

TMARKER: A free software toolkit for histopathological cell counting and staining estimation

Peter J. Schüffler^{*1,5}, Thomas J. Fuchs², Cheng Soon Ong³, Peter J. Wild⁴,
Niels J. Rupp⁴, and Joachim M. Buhmann^{1,5}

¹ Institute for Computational Science, ETH Zurich, Switzerland

² California Institute of Technology, Pasadena, CA 91125, USA

³ National ICT Australia

⁴ Institute of Pathology, University Hospital Zurich, Switzerland

⁵ Competence Center for Systems Physiology and Metabolic Diseases, ETH Zurich

Abstract. Pathological tissue rating often involves manual cell counting and percentage staining estimation of cancerous cells. We present in this paper a novel, free, and open-source software toolkit that connects already available workflows for computational pathology and immunohistochemical tissue assessment with modern active learning algorithms from machine learning and computer vision. This new connection together with a user-friendly Java GUI enables comprehensive computational assistance in the pathological tissue rating. We introduce the platform independent program and show on a test dataset of human renal clear cell carcinoma and rabbit prostate carcinoma the validity of the used algorithms.

Keywords: Pathology, Staining Estimation, Nuclei Detection, Color Deconvolution, Superpixel Classification, Segmentation

1 Introduction

Accurate immunohistochemical staining estimation of human tissue plays a crucial role in various kinds of clinical applications and medical research. One example is the estimation of MIB-1 for assessing the proliferation factor of renal clear cell carcinoma or prostate cancer from a biopsy [1], where pathologists observe morphological and immunohistochemical (IHC) characteristics of a given tissue. IHC properties can be observed by staining thin tissue slices attached on a glass plate with a protein specific antibody that is linked to a dye. Thus, cell nuclei in the tissue layer that express this protein (“positive” for this protein) will exhibit e.g. a dark brown color. Other cells (“negative”) will lose the antibody in a subsequent washing procedure. Since the whole tissue slice is unspecifically colored with hematoxylin, all negative cell nuclei show a light blue color.

* Corresponding Author, peter.schueffler@inf.ethz.ch

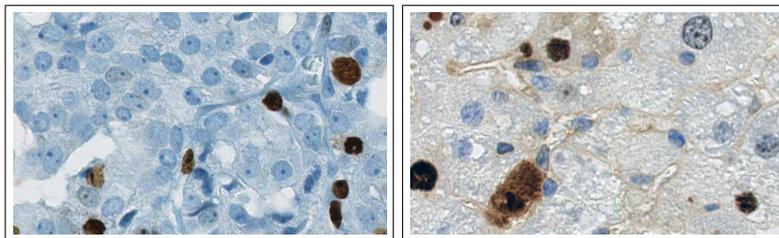


Fig. 1. Two parts of typical images on which cell nuclei are to be counted and staining percentage is to be estimated. Blue spots are unstained nuclei, brown spots are stained nuclei. Only cancer cells are relevant for counting.

In medical research, the IHC estimation is often done on tissue micro arrays (TMA). The enormous advantage of TMAs is the fact that the tissue samples from many patients of a study cohort are stained simultaneously on one micro array, such that the experimental settings do not change between the different patients. The scientific goal is then the screening for differences of protein expression patterns in different patient groups. Subsequently, such proteins might serve as biomarkers for the respective patient groups.

Pathologists generally perform IHC estimation by eye with light microscopy or high resolution scans (see Fig. 1). For a typical image, two relevant questions for these types of problems are: (i) How many malignant or benign cells are present in the image? (ii) How many nuclei **of the cancer cells** express the protein under consideration?

These questions can be formulated as problems in medical image processing and computer vision. Note, that there is a difference between malignant or benign cells and stained or unstained cells. The staining of the nuclei only reveals presence of protein or absence. The distinction of malignant and healthy cells is considerably more difficult and relies on parameters like size, shape and morphology of the cell nuclei but not necessarily on the IHC staining [2].

1.1 Motivation

High throughput IHC staining estimation of tissue images poses several challenges in practice. By nature, IHC stained images are much more difficult to analyze than e.g. immunofluorescence images (IF), where cell nuclei can be separated much more easily from the comparatively homogeneous background. Further, in contrast to IF, the morphological structures of the tissue is much better visible in IHC images. In addition to disrupted stained cells, cell compartments and the extracellular matrix, these structures might disturb the perception of a nucleus. In the manual grading, this leads to highly subjective and biased cell counts, even by trained human experts. Experiments have shown, that manual

counting of cells is very subjective and not easily reproducible [3]. Further, the manual assessment of TMAs is very time consuming, prone to error and expensive. To this end, several approaches have been presented to assist or automate the whole estimation process ([2], [4]). These approaches have shown promising results with respect to nuclei detection accuracy and their performances are comparable to trained pathologists. However, while there exist a number of general purpose medical imaging toolkits and commercial solutions, to the best of our knowledge, no software package especially tailored to nuclei counting and nuclear IHC staining estimation of human tissue is available publicly and free of charge for the research community.

1.2 Own Contribution

To provide medical researchers and clinical pathologists a software package to alleviate the aforementioned tasks, we present TMARKER, a user-friendly, freely available and platform independent toolkit to assist in cell nuclei counting and staining estimation of IHC stained tissue on whole tissue slides and TMAs (see Fig. 2). The software aims to fit following needs:

- Semiautomatic, reproducible, and fast cell nuclei detection and counting in a given set of IHC stained images,
- automatic classification of cells into malignant and benign (on the basis of the frameworks presented in [2] and [4]),
- platform independence, free availability, user-friendly Java webstart GUI.

The program is implemented in Java v1.6 and can be executed on any client with a Java virtual machine. It is publicly available at <http://www.comp-path.inf.ethz.ch>.

In Section 2, we report algorithmic details, as well as implementation details. In Section 3, we describe the results from several experiments, which demonstrate a performance comparable to trained pathologists. Finally, we discuss the utility of the software and the achieved results and give prospects to further developments in Section 3.

2 Methods

2.1 Superpixels and Active Learning based Nucleus Detection and Classification

Superpixels. In the scope of human tissue images, superpixels can be used to segment cell nuclei as well as other structural compartments. For this purpose, the size of superpixels should roughly cover a typical size of a nucleus. TMARKER finds the correct size and number of superpixels by $n = (w * h) / (4 * r^2)$, where w and h are image width and height and r is the nucleus radius. For superpixel formation, we use an adapted and natively in Java implemented version of the simple linear iterative clustering (SLIC) superpixel algorithm as introduced by [5].

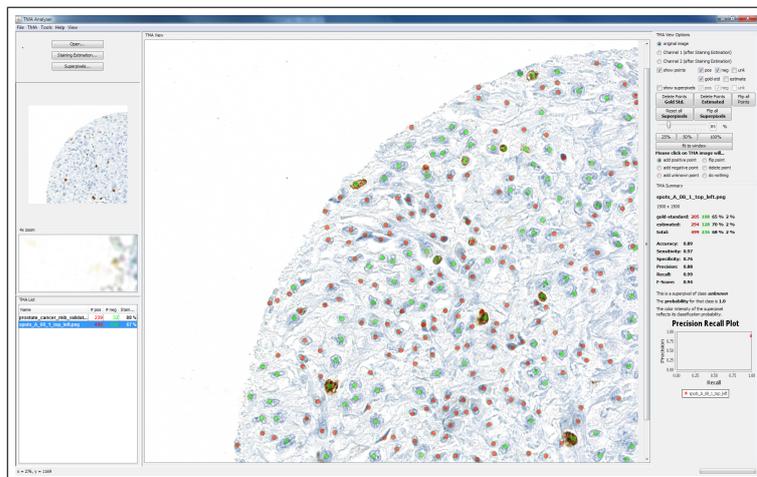


Fig. 2. Screenshot of TMARKER with a TMA image of MIB-1 stained renal clear cell carcinoma. Detected cancerous and benign nuclei are marked with red cancerous and green benign points.

Nucleus Detection. For the detection of cell nuclei, the whole tissue image is partitioned into superpixels. The superpixels are then considered the samples for training a classifier. From each superpixel, a feature vector based on the underlying image properties is calculated. We implemented three feature extractors, which proved valuable for histology in the past: color histograms (3×16 bins), local binary patterns (LBP) [6] (size 256), and pyramid histograms of oriented gradients (PHOG) [7] (size 338). The features are concatenated resulting in a feature vector of size 642. The default classifier in TMARKER is a random forest [8] (WEKA package, [9]), although support for Support Vector Machines (SVM) [10] and Bayesian Networks [11] is implemented. Based on the labels provided by the pathologist, the classifier learns to discriminate between superpixels which represent a nucleus (foreground) and superpixels belonging to the background. The foreground superpixels are subjected to the nucleus classification.

Nucleus Classification. After the detection of the cell nuclei, the goal is to classify them into malignant (cancerous) and benign (normal). To this end, we are using the same feature vector as before, but the classifier is now trained only on cell nuclei labeled by the domain expert. The superpixels corresponding to detected nuclei from the step before are hence classified into malignant and benign. These classification are visualized on the histological image so that the pathologist is able to correct and retrain the classifier (see Fig. 3).

Active Learning. Active learning describes a learning method in which the learner is able to choose the (most informative) training samples [12]. TMARKER

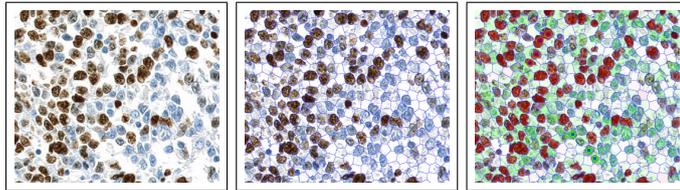


Fig. 3. Superpixel algorithm. **Left:** Part of the original image. **Middle:** The image is segmented into superpixels. **Right:** Superpixels are classified into red and green superpixels (positive and negative). As training set serve the labels of the user (red and green circles). The color intensity reflects the classification probability.

provides a simple active learning algorithm which we call “Semiautomatic Labeling”. The user starts labeling the image with positive (i.e. cancerous) and negative (i.e. benign) nuclei. Optionally, the background of the image can also be labeled in addition. With each label provided by the user, TMARKER re-trains the classifier and updates the visualization of the classification results for all superpixels. By iterating this process, the user improves the classification results by continually labeling the image. For the active learning, the user labels superpixels with low classification confidence (high uncertainty) and thus improving the discriminative classifier at the decision boundaries. The algorithm can be trained over several images, to cover the larger variance among different specimens. Once trained, the classifier can be applied to any new images in a high-throughput manner.

Validation. Within the framework, it is important to distinguish between gold-standard nuclei (GS), which were labeled and classified by the user, and estimated nuclei (ES) identified by the system. To evaluate the performance of the presented algorithms, we calculate the match statistics between GS and ES. Two points with the distance d are matched to each other, if $d \leq 2r$, where r is the nucleus radius. Based on this distance, *precision*, *recall* and *F-Score* are measured. Subsequently, the detected nuclei are classified into malignant and benign and *sensitivity*, *specificity* and overall classification *accuracy* are calculated.

2.2 Color Deconvolution based Nucleus Detection

In cases when staining estimation is performed without nucleus classification (i.e. the cells are homogeneous on the image), color deconvolution provides a fast alternative to the superpixel approach. Color deconvolution enables nuclei detection based on the method presented in [13]. The image is deconvolved into separate color channels (e.g. hematoxylin channel and DAB channel), which are smoothed with a Gaussian blur filter and subsequently screened for local intensity maxima (see Fig. 4). These steps are performed with ImageJ for Java [14]. A few parameters are needed for local maxima detection: The radius r of

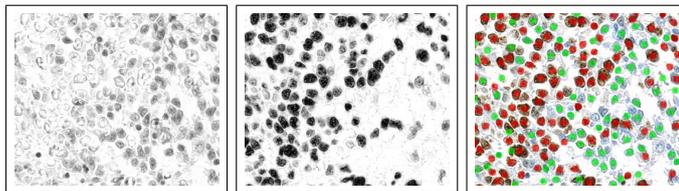


Fig. 4. Color Deconvolution: **Left:** The hematoxylin channel image. **Middle:** The DAB channel image. **Right:** Found nuclei based on the intensities on the two channels and the nucleus radius r .

cell nuclei describing the size of the local environment, and an intensity threshold t per channel, above which a local maximum is accepted. Since these parameters vary between experimental protocols for the preparation of tissue, TMARKER provides visual assistance to select the parameter values. Interactively, the user gets immediate feedback on changes to the parameters.

3 Results & Discussion

Superpixels and Active Learning. We present a new software toolkit called TMARKER which is suitable for nucleus classification (malignant/benign vs. stained/unstained). TMARKER uses a superpixel based approach for classification. We show in Fig. 5 (left) that superpixels are suitable for image segmentation and classification of histopathological images. The quantitative detection accuracy (92%) and classification accuracy (64%) touches the range of the interpathologists error (97%/74%) even at such a difficult problem. Thus, the detections and classifications of cell nuclei are comparable to these of pathologists. Moreover, we show in Fig. 5 (right) that active learning approach profits from user input especially on borderline cases. The classification accuracy saturates already after 160 labels with systematic user input, instead of 360 user interactions with randomly labeling.

Color Deconvolution. The nucleus detection and staining estimation with color deconvolution provides facilitated parameter settings but no classical machine learning influence. If the pathology goal is not dependent on nucleus type, or the nucleus types are known for the given image set, this method is a fast alternative to the more comprehensive classification. As shown in Fig. 6, TMARKER achieves a reproducible precision and recall in nucleus detection. The performance hereby is still comparable to two individual pathologists, as measured by their inter-precision and inter-recall.

TMARKER is a free software with high potential in cell counting and staining estimation of pathological IHC stained tissue images. A major advantage of TMARKER is the high reproducibility of competitive cell counts. A fast way for

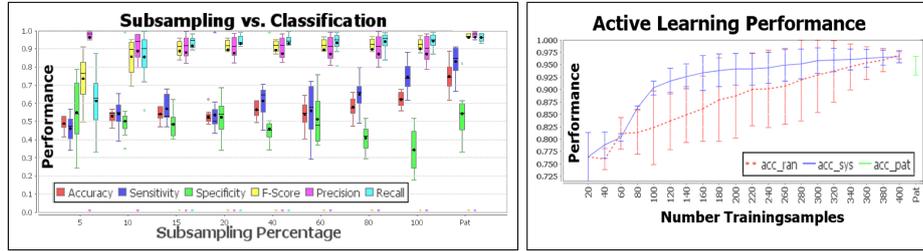


Fig. 5. Left: Performance plot for nucleus detection and classification via superpixels. Depicted are precision, recall and F-score for the nucleus detection, as well as sensitivity, specificity and accuracy for the nucleus classification. Experiments were conducted with training set sizes from 5 % to 100 % (x-axis) of all nuclei in 8 completely labeled TMA spots. Each box represents a Leave-One-Image-Out cross validation run with a SVM and a polynomial kernel. The performance stabilizes with 15% of training samples. The inter-expert performances of two pathologists is plotted last (“Pat”). **Right:** Proof of concept for the active learning approach in TMARKER. For three given TMA images, initially 20 nuclei (ten per class) were selected to train a SVM separating malignant from benign nuclei. The classification result is shown as accuracy on the y-axis. Consecutively, 20 additional nuclei were added repeatedly to the training (x-axis), such improving the classification performance. The additional nuclei are chosen at random (“acc_ran”) or systematically according to the respective lowest classification score (“acc_sys”). The systematic approach saturates much faster. The classification accuracy reaches the level of the two pathologists (“acc_pat”).

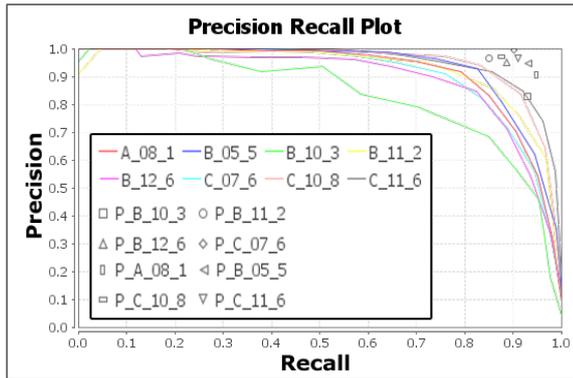


Fig. 6. Precision/Recall curve for the nucleus detection via color deconvolution with varying radius r . Each curve represents one out of 8 images of TMA spots. Higher radius r reveals less nuclei. The automatically detected nuclei were validated against the labels of a trained pathologist. For all images the inter-expert values are pointed (top left) as the precision and recall of one pathologist to match the other.

staining estimation is provided by the integrated color deconvolution method. When only relevant cells are considered for staining estimation, e.g. with distinction between malignant and benign cells, TMARKER provides modern machine learning methods for nucleus detection and classification. While the potential of TMARKER has been shown, it has to be further validated and improved on larger and different datasets.

Acknowledgments. This work was supported by the FET-Open Program (FP7) of the European Commission, under project SIMBAD grant no. 213250.

References

1. Madabhushi, A., Dowling, J., Huisman, H.J., Barratt, D.C., eds.: Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions - International Workshop, Held in Conjunction with MICCAI 2011, Toronto, Canada. Proceedings. Volume 6963 of Lecture Notes in Computer Science., Springer (2011)
2. Schüffler, P.J., Fuchs, T.J., Ong, C.S., Roth, V., Buhmann, J.M.: Computational tma analysis and cell nucleus classification of renal cell carcinoma. In: Proceedings of the 32nd DAGM conference on Pattern recognition, Berlin, Heidelberg, Springer-Verlag (2010) 202–211
3. Fuchs, T.J., Buhmann, J.M.: Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics* **35**(78) (2011) 515 – 530
4. Fuchs, T.J., Wild, P.J., Moch, H., Buhmann, J.M.: Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. In: MICCAI 2008. Pt. 2. Volume 11 of Lecture Notes in Computer Science. Springer, Berlin, DE (2008) 1–8
5. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Suesstrunk, S.: SLIC Superpixels. Technical report, EPFL, EPFL (2010)
6. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. *ECCV 2004* **2** (2004) 469–481
7. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval. CIVR '07, New York, NY, USA, ACM (2007) 401–408
8. Breiman, L.: Random forests. *Machine Learning* **45**(1) (2001) 532
9. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. *SIGKDD Explor. Newsl.* **11** (November 2009) 10–18
10. EL-Manzalawy, Y., Honavar, V.: WLSVM: Integrating LibSVM into Weka Environment. (2005) Software available at www.cs.iastate.edu/~yasser/wlsvm.
11. Friedman, Geiger, Goldszmidt: Bayesian network classifiers. *Machine Learning* **29** (1997) 131163
12. Cohn, D.A., Ghahramani, Z., Jordan, M.I.: Active Learning with Statistical Models. *Journal of Artificial Intelligence Research* **4** (1996) 129–145
13. Ruifrok, A.C., Johnston, D.A.: Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology* **23**(4) (August 2001) 291–299
14. Burger, W., Burge, M.J.: *Digitale Bildverarbeitung : Eine Einführung mit Java und ImageJ. 2., bearbeitete auflage* edn. Springer-Verlag, Berlin, Heidelberg (2005)