AUTOMATIC DETECTION OF NEURONS IN HIGH-CONTENT MICROSCOPE IMAGES USING MACHINE LEARNING APPROACHES

Gadea Mata[1]*, Miroslav Radojevic*[2]*, Ihor Smal*[2]*, Miguel Morales*[3]*, Erik Meijering*[2]*, Julio Rubio*[1]

[1]Departamento de Matemáticas y Computación, Universidad de La Rioja, Spain
[2]Erasmus University Medical Center, Rotterdam, the Netherlands
[3]Institut de Neurociències, Universitat Autònoma de Barcelona, Spain

ABSTRACT

The study of neuronal cell morphology and function in relation to neurological disease processes is of high importance for developing suitable drugs and therapies. To accelerate discovery, biological experiments for this purpose are increasingly scaled up using high-content screening, resulting in vast amounts of image data. For the analysis of these data fully automatic methods are needed. The first step in this process is the detection of neuron regions in the high-content images. In this paper we investigate the potential of two machine-learning based detection approaches based on different feature sets and classifiers and we compare their performance to an alternative method based on hysteresis thresholding. The experimental results indicate that with the right feature set and training procedure, machine-learning based methods may yield superior detection performance.

***Index Terms*—** Neuron screening, high-content analysis, feature extraction, object detection, machine learning.

1. INTRODUCTION

Neurons are excitable cells with the extraordinary ability of processing and transmitting information. In the human brain billions of neurons interconnect through axons and dendrites to form a highly complex network that constructs our perception of the external world and controls the mechanics of our actions [1]. Different parts of the brain fulfill different functions in this process and, correspondingly, the neuronal cells involved typically have different morphologies. As neuron morphology and function are intimately intertwined, altered morphology has been implicated in various brain pathologies that cause progressive loss of function, such as Alzheimer and Parkinson, but also in ageing [2]. Thus the study of neuronal cell morphology in relation to specific disease processes and the effects of drug compounds is of high importance for developing suitable drugs and therapies [3].

Partially supported by the Spanish Ministerio de Economía y Competitividad (project number MTM2014-54151-P), the University of La Rioja (project number FPI-UR-13), and the Netherlands Organization for Scientific Research (NWO project number 612.001.018).

An exciting recent development in neuroscience is the use of high-content analysis (HCA). HCA generally refers to the combination of automated acquisition and analysis of large microscopic image data sets for biological discovery and is often employed by pharmaceutical and biotechnology companies but increasingly also in academia and research institutes [4]. In view of the vast amounts of data and the desire to eliminate possible human bias, automation is a key requirement in HCA, and thus the analysis methods must be highly robust and reliable. Although challenging, HCA is now used also in basic neuroscience research [5–7], and various image analysis pipelines have already been developed for neuron quantification in high-content image data [8–11].

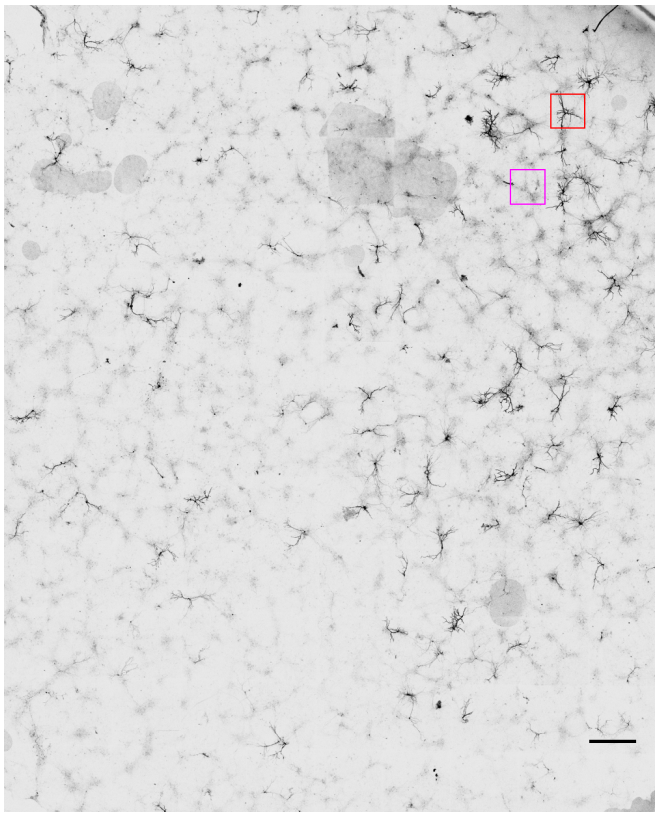
The first step in an HCA pipeline for neuron screening is to detect image regions of interest containing neurons as opposed to background or irrelevant structures (Fig. 1). Commonly this is done by image prefiltering (denoising, illumination correction, contrast enhancement) followed by some form of intensity-based thresholding. However, the images often contain debris and other artifacts that are larger or more complex than standard prefiltering techniques can eliminate, and thus more sophisticated solutions are needed.

Here we investigate the potential of machine-learning based approaches for automatic detection of neurons in microscopic images for HCA. Specifically, we implement and compare two approaches based on different feature sets and classifiers, and compare their performance to each other and an intensity-based detection method, using expert manual annotation as the gold standard. We demonstrate that with the right feature set and training procedure, machine learning can indeed improve neuron detection.

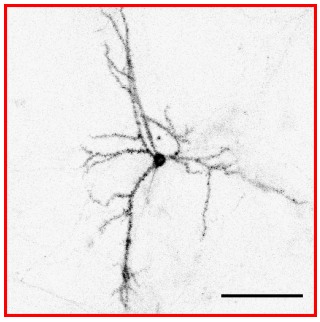
2. METHODS

2.1. Image Data and Annotation

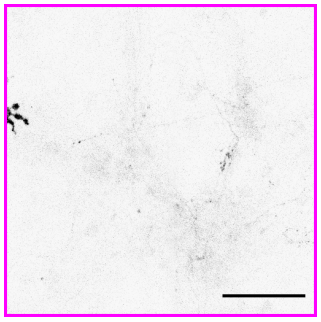
For this study we used rat hippocampal neurons in culture since their in-vivo morphological features are well conserved in culture conditions (typically a pyramidal cell soma with a complex dendritic tree) [12] and the relation of the hippocampus with higher brain functions such as learning and



(a) Example high-content image. Scale bar: 800 px ($\approx 600\,\mu$m)



(b) Neuron region.
Scale bar: 130 px ($\approx 100\,\mu$m)



(c) Background region.
Scale bar: 130 px ($\approx 100\,\mu$m)

Fig. 1. Example of a high-content microscope image (a) and regions containing a neuron (b) and background (c). Image intensities are inverted here for displaying purposes.

memory are well established [13]. Two-dimensional (2D) images were acquired with a Leica SP5 automated confocal microscope (20× lens) using the Matrix modules. Each image was about 10,000 × 12,000 pixels (covering $\approx 70\,\text{mm}^2$ of culture dish) and contained on the order of 40 transfected (with the fluorescent protein GFP) neurons (Fig. 1). Specimens usually have about 100 neurons but more than half are not imaged as they are in different optical planes or close to the borders of the dish. Eight images (total size $>$ 1 GB) were used in our experiments. An expert neurobiologist manually marked all regions in these images that contained neurons (in our case, 409 neurons were marked). Through this we learned that neuron regions in our images are typically 500 × 500 pixels. From the non-neuron regions in the images we randomly sampled 1,000 patches of this size to serve as negative examples (background) for training the machine-learning methods.

2.2. Intensity Based Neuron Detection

Considering that transfected neurons ideally have clearly different intensities than the background, as a first approach we used only intensity and connectivity information to detect neurons in the images. The algorithm is based on hysteresis thresholding [14]. It starts with a breadth-first search for image pixels with intensity above a global user-defined threshold. This threshold is chosen sufficiently high to include pixels with very high probability of being part of a neuron (true positives), and very low probability of being background (false positives), while likely resulting in many missed neuron pixels (false negatives). The latter are largely added in a second round, in which segmented pixels from the first round are taken as seeds in a depth-first search to find connected pixels with intensity above a second, lower user-defined threshold. This approach is sensible due to the way cells are dyed in the experiments and allows for "graceful degradation" of intensity within neuronal image structures. The algorithm was implemented as a plugin for ImageJ [15] and is referred to as LocationJ in the sequel.

2.3. Machine-Learning Based Neuron Detection

As an alternative to intensity based neuron detection we considered two machine-learning based approaches. Detection was achieved by classification of image patches as positive (containing neuron structure) or negative (containing background) using features computed from these patches. The first approach was based on computing an extensive list of image features using the WND-CHARM library [16]. Specifically, over 1,000 features were computed, including many different types of polynomial decompositions, high contrast features, pixel statistics, and texture descriptors. Since not all features may be equally relevant, in the training step a Fisher discriminant score [17] was computed for each, which allowed building a ranked preference list of features. In the testing phase of the experiments we used the top-15% features according this score. Classification of patches based on these features was done using a weighted neighbor distances (WND) classifier [16]. The second approach was based on computing the scale-invariant feature transform (SIFT) [18], resulting in keypoint descriptors of the image scale-space that are highly distinctive and invariant to image scaling, rotation, a range of affine distortions, and robust to noise and illumination change. Classification of patches based on these features was done using a bag-of-words class model [19] and the Naive-Bayes (NB) classifier from the WEKA toolkit [20].

2.4. Experimental Procedure

The performance of the methods was initially evaluated directly on the expert annotated neuron patches (positives) and sampled background patches (negatives) as described above. A random selection of 75% of the patches from each class was used for training of the machine-learning based detectors while the remaining 25% was used for testing. With LocationJ, a patch was classified as positive if it contained a connected segment of at least 15 pixels. Given the classification output and the annotation, the number of true positives (TP), false positives (FP), and false negatives (FN) could automatically be determined for each method, from which the classification performance was quantified in terms of recall (TP/(TP+FN)) and precision (TP/(TP+FP)).

When using the machine-learning based detectors in practice, the patches to be classified must first be sampled from the image. Since it is computationally very costly to exhaustively check all possible patch locations in an image, in our application 1,000 patches (of 500×500 pixels as motivated above) are sampled uniformly from the image, which suffices to capture the true neuron regions (typically only several dozens). Thus in a second experiment we tested the performance of the detectors when using the same patch size as before but now using this sampling scheme. In this case the locations of the considered patches typically did not match exactly with those of the annotated patches. Therefore patches classified as positive were further examined: if a positive patch overlapped less than 20% with any annotated neuron patch, it was declared a false positive, and all positive patches overlapping 20% or more with the same annotated neuron patch were considered a unique (single) true positive match. The 20% threshold may seem rather conservative but additional experiments (not reported here) showed us that higher percentages did not consistently improve the results.

Due to the sampling, many patches in the second experiment potentially contained small portions of neuron structures, which to some degree caused a mismatch with the initial training set. Therefore we expected the performance of the methods to be lower in the second experiment as compared to the first. To improve performance we implemented a bootstrapping approach, where the classifiers were retrained in a second round using as negative examples only the false positives from the first round, while continuing using the true positive examples from the expert annotation. The machine-learning detectors without (versus with) using bootstrapping are referred to as WND-CHARM-A and NB-SIFT-A (versus WND-CHARM-B and NB-SIFT-B).

3. RESULTS

The results of the first experiment are presented in Table 1. Although WND-CHARM-A showed perfect recall, its precision was lower than LocationJ, indicating that it produced more false positives and the training set did not well-reflect possible background variability. Precision was drastically improved by using the bootstrapped variant, WND-CHARM-B, with still near-perfect recall. Bootstrapping improved the performance of NB-SIFT in terms of both measures, resulting in a better recall than with LocationJ but a slightly lower precision. The negative effect of patch sampling on the performance of all methods is seen from the results of the second experiment in Table 2. As anticipated, in virtually all cases both the recall and the precision was considerably lower than in the first experiment. The differences between the two measures are also much higher, with the precision being substantially lower than the recall in all cases, indicating an increase of false positives. Of the machine-learning approaches only NB-SIFT-B performed better than LocationJ in terms of both measures in these experiments. Example detection results from NB-SIFT-B are shown in Fig. 2.

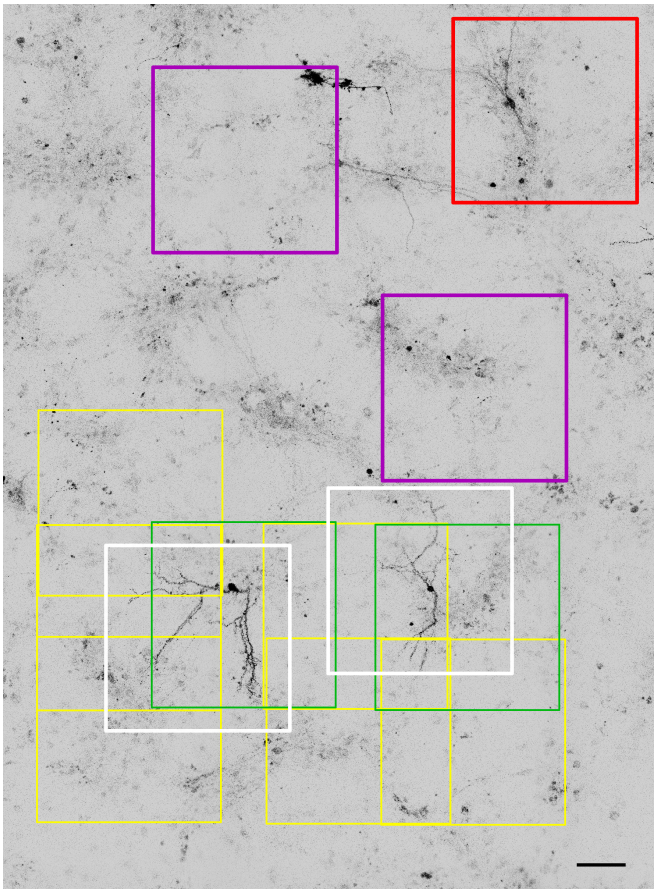


Fig. 2. Example detection result using NB-SIFT-B. Shown are part of a high-content microscope image of neurons and the borders of various patches overlaid in color coding: an expert annotated neuron region missed by the detector (red), two annotated regions found by the detector (white), true-positive patches (yellow) ignored by the detector in favor of the best overlapping patch (green), and false-positive patches (magenta). Scale bar: 130 px ($\approx 100\,\mu$m).

Method	Recall	Precision
LocationJ	0.80	0.86
WND-CHARM-A	1.00	0.62
WND-CHARM-B	0.97	0.82
NB-SIFT-A	0.74	0.46
NB-SIFT-B	0.88	0.83

Table 1. Results of the first experiment.

Method	Recall	Precision
LocationJ	0.64	0.50
WND-CHARM-A	0.89	0.19
WND-CHARM-B	0.66	0.44
NB-SIFT-A	0.60	0.26
NB-SIFT-B	0.94	0.57

Table 2. Results of the second experiment.

4. DISCUSSION

In this paper we have investigated the problem of neuron detection in high-content microscope images from screening experiments in neuroscience. We have evaluated the performance of two machine-learning based approaches using different feature sets and classifiers in comparison with a method based on hysteresis thresholding of intensity information only. The results showed that machine-learning approaches may perform superiorly. However this improvement does not come lightly but requires careful consideration of the ingredients. Of the two machine-learning based approaches considered, the best results were obtained with NB-SIFT in combination with a bootstrapping procedure for the training stage, even though WND-CHARM takes into account a much wider variety of image features. This suggests that selection of the right feature set as well as the right training set is crucial to achieve good detection performance. In future work we aim to perform a more detailed study of the effects of different types of features in combination with different types of classifiers and training procedures. The ultimate goal of our research is to develop fully automatic methods for studying the structural and functional effects of mutations and of new drugs in neurological diseases. In addition to detection and localization of neuron regions in high-content images this requires accurate neuron reconstruction and analysis of the complexity of the dendritic arborizations and spine properties. To this end higher-resolution scans of the specimen are needed in those areas where neurons are found. The use of reliable neuron detection methods will greatly improve the efficiency of such experiments. This underscores the importance of further reducing false-positive detections and will be further explored in future work.

5. REFERENCES

[1] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, McGraw-Hill, New York, NY, 4th edition, 2000.

[2] J. van Pelt, A. van Ooyen, and H. B. Uylings, "The need for integrating neuronal morphology databases and computational environments in exploring neuronal structure and function," *Anat. Embryol.*, vol. 204, no. 4, pp. 255–265, Oct. 2001.

[3] E. Meijering, "Neuron tracing in perspective," *Cytom. A*, vol. 77, no. 7, pp. 693–704, July 2010.

[4] X. Xia and S. T. C. Wong, "Concise review: a high-content screening approach to stem cell research and drug discovery," *Stem Cells*, vol. 30, no. 9, pp. 1800–1807, Sep. 2012.

[5] M. Dragunow, "High-content analysis in neuroscience," *Nat. Rev. Neurosci.*, vol. 9, no. 10, pp. 779–788, Oct. 2008.

[6] J. L. Anderl, S. Redpath, and A. J. Ball, "A neuronal and astrocyte coculture assay for high content analysis of neurotoxicity," *J. Vis. Exp.*, vol. 5, no. 27, pp. 1173, May 2009.

[7] N. M. Radio, "Neurite outgrowth assessment using high content analysis methodology," *Meth. Mol. Biol.*, vol. 846, pp. 247–260, Jan. 2012.

[8] P. Vallotton, R. Lagerstrom, C. Sun, M. Buckley, D. Wang, M. D. Silva, S.-S. Tan, and J. M. Gunnersen, "Automated analysis of neurite branching in cultured cortical neurons using HCA-Vision," *Cytom. A*, vol. 71, no. 10, pp. 889–895, Sep. 2007.

[9] Y. Zhang, X. Zhou, A. Degterev, M. Lipinski, D. Adjeroh, J. Yuan, and S. T. C. Wong, "A novel tracing algorithm for high throughput imaging: screening of neuron-based assays," *J. Neurosci. Meth.*, vol. 160, no. 1, pp. 149–162, Feb. 2007.

[10] C. Wu, J. Schulte, K. J. Sepp, J. T. Littleton, and P. Hong, "Automatic robust neurite detection and morphological analysis of neuronal cell cultures in high-content screening," *Neuroinform.*, vol. 8, no. 2, pp. 83–100, June 2010.

[11] P. Charoenkwan, E. Hwang, R. W. Cutler, H.-C. Lee, L.-W. Ko, H.-L. Huang, and S.-Y. Ho, "HCS-Neurons: identifying phenotypic changes in multi-neuron images upon drug treatments of high-content screening," *BMC Bioinform.*, vol. 14, no. S16, pp. S12, Oct. 2013.

[12] K. Goslin, H. Asumussen, and G. Banker, "Rat hippocampal neurons in low-density culture," in *Culturing Nerve Cells*, Cambridge, MA, 1998, pp. 339–370, The MIT Press.

[13] L. R. Squire, "Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans," *Psych. Rev.*, vol. 99, no. 2, pp. 195–231, April 1992.

[14] J. Canny, "A computational approach to edge detection," *IEEE Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[15] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Meth.*, vol. 9, no. 7, pp. 671–675, July 2012.

[16] N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley, and I. G. Goldberg, "WND-CHARM: multi-purpose image classification using compound image transforms," *Pattern Recognit. Lett.*, vol. 29, no. 11, pp. 1684–1693, Aug. 2008.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[19] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering objects and their location in images," in *Tenth IEEE International Conference on Computer Vision*, 2005, vol. 1, pp. 370–377.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Expl.*, vol. 11, no. 1, pp. 10–18, July 2009.