

# Cell nuclei detection for computerized cancer diagnosis based on stochastic geometry and deep learning

Józef Korbicz and Marek Kowal

University of Zielona Góra,  
Institute of Control and Computation Engineering,  
Poland  
J.Korbicz@issi.uz.zgora.pl

DECT-2018  
Kharkov, Ukraine

## Object segmentation in cytological microscopic images using stochastic geometry methods

**University of Zielona Góra, Poland**  
**Institute of Control and Computation Engineering:**

Prof. Józef Korbicz - head  
Prof. Andrzej Obuchowicz  
Dr. Marek Kowal  
Michał Żejmo  
Marcin Skobel



**University Hospital in Zielona Góra, Poland**  
**Department of Pathomorphology:**

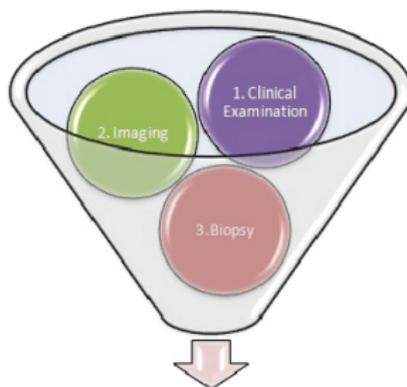
Dr. Roman Monczak  
Sylwia Magalas  
Anna Kamińska

- **Introduction to cancer cytodiagnostics**
- **Deep learning and stochastic geometry for nuclei detection**
- **Experimental results**
- **Conclusions**

# Introduction to cancer cytodiagnosics

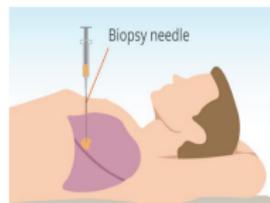
## Triple test:

- **clinical examination**
- **imaging**
  - magnetic resonance
  - computer tomography
  - ultrasound
- **biopsy**

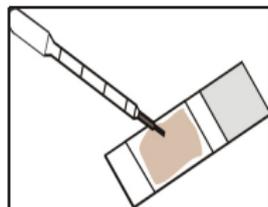


Cancer diagnosis

- A modern cancer diagnostic is based heavily on cytological tests,
- Cytological material is obtained directly from the tumor by biopsy,
- The collected material is examined by the pathologist under a microscope to determine the prevalence of cancer cells.



obtaining cytological material



staining and fixing



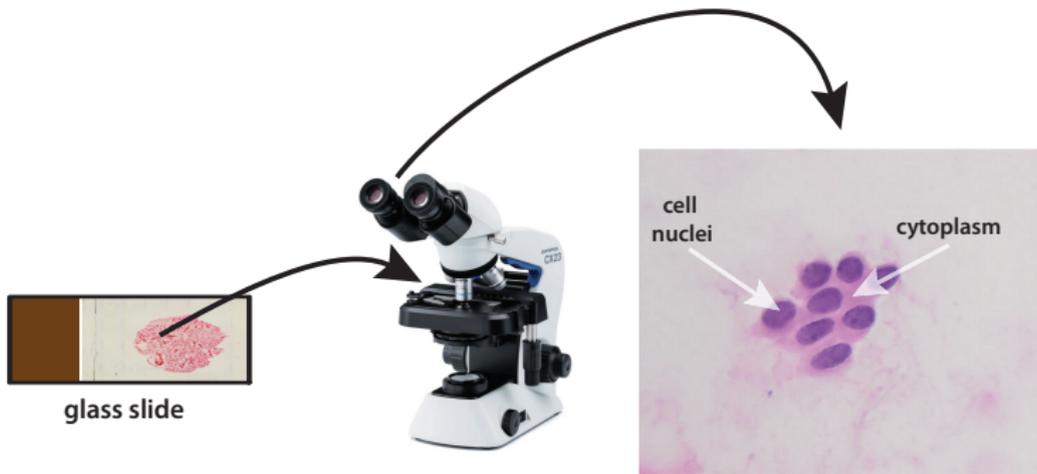
visual analysis



diagnosis

- **Pathologist usually analyse:**

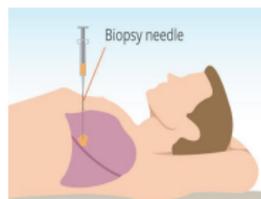
- morphometric features of cell nuclei (size of nuclei, shape of nuclei, etc.),
- colorimetric and textural features of nuclei,
- spatial distribution of nuclei.



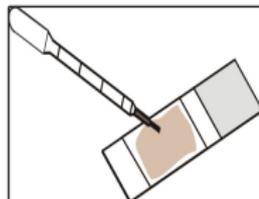
- **Analysis of cytological samples is very time-consuming process,**
- **Pathologists spend a lot of time on tedious inspection of samples under the microscope,**
- **Many scientific publications report that intra- and inter-observer variation in cytology diagnosis is substantial,**
- **Cytological diagnostics can be facilitated and speeded up by using automatic image analysis methods,**
- **Early cancer diagnosis increases the chances of survival of patients.**

# Computerized cytology

- The first two steps are the same as in classical cytology,
- Glass slides are digitized into virtual slides using special scanners,
- Pathologists analyse cytological samples on the computer screen,
- Automatic algorithms can be used to analyze virtual slides.



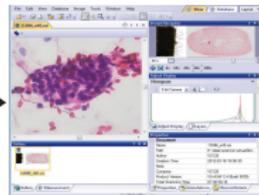
biopsy



staining and fixing



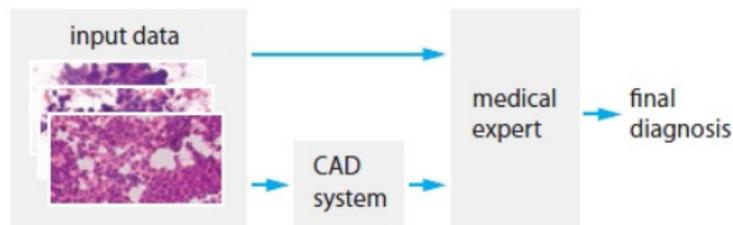
digitizing slide



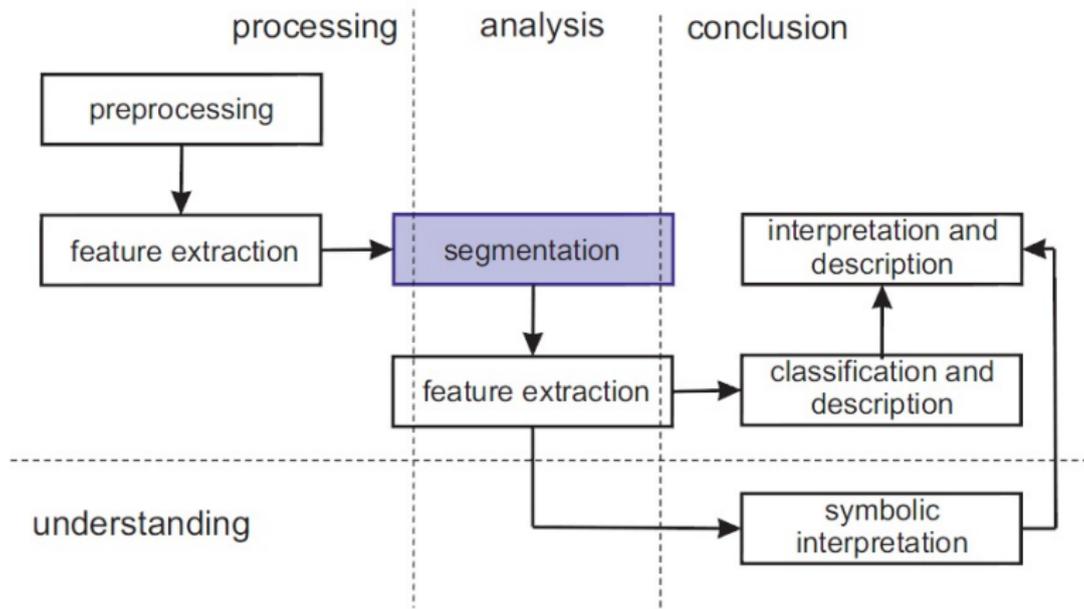
automatic image analysis

# Computer-Aided Diagnosis (CAD) in cytology

- **CAD systems can assist pathologist by:**
  - automatic cell nuclei detection,
  - extracting morphometric features of cells and nuclei,
  - classifying tumors using machine learning algorithms,
  - discovering new diagnosis rules, invisible to the naked eye.
- **CAD systems are not routinely used in cytology because there is still a need for improvement in this area.**

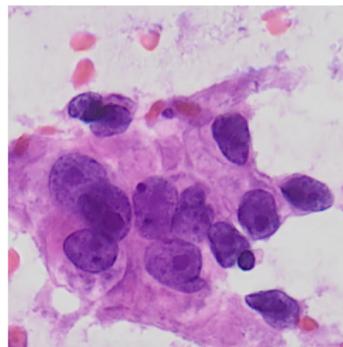
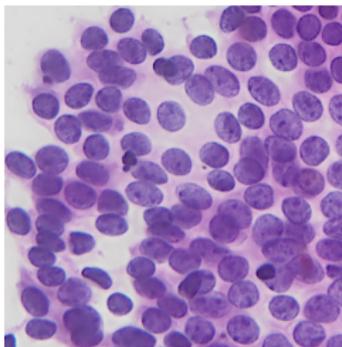
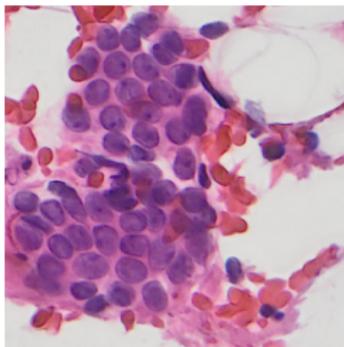


# Overview of automatic diagnosis system



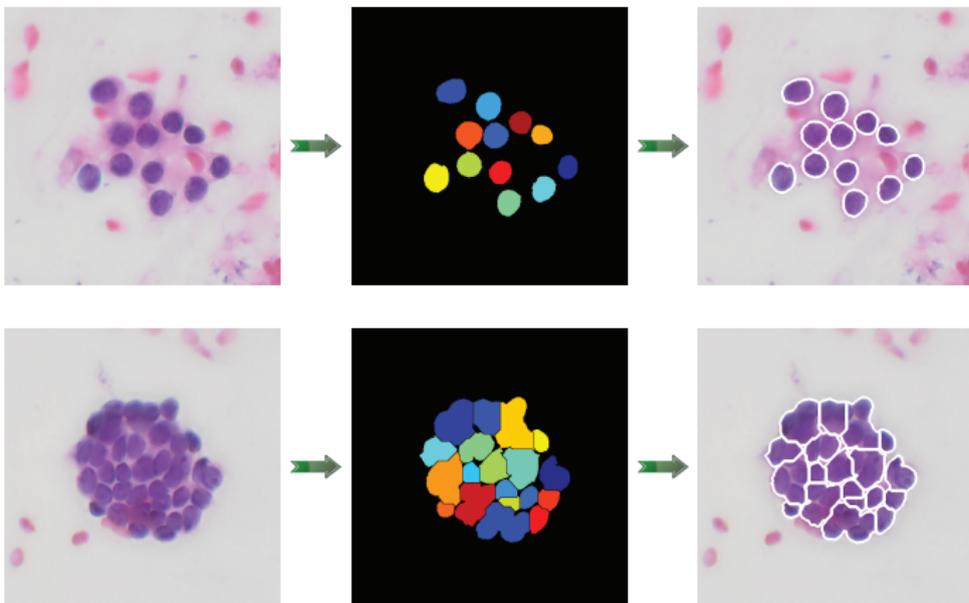
# Challenges of nuclei segmentation/detection

- Nuclei very often overlap, create clumps. This creates ambiguities that make understanding of scene hard,
- Inter-lab variation and intra-lab variation in staining intensity,
- Samples can contain a lot of irrelevant objects (red blood cells, lymphocytes, cytoplasm, fat cells, etc.),
- Obtaining learning data requires a tedious, manual segmentation of nuclei,
- Huge information to process (typical size of virtual slide: 20GB).



# Classical segmentation methods (easy vs hard cases)

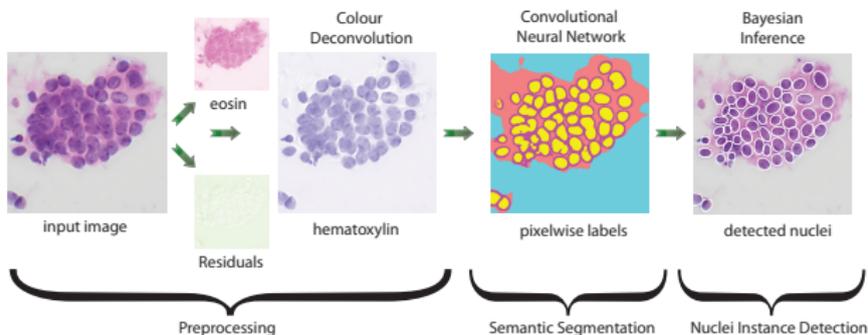
- Classical nuclei detection/segmentation methods like **watershed transform** or **active contours** work well for easy cases,
- Unfortunately, these methods cannot deal with more complex scenes where a lot of nuclei overlap.



# Deep learning and stochastic geometry for nuclei detection

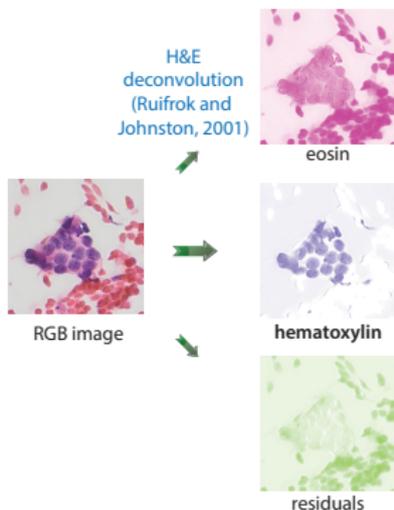
# Nuclei detection using deep learning and stochastic geometry - overview

- The input image is preprocessed by the **color deconvolution procedure** to extract the intensity of a blue dye (hematoxylin),
- **Convolutional neural network** is trained to make semantic segmentation of intensity image of hematoxylin,
- **Bayesian inference** is applied to detect nuclei,
- The main novelty of the presented approach is in the use of **high-level image analysis** to interpret and understand the results of semantic segmentation.



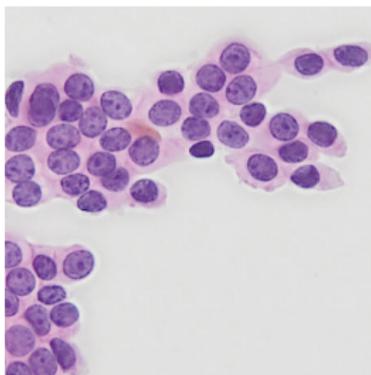
# Preprocessing

- Cytological preparations are dyed using **hematoxylin** and **eosin**,
- Hematoxylin is mainly absorbed by nuclei and eosin by cytoplasm. As a result, nuclei have blue color and cytoplasm is red,
- Cancer diagnosis is based on nuclei structures, cytoplasm should be removed,
- Nuclei structures deposit eosin to some extent and absorption spectra of hematoxylin and eosin are overlapping in RGB space,
- We can use **H&E deconvolution algorithm** to extract the contribution of hematoxylin.

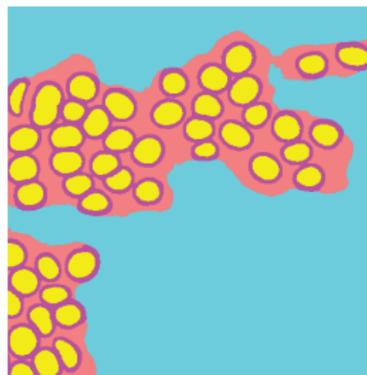


# Semantic segmentation

- **Semantic segmentation** is an algorithm that associates a label or category with every pixel in an image,
- It is used to partition a digital image into multiple segments (sets of pixels that share certain, similar characteristics),
- It is typically used to locate objects and boundaries (lines, curves, etc.) in images,
- **Semantic segmentation** is very often realized using **Convolutional Neural Networks (CNN)**.



Semantic  
segmentation



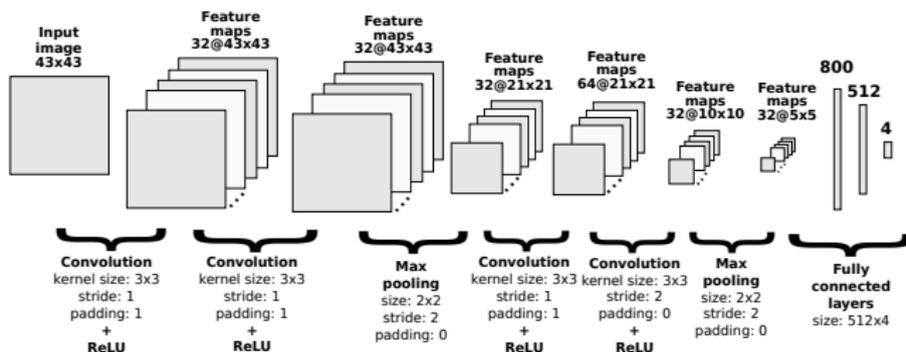
labels: nuclei, cytoplasm  
nuclei edges and background

# Convolutional Neural Networks (CNN)

- In recent years, **CNN** has gained a lot of popularity as a tool for **image segmentation** and **object detection and recognition**,
- The main advantage of this approach is that **CNN** learns from training data a hierarchy of filters to extract **invariant features** to represent an image,
- This approach has proven to be more accurate for **semantic segmentation** than methods based on features engineered by hand,
- To train **CNN**, we need very reach set of training data,
- **CNN** often encounter problems with separating overlapped or touched objects (eg. overlapped nuclei),
- We propose to overcome this problem by post-processing the results of semantic segmentation using **stochastic geometry** method.

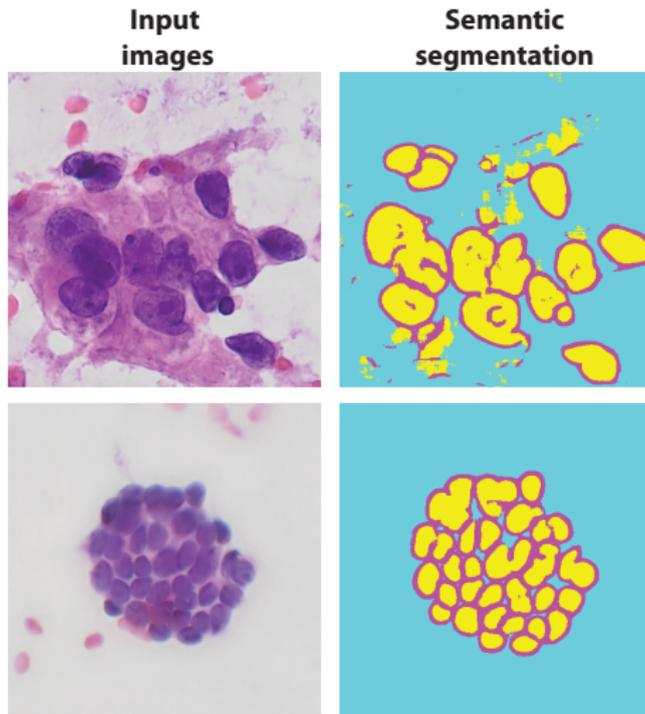
# Structure of CNN

- Typical **CNN** is usually comprised of few **convolutional layers**, combined with **pooling layers** and ended by at least one **fully connected layer**,
- **Convolutional layer** is a core part of CNN, composed of a set of learnable filters. Each filter extracts different features from the input image. Filter parameters (weights) are tuned during the learning procedure,
- **Pooling layer** is used to progressively reduce the spatial size of the input in order to extract higher level features. Spatial size reduction is usually done by max pooling,
- **Fully-connected layer** is at the end of CNN and is connected to all activations in the previous layer. The input of this layer is one-dimensional feature vector. The task of this layer is to capture the complex relationships between high-level features and output labels.

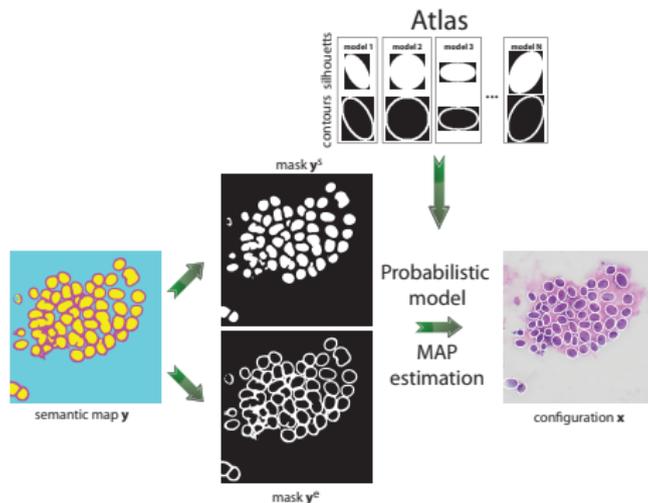


# Semantic segmentation - ambiguities

- The results of semantic segmentation can be ambiguous due to occlusions and inaccurate segmentation,
- We require a high-level image analysis algorithm to resolve such ambiguities.



# Resolving ambiguities - Bayesian object detection



## Problem formulation

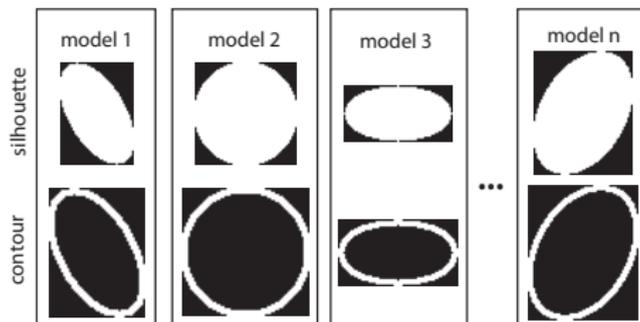
Given the atlas of nuclei models, we must infer which configuration  $x = \{x_1, x_2, \dots, x_n\}$ , fit to the semantic map  $y$ . By adopting the Bayesian framework, we can formulate the problem in the following form:

$$p(x | y) \propto f(y | x)p(x),$$

where data term  $f(y | x)$  evaluates the consistency of configuration  $x$  respect to data  $y$  and a priori term  $p(x)$  reflects some constraints on interactions between nuclei models.

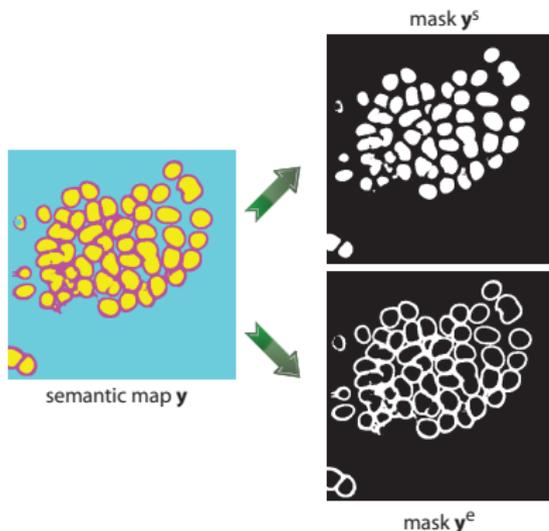
# The structure of atlas

- The atlas is a set of nuclei models with arbitrary shape and size,
- Each nuclei model is defined by means of a silhouette and a contour,
- Typically, nuclei models are limited to simple shapes such as discs or ellipses,
- Algorithm can freely choose models from the atlas and place them in any place on semantic map to approximate the semantic map accurately.



# Semantic map

- Semantic map is generated by CNN,
- Three labels are distinguished: nuclei interior, nuclei edge and background,
- CNN is also able to extract cytoplasm but here it is treated as background,
- Semantic map is decomposed into two binary maps  $y^s$  and  $y^e$ , which represent nuclei interior and nuclei edge respectively.



## Data term

- The task of the data term  $f(\mathbf{y}|\mathbf{x})$  is to evaluate how well configuration  $\mathbf{x}$  fit the semantic map  $\mathbf{y}$ . Based on the fact that semantic map consists of two independent components: binary silhouette map  $\mathbf{y}^s$  and binary edge map  $\mathbf{y}^e$ , it was decided to decompose data term in the following way:

$$f(\mathbf{y} | \mathbf{x}) = f(\mathbf{y}^s, \mathbf{y}^e | \mathbf{x}) = f(\mathbf{y}^s | \mathbf{x})f(\mathbf{y}^e | \mathbf{x}).$$

- To get the final model that will be used for nuclei detection, we must define data terms  $f(\mathbf{y}^s | \mathbf{x})$  and  $f(\mathbf{y}^e | \mathbf{x})$  for pixel-wise level of granularity:

$$f(\mathbf{y}^s | \mathbf{x}) = \prod_{t \in S(\mathbf{x})} B(y_t^s; p_s) \prod_{t \in T \setminus S(\mathbf{x})} B(y_t^s; q_s),$$

$$f(\mathbf{y}^e | \mathbf{x}) = \prod_{t \in C(\mathbf{x})} B(y_t^e; p_c) \prod_{t \in T \setminus C(\mathbf{x})} B(y_t^e; q_c),$$

where  $T$  is a pixel lattice of the size of the input image,  $S(\mathbf{x}) = \bigcup_{i=1}^n S(x_i)$  is a silhouette of the configuration  $\mathbf{x}$  and  $S(x_i)$  is a silhouette of the single object  $x_i$ ,  $C(\mathbf{x}) = \bigcup_{i=1}^n C(x_i)$  is a contour of the configuration  $\mathbf{x}$  and  $C(x_i)$  is a contour of the single object  $x_i$ ,  $B(\cdot)$  is a Bernoulli probability mass function (pmf),

- Bernoulli pmfs are used to model data terms  $f(\mathbf{y}^s | \mathbf{x})$  and  $f(\mathbf{y}^e | \mathbf{x})$  for pixel-wise level,
- Bernoulli distribution was chosen intentionally to not to overcomplicate the model but the other distributions are also applicable.

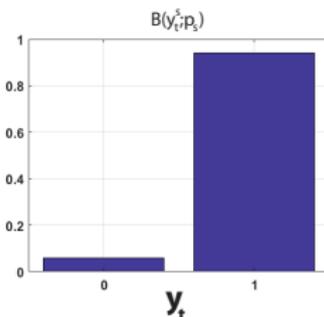
## Bernoulli pmf $B(y_t^s; p_s)$

- $B(y_t^s; p_s)$  is a Bernoulli pmf, which describes the distribution of  $y_t^s$  inside the silhouette  $S(x)$ :

$$B(y_t^s; p_s) = \begin{cases} 1 - p_s & \text{if } y_t^s = 0, \\ p_s & \text{if } y_t^s = 1, \end{cases}$$

where  $y_t^s$  describe the value of  $t$ -th pixel of binary silhouette map and  $S(x)$  is a binary map which describe silhouette of configuration of nuclei models given by  $x$ ,

- The value of  $p_s$  is chosen arbitrary e.g.  $p_s = 0.95$ ,
- Such large value of  $p_s$  says that the probability of the scenario that given pixel  $y_t$  belongs to **nuclei interior** on the semantic map and at the same time belongs to configuration silhouette  $S(x)$  is very likely,
- On the other hand if the given pixel  $y_t$  does not belong to **nuclei interior** but belongs to configuration silhouette  $S(x)$  then such scenario is very unlikely  $1 - p_s$ .



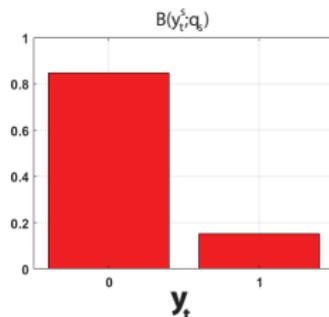
## Bernoulli pmf $B(y_t^s; q_s)$

- $B(y_t^s; q_s)$  is a Bernoulli pmf, which describes the distribution of  $y_t^s$  outside the silhouette  $T \setminus S(x)$ :

$$B(y_t^s; q_s) = \begin{cases} q_s & \text{if } y_t^s = 0, \\ 1 - q_s & \text{if } y_t^s = 1, \end{cases}$$

where  $y_t^s$  describe the value of  $t$ -th pixel of binary silhouette map and  $S(x)$  is a binary map which describe silhouette of configuration of nuclei models given by  $x$ ,

- The value of  $q_s$  is chosen arbitrary e.g.  $q_s = 0.87$ ,
- Such large value of  $q_s$  says that the probability of the scenario that given pixel  $y_t$  belongs to **background** on the semantic map and at the same time does not belong to configuration silhouette  $S(x)$  is very likely,
- On the other hand if the given pixel  $y_t$  does not belong to **background** and does not belong to configuration silhouette  $S(x)$  then such scenario is very unlikely  $1 - q_s$ .



## Bernoulli pmfs $B(y_t^e; p_c)$ and $B(y_t^e; q_c)$

- Bernoulli pmfs  $B(y_t^e; p_c)$  and  $B(y_t^e; q_c)$  are defined in a similar way to the previously presented distributions:

$$B(y_t^e; p_c) = \begin{cases} 1 - p_c & \text{if } y_t^e = 0, \\ p_c & \text{if } y_t^e = 1, \end{cases}$$

$$B(y_t^e; q_c) = \begin{cases} q_c & \text{if } y_t^e = 0, \\ 1 - q_c & \text{if } y_t^e = 1, \end{cases}$$

- These distributions evaluate the probability of scenarios that pixel belongs or not to the edge in the semantic map  $y^e$  and at the same time it belongs or not to the contour  $C(x)$  of the configuration  $x$ .

# Bayesian object detection - details

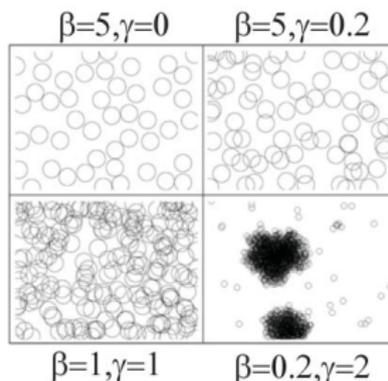
## A prior term

- A prior term  $p(\mathbf{x})$  play a role of a **regularizer**, which penalize potential configurations of nuclei models which contain too many overlapping objects:

$$p(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{u(\mathbf{x})},$$

where  $u(\mathbf{x})$  is the number of pairwise overlaps in the configuration  $\mathbf{x}$ ,  $n(\mathbf{x})$  is the number of objects,  $\gamma$  is the parameter controlling interactions between disks and  $\beta$  is the constant, which can be interpreted as the cost of adding object to the configuration,

- For  $\gamma < 1$ , model exhibits repulsion between objects, thus number of overlapping objects is low, for  $\gamma = 1$  we receive the stochastic process with Poisson distribution (objects are located in a given area with an even density), for  $\gamma > 1$  model exhibits attraction between objects and process usually generates more overlaps than respective Poisson process, objects tend to create clusters but such stochastic process is usually undefined.



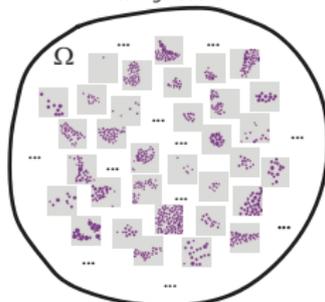
## Optimization

- Given a probabilistic model  $f(\mathbf{y} | \mathbf{x})p(\mathbf{x})$ , the problem of finding configuration  $\mathbf{x}$  of nuclei models that would be consistent with the semantic map  $\mathbf{y}$  can be formulated as a **maximum a posteriori estimation (MAP)**:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} f(\mathbf{y} | \mathbf{x})p(\mathbf{x}),$$

- By sampling configurations from  $\Omega$  using a posteriori distribution  $f(\mathbf{y} | \mathbf{x})p(\mathbf{x})$  we could solve the MAP problem,
- Unfortunately direct sampling from a posteriori distribution is numerically and analytically intractable,
- One way to overcome this problem is to use **Iterated Conditional Modes (ICM)** for optimization,
- ICM** was originally proposed by Besag to reconstruct a true scene from a noisy image,
- Later, it was shown that the method can be adapted to deal with **Marked Point Process (MPP)**,
- Random element defined on  $\Omega$  can be viewed as **MPP**.

$\Omega$  – a set of all potential configurations  $\mathbf{x}$



## Iterated Conditional Modes

- **ICM** starts with the provisional configuration  $\hat{\mathbf{x}}_0$  and then begins updating it by introducing some minor changes to the current configuration (e.g. adding , shifting or removing nuclei model),
- Candidate configurations are scanned in some predefined order. If candidate configuration  $\mathbf{x}_c$  is better than the current one  $\hat{\mathbf{x}}_k$  then algorithm accepts it as the new current configuration  $\hat{\mathbf{x}}_{k+1} = \mathbf{x}_c$ ,
- Such approach ensures that probability of the current configuration never decreases at any stage and eventual convergence is guaranteed. However, the algorithm can be stuck in the nearest local maxima.

$$\hat{\mathbf{x}}_0 = \emptyset$$

$$k = 0$$

**repeat**

*generate the list of candidate configurations  $X_{k+1}$*

**for all**  $\mathbf{x}_c \in X_{k+1}$  **do**

$$w(\mathbf{x}_c) = \log \left( \frac{f(\mathbf{y}|\mathbf{x}_c)p(\mathbf{x}_c)}{f(\mathbf{y}|\hat{\mathbf{x}}_k)p(\hat{\mathbf{x}}_k)} \right)$$

**end for**

*find configuration  $\bar{\mathbf{x}}_c$  which maximizes  $w(\mathbf{x}_c)$*

**if**  $w(\bar{\mathbf{x}}_c) > 0$  **then**  $\hat{\mathbf{x}}_{k+1} = \bar{\mathbf{x}}_c$

**end if**

$$k := k + 1$$

**until**  $w(\bar{\mathbf{x}}_c) \leq 0$  **or**  $k > n$ .

## Iterated Conditional Modes

- Computational burden of **ICM** is very high due to iterating over huge space of candidate solutions,
- The crucial and computationally expensive step of the algorithm determines if the candidate configuration  $\mathbf{x}_c$  is better than the current one  $\hat{\mathbf{x}}_k$  based on **log posterior probability ratio**:

$$\begin{aligned}w(\mathbf{x}_c) &= \log \left( \frac{f(\mathbf{y}^s | \mathbf{x}_c) f(\mathbf{y}^e | \mathbf{x}_c) p(\mathbf{x}_c)}{f(\mathbf{y}^s | \hat{\mathbf{x}}_k) f(\mathbf{y}^e | \hat{\mathbf{x}}_k) p(\hat{\mathbf{x}}_k)} \right) = \\&+ \log(\beta) (n(\mathbf{x}_c) - n(\hat{\mathbf{x}}_k)) + \log(\gamma) (u(\mathbf{x}_c) - u(\hat{\mathbf{x}}_k)) + \\&= \sum_{t \in S(\mathbf{x}_c) \setminus S(\hat{\mathbf{x}}_k)} z_s(y_t^s) - \sum_{t \in S(\hat{\mathbf{x}}_k) \setminus S(\mathbf{x}_c)} z_s(y_t^s) + \\&+ \sum_{t \in C(\mathbf{x}_c) \setminus C(\hat{\mathbf{x}}_k)} z_c(y_t^e) - \sum_{t \in C(\hat{\mathbf{x}}_k) \setminus C(\mathbf{x}_c)} z_c(y_t^e),\end{aligned}$$

where

$$z_s(y_t^s) = \log B(y_t^s; p_s) - \log B(y_t^s; q_s),$$

$$z_c(y_t^e) = \log B(y_t^e; p_c) - \log B(y_t^e; q_c).$$

- The problem is mitigated by **Graphics Processing Unit GPU** implementation of **ICM**.

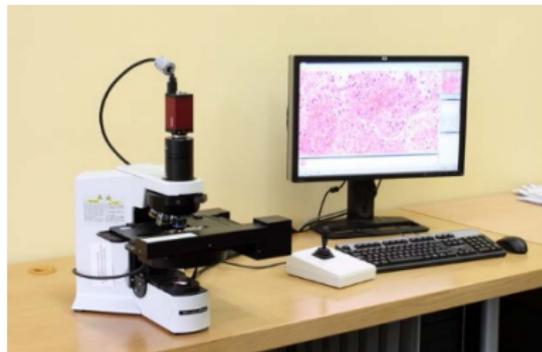
# Experimental results

## Equipment

- Olympus VS120

## Characteristics of the base

- number of cases (slides): 50
- malignant: 25
- benign: 25
- file format: VSI
- image size:  $200,000 \times 100,000$  px
- focus:  $40\times$

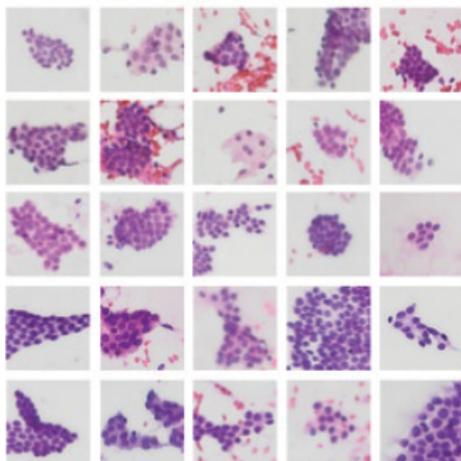


All cancers were histologically confirmed and all patients with benign disease were either biopsied or followed for a year

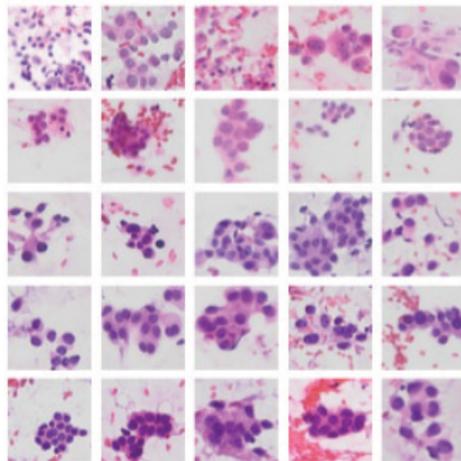
# The material on which the research was carried out

- Proposed methods were tested on **90** fragments (size: **500px × 500 px**) of virtual slides of **breast cancer**,
- All fragments were selected and manually labeled by pathologists
- **40 (20 benign, 20 malignant)** images were used to train and test **CNN**,
- **50 (25 benign, 25 malignant)** images were used to verify the effectiveness of the **stochastic geometry** approach in nuclei detection.

Benign cases



Malignant cases



- Training and testing images were cut into patches of size **43 × 43 pixels**,
- The class of each patch is assigned based on the label of central pixel. Label is read from semantic map,
- The patches were subjected to normalization and augmentation by scaling, rotating or flipping,
- Training was conducted using **stochastic gradient descent**, mini-batch was set to 256 and training process was finished after 20 epochs,
- We applied **dropout** technique to prevent the network from over-fitting,
- This allowed us to achieve the **91.33%** classification for the test set.

	Training	Testing
Number of images	20	20
Total patches	18,315,912	18,961,569
- nuclei border	1,739,200	1,968,232
- nuclei center	5,105,059	5,788,305
- cytoplasm	5,588,398	5,770,039
- background	5,883,255	5,434,993

# Evaluation of segmentation results

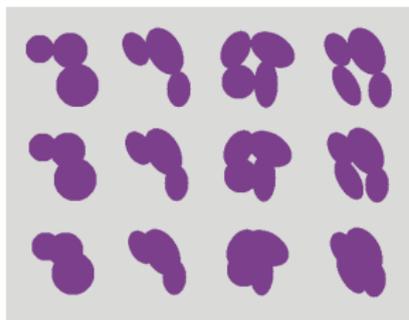
- **Hausdorff distance** was used to measure the similarity between reference nuclei and detected nuclei:

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\},$$

- For each reference nucleus closest segmented nuclei was determined,
- Reference nucleus can be paired with the nearest segmented nucleus if their **Hausdorff distance** is below predefined **threshold**  $T_{HD}$  (such case is classified as **true positive (TP)**),
- The Hausdorff distance between reference nucleus and nearest segmented nucleus is above the predefined threshold (such case is classified as **false negative (FN)**),
- Segmented nucleus can stay without corresponding reference nucleus (such case is classified as **false positive (FP)**),
- The effectiveness of the stochastic geometry was compared with the effectiveness of **marker-controlled watershed (MCW)** which is the state of art method for nuclei detection,
- Both methods were processing the same semantic maps generated by prepared **CNN model**.

# Sample results for artificially generated test image

**Input image**



**Stochastic geometry**



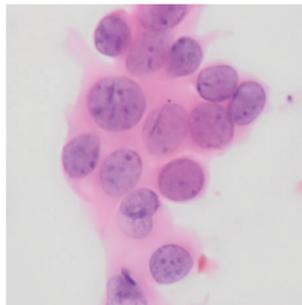
**Marker-controlled watershed**



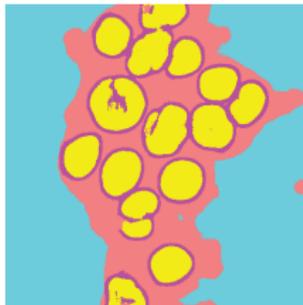
- Our approach is able to infer the structure of clustered objects even for highly overlapping cases,
- Marker-controlled watershed is not able to separate objects for difficult cases.

# Sample results - breast cancer 1

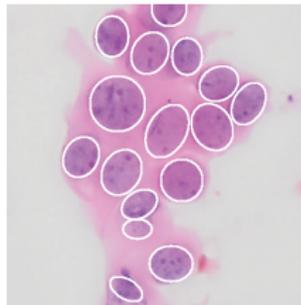
Input image



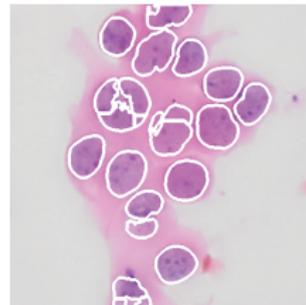
CNN  
(semantic segmentation)



Stochastic  
geometry

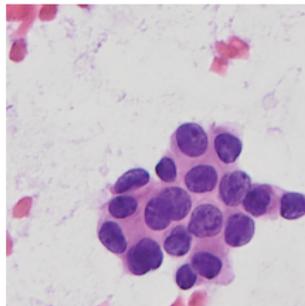


Marker-controlled  
watershed

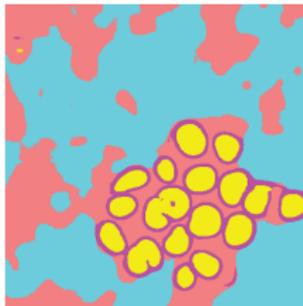


# Sample results - breast cancer 2

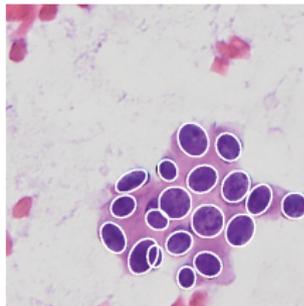
Input image



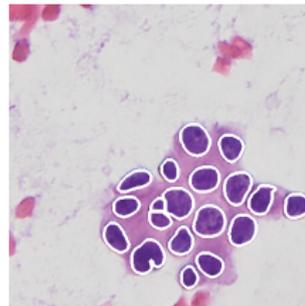
CNN  
(semantic segmentation)



Stochastic  
geometry

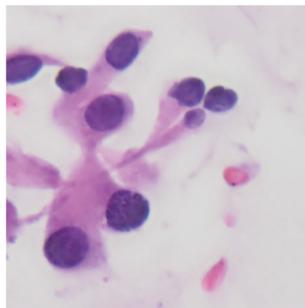


Marker-controlled  
watershed

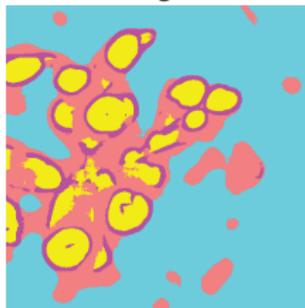


# Sample results - breast cancer 3

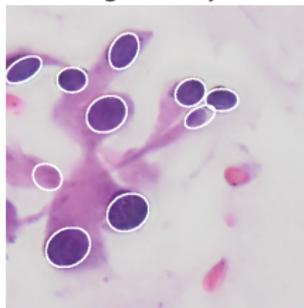
Input image



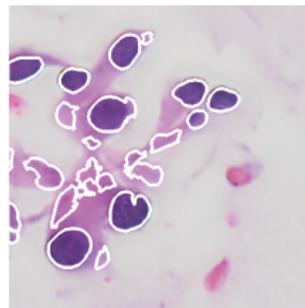
CNN  
(semantic segmentation)



Stochastic  
geometry



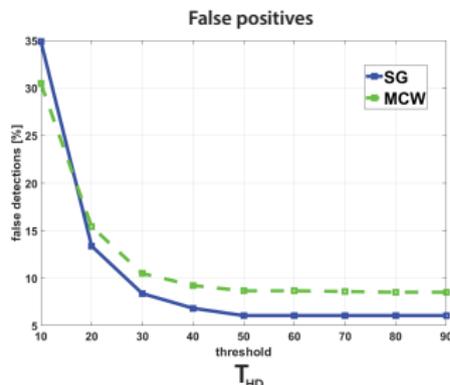
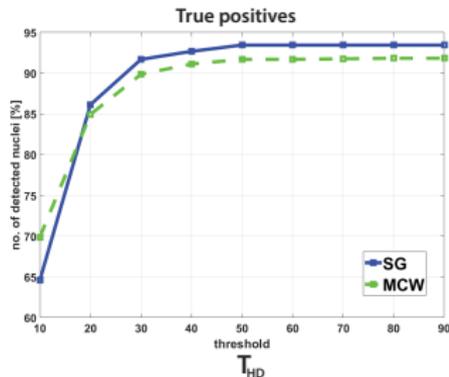
Marker-controlled  
watershed



# Nuclei detection accuracy - summary results

- Number of **true positives (TP)** and **false positives (FP)** detections for various values of threshold  $T_{HD}$ .

$T_{HD}$	TP		FP	
	SG	MCW	SG	MCW
10	918	993	496	433
20	1224	1207	190	219
30	1303	1277	119	149
40	1317	1295	97	131
50	1328	1303	86	123
60	1328	1303	86	123
70	1328	1304	86	122
80	1328	1305	86	121
90	1328	1305	86	121



- We proposed a hybrid approach, which employs CNN to prepare semantic segmentation and then stochastic geometry to detect nuclei,
- The results of the experiments carried on have shown that proposed approach outperforms the state-of-the-art method marker-controlled (MCW) watershed,
- Our method has natural ability to model overlapping objects and thus to detect the parts of the objects that are not visible,
- MCW, in this case, can detect only the part of the object, which is not occluded,
- Despite the lack of elasticity of SG in the description of nuclei shapes (only ellipses and disks), the method achieved higher detection accuracy than MCW due to the ability to model and detect overlapping nuclei,
- In future work, we are going to enrich the atlas of nuclei models with more complex shapes.

**Thank you**