You Should Use Regression to Detect Cells

Philipp Kainz[1,⋆], Martin Urschler[2,3], Samuel Schulter[2],
Paul Wohlhart[2], and Vincent Lepetit[2]

[1] Institute of Biophysics, Medical University of Graz, Austria
[2] Institute for Computer Graphics and Vision,
BioTechMed, Graz University of Technology, Austria
[3] Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria

Abstract. Automated cell detection in histopathology images is a hard problem due to the large variance of cell shape and appearance. We show that cells can be detected reliably in images by predicting, for each pixel location, a monotonous function of the distance to the center of the closest cell. Cell centers can then be identified by extracting local extremums of the predicted values. This approach results in a very simple method, which is easy to implement. We show on two challenging microscopy image datasets that our approach outperforms state-of-the-art methods in terms of accuracy, reliability, and speed. We also introduce a new dataset that we will make publicly available.

1 Introduction

Analysis of microscopy image data is very common in modern cell biology and medicine. Unfortunately, given the typically huge number of cells contained in microscopic images of histological specimen, visual analysis is a tedious task and can lead to considerable inter-observer variability and even irreproducible results because of intra-observer variability [1].

Automated cell detection and segmentation methods are therefore highly desirable, and have seen much research effort during the previous decades [2]. In histological image analysis, one of the main problems is to count how many cells are present in the captured images, and many automatic methods have already been proposed for this task [3–8]. Some methods are based on simple contour-based cell models [5] or leverage shape and appearance priors [7] in a global optimization strategy. While reasonable success in cell detection may be achieved using conventional image processing, for example based on local symmetry features [4] or using normalized cuts and spectral graph theory to segment cells [3], recently learning based approaches have proven to achieve state-of-the-art results on detection benchmarks like [9]. On this benchmark, the work of [6]

⋆ The authors are grateful to Nasir Rajpoot and Mahonar Kuse for providing the cell data set from the ICPR 2010 Pattern Recognition in Histopathological Images contest, and to Martin Asslaber for providing the bone marrow histology images.

currently outperforms other approaches. It is based on extracting a large number of candidate maximally stable extremal cell regions using the MSER detector [10], which are pruned using several, increasingly complex classifiers based on structured SVMs (SSVM). Other approaches apply a classifier densely over the input images in a sliding window fashion [11] or learn regions revealed by the SIFT [12] keypoint detector [8]. A related, but different problem is to count cells without explicitly detecting them by estimating their density [13, 14]. However this approach does not produce the locations of the cells, which are important for example to perform cell type recognition.

We consider here an alternative approach to cell detection. Our method is inspired by the recent [15], which considers the extraction of linear structures in images: Instead of relying on an *ad hoc* model of linear structures such as neurons [16] or a classifier [17], [15] proposes to predict, for each pixel of the input image, a function of the distances to the closest linear structure in a regression step. The local maximums of the predicted function can be extracted easily and correspond to the desired linear structures.

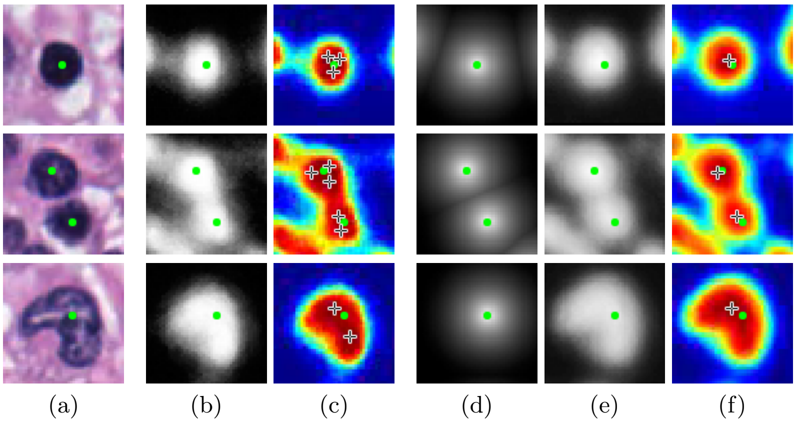

Fig. 1. Comparing classification and regression for cell detection. (a) Several patches of input images from our bone marrow dataset, centered on one or two cells. First row: one fully stained nucleus, second row: two closely located nuclei, third row: one nucleus of anisotropic shape and non-uniform staining. Green dots indicate ground truth annotation of the cell centers. (b) Probability maps provided by a classifier applied to these patches, and (c) the local maximums of these probability maps. They exhibit many local maximums – indicated by crosses – around the actual cell centers while only one maximum is expected. (d) The expected score map that the regressor should predict and (e) the actual predictions. (f) The local maximums of these predictions correspond much better to the cell centers and do not suffer from multiple responses.

As depicted in Fig. 1, we show in this paper that this approach transfers to cell detection, and actually outperforms state-of-the-art approaches over all the standard metrics: Using a standard regression Random Forest [18], we predict for each image location a function of the distance to the closest cell center. We can then identify the cell centers by looking for local maximums of the predictions.

We evaluate our approach on two challenging datasets, illustrated in Fig. 2. For both datasets, the goal is to predict the center of the cells. The first dataset we consider is from the ICPR 2010 Pattern Recognition in Histopathological Images contest [9], consisting of 20 100 × 100 pixel images of breast cancer tissue. We also introduce a new dataset BM, containing eleven $1,200 \times 1,200$ pixel images of healthy human bone marrow from eight different patients.

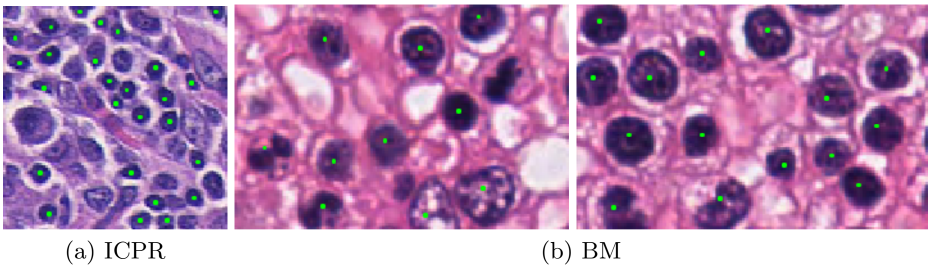

Fig. 2. Dataset samples: (a) Breast cancer tissue from the ICPR 2010 contest [9] dataset (20× magnification), (b) bone marrow tissue (cropped from full images at 40× magnification). The green dots denote ground truth locations of the cell nuclei.

2 Learning to Localize Cells

Our approach to cell detection is to predict a score map based on the Euclidean distance transform of the cell centers in a given input image. This score map is computed such that each pixel value encodes its distance to the nearest cell center and, ideally, the local extremums correspond to the center of the cells to be detected. The prediction of the score map is performed using a regression method trained from cell images and ground truth cell locations.

2.1 Defining the Proximity Score Map

As shown in Fig. 1(a), our approach is based on statistical learning and relies on a set of annotated images for training. An expert labeled the centers of the cell nuclei in each of these images.

The standard classification approach applied to the cell detection problem would consist of training a classifier to predict whether the center pixel of an input image patch is the center of a cell or is located in the background. Unfortunately, as shown in Fig. 1(c), this often results in multiple peaks for cell nuclei and hence in multiple detections corresponding to a single cell. One option is to apply post-processing to group multiple detections into a single one, for example by smoothing the output of the classifier for each image location with a Gaussian kernel to merge multiple peaks into a single one. However, such a strategy may merge together the responses actually created by multiple cells.

We propose here a better, more principled approach. Inspired by [15], we exploit additional context during the learning phase and define a smooth, continuous prediction target function $d(\mathbf{x})$ expressing the proximity of each pixel $\mathbf{x}$ to the ground truth cell centers, such as the ones shown in Fig. 1(d). This therefore shifts the task from binary classification to regression of the continuous proximity score map.

A straightforward way of defining our proximity score map $d(\mathbf{x})$ is to take the inverted Euclidean distance transform $\mathcal{D}_C$ of the set $C = \{\mathbf{c}_i\}$ of ground truth cell centers: $d(\mathbf{x}) = -\mathcal{D}_C(\mathbf{x})$. However, this approach produces high proximity scores even in background areas. Additionally, it forces the regression model to predict varying scores for different regions of the background. Moreover, cell centers are not well-defined, which exacerbates the learning problem.

Hence, it is better to predict a function of the distance transform that is flat on the background and has better localized, distinctive peaks at cell centers [15]:

$$d(\mathbf{x}) = \begin{cases} e^{\alpha\left(1-\frac{\mathcal{D}_C(\mathbf{x})}{d_M}\right)} - 1 & \text{if } \mathcal{D}_C(\mathbf{x}) < d_M \\ 0 & \text{otherwise} \end{cases} , \tag{1}$$

where α and d_M control the shape of the exponential function and $\mathcal{D}_C(\mathbf{x})$ is the Euclidean distance transform of the cell centers. In practice, we select d_M such that the maximum width of peaks in $d(\mathbf{x})$ corresponds to the average object size to be detected. We used $\alpha = 5$, $d_M = 16$ for ICPR, and $\alpha = 3$, $d_M = 39$ for BM.

Our goal is now to learn a function g that predicts $d(\mathbf{x})$ given an image patch $I(\mathbf{x})$: By applying g over each $I(\mathbf{x})$, we obtain an entire proximity score map. This is detailed in the next subsection.

2.2 Training and Evaluating a Regression Model

Many options are available for learning function g, and we opted for standard regression Random Forests [18], because they are fast to evaluate, were shown to perform well on many image analysis problems, and are easy to implement.

Instead of directly relying on pixel intensities, we apply the forest on image features extracted from input patches. We use 21 feature channels: RGB channels (3), gradient magnitude (1), first and second order gradients in x- and y-directions (4), Luv channels (3), oriented gradients (9), and histogram equalized gray scale image (1). The split functions in the nodes of the forest include single pixel values, pixel value differences, Haar-like features, and a constrained pixel value difference, where the second patch location for difference computation was chosen within a distance of 10 pixels clamped at the image patch borders. For all the split functions but single pixel values, we randomly select whether to use the values for the same feature channel or across feature channels.

In all experiments, each split was optimized on a random subset of 200 training samples with 1000 random tests and 20 thresholds each. Splitting stops once either maximum tree depth or minimal number of 50 samples per node is reached.

2.3 Detecting the Cells from the Proximity Score Map

Once the forest g is trained, it can predict a proximity score map for unseen images. By construction, the local maximums in this map should correspond to the cell centers. We therefore apply non-maximum suppression, where maximums below a certain threshold κ are discarded. As will be shown, varying κ facilitates optimization of either precision or recall, depending on the task.

3 Experimental Results

We first describe the datasets and protocol used for the evaluation. Then, we provide comparisons of our approach to the current state-of-the-art method of Arteta *et al.* [6], and a standard classification Random Forest based on the same image features as the proposed regression forest.

3.1 Datasets

The ICPR dataset consists of 20 100×100 pixel microscopy images of breast cancer tissue [9] (ICPR). We also introduce a new dataset BM containing eleven $1,200 \times 1,200$ pixel images of healthy human bone marrow from eight different patients[4]. Tissue in both datasets was stained with Hematoxylin and Eosin.

For our BM dataset, all cell nuclei were labeled as *foreground* by providing the location of the center pixel as dot annotation. Debris and staining artifacts were labeled as *background*. Ambiguous parts, for which cell nuclei were not clearly determinable as such, were labeled as *unknown*. Nevertheless, all ambiguous objects are treated as foreground, since the detection method proposed in this work is supposed to identify these objects as candidates for subsequent classification. The resulting $4,205$ dot annotations cover *foreground* and *unknown* labels.

3.2 Model Evaluation

To decide if a detection actually corresponds to a cell center, we consider a distance threshold ξ. If the distance between a detection and a ground truth annotation is less or equal ξ, we count the detection as true positive (TP). If more than one detection falls into this area, we assign the most confident one to the ground truth location and consider the others as false positives (FP). Detections farther away than ξ from any ground truth location are FP, and all ground truth annotations without any close detections are false negatives (FN).

Accuracy is evaluated in terms of precision ($= TP/(TP + FP)$), recall ($= TP/(TP + FN)$), F1-score, average Euclidean distance and standard deviation $\mu_d \pm \sigma_d$ between a TP and its correctly assigned ground truth location, as well as the average absolute difference and standard deviation between number of ground truth annotations and detections $\mu_n \pm \sigma_n$. We report results in this section computed with forests composed of 64 trees and a maximum tree depth of 16, an optimal complexity determined in leave-one-out cross validation (LOOCV) on the more complex BM dataset.

[4] The dataset is available from `https://github.com/pkainz/MICCAI2015/`.

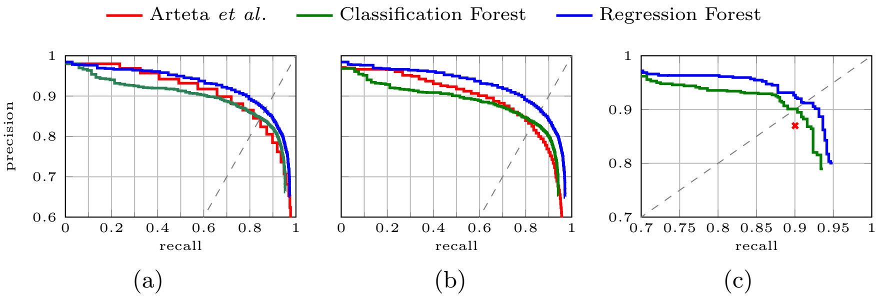

Fig. 3. Precision-recall curves on the two datasets, obtained by varying the threshold κ on the confidence and recording precision and recall of the resulting detections wrt. the ground truth. (a,b) LOOCV on the *BM dataset*. In (a) the threshold ξ defining the maximum accepted distance between a detection and the ground truth was set to $\xi = 16$, whereas in (b) it was tightened to $\xi = 8$. With the looser bound in (a) the performance of Arteta *et al.* is insignificantly lower than ours (AUC 89.9 vs. 90.7). However, in (b) it drops considerably (AUC 86.9), while for our regression method the curve stays the same (AUC 90.5), demonstrating its capability to localize the cells much more precisely. (c) On the *ICPR benchmark dataset* both classification and regression outperform the current state-of-the-art detector.

Figs. 3(a,b) show precision-recall evaluations on the BM dataset: each curve was obtained in LOOCV by varying κ. Additionally, we assessed the method of Arteta *et al.* [6] in a LOOCV and compared performance measures. Most prominently, the maximum distance threshold between ground truth and detection ξ is responsible for the localization accuracy. In Fig. 3(a), we moderately set $\xi = 16$ and observed that the area under the curve (AUC) is only slightly lower than ours: 89.9 vs. 90.7. As soon as we tighten $\xi = 8$, their AUC measure drops considerably to 86.9, whereas our regression method exhibits the same performance (90.5). This, and the consistent shape of the regression curves strongly indicate our method's superiority over the current state-of-the-art in terms of localization accuracy. Further, by allowing a smaller value of ξ, a more rigorous definition of TP detections is enabled, thus resulting in increased detection confidence. The achieved F1-score on the BM dataset is 84.30 ± 3.28 for Arteta *et al.* [6] vs. 87.17 ± 2.73 for our regression approach.

To assess the stability of our method, which relies on random processes during feature selection, we performed ten independent runs on a predefined, fixed train-test split of the BM dataset. We trained on the first eight and tested on the last three images and achieved a stable F1-score of 88.05 ± 0.06.

Table 1 shows results on the ICPR benchmark dataset [9]. Both our regression approach and the standard classification outperform [6] over all standard metrics, cf. Fig. 3(c). Although [6] state performance values, no value for ξ is mentioned. Given our previous definition, we use a rather strict $\xi = 4$ for both, regression and classification forests. Nevertheless, [6] must have used a value $\xi > 4$ in order to match the numbers reported for $\mu_d \pm \sigma_d$ in Table 1.

Table 1. Performance comparison on the ICPR benchmark dataset [9]. F1-scores for regression and classification were averaged over ten independent iterations and metrics were obtained with $\xi = 4$. Regression outperforms all previously reported competing methods on the standard metrics.

Method	Prec.	Rec.	F1-score	$\mu_d \pm \sigma_d$	$\mu_n \pm \sigma_n$
Regr. Forest	**91.33**	**91.70**	**91.50**	$\mathbf{0.80 \pm 0.68}$	3.26 ± 2.32
Class. Forest	90.66	89.72	90.18	0.86 ± 0.68	4.04 ± 2.26
Arteta *et al.* [6]	86.99	90.03	88.48	1.68 ± 2.55	$\mathbf{2.90 \pm 2.13}$
Bernardis & Yu [3]	-	-	-	2.84 ± 2.89	8.20 ± 4.75
LIPSyM [4]	70.21	70.08	70.14	3.14 ± 0.93	4.30 ± 3.08

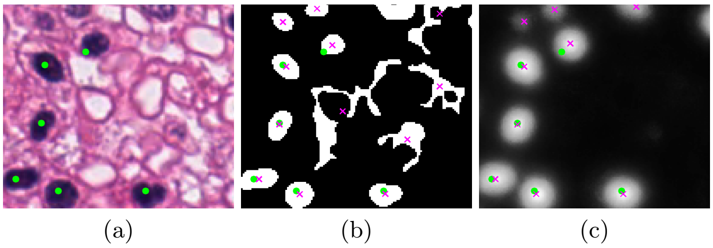

Fig. 4. Illustration of detection hypotheses on BM data (a): Our regression based detector (c) proposes much more accurate and reliable cell center hypotheses than the detector of Arteta *et al.* [6], shown in (b). Magenta crosses denote proposed cell centers, green dots are ground truth locations. While Arteta *et al.* need a separate classifier to post-process these hypotheses, a simple threshold on the detection confidence is sufficient to achieve the reported improved results for our method.

A qualitative illustration of the detector hypotheses on a BM image is depicted in Fig. 4. The state-of-the-art method [6] proposes many cell center hypotheses in a clear background region, where our regression method did not produce any proximity scores at all. Final localization is determined by post-processing and hence reliable hypotheses are beneficial for high accuracy.

We also compared the computation times on a standard Intel Core i7-4470 3.4GHz workstation. [6], the best performing method after ours, needs 3.6 hours for training on ten images of the BM dataset. Testing on a single BM image lasts around 60 seconds. In contrast, our regression method takes only 1.5 hours of training, and only 15 seconds for testing on $1,200 \times 1,200$ pixel images.

4 Conclusion

We showed that using a simple regression forest to predict a well-chosen function over the input images outperforms state-of-the-art methods for cell detection in histopathological images: Our approach is easy to implement and 4× faster than the method of [6], while being more reliable and accurate.

References

1. Al-Adhadh, A.N., Cavill, I.: Assessment of cellularity in bone marrow fragments. J. Clin. Pathol. **36** (1983) 176–179
2. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: a review. IEEE Rev. Biomed. Eng. **2** (2009) 147–171
3. Bernardis, E., Yu, S.X.: Pop out many small structures from a very large microscopic image. Med. Image Anal. **15**(5) (2011) 690–707
4. Kuse, M., Wang, Y.F., Kalasannavar, V., Khan, M., Rajpoot, N.: Local isotropic phase symmetry measure for detection of beta cells and lymphocytes. J. Pathol. Inform. **2**(2) (2011)
5. Wienert, S., Heim, D., Saeger, K., Stenzinger, A., Beil, M., Hufnagl, P., Dietel, M., Denkert, C., Klauschen, F.: Detection and Segmentation of Cell Nuclei in Virtual Microscopy Images: A Minimum-Model Approach. Sci. Rep. **2** (2012) 1–7
6. Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A.: Learning to Detect Cells Using Non-overlapping Extremal Regions. In: MICCAI. (2012) 348–356
7. Nosrati, M.S., Hamarneh, G.: Segmentation of cells with partial occlusion and part configuration constraint using evolutionary computation. In: MICCAI. (2013) 461–468
8. Mualla, F., Schoell, S., Sommerfeldt, B., Maier, A., Steidl, S., Buchholz, R., Hornegger, J.: Unsupervised Unstained Cell Detection by SIFT Keypoint Clustering and Self-labeling Algorithm. In: MICCAI. (2014) 377–384
9. Gurcan, M.N., Madabhushi, A., Rajpoot, N.: Pattern recognition in histopathological images: An ICPR 2010 contest. In: Recognizing Patterns in Signals, Speech, Images and Videos. (2010) 226–234
10. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust Wide Baseline Stereo From Maximally Stable Extremal Regions. In: BMVC. (2002) 384–393
11. Quelhas, P., Marcuzzo, M., Mendonça, A.M., Campilho, A.: Cell nuclei and cytoplasm joint segmentation using the sliding band filter. IEEE Trans. Med. Imaging **29**(8) (2010) 1463–1473
12. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. **60**(2) (2004) 91–110
13. Lempitsky, V., Zisserman, A.: Learning to Count Objects in Images. In: NIPS. (2010) 1324–1332
14. Fiaschi, L., Nair, R., Koethe, U., Hamprecht, F.: Learning to Count with Regression Forest and Structured Labels. In: ICPR. (2012) 2685–2688
15. Sironi, A., Lepetit, V., Fua, P.: Multiscale Centerline Detection by Learning a Scale-Space Distance Transform. In: CVPR. (2014)
16. Zhou, Z.H., Jiang, Y., Yang, Y.B., Chen, S.F.: Lung cancer cell identification based on artificial neural network ensembles. Artif. Intell. Med. **24** (2002) 25–36
17. Yin, Z., Bise, R., Chen, M., Kanade, T.: Cell segmentation in microsopy imagery using a bag of local Bayesian classifiers. In: ISBI. (2010) 125–128
18. Breiman, L.: Random Forests. Mach. Learn. **45** (2001) 5–32