

Università di Bologna

CdL in Informatica

Teorie e tecniche del riconoscimento - Pattern recognition

AA 2006/2007

Advanced Machine Learning Techniques for Digital Mammography

Lesson 1

26 April 2007

Matteo Roffilli

`roffilli@csr.unibo.it`

<http://www.cs.unibo.it/people/phd-students/roffilli/>

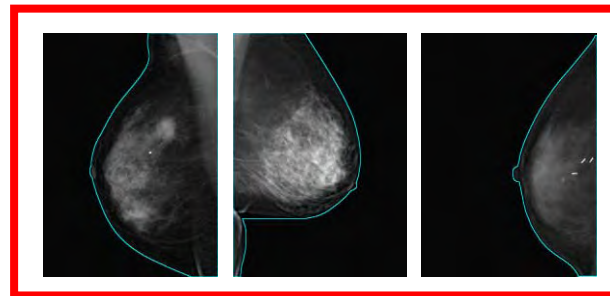
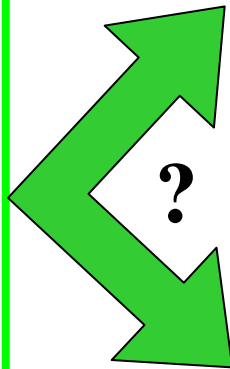
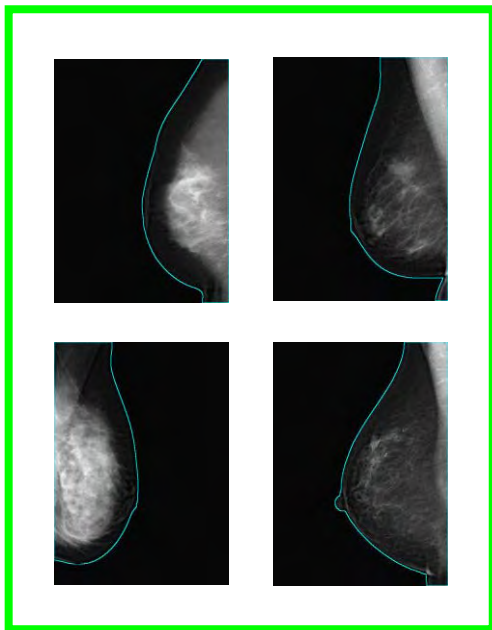
Computer Aided Detection

CAD GOAL: to aid the radiologist in detecting tumoral masses

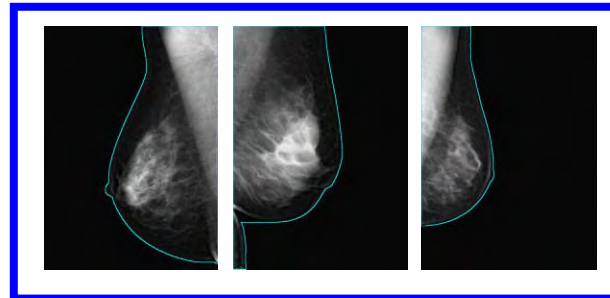
1. to classify unknown images in 2 classes: diseased or healthy
2. to locate the lesion

REQUIREMENT: to find all lesions without prompting false signals

unknown

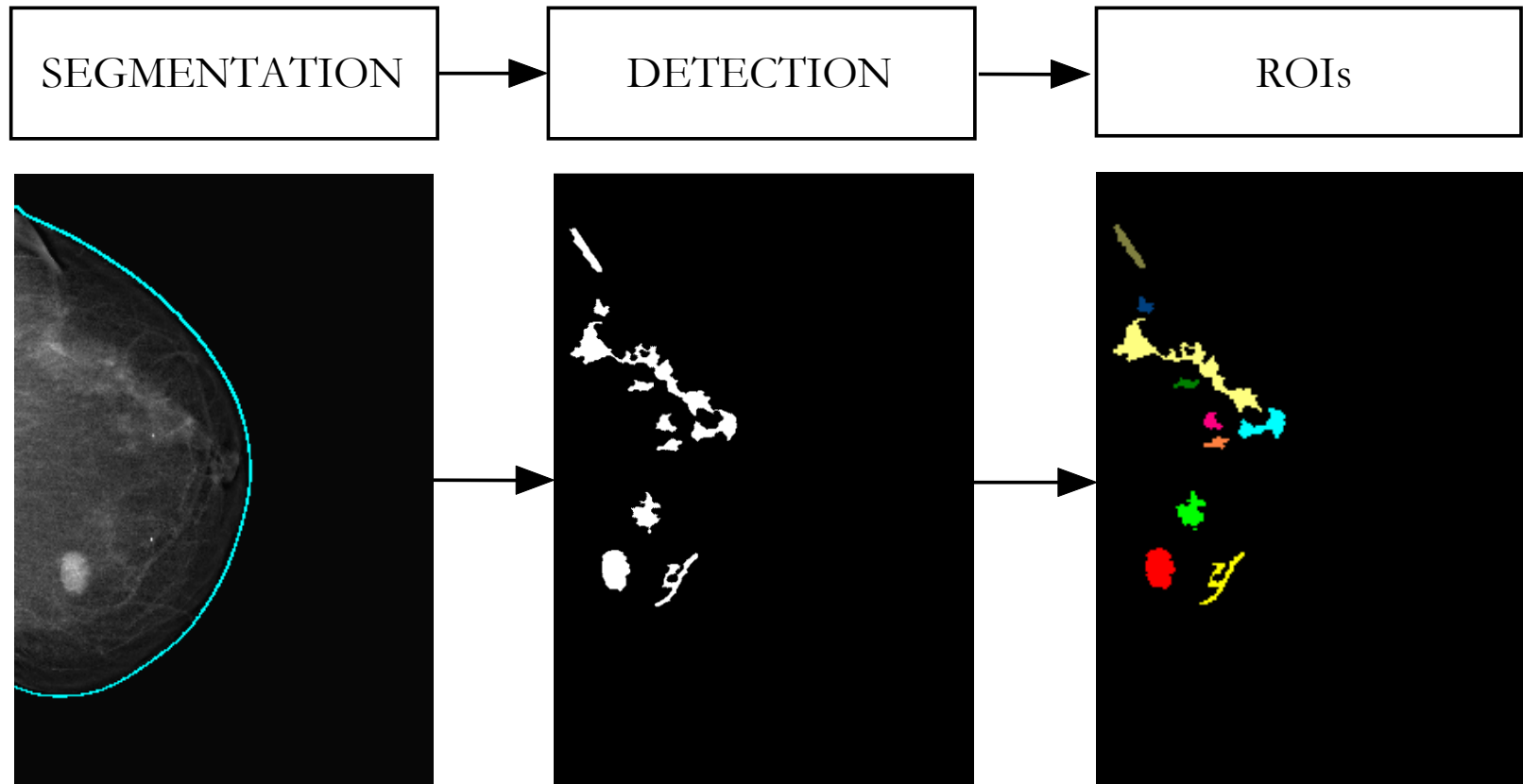


diseased



healthy

State of the art: Detection



1. Based on appearance models (external knowledge)
2. Segmenting borders is a difficult task and some difficult masses are lost!
3. About **10-50** Regions Of Interest (ROIs)

State of the art: Classification

Data representation Feature extraction



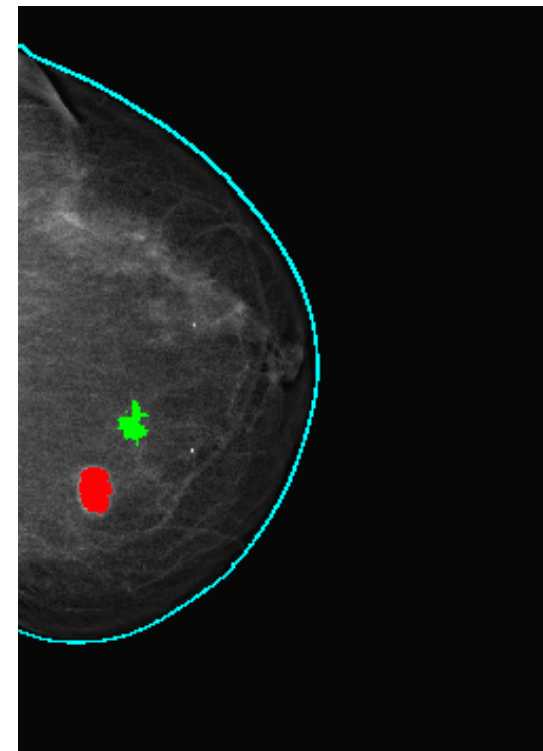
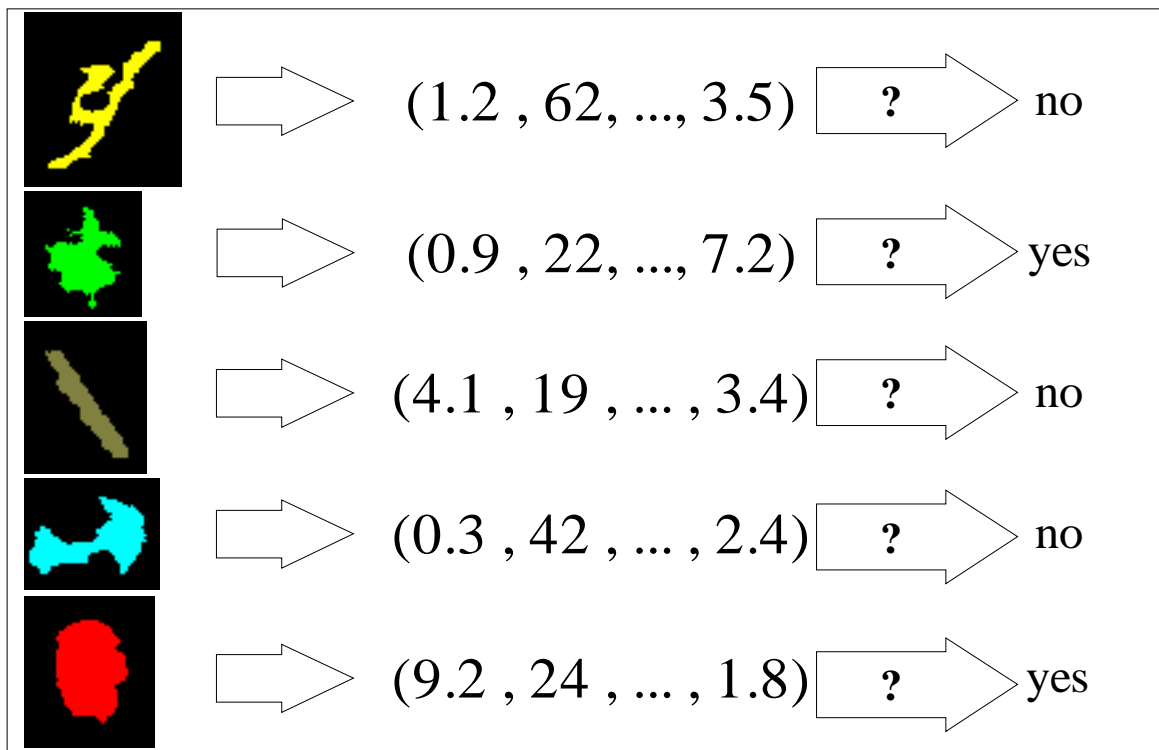
- Area
- Perimeter
- Size
- Intensity
- Shape
- ...

About **10-15** features

Classification

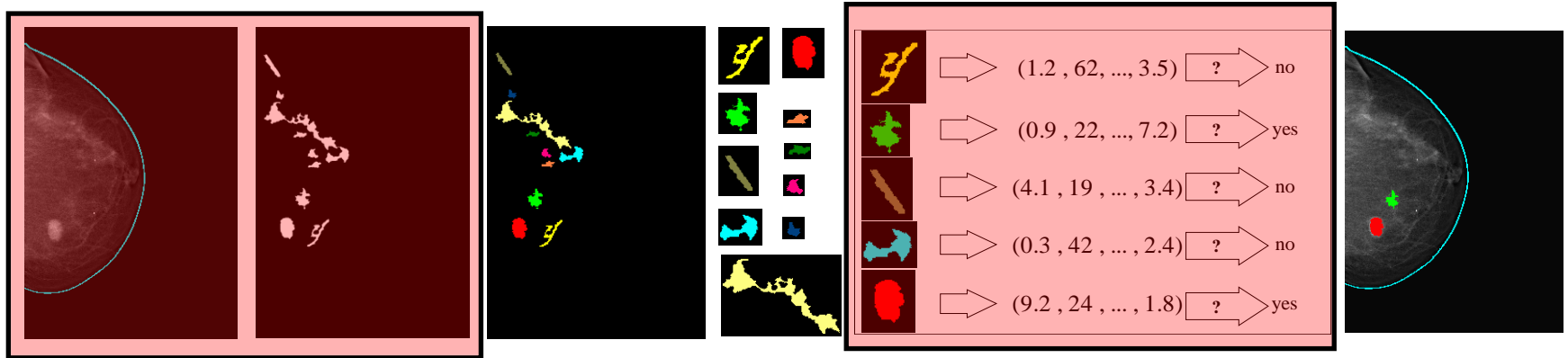
- ANN
- RBF
- Bayesian Networks
- Decision Tree
- Hand-made classifiers

State of the art: Result



ROIs classified as positive are prompted on the original image

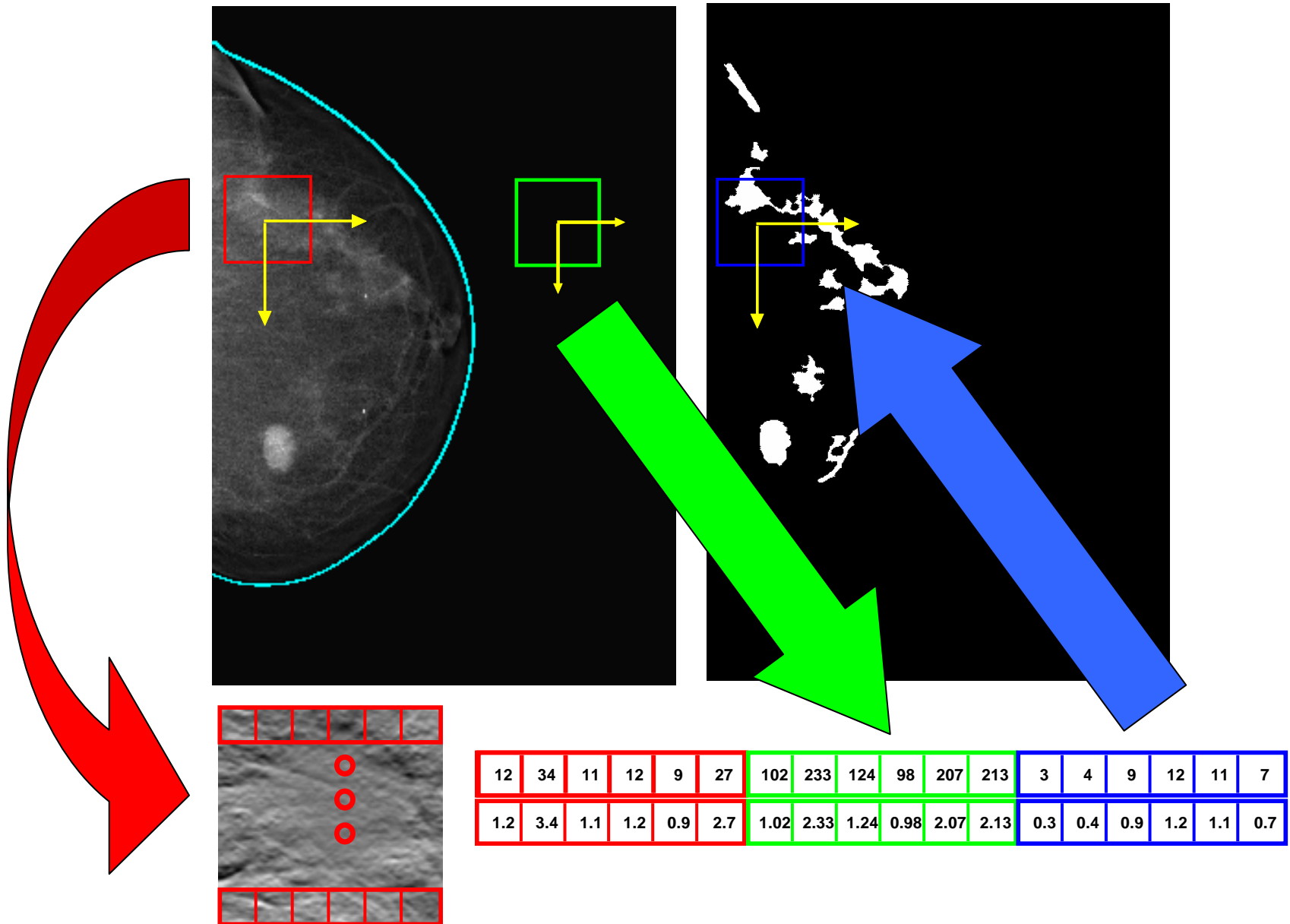
Computational overview



- Red blocks produce about 80% of computational cost.
- How can we improve the performance?

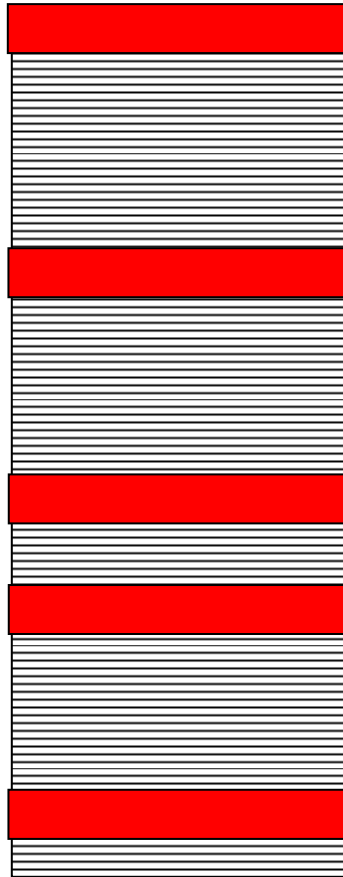
Efficient exploitation of memory/CPU bus!!!

Convolution filter



Memory access

Not efficient



Matrix `float**`

Efficient



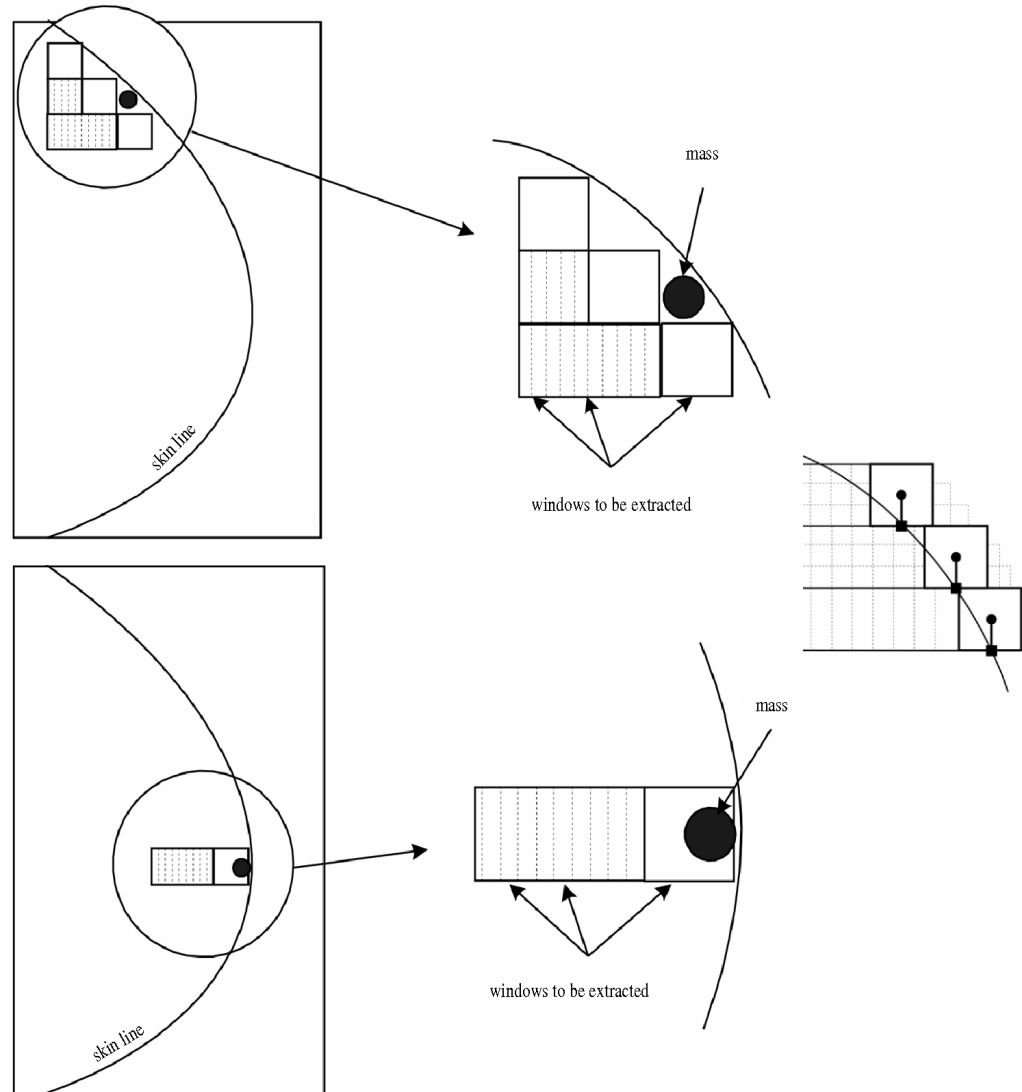
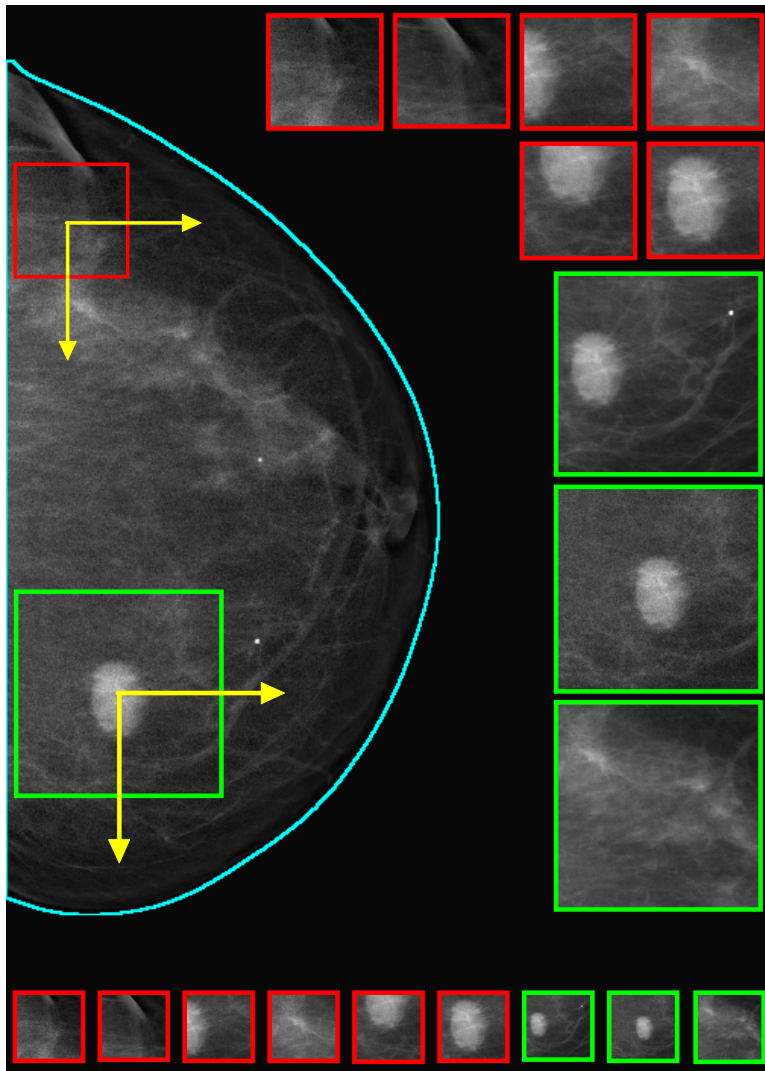
Array `float*`

Contributions

The novel contributions of this work are mainly three:

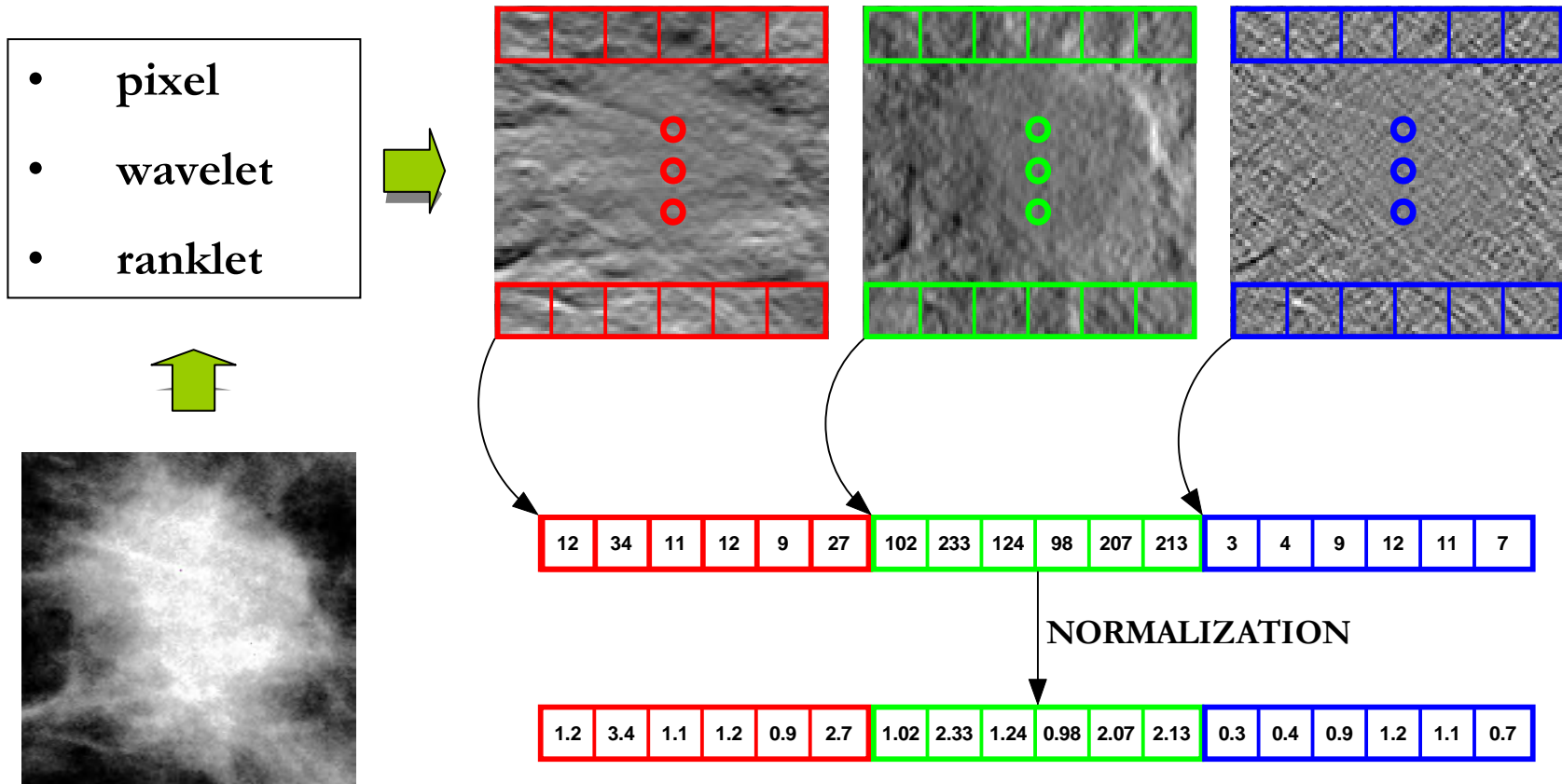
1. the **detection** step is performed without the use of external knowledge
2. the **feature extraction** step is avoided
3. SVM and RVM are used as **classifiers** for the classification step

Detection



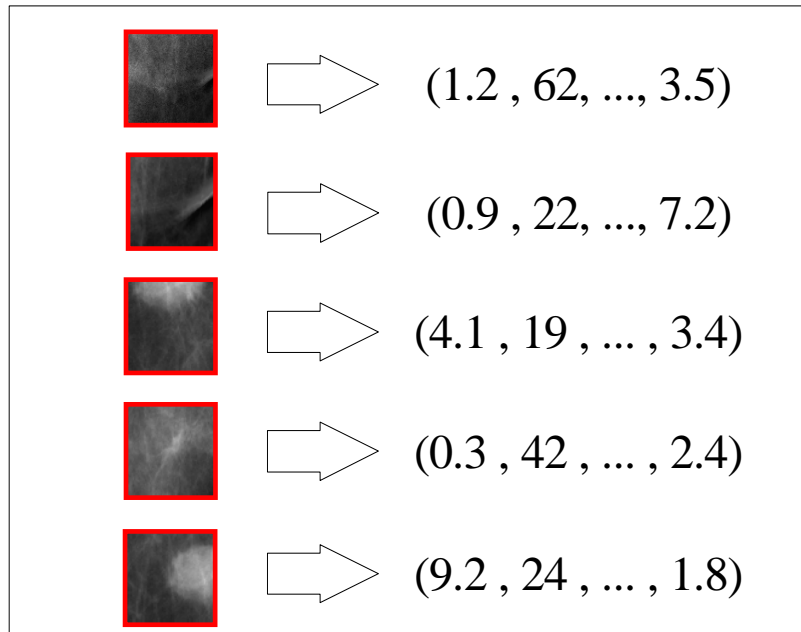
This detection produces about **100.000** ROIs

Data representation



This vector identifies a point in a ***n*-dimensional space** ($n \sim 4000$)
Each element of the vector is a **feature**

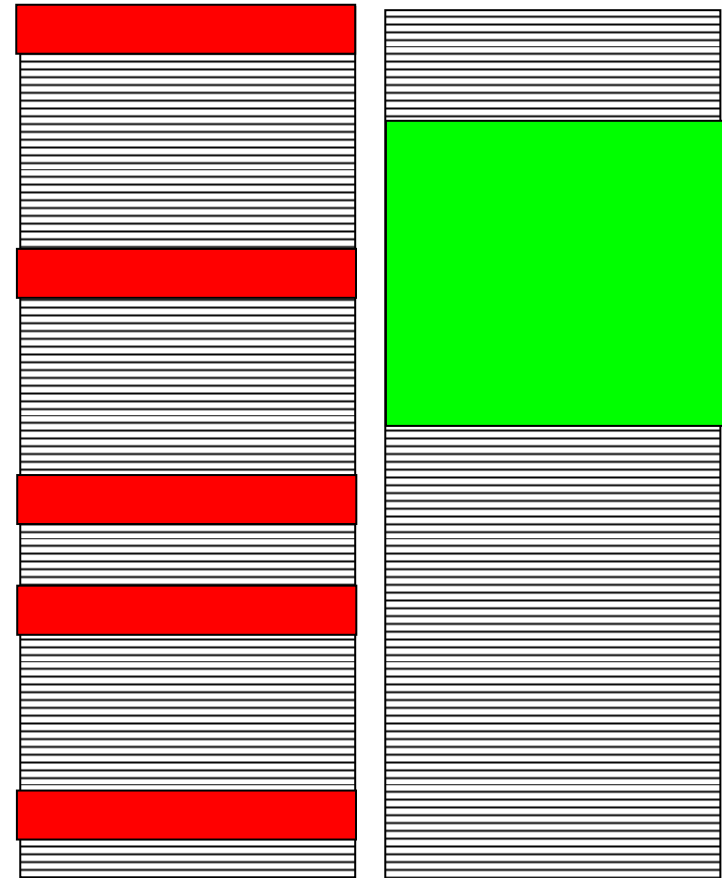
Data representation storage



~ 100'000 * 4000 float values
~ $4*4*10^8$ bytes
~ 1.50 Giga bytes

Not efficient

Efficient

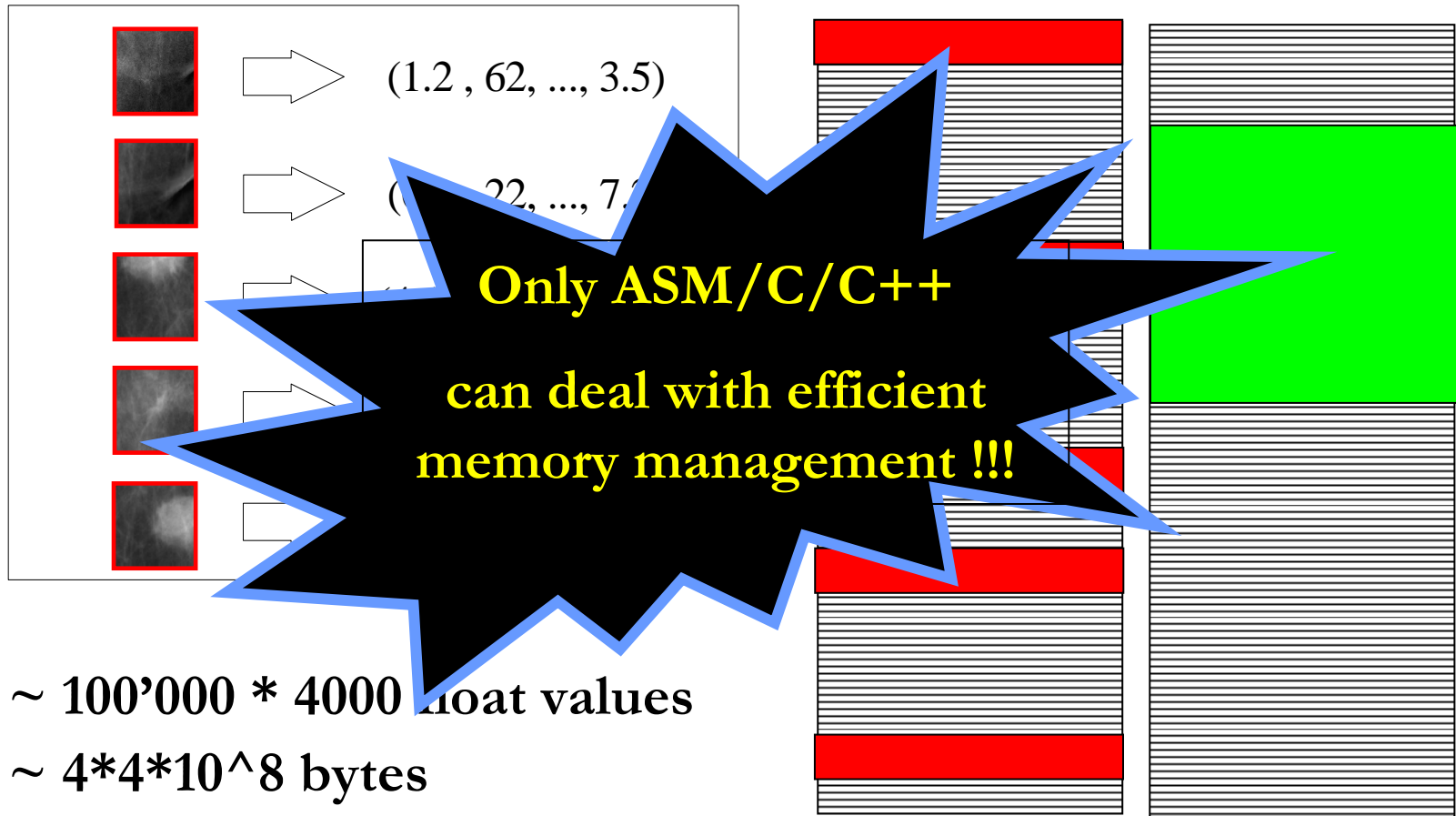


Matrix `float**` Array `float*`

Data representation storage

Not efficient

Efficient



~ 100'000 * 4000 float values

~ $4 \times 4 \times 10^8$ bytes

~ 1.50 Giga bytes

Matrix `float**` Array `float*`

Next lesson...

