that their information is complementary.

Even though the results show the importance of using a **multiview** approach at the lesion detection level, future research is needed on how to integrate the two losses together in a single training scheme. Also, it is desirable to have an automatic algorithm that does not require an *anchor*. In this context, it is yet to be shown the value of a **multiview** approach when there are only candidates for view, rather than a "known" positive.

4 Conclusions

Breast cancer is a considerable burden on patients worldwide. The development of more accurate CAD systems can help reduce this burden through earlier and more accurate diagnosis. The development of CNNs has allowed an increase in accuracy. However, current methods could be further improved by better integrating the information of the two views. This work demonstrates this potential by showing that a model that uses the opposite view's appearance can outperform a naive baseline in lesion detection. Future research should focus on how to translate this potential to a more realistic/less controlled scenario.

5 Acknowledgements

The project TAMI - Transparent Artificial Medical Intelligence (NORTE-01-0247-FEDER-045905) leading to this work is co-financed by ERDF - European Regional Fund through the Operational Program for Competitiveness and Internationalisation - COMPETE 2020, the North Portugal Regional Operational Program - NORTE 2020 and by the Portuguese Foundation for Science and Technology - FCT under the CMU - Portugal International Partnership. This work is also financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within PhD grant number *SFRH/BD/136274/2018*. The authors would also like to aknowledge NVIDIA for their generous donation of a TitanX gpu.

References

[1] Donald A. Berry, Kathleen A. Cronin, Sylvia K. Plevritis, Dennis G. Fryback, Lauren Clarke, Marvin Zelen, Jeanne S. Mandelblatt, Andrei Y. Yakovlev, J. Dik F. Habbema, and Eric J. Feuer. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine*, 353(17):1784–1792, 2005. doi: 10.1056/NEJMoa050518. PMID: 16251534.

[2] Carol E. DeSantis, Jiemin Ma, Mia M. Gaudet, Lisa A. Newman, Kimberly D. Miller, Ann Goding Sauer, Ahmedin Jemal, and Rebecca L. Siegel. Breast cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(6):438–451, 2019. doi: 10.3322/caac.21583.

[3] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D.M. Parkin, M. Piñeros, A. Znaor, and F. Bray. Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods. *International Journal of Cancer*, 144(8):1941–1953, 2019. doi: 10.1002/ijc.31937.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 0028-0836. doi: 10.1038/nature14539.

[5] Rebecca Lee, Francisco Gimenez, Assaf Hoogi, Kanae Miyake, Mia Gorovoy, and Daniel Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*, 4:170177, 12 2017. doi: 10.1038/sdata.2017.177.

[6] Thomas Schaffter, Diana S. M. Buist, Christoph I. Lee, Yaroslav Nikulin, Dezső Ribli, Yuanfang Guan, William Lotter, Zequn Jie, Hao Du, Sijia Wang, Jiashi Feng, Mengling Feng, Hyo-Eun Kim, Francisco Albiol, Alberto Albiol, Stephen Morrell, Zbigniew Wojna, Mehmet Eren Ahsen, Umar Asif, Antonio Jimeno Yepes, Shivanthan Yohanandan, Simona Rabinovici-Cohen, Darvin Yi, Bruce Hoff, Thomas Yu, Elias Chaibub Neto, Daniel L. Rubin, Peter Lindholm, Laurie R. Margolies, Russell Bailey McBride, Joseph H. Rothstein, Weiva Sieh, Rami Ben-Ari, Stefan Harrer, Andrew Trister, Stephen Friend, Thea Norman, Berkman Sahiner, Fredrik Strand, Justin Guinney, Gustavo Stolovitzky, , and the DM DREAM Consortium. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Network Open*, 3(3):e200265–e200265, 03 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.0265.

[7] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

*/