Contextual grouping in a concept: a multistage decision strategy for EM segmentation

Erhan Bas[1], Mustafa G. Uzunbas[2], Dimitris Metaxsas[2], and Eugene Myers[1]

[1]Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA
[2]Rutgers University, Computer Science Dept., New Jersey,USA

Abstract. In order to extract useful information from EM images, such as segmentation, it is compulsory to characterize regions structurally, as well as contextually. For that reason, we propose a multistage decision mechanism that utilizes underlying differential geometric properties of objects in a biologically inherited framework. Consequently, we start with an initial feature selection procedure to select most relevant features to characterize distinct regions, such as membrane, cytoplasm and outliers. Similar to a topographic map, a random-forest classifier is employed to highlight mountain ridge like structures, e.g. membranes as well as plateaus, e.g. cytoplasm. In order to extract the underlying geometry of structures on this topographic map, especially membrane like structures, principal surface analysis is utilized. This unsupervised technique returns highly sparse yet accurate low dimensional representation of the data and especially characterizes membrane like regions. A task specific, second stage decision mechanism is employed to distinguish contextually different mitochondria and cell boundary membranes. This second stage learning/decision mechanism is based on the appearance, the initial topographic map with its low dimensional reconstruction and expert supervision on different types of membranes. Initial results on individual EM slices indicate that the proposed approach can successfully segment objects with minimal expert supervision and can potentially form a basis for a larger scale volumetric data interpretation.

1 Introduction

With the recent developments in serial section transmission electron microscopy (ssTEM) technology, neuroscientists now have the chance to work on very large scale and high resolution volumes towards understanding the functionality of neuronal structures. ssTEM imaging is the first step towards obtaining a complete neuron wiring diagram and provides potential of bringing a huge impact on the understanding of the whole nerve system by providing sufficiently high (synaptic) resolution in tractable amount of time.

Currently, there are remarkable progress in detection, segmentation and reconstruction of neuronal arbors; however, the state-of-the-art methods are not completely sufficient in practice since the accuracy requirement is extremely high. The whole idea behind reconstructing the neuronal circuit relies on the accurate detection and segmentation of branchings and mergers of neuronal structures, thus small merge or split errors would make the results useless [1]. In 2D stack of ssTEM images, in order to reconstruct neuronal arbors, one requires to identify the boundary (cytoplasmic membrane) that encapsulates adjacent neurons. Identification of boundary from EM images is not a trivial task due to artifacts related to the chemical fixation process in the tissue preparation and the limitation in the section thickness as a result of anisotropic

resolution [Albert Cordona, personal communication]. Extracting the correct boundary of interest might need grouping relevant boundary components together which are scattered on image domain.

In the presence of weak, cluttered or even no local boundary evidence, supervised learning methods have been increasingly popular as a means of achieving more accurate boundary detection [2], [3], [4], [5]. In [2], a multi layer convolutional neural network is utilized to classify pixels as foreground and background. The network presents two critical properties such that the classification filter for each layer is obtained directly from the data, and the multiple convolutions throughout the layers of the network provide an indirect filter effects. In their work the neural network contains huge number of parameters and therefore is computationally intensive and requires very large training set.

To reduce computation time and the memory consumption of optimization methods employed on affinity graphs (in the presence of large scale data as in EM images), sparsification of the affinity graph has been used as a solution. In these approaches the nodes of the graph represents super-pixels (small regions) and the affinity is defined as the relationships between adjacent super-pixels. For instance, in [3], a hierarchical segmentation procedure based on local statistical learning and topology-preserving grouping is proposed. In a two step hierarchal scheme, first, image derived features for each pixel in the pixels neighborhood is computed and then transformed into boundary map via random forest classifier and an over-segmentation (super pixels) is obtained via watershed transform on the basis of the classification scores. Next, and finally, adjacent regions (super pixels) are merged with a second random forest classifier which is trained on super-pixels with a set of features in order to segment target regions.

There have also been some efforts on bringing high level contextual information into consideration for accurate boundary delineation. Graph cut segmentation algorithm is one of the global optimization based solver of the affinity graphs via which, both local, nonlocal and contextual constraints can be incorporated into optimization. For example, in [6], the flux of the gradient vector field has been incorporated into graph cut approach as a solution to prevent gaps in segmentation of thin and elongated boundaries. However, the gradient flux is reported to introduce a large amount of false positives when the image gradient is remarkably high at the undesired image regions in addition to the target segmentation borders. As a more specific work to ssTEM images, we refer [7] as an instance which uses the flux of gradient vector field in segmentation of cytoplasmic membranes. In [8], the probability output of a random forest classifier is used in a regular graph cut energy (cost) function. Choosing correct weighting parameters which balances the effect of different energy terms is also task specific.

As stated above, training a classifier based on the prior knowledge of the shape and the appearance of the membranes is a common approach to highlight cell boundaries. Such a classifier is ideally expected to highlight only cytoplasmic membrane regions; however, notice that, other elements in a cell such as vesicles or mitochondria have their own inner/outer membranes, thus make the recognition task more challenging. For that reason, in this paper, we propose a framework composed of two stage decision mechanisms that can identify accurate cytoplasmic membranes (we will call this as the cell boundary in the rest of the manuscript). In the first stage, we highlight potential cell boundary candidates, whereas in the second stage such candidates are interpreted into meaningful elements based on their relative spatial distribution and appearance.

2 Method

In this section, we give the details of our approach and also summarize the overall flow via diagram shown in Fig. 2. Briefly, our approach is composed of training and testing sessions. In the training, two classifiers are built. The first classifier is employed on

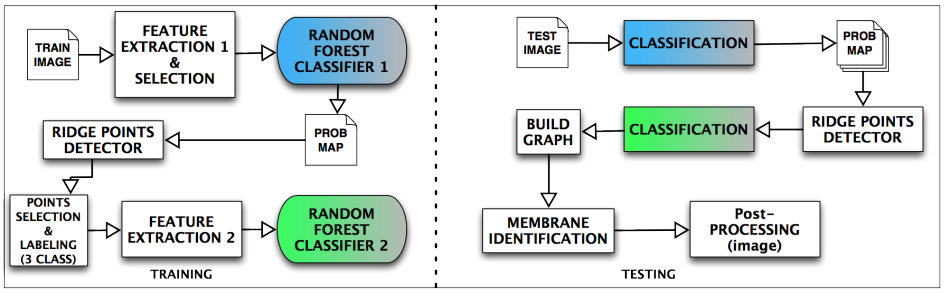


Fig. 1. The block diagram of our workflow with its training and test parts.

the pixel domain of the image (see Section 2.1) while the second is on the intrinsic membrane manifold (see Section 2.3). Having the two classifiers prepared, given a test image, first a membrane detection (Section 2.1), then the ridge point detection for low rank representation of the data (Section 2.2), graph construction from low rank representation and the membrane identification on this graph (Section 2.3) is applied consecutively. Finally, morphology based post processing is utilized to obtain the ultimate segmentation.

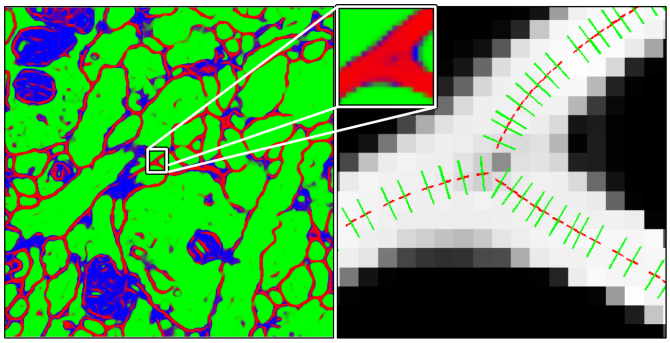
2.1 Membrane Detection

In order to highlight membrane like structures in EM slices, we used random forest classifier [8]. We start with a feature selection procedure to select the most relevant features with respect to the groundtruth on a training set. Initial pool of features is a mixture of appearance, e.g. edge, min/max intensity and structural, e.g. curvature, filtering responses evaluated at different scales (in total 243 distinct features). We estimated the mutual information of a feature [9] with groundtruth and selected the first 4 features (3 different scales of anisotropic diffusion and Gabor filter response). The total number of the features are arbitrarily selected based on the visual inspection.

In order to extract accurate boundary from EM images, one needs to distinguish cell boundary from the rest of the elements. However, note that, any element that resamples the cell membrane in terms of appearance or structure will also be miss classified. Moreover, cell membranes are not always smooth and might take arbitrary shapes, especially around synapses. Consequently, we approach the recognition task as a three class problem where we highlighted *i*) cell boundary, *ii*) cytoplasm, and *iii*) ambiguity regions in a training set. Ambiguities are locations where other membrane types such as mitochondria inner/outer membrane or synapse locations are present. In fact, the major challenge in boundary decision is to develop efficient and accurate algorithms for large volumes of data to categorize such ambiguity regions into cell membrane or cytoplasm. For that reason, we calculate the low rank membrane representation that accurately governs the underlying biological structure, yet provides sparse representation of the data.

2.2 Low rank membrane representation

Fig. 2-a shows the result of the random-forest classifier on a sample test image. Each color depicts the probability of having one class where red, green and blue represent cell membrane, cytoplasm and ambiguity regions correspondingly. In order to sparsely reconstruct a low rank representation of the data, we use the membrane probability map (red channel) and estimate the intrinsic structure of the membranes. Unlike previous approaches that partition the probability map into arbitrary regions, e.g.



(a) (b)

Fig. 2. (a) Result of 1st stage random forest classifier on a test image. (b) A closer look to the detected point locations on the ridge of the cell membrane probability map. Note that the probability map is just the red channel of image in (a). Unit vectors in the tangent space (red) and the vectors (green) that are orthogonal to the tangents are also rendered.

super-voxels, our sparse reconstruction of the data inherits the differential geometric properties of the cell tissue.

In order to obtain a sparse representation of the data, we utilized nonparametric principal curve projections [10, 11]. Intuitively, our goal is to find samples on the ridge of the probability map that is sufficiently sparse to efficiently analyze the data, yet dense enough to estimate the embedded underlying structure, i.e. curve. In fact, cell membranes are 2D surfaces embedded in 3D volume, however, we restrict ourself to the estimation of curves in 2D slices due to the high anisotropy in the data. For that purpose, we used the first and second derivatives of a ridge regression function jointly in order to find the ridge locations where gradient becomes orthogonal to the maximal curvature direction[1]. Fig. 2-b shows the overlay of the detected point locations on the ridge of the cell membrane probability map.

2.3 Membrane Identification

A crucial recognition step in membrane recognition is to analyze the data in the context and to classify ambiguities in the data into correct elements (cell membrane/cytoplasm). Pairwise similarities with the neighboring elements, appearance, and shape are necessary features to decide if an ambiguous location in the data is cell boundary or not. In order to identify different types of membranes and uncertainty regions, we classify the projected membrane points into three: isolated (L_1), transition (L_2) and inner samples (L_3) based on their cross-section probability profiles. This grouping is done via second stage random forest classifier which is trained on the probability profile along the normal direction of the tangent space of each projection point. Fig. 3 shows common uncertainty regions (dense blue regions) where the recognition task is not trivial. Fig. 3-a displays the 3 channel probability map obtained from the first stage random forest classifier. Orthogonal profiles of the probability values across isolated, transition and inner (cell, mitochondria outer and inner membranes with their associated numbers respectively) are overlaid on the probability map on the left. Similarly, membranes in the transition and inner regions of the synapse location are depicted on the right. In general, membranes have distinct profiles based on the spatial location in the cell. An isolated sample lies between cytoplasmic region, whereas samples in the

[1] Maximal curvature direction is defined as the eigenvector of the local Hessian matrix that has the largest absolute eigenvalue. For details see [10].

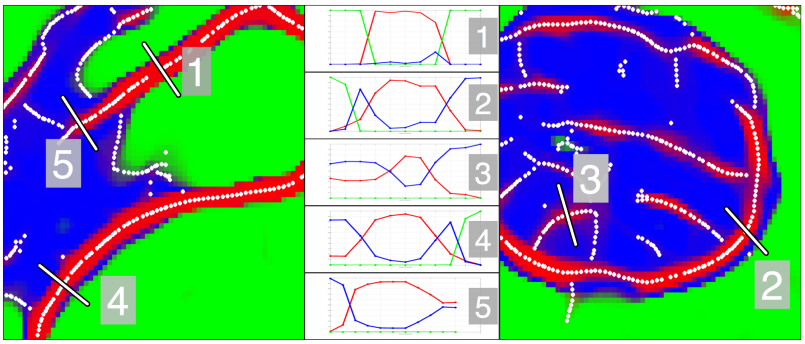


Fig. 3. Illustration of feature vectors for the three classes (mitochondria - synapsis borders, cell membrane and mitochondria - synapsis inside) of projection points.

mitochondria or synapse are encapsulated by blue channel. Lastly, transitional regions are boundary between these two.

Motivated by the aforementioned observations, we formulate the membrane identification problem as a competitive label propagation, operating over a graph network which is well matched with our sparse representation.

Let $\mathcal{V}$ be the set of projection points $L := L_1 \bigcup L_2 \bigcup L_3$ and E be the pairwise edges between them with constructed sparse graph $\mathcal{G} = \langle \mathcal{V}, E \rangle$. Edges between samples are determined by constrained Delaunay triangulation, where any edges that passes through high probability cytoplasmic regions are deleted from the graph. In the proposed competitive label propagation process, two classes: isolated (teal) and inner (yellow) propagate in the network and compete each other to occupy the transition (brown) class nodes. Starting from an initial state $L^{t=0} = \bigcup L_i^{t=0}$, $i = 1 \dots 3$ and $\Phi(t=0)$ the iterative competition model grows a dynamic front to span all unlabeled nodes ($j \in L_3$) in the graph to cover the whole data. The decision of an unclaimed label at iteration t is defined as the following optimization:

$$\begin{aligned} \tilde{L}_3^t(j) \quad & \Longleftarrow \quad \arg\max_k \big\{ |(\mathcal{N}_{k,t}(j))| \big\}, \quad k \in 1, 2 \\ \text{subject to} \quad & \mathcal{N}_{k,t}(j) \in L_1 \cup L_2 \\ & j \cap \mathcal{N}_{k,t}(j) \neq \emptyset \\ & \Theta(\mathcal{N}_{k,t}(j)) < \Phi(t) \end{aligned} \tag{1}$$

where optimization is formulated as majority voting in the graph. Here, $|(\mathcal{N}_{k,t}(j))|$ is the cardinality of the neighborhood of j having class label k at iteration t. First and second constraints indicate the propagating front of the competitive voting and the third constraint indicates the current state decision level $\Phi(t)$ at iteration t. In our competitive voting model, state decision level is a monotonically non-decreasing function and is a measure of anisotropy in the nodes. More clearly, as the front grows and occupies, unlabeled nodes that are intrinsically similar to the labeled instances are ideally favored. Estimated low rank representation of the data is utilized to align the tangent space of an unlabeled instance with a labelled one, hence highlight the local anisotropy in the graph. In order to align samples, dissimilarity/divergence of samples from the underlying structure is calculated as the total angle, $\Theta()$ between the aligned tangential spaces. The total angle is calculated as the sum of the interior angles between the tangential vectors and the edge between nodes. At a given decision level $\Phi(t)$, front continues to grow as long as there exits samples in the feasible set of the above optimization. After a single pass through all samples in L_3, $\Phi(t)$ level is relaxed with an arbitrarily selected percentage until iterative search process described above is finished.

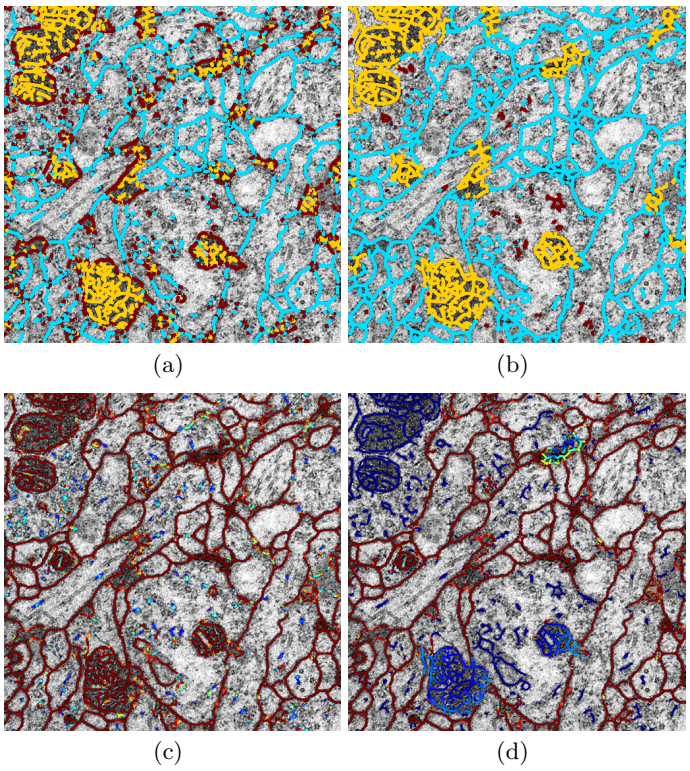
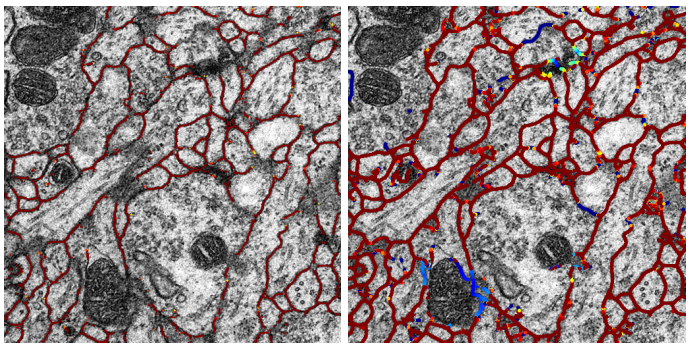


Fig. 4. (a) A classification result via second-stage random forest classifier based on the features shown in Fig. 3. (b) Projection points with their label colormap. Points classified as membrane are shown in blue and non-membrane in yellow. Red points are discarded since they have no valid connection to other points on the constrained map. (c) Projection points rendered with their initial affinity colormap. (d) Projection points rendered with their updated affinities after affinity propagation and False Positive removal.

Affinity Propagation and False Positive Removal Ideally, the competitive label propagation method is expected to return two distinct clusters (cell boundary and non-cell boundary) that span the whole graph. However, due to the complexity and difficulties of the data as mentioned before, there might exist false positively (non-cell boundary) labelled nodes although they belong to the cell boundary (see Fig. 4-b).

Having the two groups of nodes, the nodes that are actually parts of the cell boundary has to be distinguished from the non-cell boundary label. For this purpose, affinity values are assigned for each labeled points (see Fig. 4-c) based on the result of the first stage random forest classifier which highlights regions that are likely to be membrane. From this point, we target identifying non-cell boundary by employing a region growing type of algorithm which is run on the connectivity graph of the nodes that was built previously as defined in Section 2.3. Before running the region growing, the affinity values for the nodes that are classified as non-cell boundary have to be updated since the affinities on mitochondria boundaries are very similar to cell boundaries. Thus, the affinities are updated such that total affinity level difference between the true positives and false positives are increased. In the update scheme, new affinities are computed based on a descriptor that is function of both high level shape information e.g. curvature and solidity, of the underlying regions of the false positive labeled points. In Fig. 4-c and d, the merit of this update scheme can be observed clearly.



(a) (b)

Fig. 5. (a) Initial set of nodes that are chosen above a very large affinity threshold. Note that points are rendered with affinity colormap. (b) Final cell boundary points obtained after region growing on the connectivity graph.

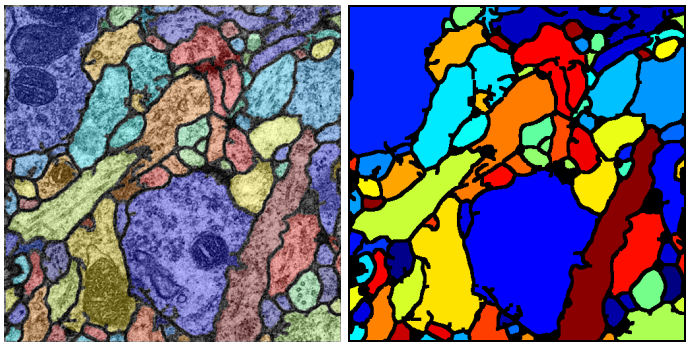


Fig. 6. Final segmentation results for an example section from the test set.

Region growing starts from an initial set of nodes (see Fig 5-a) which are selected as the nodes having affinity greater than a very large threshold. Final cell boundary points are obtained as shown in Fig. 5-b. Given the final cell boundary points, segmentation of encapsulated neurons with morphological post processing steps is straightforward and out of the scope of this paper.

3 Results

As part of the "ISBI'12 Segmentation of neuronal structures in EM stacks" challenge, the method was evaluated on 30 sections from 2d stack of ssTEM images of Drosophila first instar larva ventral nerve cord (VNC). The microcube measures 2 x 2 x 1.5 microns approximately, with a resolution of 4x4x50 nm/pixel (see [12] for the details of the data set). We trained our classifiers on arbitrarily chosen 20 sections out of 30 training images. Our results were compared to the groundtruth using the evaluation program provided by the organizers. In order to evaluate the performance, Minimum Splits and Mergers Warping error, Rand error and the Pixel error were used (see [13] for the details of evaluation metrics). The performance of our method was reported as: $\%16.2303, \%0.1613$ and $\%10.939$ rates for Rand error, Warping error and Pixel error, respectively for all images in the test set. In Fig. 6, an example segmentation result on a section from the test set is shown.

4 Conclusion

In this paper, we have underlined many challenging facts about the reconstruction of neuron structures from ssTEM images and proposed a two stage decision mechanism for segmentation of ssTEM images by using both low level (differential geometric) and high level contextual properties of biological elements. We defined a membrane identification problem which can be solved over a sparse connectivity graph. We proposed an iterative competition based label propagation method for the membrane identification. Demonstrated results on 30 sections of the test data of ISBI'12 EM Segmentation Challenge look promising for future research directions. Future work includes extending current 2D label propogation into 3D with global optimization extensions.

References

1. Jain, V., Seung, H.S., Turaga, S.C.: Machines that learn to segment images: a crucial technology for connectomics. Current Opinion in Neurobiology **20**(5) (2010) 653–666
2. Jain, V., Murray, J., Roth, F., Turaga, S., Zhigulin, V., Briggman, K., Helmstaedter, M., Denk, W., Seung, H.: Supervised learning of image restoration with convolutional networks. ICCV (2007) 1–8
3. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.: Segmentation of sbfsem volume data of neural tissue by hierarchical classification. Proceedings of the 30th DAGM symposium on Pattern Recognition (2008) 142–152
4. Jurrus, E., Paiva, A.R.C., Watanabe, S., Anderson, J.R., Jones, B.W., Whitaker, R.T., Jorgensen, E.M., Marc, R.E., Tasdizen, T.: Detection of neuron membranes in electron microscopy images using a serial neural network architecture. Medical Image Analysis **14**(6) (2010) 770–83
5. Turaga, S.C., Murray, J.F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., Seung, H.S.: Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Comput **22**(2) (2010) 511–38
6. Kolmogorov, V., Boykov, Y.: What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. ICCV 2005 **1** (2005) 564 – 571 Vol. 1
7. Vu, N., Manjunath, B.: Graph cut segmentation of neuronal structures from transmission electron micrographs. ICIP (2008) 725–728
8. Kaynig, V., Fuchs, T., Buhmann, J.: Neuron geometry extraction by perceptual grouping in sstem images. CVPR (2010) 2902–2909
9. Peng, H.P.H., Long, F.L.F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(8) (2005) 1226–1238
10. Bas, E., Erdogmus, D.: Principal curves as skeletons of tubular objects - locally characterizing the structures of axons. Neuroinformatics **9**(2-3) (2011) 181–191
11. Ozertem, U., Erdogmus, D.: Locally defined principal curves and surfaces. J. Mach. Learn. Res. **12** (2011) 1249–1286
12. Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., Hartenstein, V.: An integrated micro- and macroarchitectural analysis of the ¡italic¿drosophila¡/italic¿ brain by computer-assisted serial section electron microscopy. PLoS Biol **8** (10 2010) e1000502
13. Jain, V., Bollmann, B., Richardson, M., Berger, D., Helmstaedter, M., Briggman, K., Denk, W., Bowden, J., Mendenhall, J., Abraham, W., Harris, K., Kasthuri, N., Hayworth, K., Schalek, R., Tapia, J., Lichtman, J., Seung, H.: Boundary learning by optimization with topological constraints. CVPR (2010) 2488–2495