Efficient Region-Based Classification for Whole Slide Images

Grégory Apou[1], Benoît Naegel[1], Germain Forestier[2], Friedrich Feuerhake[3], and Cédric Wemmert[1]

[1] ICube, University of Strasbourg, 300 bvd Sébastien Brant, 67412 Illkirch, France
[2] MIPS, University of Haute Alsace, 12 rue des Frères Lumiere, 68093 Mulhouse, France
[3] Institute for Pathology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany

Abstract. For the past decade, new hardware able to generate very high spatial resolution digital images called Whole Slide Images (WSIs) have been challenging traditional microscopy. But the touted potential for automation is hindered by the large size of the files, possibly tens of billions of pixels. We propose a fast segmentation method coupled with an intuitive multiclass supervised classification that captures expert knowledge presented as morphological annotations to establish a cartography of a WSI and highlight biological regions of interest. While our primary focus has been the development of a proof of concept for the analysis of breast cancer WSIs acquired after chromogenic immunohistochemistry, this method could also be applied to more general texture-based problems.

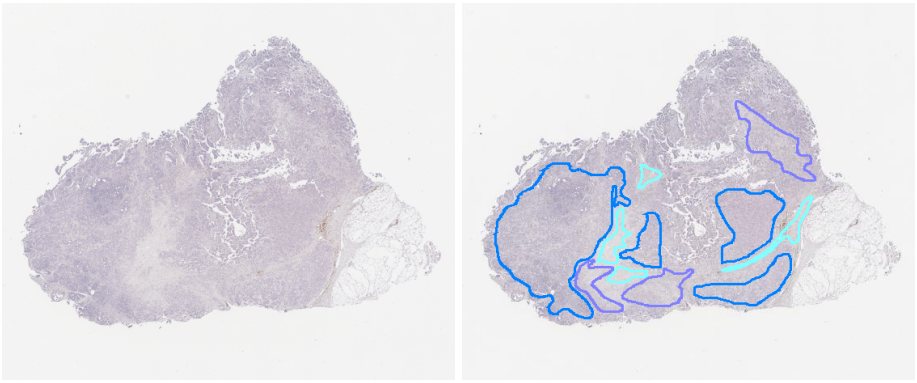
Keywords: Whole Slide Images, Biomedical Image Processing, Segmentation, Classification

1 Introduction

In recent years, the advent of digital microscopy deeply modified the way certain diagnostic tasks are performed. While the initial diagnostic assessment and the interpretation histopathological staining results remain a domain of highly qualified experts, digitization paved the way to semi-automated image analysis solutions for biomarker quantification and accuracy control. With the expected increase of the number and quality of slide scanning devices, pathologists are facing the challenge to integrate complex sets of relevant information, partially based on conventional morphology, and partially on molecular genetics and computer-assisted readout of single immunohistochemistry (IHC) parameters [4].

Despite their potential for automation to help reduce bias [15], the integration of WSIs in routine diagnostic workflows in the clinical setting is not straightforward [3]. Indeed, these images can contain hundreds of millions or even billions of pixels, causing practical difficulties for the storage, transmission, visualization and processing by conventional algorithms in a reasonable time. Figure 1

presents an example of a "small" WSI of 18000 by 15000 pixels. Moreover, this new technology is still perceived as ineffective by pathologists who are more familiar with the use of classical light microscopy.



(a) (b)

Fig. 1: Example of a Whole Slide Image: (a) raw; (b) manually annotated by an expert : the regions are outlined, each color represents a class; excluded regions (dark green) are given as examples, which is why the image is not fully annotated.

This paper presents a new approach, based on an object-oriented analysis (segmentation, classification) to establish an automatic cartography of WSI. The main objective is to propose a decision support tool to help the pathologist to interpret the information contained by the WSI.

Compared to former works on WSI analysis, our contributions are: (i) an efficient computational framework enabling the processing of WSI in reasonable time, (ii) an efficient texture descriptor based on quantized color histograms and (iii) a multiclass supervised classification based on expert annotations allowing a complete cartography of the WSI.

The paper is organized in 4 sections. First, existing approaches to analyze WSI are presented (Sect. 2), followed by the different steps of the method (Sect. 3). Then, experiments on WSI of breast cancer samples are described to evaluate the benefits of this approach (Sect. 4). Finally, we conclude and present some perspectives (Sect. 5).

2 Related work

As optical microscopy image analysis is a specific field of image analysis, a great variety of general techniques to extract or identify regions already exists. The main distinctive characteristic of the whole slide images (WSI) is their very large size, which makes impossible the application of number of conventional processing, despite their potential interest.

Signolle and Plancoulaine [14] use a multi-resolution approach based on wavelet theory to identify the different biological components in the image, according to their texture. The main limitation of this approach is its speed: about 1 hour to analyze a sub-image of size 2048×2048 pixels, and several hundred hours for a complete image (60000×40000 pixels).

To overcome this drawback, several methods have been developed to avoid the need for analyzing entire images at full resolution. Thus, Huang et al. [7] noted that, to determine the histopathological grade of invasive ductal breast cancer using a medical scale called *Nottingham Grading System* [1,15], it is important to detect areas of "nuclear pleomorphism" (i.e. areas presenting variability in the size and shapes of cells or their nuclei), but such detection is not possible at low resolutions. So, they propose a hybrid method based on two steps: (i) the identification of regions of interest at a low resolution, (ii) multi-scale algorithm to detect nuclear pleomorphism at a high resolution in the regions of interest identified previously. In addition, through the use of GPU technology, it is possible to analyze a WSI in about 10 minutes, which is comparable to the time for a human pathologist.

Indeed, the same technology is used by Ruiz [12] to analyze an entire image (50000×50000) in a few dozen seconds by splitting the image into independent blocks. To manage even larger images (dozen of gigapixel) and perform more complex analyzes, Sertel [13] uses a classifier that starts on low-resolution data, and only uses higher resolutions if the current one does not provide a satisfactory classification. In the same way, Roullier [11] proposed a multi-resolution segmentation method based on a model of the pathologist activity, starting from the coarsest to the finest resolution: each region of interest determined at one resolution is partitioned into 2 at the higher resolution, through a clustering performed in the color space. This unsupervised classification can be performed in about 30 minutes (without parallelism) on an image of size 45000×30000 pixels.

More recently, Homeyer [6] used supervised classification on tile-based multi-scale texture descriptors to detect necrosis in gigapixel WSI in less than a minute. While tessellating an image with square tiles is simple and effective, the resulting contours can exhibit a lack of smoothness. In order to address that, our method combines supervised classification with a fast segmentation method close to the superpixel lattices described by Moore [9].

The characteristics of our method compared to others are summarized in Table 1.

3 Method

To achieve a fast and efficient classification of whole slide images, we propose a methodology enabling to partition the initial image in relevant regions. This approach is based on a superpixel segmentation algorithm and on a supervised classification using a textural characterization of each region.

Table 1: Comparison of our method with some existing ones. H&E means Hematoxylin and Eosin, a widely used staining.

Method	Pixels	Coloration	Classes	Performance	Parallelism
[12]	10^9	H&E	2 (supervised)	145 s (GeForce 7950 GX2)	GPU
[14]	10^9	H, DAB	5 (supervised)	>100 h (Xeon 3 GHz)	Unknown
[13]	10^{10}	H&E	2 (supervised)	8 min (Opteron 2.4 GHz)	Cluster of 8 nodes
[7]	10^9	H&E	4 (supervised)	10 min (GeForce 9400M)	GPU
[11]	10^9	H&E	5 (unsupervised)	30 min (Core 2.4 GHz)	Parallelizable
[6]	10^9	H&E	3 (supervised)	<1 min (Core 2 Quad 2.66 GHz)	Unknown
Proposed method	10^8	H, DAB, PRD	6 (supervised)	10 min (Opteron 2 GHz)	Parallelizable

3.1 Overview

The proposed method relies on two successive steps: (i) image segmentation into segments or "patches"; (ii) supervised classification of these segments. The main challenge of the segmentation step is to provide a relevant partioning of the image in an efficient way due to the large size of whole slide images. To cope with this problem, we propose to partition the image by using a set of horizontal and vertical optimal paths following image high gradient values. We make the assumption that the distribution of biological objects that a pathologist would use to produce a decision can be described by a textural representation of a region. Thus, the classification step is based on a textural approach where each region is labelled according to its texture description.

A training set of texture descriptors is computed from a set of manually annotated images enabling a supervised classification based on a k-nearest neighbor strategy. The method overview is illustrated in Fig. 2 and Fig. 3.

3.2 Segmentation

Let $f : E \to V$ be a 2D discrete color image defined over a domain $E \subseteq \mathbb{Z}^2$ with $V = [0, 255]^3$. Let f_i denotes the scalar image resulting of the projection of f on its i^{th} band. We suppose that E is endowed with an adjacency relation. A path is a sequence of points $(p_1, p_2, \ldots, p_n)$ such that, for all $i \in [1, \ldots, n-1]$, p_i and p_{i+1} are adjacent points. Let W, H be respectively the width and the height of f. The segmentation method is based on two successive steps.

First, the image f is partitioned into W/S vertical and H/S horizontal strips, with $S > 1$ an integer controlling the width of a strip.

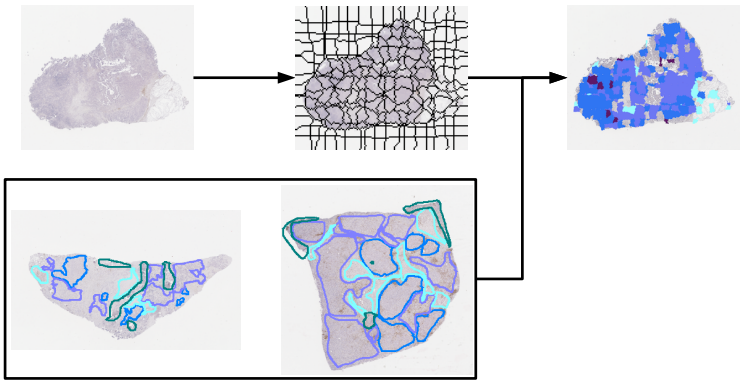


Fig. 2: Method overview. Left: image under analysis, image partitioning and regions classification. Right: manually annotated images allowing the construction of the training set.

Second, a path of optimal cost is computed from one extremity of the strip to the other in each image strip. The cost function is related to the local variations, hence favoring the optimal path to follow high variations of the image. More precisely, the local variation of f in the neighborhood of p is computed as:

$$g(p) = \max_{q \in N(p)} d(f(p), f(q)), \tag{1}$$

where d is a color distance, and $N(p)$ the set of points adjacent to p. In our experiments we used $d(a,b) = \max_i |a_i - b_i|$ (L_∞ norm) and $N(p) = \{q \mid \|p-q\|_\infty \leq 1\}$ (8-adjacency).

The global cost associated to a path $(p_i)_{i \in [1 \ldots n]}$ of length n is defined as:

$$G = \sum_{i=1}^{n} g(p_i) \tag{2}$$

From an algorithmic point of view, an optimal path maximizing this summation can be retrieved using dynamic programming [8] in linear time with respect to the number of points in the strip, hence requiring to scan all image pixels at least once. To speed up the process, a suboptimal solution is computed instead by using a greedy algorithm: starting from an arbitrary seed at an extremity of the strip, the successive points of the path are added by choosing, in a local neighborhood, the point q where $g(q)$ is maximal. By doing so, the values of g are computed on the fly, only for the pixels neighboring the resulting path.

Some notable properties of this algorithm are: (i) its speed due to the fact that not all pixels need to be processed, (ii) a low memory usage even for large images because the only required structures are the current strip and a binary mask to store the result, (iii) its potential for parallelization, because all strips in a given direction can be processed independently.

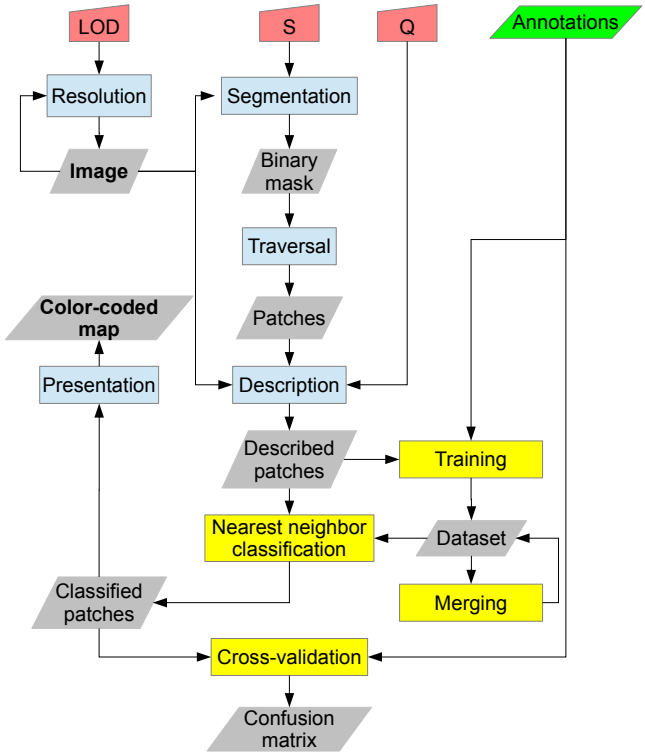


Fig. 3: Data structures and processes. Red boxes represent parameters, green box represents expert annotations (ground truth). Blue boxes are image-related operations, yellow boxes are data mining-related. Gray boxes are data structures input, exchanged, modified and output by the processes.

Fig. 4 illustrates the steps of the segmentation method and Fig. 5 gives an example of the end result.

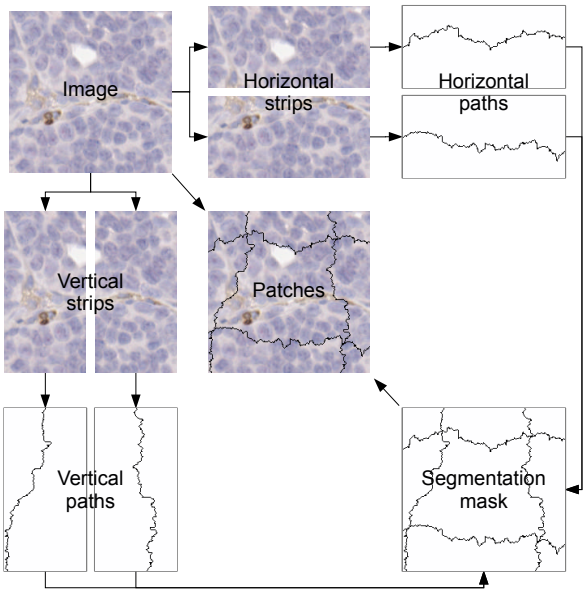


Fig. 4: Path computation in horizontal and vertical strips, leading to an image partition.

3.3 Training

To create a training base, the reference images are segmented into patches. Using the expert annotations, each patch is associated with a class or label. Then, by computing a texture descriptor for each patch, we can create an association between a texture and a group of labels. A texture can have several labels if it is present in regions of different classes. As a result, the training base can be modeled as a function $B : T \to G$ where T is the set of texture descriptors and $G = \mathcal{P}(L)$ is the power set of all labels L.

When $|B(t)| \neq 1$, the texture t is ambiguous. Section 4.2 describes how to measure this phenomenon and thus quantify the validity of the model. In order to perform the classification, all $B(t)$ must be singletons. To that end, B is updated so that ambiguous textures are classified as excluded elements.

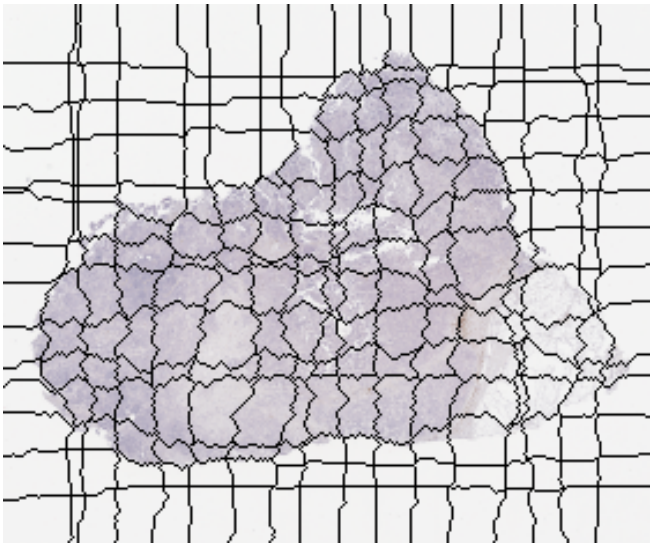


Fig. 5: Example of image partitioning based on our algorithm.

3.4 Classification

Some authors use distributions of descriptors to describe textures [10]. The chosen descriptors can be arbitrarily complex, and, as a starting point, we decided to use simple color histograms that are functions $H : V \rightarrow [0, 1]$ that associate each pixel color to its frequency in a given patch. The ability of histogram to discriminate between textures is illustrated in Fig. 6. Given the fact that all images were obtained using the same process and equipment with the same settings, no image preprocessing was deemed necessary.

In order to perform a supervised multi-class classification, we opted for a one nearest neighbor classification because of its simplicity (no assumption needs to be made on the distribution of the textures in the descriptor space) and the sufficient amount of training samples available (which turned out to be a bit excessive with sometimes up to several millions of elements). For this kind of classification, we measure the distance between histograms using the euclidean metric:

$$d\left(h_1, h_2\right) = \sqrt{\sum_v \left(h_2\left(v\right) - h_1\left(v\right)\right)^2} \tag{3}$$

This choice of metric is arbitrary and will serve as a reference for future work.

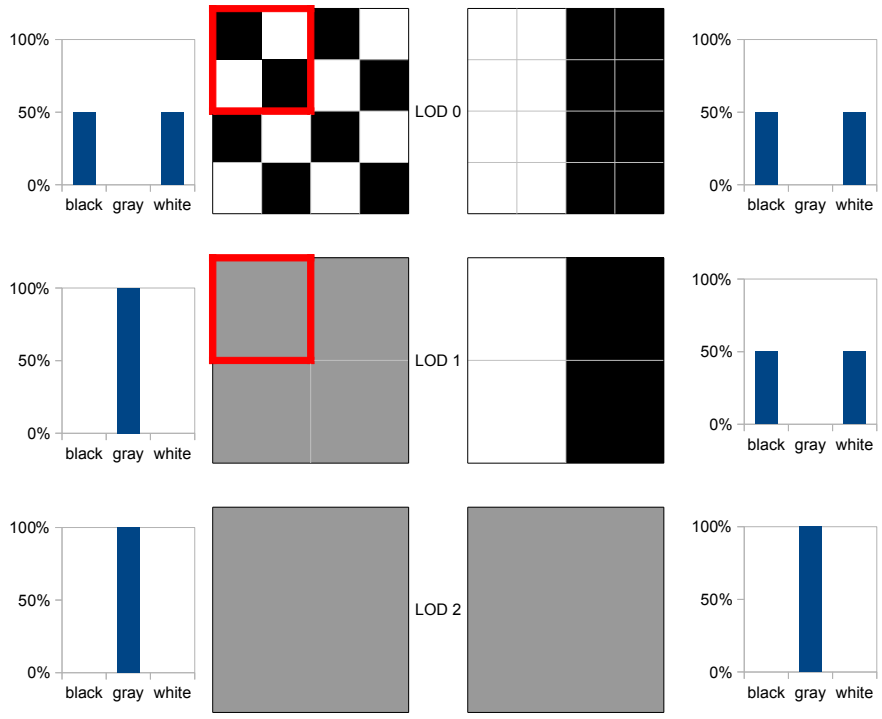


Fig. 6: The most common argument against the use of histograms for texture characterization can be visualized in the top row : the two presented images cannot be distinguished by their histograms. But by performing a subsampling operation to reduce the resolution (example outlined in red), local constraints are introduced and the middle row shows that the two images can now be distinguished by their histograms. This ability disappears in this example after further subsampling, suggesting the possible existence of an optimal level of resolution to characterize a given set of textures with their histograms.

4 Experiments and results

4.1 Data

Unlike most of the work published on the subject, our images are obtained using a double-staining process [16]. More precisely, we used formalin-fixed paraffin-embedded breast cancer samples obtained from Indivumed ®, Hamburg, Germany. Manual immunohistochemistry staining was performed for CD8 or CD3 and Perforin, and antibody binding was visualized using 3,3-diaminobenzidine tetrahydrochloride (DAB, Dako, Hamburg, Germany) and Permanent Red (PRD, Zytomed, Berlin, Germany). Cell nuclei were counterstained with hematoxylin before mounting. As a result, cancerous (in this type of tumor predominantly large) and noncancerous (predominantly small) cell nuclei appear blue (hematoxylin). The chromogenic labeling of the lymphocyte lineage markers CD3 and CD8 results in brown (DAB) staining of cell membranes and/or cytoplasm; the antibody labeling for perforin is visualized in red (PRD) as granular cytoplasmic color dots. Due to the small size of cells and sectioning effects, the blue nuclei may sometimes be covered by brown or red color.

For our experiments, 7 whole slide images ranging from $1 \cdot 10^8$ to $5 \cdot 10^8$ pixels have been annotated by a pathologist using 6 simplified labels (Fig. 7): invasive tumor (defined as predominantly solid formations), invasive tumor (simplisticly defined as less coherent tumor cell groups diffusely infiltrating pre-existing tissues), intersecting stromal bands (defined as the non-malignant mesenchymal tissue component regardless whether pre-existing, or induced by tumor growth), DCIS (Ductal Carcinoma In Situ; in a simplified manner this class encompasses real ductal invasive carcinoma in situ, and invasive tumor accidentally growing within ductal structures), non-neoplastic glands and ducts, and edges and artifacts to be excluded.

The annotations do not explicitly provide quantifiable cell characteristics that could be used to design a medically relevant cell-based region identification. Instead, they take the form of outlines that may or may not match visual features (Fig. 1). Even though some of these seem obvious, like the background, it is not easy for an untrained eye to establish a set of intuitive rules that would explain the expert's opinion, even after trying some simple visual filters (quantization, thresholding). Moreover, the classes are not uniformly represented in our data (Fig. 8): while excluded elements are described by only a handful of annotations, they actually account for the majority of the area of the images, especially because of the background; on the other hand, ductal structures constitute a minority and are sometimes completely missing.

Nonetheless, the delimited regions appear to exhibit a texture-related behavior, and we can use that to decide on a model: a delimited region is made of a set of patches that can be identified by their texture, and delimited regions of the same class share the same set of textures. Thus, by partitioning the image into patches and labeling each patch based on its texture, we can draw a color-coded map like in Fig. 9.

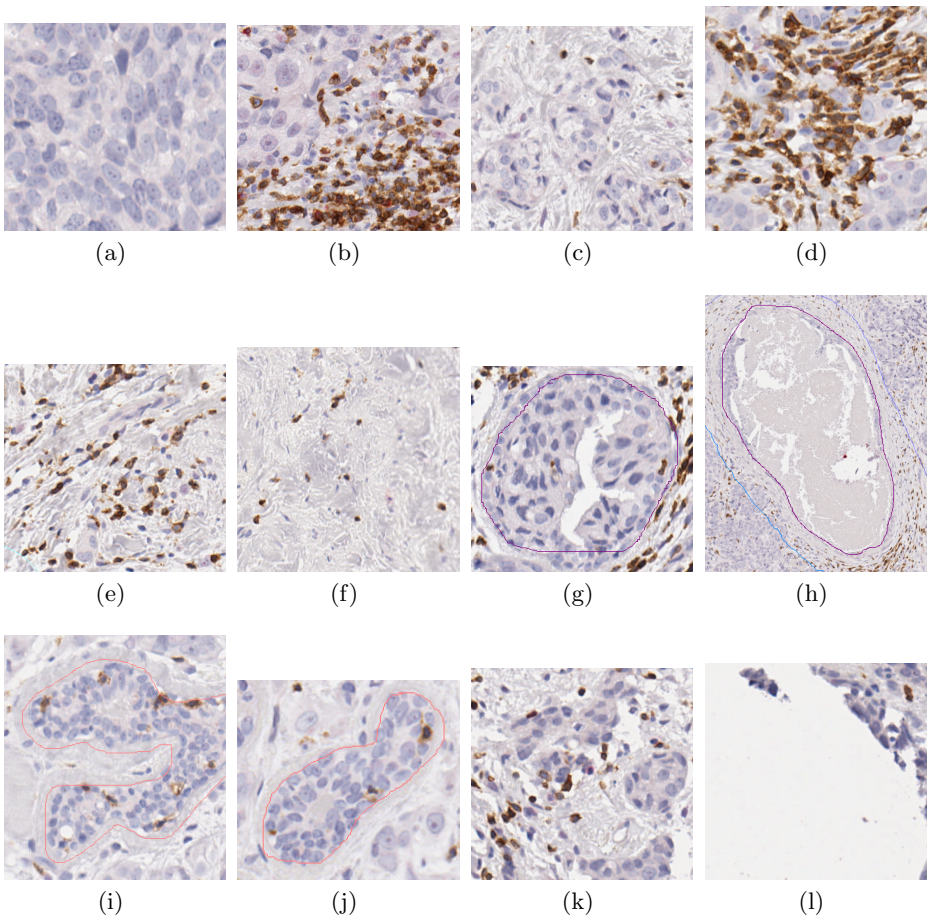


Fig. 7: (a,b) Annotation: Invasive tumor (solid formations).
Description: High concentration of cancerous cells.
Note: All classes can contain foreign objects, such as the brown lymphocytic infiltrate that can be seen in (b), and sometimes the same objects (foreign or not) can be seen in several classes (compare with c,d).
(c,d) Annotation: Invasive tumor (diffusely infiltrating pre-existing tissues).
Description: Cancerous cells disseminated in noncancerous tissue.
(e, f) Annotation: Intersecting stromal bands.
Description: Connective tissue.
(g, h) Annotation: DCIS and invasion inside ductal structures.
Description: A Ductal Carcinoma In Situ refers to cancer cells within the milk ducts of the breast; the simplified definition used here encompasses "real" ductal invasive carcinoma in situ, in the sense of a ductal proliferation respecting the anatomical structure of ducts, and invasive tumor components growing withing pre-existing ducts; some examples may include central necroses as shown in (h).
(i, j) Annotation: Nonneoplastic glands and duct.
Description: Noncancerous structures.
(k, l) Annotation: Edges and artefacts to be excluded.
Description: Nonbiological features (background, smudges, bubbles, blurry regions, technician's hair, ...) and damaged biological features (borders, defective coloration, missing parts, ...).

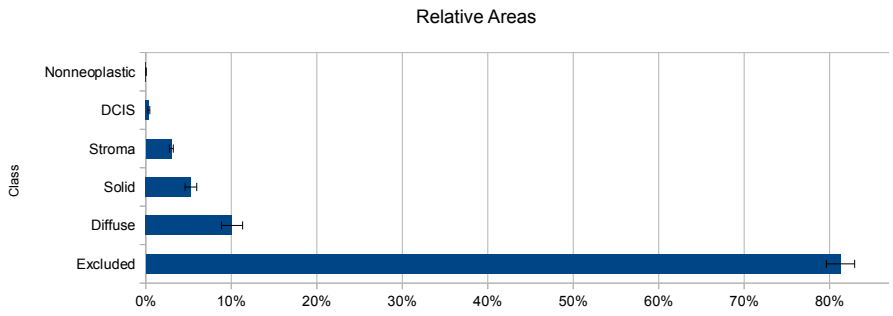


Fig. 8: Relative areas of all the classes: average for each training set. Standard deviation is given as horizontal bars.

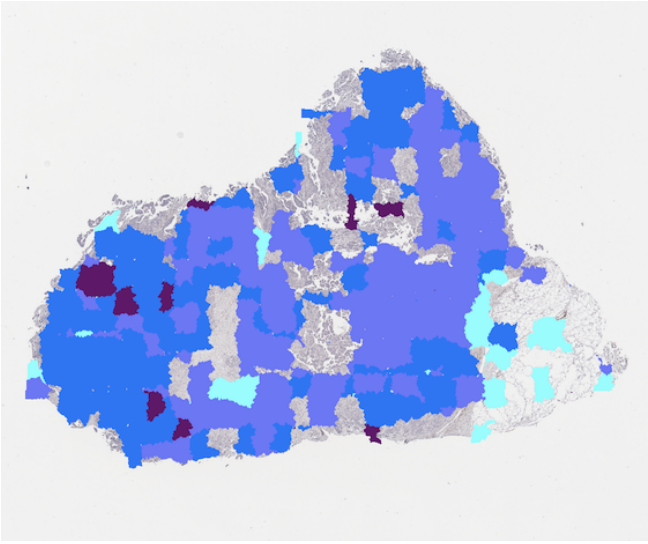


Fig. 9: Classification map obtained by the presented method; the colors match those used by the expert, except for the excluded regions which are left untouched.

4.2 Model Evaluation

The method is evaluated with a leave-one-out cross-validation involving all the annotated images: for each image (in our set of 7), a training base is created with the other 6. All the values given in the rest of this article are obtained by averaging the values from 7 experiments.

The quality of the model can be measured by computing the certainty of the training base for each label l:

$$C(l) = \frac{|B^{-1}(\{l\})|}{|\{t \in T : l \in B(t)\}|} \tag{4}$$

When the certainty is 100%, it means that the only group containing the label is a singleton, and so the textures can be used to uniquely characterize the corresponding class. On the other hand, a certainty of 0% means that the textures are too ambiguous for a one-to-one mapping.

Since a human pathologist uses a multi-resolution approach [11], a whole slide image is typically provided as a set of images corresponding to different magnifications that can be used by visualization software to speed up display. But they restrict the systematic study of the impact of the resolution level, and can also cause additional degradation due to lossy compression. So, in order to determine the information available at each level of detail (LOD), we compute for each image I a pyramid defined by:

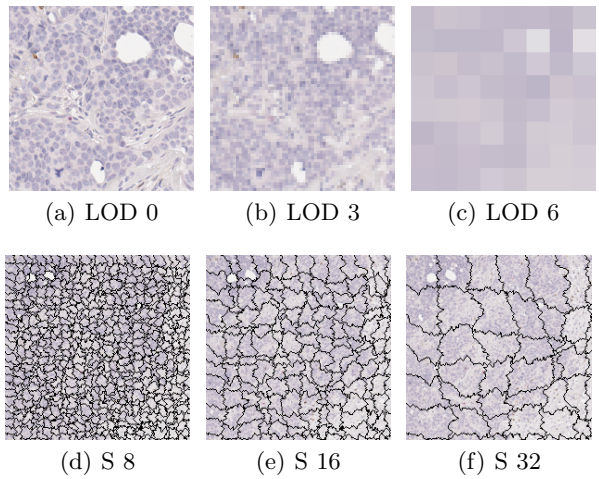
$$I_{LOD}(x, y) = \frac{1}{4} \sum_{(i,j)\in\{0,1\}^2} I_{LOD-1}(2x + i, 2y + j) \tag{5}$$

The original image is at LOD 0 (Fig. 10). It can be observed that high resolution is correlated with high data set certainty for the chosen texture descriptor (Fig. 11).

Ideally, the segmentation algorithm should create patches of the right size, so that each patch would contain just enough information to identify a class-characteristic texture. Instead, we will assume the existence of a common texture scale that applies to all classes: the segmentation parameter S (Fig. 10). At high resolution, a texture described by its color histogram can help identify a class with very little doubt (Fig. 11). But at lower resolutions, larger values of S increase the ambiguity of the texture description, because the patches become large enough to contain multiple textures from adjacent regions of different classes.

Finally, despite the use of sparse structures, a one nearest neighbor classification using color histograms requires large amounts of memory and processing time. A simple yet effective technique to mitigate this issue is to use a quantization scheme where the values used as histogram keys are $2^Q \lfloor \frac{v}{2^Q} \rfloor$ instead of v, and Q is the quantization parameter. Less memory is required because textures with close descriptors are merged. As Fig. 11 shows, a mild quantization ($Q \leq 4$) barely affects the certainty of the training base.

By considering only this measure of the training base quality, we would expect to get the best results with high resolution, small patches and minimal



(a) LOD 0 (b) LOD 3 (c) LOD 6

(d) S 8 (e) S 16 (f) S 32

Fig. 10: (a-c) Visualization of the effect of the resolution parameter LOD on pixel data. (d-f) Visualization of the effect of the segmentation parameter S on pixel data.

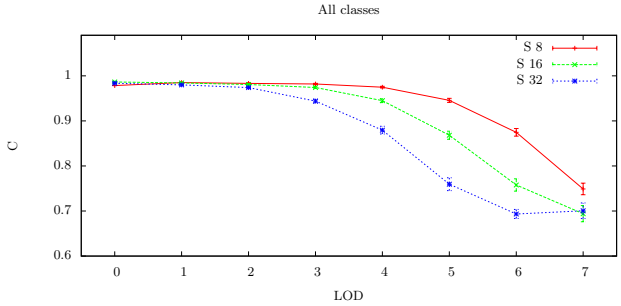
quantification. But we experimentally determined that the configuration (LOD 4, Q 4, S 32) yielded the best overall outcome when plotting the data in ROC space (Fig. 11). The discrepancy between high training base certainty and lesser classification results can have several causes, explored in the next section.

4.3 Classification results

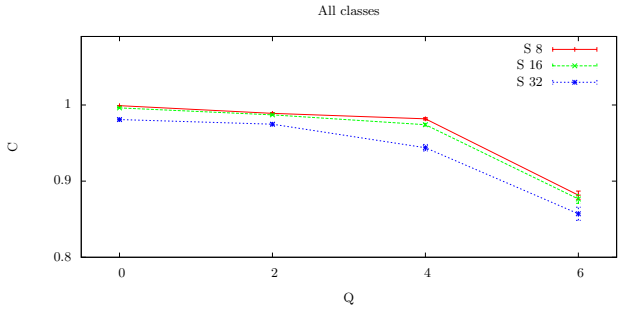
By observing the confusion matrix for a chosen set of parameters (table 2), we can see that the class of the excluded regions is the only one to be adequately detected.

The structures of the DCIS class are difficult to identify because of the ambiguity between seemingly ductal structures filled with cancerous cells, but also due to their relative rarity in the current series of images. The latter holds also true for the class of non-neoplastic ductal structures, which were rare in the current series of WSIs as well. As expected for a highly differentiated epithelial structure, the glands and ducts of the pre-existing mammary gland tissue were more prone to textural characterization, but more examples would be needed for confirmation.

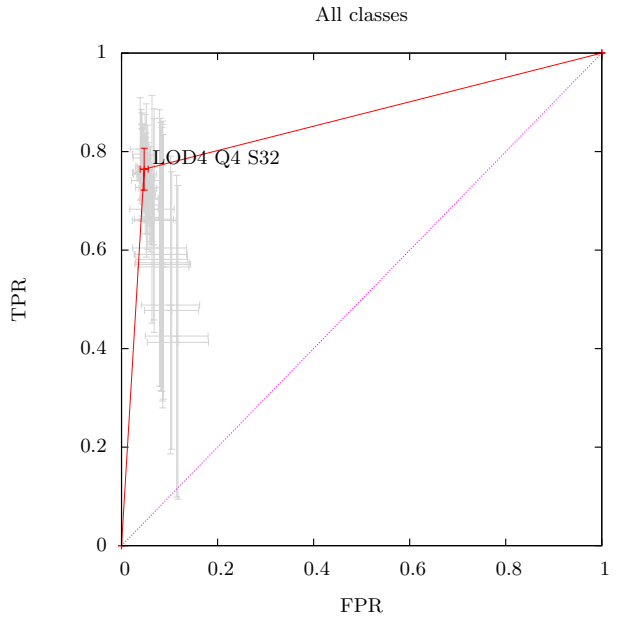
The remaining 3 classes illustrate some limitations of the model. The "stroma" class is detected as "excluded regions", "stroma" and "diffuse invasion". As it turns out, "diffuse invasion" means that textures corresponding to cancerous cells are mixed with textures corresponding to stroma. This mixing creates conflicts which are resolved by assigning the "excluded regions" class to the ambiguous textures (section 3.3). The same phenomenon explains why both "solid formations" (regions of high cancerous cell density) and "diffuse invasion" (regions of low cancerous cell density) are detected as a mixture of "excluded regions",



(a)



(b)



(c)

Fig. 11: (a) Visualization of the effect of the resolution parameter LOD on the overall certainty of the training base for different values of S and Q = 4. Standard deviation is given as vertical bars.
(b) Visualization of the effect of the quantization parameter Q on the overall certainty of the training base for different values of S and LOD = 4. Standard deviation is given as vertical bars (barely visible because they are small).
(c) ROC points (light gray) for each parameterization are obtained by merging experimental data points for all the classes. For clarity, a convex curve (red) has been synthesized from these points with one point that stands out; such a synthetic curve makes sense because a classifier can be built for any interpolated point [2]. Standard deviation is given as horizontal and vertical bars.

"solid formations" and "diffuse invasion". Regarding the distinction between "solid formations" and "diffuse invasion", it is important to note that the current annotation is by definition a preliminary one based on few samples that needs to be iteratively improved in further studies. There is a poorly defined range of variation between what descriptive reports would consider "diffusely infiltrating" versus "solid", or "coherent" growth, and this variability will even increase with inclusion of breast cancer subtypes other than "unspecific type" (synonymous with the former designation "ductal invasive"). One possible workaround for this problem could be to merge the two classes into one "tumor are" class, which could result in a much higher detection rate of up to 90%. Another way to address the challenge of tumor heterogeneity is to investigate a broader range of samples and work on a closer approximation towards widely accepted annotations concordant between experienced pathologists.

But the major underlying problem is that the model is flawed: a texture unit defined by one patch is not enough to identify a class. The certainty of the training bases is high because, unexpectedly, simple color histogram are not only strong enough to describe such texture units, but they also capture small variations that can almost identify the patches themselves, hence the large size of the training bases (see next section). We conjecture that if we were to ignore these variations, we would obtain a small set of textures (maybe a few dozens) that could be used to identify regions like a pathologist does. With that in mind, the confusion matrix could suggest that regions corresponding to stroma, diffuse invasion and solid formations are made up of the same set of textures (as described by color histograms of patches), but they differ by the proportions of these textures; that phenomenon could be quantified by measuring their local density distributions.

Another clue in support of that conjecture is the specificity data (table 3). While the presence of a given texture is not always enough to identify only one given class, its presence might still be a necessity and thus its absence can reliably be used to exclude some classes, especially for stroma (sparsely populated regions with a "sinuous" appearance) and solid formations (dense clumps of large cancerous cell nuclei). The specificity for the excluded regions is lower because of its role as "default class" to resolve ambiguities in our current method, as explained previously. At this time, we believe that we don't have enough data on DCIS and nonneoplastic objects to draw a definite conclusion on these classes.

Additional experiments were performed on small computer generated texture mosaics available online as part of the Prague Texture Segmentation Datagenerator and Benchmark [5]. On the normal color set comprising 89 textures in 20 images, the parameterization (LOD 0, Q 5) with square patches of size 16x16 yielded a false positive rate of $0.2\% \pm 0.2$ and a true positive rate of $79.8\% \pm 15.6$. At the time of testing, the benchmark website attributed to the results a "correct segmentation" score of 55.44, corresponding to the 16th place. A detailed summary indicated that the quality of the classification depends on the nature of the texture, with textile (Fig. 12) and wood being better characterized than glass and plants. It is interesting to note that despite $(0.2\%, 79.8\%)$ being a seemingly

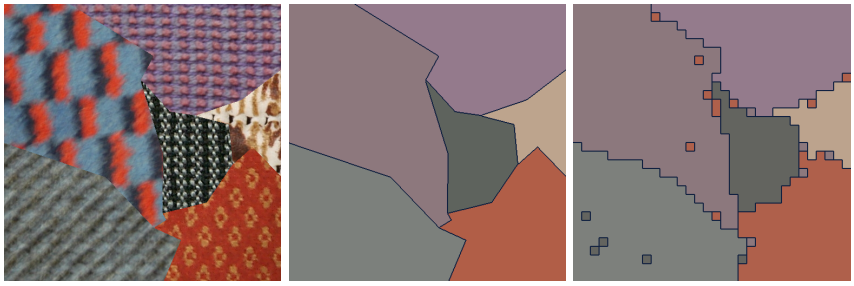
Table 2: Confusion matrix for a particular parameterization (LOD 4, Q 4, S 32). Results are given as mean and standard deviation computed from 7 values.

	DCIS	Excluded regions	Stroma	Solid formations	Nonneoplastic objects	Diffuse invasion
DCIS	**57**% ± **49**	29% ± 35	2% ± 5	4% ± 4	0.1% ± 0.2	8% ± 10
Excluded regions	0.3% ± 0.4	**85**% ± **3**	3% ± 1	3% ± 2	0.2% ± 0.2	9% ± 3
Stroma	1% ± 3	29% ± 10	**31**% ± **10**	4% ± 3	0.1% ± 0.2	25% ± 10
Solid formations	0.9% ± 2	19% ± 14	0.5% ± 0.6	**44**% ± **32**	0.0% ± 0.1	35% ± 24
Nonneoplastic objects	0% ± 0	11% ± 20	0.6% ± 1	6% ± 14	**71**% ± **45**	12% ± 18
Diffuse invasion	0.2% ± 0.2	37% ± 12	3% ± 3	11% ± 9	2% ± 3	**47**% ± **15**

Table 3: Sensitivity and specificity for a particular parameterization (LOD 4, Q 4, S 32). Results are given as mean and standard deviation computed from 7 values.

Class	Sensitivity	Specificity
DCIS	57% ± 49	99.6% ± 0.6
Excluded regions	85% ± 3	67% ± 13
Stroma	31% ± 10	97% ± 2
Solid formations	44% ± 32	96% ± 3
Nonneoplastic objects	71% ± 45	99.8% ± 0.2
Diffuse invasion	47% ± 15	89% ± 4

good ROC point, the segmentation score is only 55.44, due to the distribution of the textures in the images.



(a) (b) (c)

Fig. 12: Partial application of the method to an image of a texture segmentation benchmark available online (http://mosaic.utia.cas.cz): (a) original image; (b) ground truth; (c) actual result.

4.4 Performance

The experiments were run on an AMD Opteron 2 GHz with 32 Gb of memory. The segmentation step takes at most a few minutes even on large images (less than 2 minutes for a gigapixel image with our current sequential Java implementation). But, as shown on Fig. 13, the main bottleneck of the method is the size of the training bases, mostly because of the time needed to search a nearest neighbor. So far, this has prevented us from testing our current algorithm with high resolutions but we have verified that capping the training base size to 10000 elements (which is still large) can bring down the computing time to less than 2 hours for the highest resolution. That being said, our current results suggest that lower resolutions may already have enough information to completely analyze the image.

5 Conclusion and future work

The advent of whole slides images is a great opportunity to provide new diagnostic tools and to help pathologists in their clinical analyses. However, it also comes with great challenges, mainly due to the large size of the images and the complexity of their content. To achieve a fast and efficient classification of the images, we proposed in this paper a methodology enabling to partition the initial image in relevant regions. This approach is based on an fast segmentation algorithm and on a supervised classification using a textural characterization of

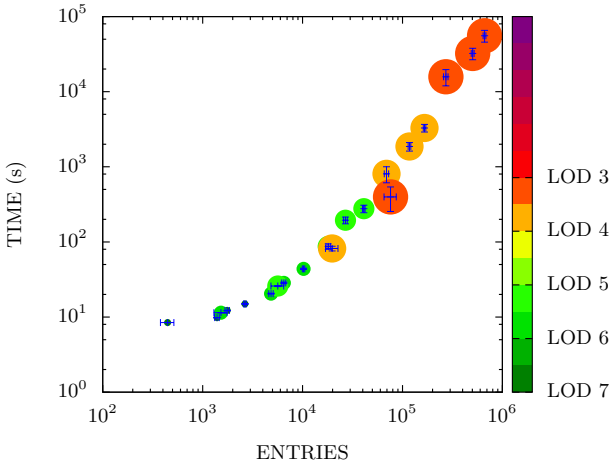


Fig. 13: Mean sequential calculation time by image according to the size of the training base. Each point corresponds to a setting (LOD, Q) for a segmentation of size 32 pixels. The colored discs symbolize the image size, related to the parameter LOD. Some experiments at high resolution were not performed because of excessive time and memory requirements. Standard deviation is given as horizontal and vertical bars.

each region. We carried out experiments on 7 annotated images and obtained promising results.

In the future, we will extend the level of detail for annotations and increase the range of tumor variability in order to identify biologically relevant structures more precisely. We are confident that accurate automated detection of clinically relevant regions of interest in cancer-related WSIs for subsequent in-depth analysis is a key contribution to the development of novel tools for biomarker discovery and validation. Increasing speed and accuracy of digital pathology workflows will support the implementations of automated analysis modules into the diagnostic work-up and thereby help to improve cancer therapy directed against targets that are detectable in tissue biopsies.

References

1. Elston, C.W., Ellis, I.O.: Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: Experience from a large study with long-term follow-up. Histopathology (1991)
2. Fawcett, T.: An introduction to roc analysis. Pattern Recognition Letters (2006)
3. Ghaznavi, F., Evans, A., Madabhushi, A., Feldman, M.: Digital imaging in pathology: Whole-slide imaging and beyond. Annual Review of Pathology (2013)
4. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: a review. IEEE Reviews in Biomedical Engineering (2009)

5. Haindl, M., Mikeš, S.: Texture segmentation benchmark. Proceedings of the 19th International Conference on Pattern Recognition (2008)
6. Homeyer, A., Schenk, A., Arlt, J., Dahmen, U., Dirsch, O., Hahn, H.K.: Practical quantification of necrosis in histological whole-slide images. Computerized Medical Imaging and Graphics (2013)
7. Huang, C.H., Veillard, A., Roux, L., Loménie, N., Racoceanu, D.: Time-efficient sparse analysis of histopathological whole slide images. Computerized Medical Imaging and Graphics (2010)
8. Montanari, U.: On the optimal detection of curves in noisy pictures. Communications of the ACM (1971)
9. Moore, A.P., Prince, S.J.D., Warrell, J., Mohammed, U., Jones, G.: Superpixel lattices. IEEE Conference on Computer Vision and Pattern Recognition (2008)
10. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition (1996)
11. Roullier, V., Lézoray, O., Ta, V.T., Elmoataz, A.: Multi-resolution graph-based analysis of histopathological whole slide images: application to mitotic cell extraction and visualization. Computerized Medical Imaging and Graphics (2011)
12. Ruiz, A., Sertel, O., Ujaldon, M., Catalyurek, U., Saltz, J., Gurcan, M.: Pathological image analysis using the gpu: Stroma classification for neuroblastoma. 2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007) (2007)
13. Sertel, O., Kong, J., Shimada, H., Catalyurek, U.: Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. Pattern Recognition (2009)
14. Signolle, N., Plancoulaine, B., Herlin, P., Revenu, M.: Texture-based multiscale segmentation: Application to stromal compartment characterization on ovarian carcinoma virtual slides. Image and Signal Processing (2008)
15. Tavassoli, F.A., Devilee, P.: Pathology and Genetics of Tumours of the Breast and Female Genital Organs. IARCPress (2003)
16. Wemmert, C., Krüger, J., Forestier, G., Sternberger, L., Feuerhake, F., Gançarski, P.: Stain unmixing in brightfield multiplexed immunohistochemistry. IEEE International Conference on Image Processing (2013)