Mass Detection Using Deep Convolutional Neural Network with Java

Basel Alyafi, Zafar Toshpulatov, Fahad Khalid

Abstract—**In this work, transfer learning is exploited to finely tune part of pre-trained Deep Convolutional Neural Networks so they are able to classify mammographic windows (454 X 454 pixels) as benign or malignant. Another approach is to tune from scratch well-known NN architectures; namely: AlexNet and LeNet. Both methods used the INbreast data set as input to the network after applying pre-processing steps. Transfer learning showed quit better results than that from ordinary tuning.**

Keywords*—*Mammography, mass, CAD systems, fine tuning, Neural Networks, Deep Convolutional Neural Networks.

I. Introduction

In 2018, it is estimated to have 266,120 new cases of women invasive breast cancer in addition to 63,960 non-invasive cases [1]. After cardiovascular diseases, breast cancer is the leading tumor among women in Italy [2], whereas in the European Union it is responsible for one in every six deaths from cancer in women [3]. In order to control this alarming mortality rate associated with breast cancer, mass screening is recommended by the medical community world wide. Mammography is a widely-used X-ray imaging modality for breast cancer screening as it has the capability to detect various types of unusual lesions such as masses and micro-calcification. Even though this is the first step towards cancer identification, the problem to be highlighted is the manual detection of the mass, hence leading to many false positives along with miss matches. For the purpose of improving the detection accuracy and assisting the physicians and radiologists in this tedious task, Computer Aided Diagnosis (CAD) systems have been developed. The CAD technology acts as a Second opinion which has managed to reduce the workload on specialists and improved the accuracy. Studies on the effectiveness of the use of Computer-Aided Diagnosis (CAD) systems as second opinion systems show that they can indeed help junior radiologists by increasing their sensitivity from 62% to 85% and experienced radiologists by increasing their sensitivity from 77% to 85% [4]. In mammography, a mass is a lesion, which could be projected in two different directions. The malignancy of this lesion is recognized by its shape, asymmetry and contour [5].

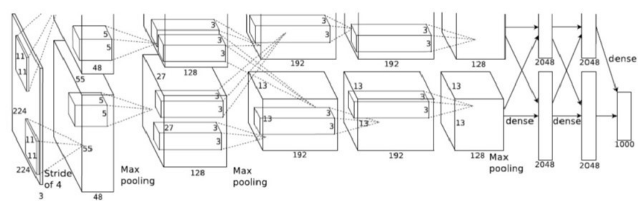
A. Problem description

To reach the goal of building a CAD system for detecting mammogram masses, Neural networks are to be trained with a decent training set. 410 digital mammograms have been taken from INbreast data set (107 with masses, 303 without masses)[6]. Those images are going to be windowed and those windows will be used to train the network. After the system fits (to some limit) the input data, it is supposed to generalize and predict the malignancy of unseen windows taken from new mammograms with reported accuracy expectations.

B. Background

1) Alex Net: AlexNet is a Deep Convolutional Neural Network for image classification that won the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, AlexNet has 8 layers. The first 5 are convolutional and the last 3 are fully connected layers. In between we also have some layers called pooling and activation. compared to 26.2% achieved by the second-best entry.

Fig. 1: AlexNet architecture



2) LeNet: The LeNet architecture is the first architecture for Convolutional Neural Networks [7]. The LeNet architecture is straightforward and small, making it perfect for teaching the basics of CNNs, it can even run on the CPU. The LeNet architecture consists of two sets of convolutional, activation, and pooling layers, followed by a fully-connected layer, activation and softmax classifier.

3) Transfer Learning: Transfer learning also known as inductive learning is a machine learning research technique that focuses on applying the knowledge gained while solving one problem to solving another related problem [8]. In practice, its not so common to train an entire convolutional Neural Network (CNN) from scratch (with random initialization), since its not an easy task to acquire a large enough data set to train a neural network. However, it is common to pre-train a Convolutional Network on a very large data set, ImageNet, which contains 1.2 million images with over 1000 categories and then use the CNN either as an initialization or a fixed feature extractor for the task of interest. It is effective to pre-train the Deep CNN (DCNN) by using very large data set, and then use the DCNN as a feature extractor for other task of which the amount of training data is too small. In this work, transfer learning by fine-tuning strategy was used to identify mass regions in mammogram

images.
Convolutional neural networks are now capable of outperforming humans on some computer vision tasks, such as classifying images. A competition-winning model for this task was the Visual Geometric Group (VGG) model by researchers at Oxford. The main contribution of VGG is to show that classification/localization accuracy can be improved by increasing the depth of CNN in spite of using small receptive fields in the layers. VGG network is characterized by its simplicity, using only 33 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier. It is currently the most preferred choice in the community for extracting features from images [9]. However, VGGNet consists of 140 million parameters, which can be a bit challenging to handle.

II. Data Preparation

Preparation of data is one of the most important parts in DCNN, obtaining an optimized result depends on relevant and smart collection of images. We used publicly available INbreast data containing 410 mammogram images. The aim was to generate 3 types of data sets depending on the quantity of images, hence, small, medium, and large data sets were created by cropping the region of interest from the mammogram. Approximately 2000 images for small data set, 6000 images for medium and 20000 for the large data set. Each data set consists of a training and test set, divided 30% for test and 70% for training set. The training and test set are further divided into positive and negative image windows, where positive window contains the mass area and negative window is the normal area.

To clearly understand the following process, it is essential to keep the following definitions in mind:

- mask image: is a binary image with white breast pixels and black pixels everywhere else, see Fig 2(a).
- ground truth image: is a binary image with white mass pixels and black pixels everywhere else, see Fig 2(b).
- mammogram image: is the original gray-scale X-ray image, see Fig 2(c).
- overlay image: is a gray image where the mass has been surrounded by a yellow contour, see Fig 2(d).

The positive and negative windows were cropped from the original mammograms images using a window size of 454 x 454 pixels. All mammogram images were scanned by moving the window from left to right and top to bottom. Different strides, positive and negative, were adopted for moving the window for small, medium, and large data sets depending on the amount of windows needed. The scanning process starts from the 10 x 10 pixels rather then the 0 x 0 pixel position, this approach was adapted as some images have 10 pixels black tiling around, and due to the fact that the negative stride is significantly large (in comparison to positive) so there is a chance of skipping the complete breast if breast size is small, otherwise. Following is the description of the method step by step:

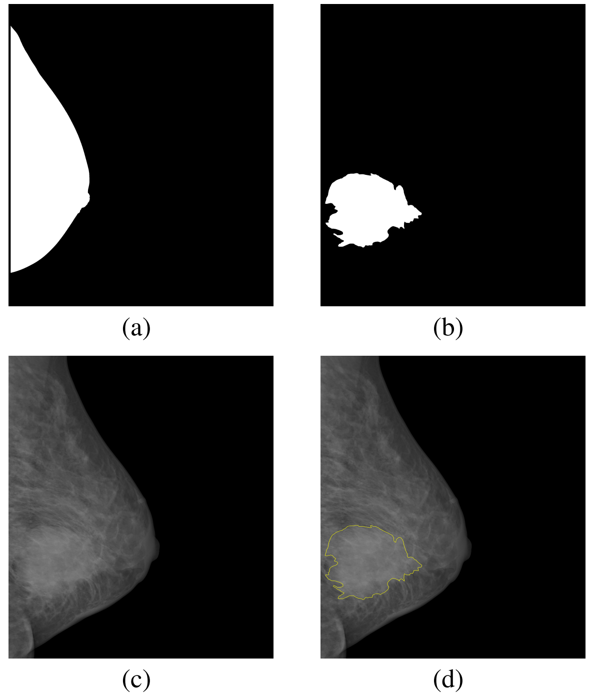


Fig. 2: a) mask image, b) ground truth image, c) mammogram image, d) overlay image

A. split images into massy and normal mammograms

This is done by comparing the ground truth images names with mammograms names (when there is a match, it means a massy image). Both groups of images with mass and without mass are divided into 70% for training set and 30% for test. Then images from both classes are merged to form one train set and one test set, see Fig 3.

B. moving window and data base size reduction

The breast area of each image is found by comparing pixels with the corresponding mask image (see Figure 2 (a) for example), while the ground truth is used to detect the mass, e.g, Figure 2 (b). To consider a window as negative, it should be free of background and mass pixels, while, a window is considered positive if the 15×15 square at the center of the window is completely from mass, See Algorithm 1.

After applying the moving window, 12000 positive and 14500 negative windows were generated. In order to reduce those amounts, the desired number of images were saved in a different folder . Assume maximum n windows are required from each mammogram image, then, if $N > n$ windows have been generated from one mammogram, one window is saved every $ceiling(N/n)$ windows. On the other hand, if the number of windows acquired from a mammogram is $m <= n$, all of them are saved. This was done to maintain the maximum Independence (minimum overlapping) between the acquired windows and to ensure that non of the mammograms has been ignored.

Algorithm 1 saving positive and negative windows mechanism

```
position = top left
while window fits inside the breast do
    if window is from massy image then
        Stride = Pos stride
        if window with no mass then
            save neg every NegStride/PosStride
        else
            if window with mass centered then
                save in pos folder
            else
                ignore
            end if
        end if
    else
        Stride = Neg stride
        save in neg folder
    end if
    position += Stride
end while
```

III. Environment

To better understand the settings under which the work has been done, the following tables show most important setup properties. The vast majority of experiments has been carried out on machine#1, see TABLE I. While on machine#2, only relatively 'low-cost'experiments were conducted. For Data preparation, Open CV in C++ version 2.4.13, was used for the sake of speed and high performance.

PROPERTY	MACHINE#1	MACHINE#2
CPU	1, 8 cores	1, 4 cores
RAM	256GB	32GB
GPU	2 TitanXp ,12GB onboard	1 TitanXp, 12GB onboard
CUDA Version	8.0.61	9.0.176
CUDNN Version	5110	7005
JAVA Version	1.8.0_171	1.8.0_171
Apache Maven Version	3.3.9	3.3.9
Linux Version	4.4.0-127-generic, amd64	4.13.0-41-generic, amd64

TABLE I: Environemnt Properties

IV. Proposed Methodology

The Deep Learning for Java examples that were used are:

- Convolution
  - Animal classification example modified to Mammogram Classification using LeNet with data augmentation (4 X flip, 2 X warp).
- Transfer Learning:
  In Transfer learning, see section I-B3, VGG16 architecture was mainly used with weight initialization (DISTRUBUTION) and activation function (SOFTMAX).
  - EditLastLayerOthersFrozen, modified to include validation step each ten iterations and the ability to accept argument list (epochs number, train folder path, test folder path, model saving path).
  - EditAtBottleneckOthersFrozen, modified to support validation each ten iterations and to accept arguments list to reduce needed building times.

For Transfer Learning architectures, the model with best accuracy is saved through consecutive iterations.

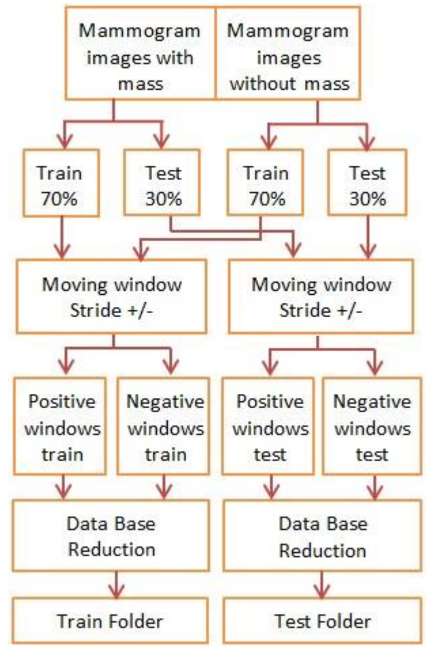


Fig. 3: Windows extraction block diagram

A. Parameters tuning

The following hyper parameters were considered:

- Epochs number: number of epochs was modified to improve the performance (Accuracy metric is considered only due to balanced classes).
- Batch size (which is the number of windows shown to the network each iteration) was changed according to the size of the data set used (small, medium, or large, discussed in section II).
- Validation portion and train test portion (train size / data set size): those were fixed during all the experiments to 80% and 20% respectively, except for the EditAtBottleneckOthersFrozen example where the validation percentage was set to 10% to reduce the processing burden.
- Feature extraction layer: for EditLastLayerOthersFrozen, the network was frozen up until the second last layer (fc2) and the last layer was tuned. Whilst, for EditAtBottleneckOthersFrozen, the network was frozen until the block_5 pool.
- Activation function: due to the wide variety of activation functions, some of them only were considered to evaluate the model, namely: LEAKYRELU and TANH ,which has been used for Bottleneck tuning and it showed better results than LEAKYRELU.

- Weights initialization: XAVIER and RELU showed a good performance, XAVIER was used for Bottleneck to improve the accuracy.

V. Results

After running dozens of experiments, best results are given in tables below, see Tables II, III, IV. First of all, the EditLastLayerOthersFrozen gave the highest Accuracy among other architectures with more than 91.9% with 7 epochs (large data set). Secondly, LeNet showed the Second best Accuracy after EditLastLayerOthersFrozen, with Accuracy 82.8% with 50 epochs (Large data set). For the Bottleneck, and because the frozen part is larger than that in EditLastLayer and LeNet, the processing time was remarkably longer and the required memory was significantly larger which caused many failures to allocate the needed memory space during experiments. One of the ways to overcome that obstacle was to decrease the batch size, so the needed memory space for every iteration. One last thing to mention is that some experiments were conducted on AlexNet, but unfortunately, results were unacceptable even after some parameter tuning.

TABLE II: EditLastLayerOthersFrozen

Theme	EditLastLayerOthersFrozen		
parameters\ data set	small	Medium	Large
Accuracy	0.8389	0.8828	0.9196
Precision	0.8399	0.8828	0.9196
Recall	0.839	0.8828	0.9196
F1 score	0.8427	0.8829	0.9197
parameters(batch size, train perc., epochs)	15-80%-5	15-80%-7	15-80%-7
AUC	0.8975	0.9422	0.9654

TABLE III: LeNet classification

Theme	Lenet classification		
parameters\ data set	small	Medium	Large
Accuracy	0.8281	0.8125	0.8281
Precision	0.8363	0.8137	0.831
Recall	0.8281	0.8125	0.8281
F1 score	0.8136	0.8065	0.8197
parameters(batch size, train perc., epochs)	32-80%-50	128-80%-50	128-80%-50

TABLE IV: EditAtBottleneckOthersFrozen

Theme	EditAtBottleNeckOthersFrozen		
parameters\ dataset	small	Medium	Large
Accuracy	0.5549	0.5011	0.5003
Precision	0.5943	0.5011	0.5003
Recall	0.5549	0.5	0.5
F1 score	0.6636	0.6676	0.6669
parameters(batch size, train perc., epochs)	8-80%-15	1-80%-15	1-80%-15

VI. Conclusion

To sum up our work and encourage future research on mass detection in light of the conducted experiments, transfer learning approach has given evidence of being used as an appropriate method for comparatively higher accuracies. Another important factor in the performance of our experiments was the preparation of valid and accurate data sets, dividing them into test and train while making sure there was no overlapping or repetition of images in the data sets. An additional keynote to take away for best accuracies larger data set is preferred but this may vary depending on factors such as the number of training features, in our case as the features were more then just a few, large data set proved to be useful. Parameter alterations and fine tuning is also an important factor in this work.

References

[1] "U.S. breast cancer statistics." [Online]. Available: http://www.breastcancer.org/symptoms/understand_bc/statistics

[2] *Italian National Institute of Statistics (2002).* Annuario Statistico Italiano, 2002.

[3] L. Eurostat, *Health statistics Atlas on mortality in the European Union*, ser. International series of monographs on physics. Luxembourg: Office for Official Publications of the European Communities, 2009.

[4] Balleyguier, Corinne, K. K, F. J, and M. S, "Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist?"

[5] M. M. F. F. P. C. H. Berment, V. Becette, "Masses in mammography: What are the underlying anatomopathological lesions?" *Journal de Radiologie Diagnostique et Interventionnelle*, vol. 95, pp. 126–135.

[6] I. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: Towards a full field digital mammographic database," *Academic Radiology*, vol. 19, pp. 236–248, 2012. [Online]. Available: http://www.inescporto.pt/ jsc/publications/journals/2012IMoreiraAcRadiology.pdf

[7] Y. Lecun, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE, pages=*.

[8] J. West, D. Ventura, and S. Warnick, "Spring research presentation: A theoretical foundation for inductive transfer."

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556