Deep Learning methods for Quality Check Algorithms in Mammography

Research Internship (April-October 2019)

Author: Baptiste Doyen
Supervisors: Pierre Fillard (*CSO*)
Yaroslav Nikulin (*Sr. Research Scientist*)

Therapixel - CentraleSupélec/MVA

6th September 2019 - Paris Saclay campus

Context: Breast Cancer, AI & Quality check

The most diagnosed and lethal cancer for women with **1/4** of total cancers and **627K** deaths worldwide[9]

- Recurrent screenings
- CAD assistance
- New generation with AI
- AI require both High Volume and High Quality Data

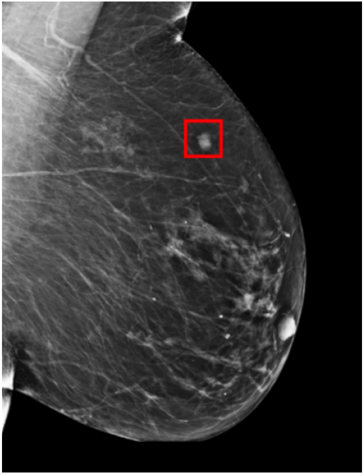


Figure 1: A malignant lesion on a Mammography

Quality Check

A crucial step for Cancer Detection algorithms.
Traditionnally performed manually but today automatized

Two main steps:

- **Pre-Quality Check:** separate screening images from non-screening images
  (global features of the image)
- **Geometry Check:** ensure that breast geometry on the mammography is correct
  (more local and constrained features of the image)

Pre-Quality Check

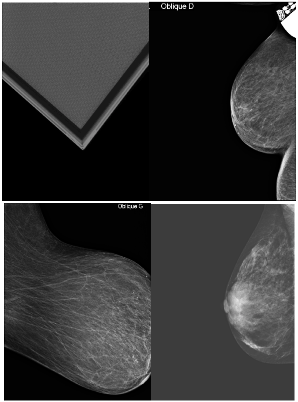


Figure 2: Non-Screening

VS.

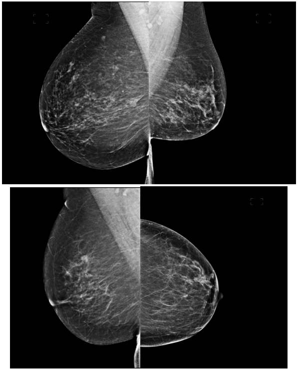


Figure 3: Screening

Geometry Check

The goal is to extract feature points from the image and check if they comply with set of rules S_{geom}

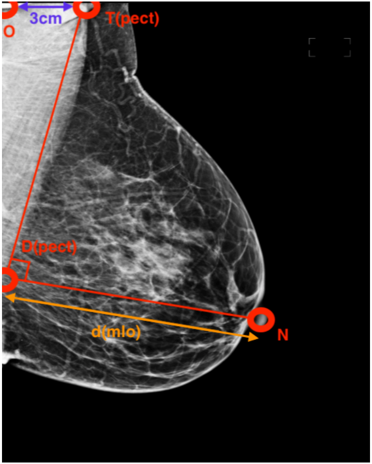


Figure 4: MLO view

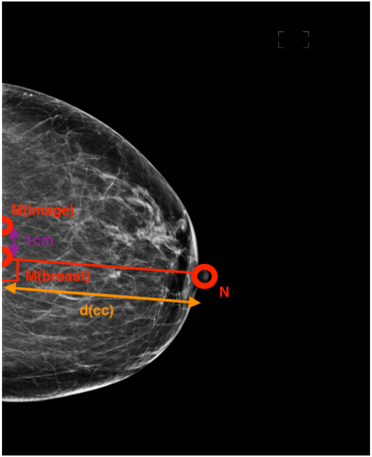
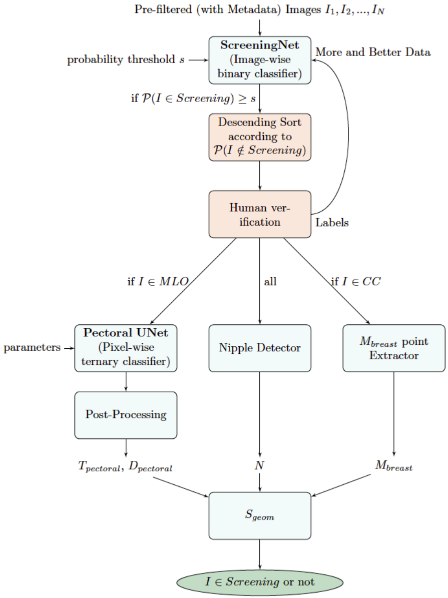


Figure 5: CC view

Decision pipeline



Decision pipeline (bis)

- In blue : automatic processes
- In orange : human assistance to improve these processes ($\sim$ anomaly detection)
- My work was mainly focused on **ScreeningNet**, **Pectoral UNet** and **Post-Processing**

Table of contents

Mammography Segmentation

- Overview
- Non Deep Learning methods
- Deep Learning methods
  - Network architecture: UNet[6] & Training
  - Data & Pre-Processing
  - Experiments, Results & Discussion

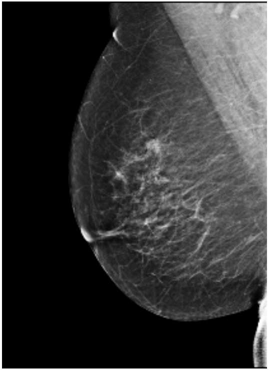
ScreeningNet

Conclusion & Future work

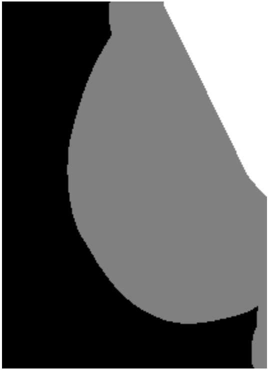
Mammography Segmentation

Overview

Visual example:



$$x_0 \xrightarrow{\boldsymbol{f}^*} y_0$$



Problem definition:

- $r, c \in \mathbb{N}$
- $x \in \mathbb{R}^{r \times c}$ a mammography image
- $y \in \{0,1\}^{r \times c \times 3}$ a Ground-Truth (GT) tensor
- ℓ a segmentation *loss* function measuring discrepancy between masks
- $\mathcal{F} \subset \left(\{0,1\}^{r \times c \times 3}\right)^{\mathbb{R}^{r \times c}}$ a function space

We look for:

$$f^* = \underset{f \in \mathcal{F}}{argmin}\, \ell(f(x), y) \tag{1}$$

Non Deep Learning methods

Test of two classical approaches:

1) Frontier estimation with Image gradient and Continuous line search

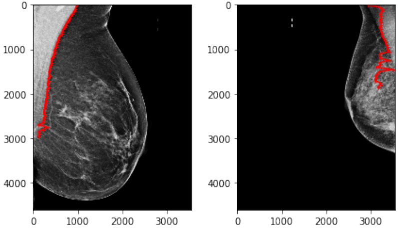


Figure 6: Success and Failure

2) FloodFill[1] algorithm

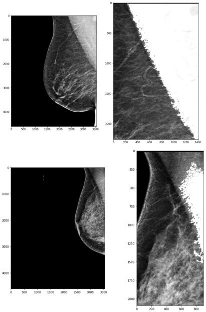


Figure 7: Success and Failure

$\Longrightarrow$ good results only on easy cases

Deep Learning methods

Method based on **learning**:
we assume $\mathcal{F} = \mathcal{F}_\Theta$ (with $\Theta \in \mathbb{R}^N$ and $N >> 1$) and we intend to learn a set of parameters Θ^* by optimizing ℓ on Θ space.

More details:

- A similar approach has been developped in 2017 by Alejandro Rodriguez-Ruiz et al.[4] (reference paper)
- Advantages:
  - extract more advanced image *features* (e.g. texture)
  - only few labelled data
- Drawbacks:
  - overfit happens and prevent *reproducibility*
  - domain *transferability*

UNet Network: f_Θ

A **symmetric** and **fully-convolutional** architecture made of:

- Contracting network: *what is there?*
- Expanding network: *where is it?*
- short and long Skip connections

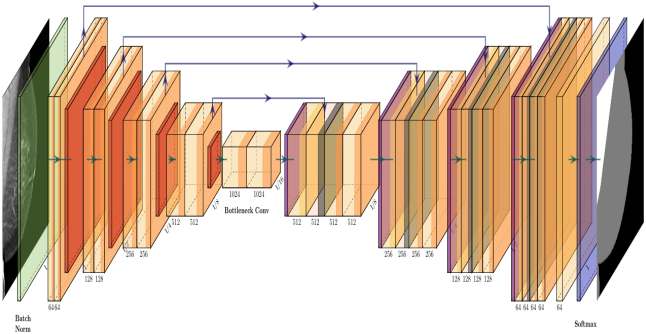


Figure 8: end-to-end architecture from Image pixels to GT labels

Loss function: ℓ

Inspired from UNet paper[6] and also from Isensee et al.[2]

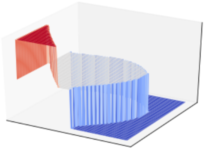
$$\ell(f_\Theta(x), y) = (1-\lambda)\underbrace{\left(-\sum_{c\in C}\sum_{i,j} w_x(i,j) y_{i,j,c} log\left(f_\Theta(x)_{i,j,c}\right)\right)}_{\mathcal{L}_{CE}} + \lambda\underbrace{\left(-\frac{2}{|C|}\sum_{c\in C}\frac{\sum_{i,j} f_\Theta(x)_{i,j,c} y_{i,j,c}}{\sum_{i,j} f_\Theta(x)_{i,j,c} + \sum_{i,j} y_{i,j,c}}\right)}_{\mathcal{L}_{Dice}} \quad (2)$$

<u>with</u>: $\left[f_\Theta(x)_{i,j,c}\right]_{i,j,c}$ output tensor, $\left[y_{i,j,c}\right]_{i,j,c}$ GT tensor, C the set of classes, λ a ponderation parameter and $[w_x(i,j)]_{i,j}$ the weight-map associated to x.

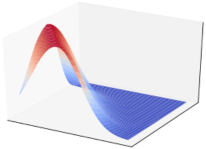
Weight-map: w_x

- Adaptation of Cross-Entropy (CE) for borders zone
- Equivalent to a 'new' GT distribution:

$$\hat{y}_{i,j} = w_x(i,j) \bullet y_{i,j} \tag{3}$$



$$y_0 \xrightarrow{\textit{\textbf{weighting}}} \hat{y}_0$$



Training procedure

- **Stochastic Gradient Descent**
  - Gradient: Backpropagation[5] to get $\nabla_\Theta \ell(f_\Theta(x), y)$
  - Optimizer: Adam[3]
- **Metrics** to assess final performances:
  - Dice Median :
    $$D\left([x_{i,j,c}], [y_{i,j,c}]\right) = \frac{2 \times \left| \left\{ (i,j), x_{i,j,c=pect} = 1 \right\} \cap \left\{ (i,j), y_{i,j,c=pect} = 1 \right\} \right|}{\left| \left\{ (i,j), x_{i,j,c=pect} = 1 \right\} \right| + \left| \left\{ (i,j), y_{i,j,c=pect} = 1 \right\} \right|}$$
  - Inter-Quartile Range (IQR)

Data & Pre-Processing

- **169** images labelled manually with software *labelme*[8] (121 training/15 validation/33 test)
- 2 main **constraints** on Data Augmentation:
  - preserve minimal pectoral presence
  - preserve spatial relative arrangement

Figure 9: From left to right *Missing pectoral*, *non-natural border* and *mirroring*

Experiments

Which model configuration performs the best ?

- **Impossible** to test all hyper-parameter combinations!

  But:

- Accurate enough *a priori* values
- Not full-independence, $batchSize \implies N_{epochs}, N_{improve}, d$
- 2 regimes ($batchSize \in \{6, 10\}$) and learnings from one transferred to the second
- We first fixed a *benchmark* model for each regime and iterations with small incremental changes one at a time.

Results

- **Best model configuration**:
  $d = 1\%$ | $batchSize = 10$ | BN=Yes | L. ReLU | Inv=No | $\lambda_{loss} = 0.5$ | Depth=5 | flips+elastic | $N_{init-filters} = 64$
- **Final results**:

Model(s)	Manufacturer	Result
Single best	Hologic	0.9647 (0.0243)
$\{Single + PP\}$ best	Hologic	0.9683 (0.0224)
Ensemble best	Hologic	0.9654 (0.0213)
$\{Ensemble + PP\}$ best	Hologic	0.9654 (0.0201)
Ref.[4] best on DBT	Siemens	**0.977 (0.0170)**
Ref.[4] best on DM	Siemens	0.974 (0.0170)
Ref.[4] best on DM	Hologic	0.947 (0.0559)

ScreeningNet

ScreeningNet

- Binary **classification** problem to ease Screening process
- **Model used**: modified VGG16[7]-like network with $\sim$ 10K images
- **Global result**: $AUC = 0.8259$ (validation of the approach)
- **More in details**: the aim is to minimize FPR (False Positive Rate). For instance, $FPR = 5\% \implies TPR = 18\%$ (still low)

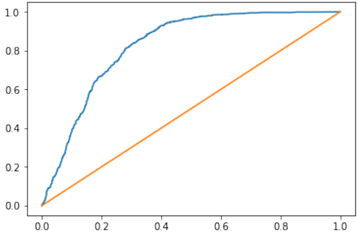


Figure 10: ROC Curve with $AUC = 0.8259$

Conclusion & Future work

Conclusion & Future work

- **Quality Check**:
  requires both non-DL & DL stand-alone expert modules
- **On Mammography Segmentation**:
  results similar to the reference paper[4].
  At first sight: same approach but some changes were necessary to achieve *reproducibility*.
  »Next steps: address domain *transferability* issue
- **On ScreeningNet**:
  general approach validated but *TPR* is still too low.
  »Next steps: More data and multi-class approach.

Future work

New framework for Mammography Segmentation

- **Motivations**: noisy/leaky/uncomplete results with UNet-only. Post-Processing is here to address this issue but can be improved.
  A learnt and thus optimized post-processing would be better.
- **SegAN[10] framework**: GAN approach with UNet as Mask Generator and a deep multi-scale concatenated network (*critic*) as Discriminator.
  Main innovation: one single loss to train both networks

$$\ell_{SegAN} = l_{MAE}\Big[f_{\Theta_{Critic}}\Big([x_{i,j}] \bullet [f_{\Theta_{UNet}}(x_{i,j})]_{c=pectoral}\Big), f_{\Theta_{Critic}}\Big([x_{i,j}] \bullet [y_{i,j,c}]_{c=pectoral}\Big)\Big] \quad (4)$$

Future work

New framework for Mammography Segmentation

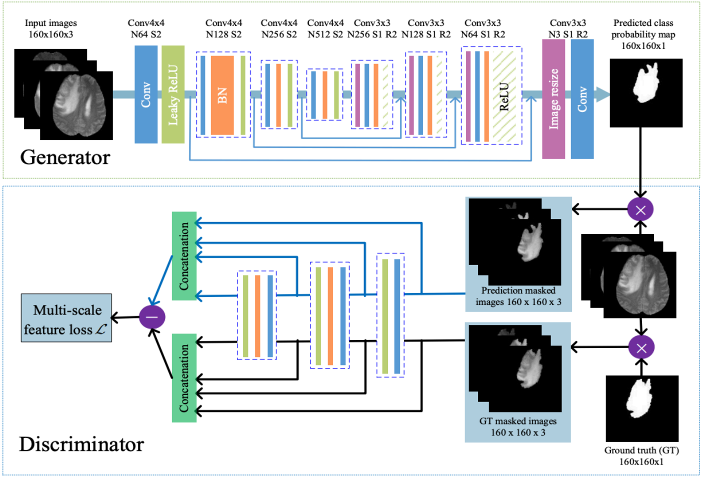


Figure 11: SegAN architecture. Image Credit to Xue Yuan et al.[10]

Thanks for your attention! :)

References

[1] FloodFill (2001). `https://en.wikipedia.org/wiki/Flood_fill`.

[2] Isensee, F. et al. (2018). nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *ArXiv e-prints*.

[3] Kingma, D. and Lei Ba, J. (2015). Adam: a method for stochastic optimization. *ArXiv e-prints*.

[4] Rodríguez-Ruiz, A. et al. (2018). Pectoral muscle segmentation in breast tomosynthesis with deep learning. *Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis*.

[5] Rojas, R. (1996). Neural Networks: A Systematic Introduction. *New York, NY, USA: Springer-Verlag New York, Inc.*

[6] Ronneberger, O. et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv e-prints*.

[7] Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large Scale Image Recognition. *International Conference on Learning Representations (ICLR)*.

[8] Wada, K. (2016). labelme: Image Polygonal Annotation with Python. `https://github.com/wkentaro/labelme`.

[9] WHO (2018). `https://www.who.int/cancer/PRGlobocanFinal.pdf`.