FLEXIBLE SYNAPSE DETECTION IN FLUORESCENCE MICROGRAPHS BY MODELING HUMAN EXPERT GRADING

Julia Herold[1], Manuela Friedenberger[2], Marcus Bode[2], Nasir Rajpoot[3],*
Walter Schubert[2], Tim W. Nattkemper[1]

[1]Applied Neuroinformatics Group, Faculty of Technology, University of Bielefeld, Germany
[2]Molecular Pattern Recognition Research Group, Institute of Medical Neurobiology,
University of Magdeburg, Germany
[3]Department of Computer Science, University of Warwick, UK

ABSTRACT

A particularly difficult task in molecular imaging is the analysis of fluorescence microscopy images of neural tissue, as they usually exhibit a high density of objects with diffuse signals. To automate synapse detection in such images, one has to simulate aspects of human pattern recognition skills to account for low signal-to-noise-ratios. We propose a machine learning based method that allows a direct integration of the experts' visual expertise who tag a low number of referential synapses according to their degree of synapse likeness. The sensitivity and positive predictive values show that by using graded likeness information in our learning algorithm we can provide an intuitively tunable tool for neural tissue slide evaluation.

Index Terms— Fluorescence microscopy, neural tissue, object detection, synapses, supervised machine learning

1. INTRODUCTION

Microscopy image analysis has recently been rediscovered as a top issue in bioinformatics [1]. However, a topic seldom dealt with, is the analysis of fluorescence microscopy images of neural tissue specimens, although such images are of high potential. By counting synapses in fluorescence microscopy images that represent large areas of the brain, unlike *e.g.* imaging with high resolution electron microscopy [2] which makes synapse detection accurate but only gives a small view on the brain, quantitative pathological measures can be derived as for instance in Alzheimer diagnosis [3]. Furthermore, by applying the Multi-Epitope-Ligand-Cartography (MELC) imaging technology [4], recently further developed as Toponome Imaging System (TIS) [5, 6], protein patterns can be extracted at each synapse position which allows for analyzing higher dimensional molecular synapse characteristics [7].

One big obstacle in analyzing fluorescence micrographs of neural tissue lies in the image characteristics. Due to the fact that fluorescence microscopy has to operate at the diffraction limited resolution of ∼200nm to image *e.g.* synapses, the micrographs exhibit a low signal-to-noise-ratio and low contrast but at the same time a great amount of small, diffuse objects. Different object detection approaches [8], which are very successful in other microscopy or fluorescence microscopy image analysis applications, are not applicable to such images. Clear object features, essential in many analysis approaches, as edges, textures, other first and second order statistics or PCA mappings, can not reliably be extracted.

Besides the problem of low signal-to-noise-ratio and low contrast, a key problem in detecting synapses in fluorescence microscopy images is to distinguish between spurious and clear staining. Objects can be inconsistently or non-specifically stained or show out of focus blur, which leads to wrong interpretations by human experts as well as by automatic detection methods. If one image is repeatedly evaluated by one expert or evaluated by different experts, this leads to significantly different results due to a low certainty of a true object [9]. This phenomenon is known as *intra-* and *inter-observer variance*. Furthermore, while high contrast synapses are usually detected with a high reproducibility, detection of lower contrast synapses greatly depends on how cautious the expert performs. Depending on the biological question, a false positive synapse is favored against false negative or vice versa, so different levels of caution need to be applied.

To automate object detection in fluorescence micrographs of neural tissue, an approach is needed that allows for tuning the caution of the detection algorithm in a way which is transparent to the biological experts and directly integrates their visual expertise.

In this paper we introduce a machine learning based synapse detection approach which accounts for the complexity of the underlying data by applying a Support Vector Machine (SVM)[10] with a nonlinear kernel function and a special postprocessing strategy. With this method, we are able to detect synapses with an accuracy comparable to human performance and superior to thresholding based approaches. Furthermore, we are able to mimic human caution in synapse detection by straight forward tuning of the detection algorithm. This is achieved mainly through the choice of the training set, which is directly understandable by the expert, but can additionally be steered by changing one parameter for the evaluation of the SVM detection.

1.1. Image acquisition

We apply the MELC/TIS imaging technology to visualize synapses in neural tissue digital micrographs. This technology is based on the fluorescence imaging technique, but enables the acquisition of fluorescence images of hundreds of distinct tags (*e.g.* antibodies) on the same tissue specimen. For this work, two mouse brain samples, fluorescence labeled with *synaptophysin*, are used to detect synaptic regions. Images are taken at a resolution of 256nm/pixel, 16bit/pixel. Thus, each synapse or synaptic region, which has a diameter of around 200nm, occupies a region smaller than a pixel. Nevertheless, due to the labeling strategy synapses appear as 3×3 - 5×5 pixels sized objects. We apply bilateral filtering [11] on each image
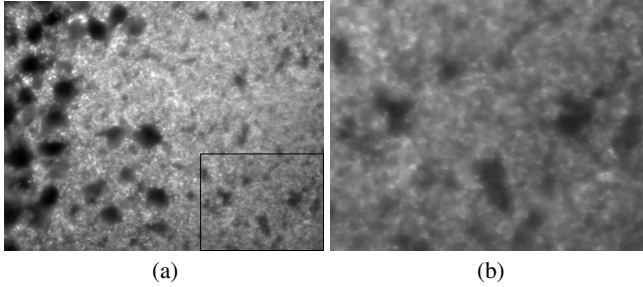

(a)   (b)

Fig. 1. Synaptophysin labeled micrographs. (a) 658×517 pixels sized original image. (b) Noise reduced sub-image of size 250×200, marked in (a) with a black rectangle.

to eliminate noise but preserve the small objects of interest. Filter parameters are manually fitted. Figure 1(a) shows one of the synaptophysin labeled micrographs of size 658×512 and 1(b) a sub-image of size 250×200. Its origin is marked in 1(a) by a black rectangle. Synapses are visible as small glowing dots.

2. SVM FOR SYNAPSE DETECTION

For synapse detection with SVMs, $N\times N$ sub-images (patches) of the input image are evaluated with respect to containing synapses (N=7). Patches containing synapses should be assigned to the positive class. Patches containing background or parts of synapses are to be classified into the negative class.

Because the SVM is a supervised machine learning approach, a representative set of patches for the SVM training is needed.

2.1. Manual image evaluation

In order to obtain a training set for SVM training and a reference set for the evaluation of the SVM classification performance, synapse positions in 250×200 pixels sized sub-images are manually labeled by three human experts (two on expert-, one on novice-level) with our own labeling software.

It is a well known problem that human experts are commonly not capable of explaining and systematically describing the object features a labeling decision is based on. Moreover, if an expert labels the same image twice, varying results are obtained in the majority of cases. In the special case of synapse marking in fluorescence labeled neural images, human experts have to label a great number of small objects with diffuse signals and low contrast, which very soon leads to mental fatigue. Analyzing a sub-image of 250×200 pixels with around 700 labeled synapses takes around 3 hours. Furthermore, computer display setups, illumination situations and a lot more of external influences can lead to a high inter- and intra-observer variance. Thus, in order to obtain a training set and to monitor the experts' performance, we record not only the synapse positions, but also one of three quality labels (A: 100% a synapse, B: quite sure a synapse, C: might be a synapse) is assigned to each synapse according to the expert's personal judgment.

We combine the three human experts to different synthetical gold experts G_X. Synapses are accepted in the gold expert when they are labeled by at least two human experts. When merging the position lists of different experts, positions p and p' are considered the same if their euclidian distance $d(p,p') < 2$. This stringent constraint is needed because synapses are very densely distributed

in the image and thus can be located very close to each other. Depending on the merged lists, we obtain different gold experts. Gold expert G_A subsumes all synapse positions which are labeled quality A by at least two experts, representing a set of 244 synapse positions. G_B (83 positions) represents all positions marked quality B by at least two experts and G_C (59) all positions marked quality C by at least two experts. In G_{AB} (417) all synapses which are labeled either quality A or B by at least two experts are combined. G_{ABC} (589) subsumes all synapses of any quality labeled by at least two experts.

2.2. Training setup

To obtain a positive training sample set S_X^+, for each synapse position in a gold experts list the gray values of the 7×7 neighborhood are written to a 49-dimensional gray value vector. In the experiments, the positive sets S_A^+ and S_{AB}^+ are used for training based on G_A and G_{AB} respectively.

Negative sample sets S_X^- are created by randomly choosing image positions p fulfilling the constraint that for any positive position p' the euclidian distance $d(p,p') > 3$. Image patches of negative and positive samples are therefore allowed to overlap.

Two balanced training sets, *i.e* each containing an equal number of negative/positive samples, are obtained (S_A and S_{AB}). A fully automatic ten fold cross validation approach is chosen for kernel parameterization.

2.3. Postprocessing

To compute synapse positions in a fluorescence micrograph, at each position p in the image its 7×7 neighborhood gray value vector is extracted and classified by a trained SVM. If the patch is classified as a synapse patch, its distance to the hyperplane, which is interpreted as the confidence value, is plotted to position p of a new matrix of the same size as the fluorescence image. We thereby obtain a new image called the *confidence map*.

Areas of high confidence correspond to synapse locations, where the highest confidence is likely to be the synapse center. In the confidence map, all positions p with a value greater or equal to a given threshold t and which are a local maximum of their 5×5 neighborhood are considered as synapses.

3. RESULTS

We analyze the generalization performance of the SVM synapse detection performance on a 250×200 pixels sized micrograph. Synaptophysin labeled fluorescence micrographs are evaluated with RBF-Kernel SVMs and S_A and S_{AB} as training sets.

3.1. Human detection performance

We use the standard ROC [12] measures *sensitivity* (SE) and *positive predictive value* (PPV) to statistically assess the performance of the human experts. SE measures the percentage of synapses which are given in the gold standard and are also detected by a classifier, here a human expert. PPV reflects the percentage of synapses detected by the classifier which are also represented in the gold standard.

As gold standard we used G_A, G_B, G_C and G_{ABC}. Thus we can analyze the humans quality in categorization and their overall detection performance. Table 1 gives an overview of the mean and standard deviation of SE and PPV for the different gold standards. It can be seen by looking at the SE and PPV values of the first three rows that there is a core set of quality A, B and C synapses which

Gold standard	SE_m	SE_{sd}	PPV_m	PPV_{sd}
G_A	83.74	10.33	77.19	12.11
G_B	69.88	13.63	35.63	6.16
G_C	66.67	23.25	24.70	5.19
G_{ABC}	81.61	17.14	81.24	10.72

Table 1. Human detection performance. Mean (m) and standard deviation (sd) (in percent) of SE and PPV achieved by human experts with respect to different gold standards.
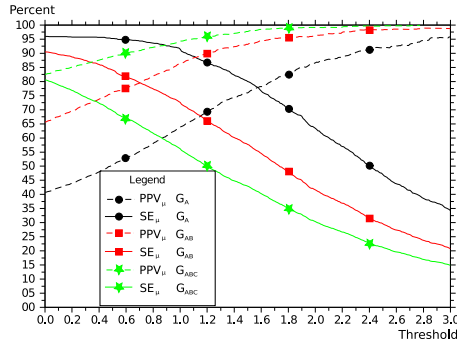

Fig. 2. Evaluation of SVM detection results. SE_μ and PPV_μ are calculated over all five runs for each threshold up to 3. Gold standards are G_A, G_{AB} and G_{ABC}.

Training set	G_A	G_{AB}	G_{ABC}
S_A	78.28, **78.85**	78.66, **64.98**	80.61, **72.45**
S_{AB}	71.23, **70.39**	80.43, **79.06**	83.09, **69.80**

Table 2. Mean SE (normal font) and mean SE for non training samples (bold face font),in percent, calculated over all five runs with individual threshold selection so that PPV $\geq$ 80% with respect to three gold standards (G_A, G_{AB}, G_{ABC}). As training sets S_A and S_{AB} are used.
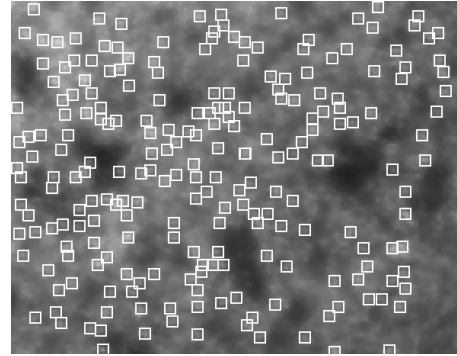

Fig. 3. Synapse detection result in synaptophysin labeled sub-image (250×200 pixels). Detected synapses are marked by a white 7×7 pixels sized box.

are labeled and categorized the same by all three human experts. The overall detection performance (last row) has a PPV of 81.24% and a SE of 81.61%.

3.2. SVM detection performance

Two strategies are applied to evaluate the SVM detection performance. Those are motivated by the strong limitation on available gold standard data. Furthermore, test and training samples are distributed across the whole sub-image to reduce illumination caused problems. This complicates the calculation of a PPV value limited to test samples. Thus, PPV and SE values are calculated on training and non training samples. Based on those results, SE values are recalculated limited to non training samples.

For detection accuracy assessment, we apply a five fold cross validation experiment. 80% of the training sets are used for training and SE and PPV are averaged over all five runs, yielding SE_μ and PPV_μ. In the confidence map evaluation we vary the threshold t to adjust the number of detected synapses. For each resulting list of synapse positions based upon a threshold t, SE_μ and PPV_μ are calculated with respect to three gold experts (G_A, G_{AB}, G_{ABC}). Figure 2 gives an overview of how SE_μ and PPV_μ develop with increasing threshold with respect to different gold standards. Training is performed on S_A.

The plotted values show that by defining a high threshold (≥ 2.28) we achieve a $PPV_\mu \geq 90\%$ with respect to G_A with a SE_μ of 55%. For the less conservative gold standards (G_{AB}, G_{ABC}) PPV_μ increases. Decreasing the threshold results in a higher SE_μ with decreasing PPV_μ. Nevertheless, for a SE_μ of 96% for G_A, still a PPV_μ of 66% and 82% with respect to G_{AB} and G_{ABC} are obtained, respectively. Thus, decreasing the threshold allows for more low quality synapses.

Optimal thresholds can vary for each individual run. To obtain the actual mean SE for a PPV of 80%, thresholds are selected individually for each SVM so that PPV $\geq$ 80% and with highest SE. Table 2 (normal font) gives an overview of the mean SEs with respect to G_A, G_{AB} and G_{ABC} if thresholds are selected individually. Table 2 shows that training with a lower quality set yields higher mean SE for low quality synapses and thus enables a selective tuning of the detection algorithm.

All runs are trained and evaluated on the same image. This gives slightly biased results, because also the training samples are included in the evaluation. To assess the performance of the classifier with respect to only non training samples, the sensitivity for non training samples is calculated for each classifier based on the previously defined individual thresholds yielding 80% PPV. Table 2 (bold face font) gives an overview of the mean SEs for non training samples with respect to the different gold experts. Again, training with high quality synapses yields an SVM optimized for detection of high quality synapses. Training with a lower quality set tunes the SVM with respect to detect low quality synapses. In most biological applications, optimizing the SVM for a detection of high quality objects is of greatest importance.

As an example, figure 3 shows one detection result for the image shown in figure 1 with a PPV of 80% and a SE of 81% (SE for non training samples: 87%). In this analysis 49 patches are falsely classified as synapses. When analyzing these patches it can be seen that most of them have synapse like appearance. When compared to G_{AB} and G_{ABC} 37 and 46 of the false positive patches are included, respectively. The remaining three are classified by one of the experts as synapses, two of them category B, one category A.

Also inter-image classification performance is measured. SVMs are trained with S_A on synaptophysin image 1(b) and tested on a

Training set	G_A (35)	G_{AB} (75)	G_{ABC} (131)
S_A	58.2 (91.4)	48.6 (82.4)	61.6 (80.2)

Table 3. SE (PPV) values in percent with respect to a PPV $\geq 80\%$ for different gold standards. The number of synapse positions in each gold standard is given in braces. Training is performed on synaptophysin image 1 with S_A, testing on synaptophysin image 2.

synaptophysin image of another tissue sample (image 2, not shown). Table 3 subsumes the mean SE and PPV values obtained for different gold standards. SEs of up to 61.6% could be achieved. As the gold expert sets are of small size, few false positives or false negatives are sufficient to greatly reduce the PPV and SE. A larger gold expert set would lead to more trustworthy values. Furthermore, as image 2 has a shifted and compressed histogram with respect to the histogram of image 1, the detection performance furthermore depends on the adaption of both histograms. The adaption performed in this study is based on shifting the histogram of image 2, so that signal values of both images more or less occupy the same gray value range. Nevertheless, by applying more sophisticated adaption strategies and/or modifying the training set, better performance should be achievable.

As synapses appear as glowing dots, which might suggest the application of thresholding approaches, we tested the performance of a basic image processing approach based on thresholding and morphological operators [13] for synapse detection. The obtained detection results are compared against G_A, G_{AB} and G_{ABC}. The best result which could be obtained is a PPV of 80.12% and a SE of 21.90% for G_{ABC}. For all other gold experts, no PPV $\geq$80% could be achieved. These results are inferior to SVM based detection.

4. DISCUSSION

The human detection results show that the evaluation of fluorescence labeled micrographs of neural tissue is of high difficulty even for human experts. Their overall detection performance has a SE of 82% and a PPV of 81%. For the quality categorization they are certain for a specific core set of synapses for each quality, but there is still variation in their personal quality estimate.

With our automated SVM based detection method we are able to achieve similar SE and PPV values as the human experts. Our acceptable PPV cutoff is set at 80%, therefore in some cases we achieve slightly lower SE values than the human experts. By selecting a representative training set, we are able to optimize the SVM to detect a specific quality class. Thus, training set selection gives one possibility to model the human experts quality assignment. A second way is given through varying the threshold for confidence map evaluation. High confidence values yield certain synapses, low confidence values uncertain synapses.

5. CONCLUSION

Our results suggest that fully automated threshold based approaches are less precise to perform segmentation because of non-homogeneous brightness and non trivial single synapse signal, which is partly caused by the high synapse density. With our machine learning approach we are able to detect nanoscale sized objects and achieve a detection performance similar to the human experts. Through varying the threshold of the confidence map evaluation and/or by adapting the training set, more or less low quality synapses are detected automatically. Thus, all further analysis can be based on a reproducible set of synapses with a tunable degree of certainty which allows biologists to analyze their data on their desired level of flexibility. Furthermore, by applying SVM based detection we give the opportunity to directly incorporate the experts' knowledge into the detection method. Detecting more than 2500 synapses in a 658×517 pixels micrograph takes less than 20 seconds on a standard PC. With respect to the human labeling of a 250×200 pixels sized micrograph which takes about 3 hours, this is an enormous time and cost reduction.

6. REFERENCES

[1] A. Gordon, A. Colman-Lerner, T.E. Chin, K.R. Benjamin, R.C. Yu, and R. Brendt, "Single-cell quantification of molecules and rates using open-source microscope-based cytometry," *Nature Methods*, vol. 4, pp. 175–81, 2007.

[2] Y. Tang, J.R. Nyengaard, D.M.G. De Groot, and H.J.G. Gundersen, "Total regional and global number of synapses in the human brain neocortex," *Synapse*, vol. 41, pp. 258–73, 2001.

[3] R.D. Terry, E. Masliah, D.P. Salmon, N. Butters, R. DeTeresa, R. Hill, L.A. Hansen, and R. Katzman, "Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment," *Annals of Neurology*, vol. 30, pp. 572–80, 2004.

[4] W. Schubert, B. Bonnekoh, A.J. Pommer, L. Philipsen, R. Böckelmann, Y. Malykh, H. Gollnick, M. Friedenberger, M. Bode, and A.W.M. Dress, "Analyzing proteome topology and function by automated multidimensional fluorescence microscopy," *Nature Biotechnol.*, vol. 24, pp. 1270–78, 2006.

[5] M. Bode and A. Krusche, "Toponome imaging system (TIS): imaging the proteome with functional resolution," *Nature Methods*, vol. 4, pp. 541–47, 2007, Application note.

[6] M. Friedenberger, M. Bode, A. Krusche, and W. Schubert, "Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures," *Nat. Protoc.*, vol. 2, no. 9, pp. 2285–94, 2007.

[7] M. Bode, M. Irmler, M. Friedenberger, C. May, K. Jung, C. Stephan, H.E. Meyer, C. Lach, R. Hillert, A. Krusche, J. Beckers, K. Marcus, and W. Schubert, "Interlocking transcriptomics, proteomics and toponomics technologies for brain tissue analysis in murine hippocampus," *Proteomics*, in press.

[8] T.W. Nattkemper, "Automatic segmentation of digital micrographs: A survey," in *Proc. of 11th World Congress on Medical Informatics (MEDINFO)*, San Francisco, USA, 2004.

[9] R. Jagoe, J.H. Steel, V. Vucicevic, N. Alexander, S. Van Noorden, R. Wootton, and J.M. Polak, "Observer variation in quantification of immunocytochemistry by image analysis," *Histochem. J.*, vol. 23, no. 11-12, pp. 541–547, 1991.

[10] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag New York, 1996.

[11] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV*, 1998, pp. 839–46.

[12] J.A. Hanley, "Reciever operating characteristic mehtodology: the state of the art," *Crit. Rev. in Diagn. Imaging*, vol. 29, pp. 307–35, 1989.

[13] R.C. Gonzalez and R.E. Woods, *Digital Image Processing (2nd Edition)*, Prentice Hall, January 2002.