

# Learning Context Cues for Synapse Segmentation in EM Volumes

Carlos Becker\*, Karim Ali, Graham Knott, and Pascal Fua

Computer Vision Lab, École Polytechnique Fédérale de Lausanne, Switzerland

**Abstract.** We present a new approach for the automated segmentation of excitatory synapses in image stacks acquired by electron microscopy. We rely on a large set of image features specifically designed to take spatial context into account and train a classifier that can effectively utilize cues such as the presence of a nearby post-synaptic region. As a result, our algorithm successfully distinguishes synapses from the numerous other organelles that appear within an EM volume, including those whose local textural properties are relatively similar. This enables us to achieve very high detection rates with very few false positives.

## 1 Introduction

New imaging technologies have been a key driver of recent advances in neuroscience. In particular, block face scanning electron microscopy (EM) can now deliver a  $4nm$  nearly isotropic sampling and produce image stacks that reveal very fine structures. Stacks such as those of Fig. 1(a) can be used to analyze the size, shape and distribution of synapses, which in turn will lead to an understanding of the connection strength between neurons and, in time, brain circuits.

Currently, analysis is carried out by manually segmenting synapses using tools such as Fiji [1]. This is not only a tedious and time consuming process but also an error-prone one. Furthermore, the need for expert knowledge and the growing size of these datasets render manual segmentation intractable and not amenable to crowd-sourcing methods. There has been great interest in automating the process. However, current synapse segmentation methods either require first finding cell membranes [2] or operate on individual 2D slices [3], thus failing to leverage the 3D structure of the data. By contrast, the recent method of [4] operates entirely in 3D. However, it does not exploit the contextual clues that allow human experts to distinguish synapses from other structures, such as endoplasmic reticula, which exhibit similar textural properties, as depicted by Fig. 1(b).

In this work, we propose an approach designed to take contextual cues into account and emulate the human ability to distinguish synapses from regions that merely share a similar texture. Thus, we significantly outperform the method of [4]. Our algorithm relies on features which compute sums of various image properties over cubes placed in an extended 3D neighborhood surrounding the voxel to be classified, as shown in Fig. 1(c). It then uses AdaBoost [5] to select the most informative ones.

---

\* This work was funded in part by the ERC MicroNano Grant

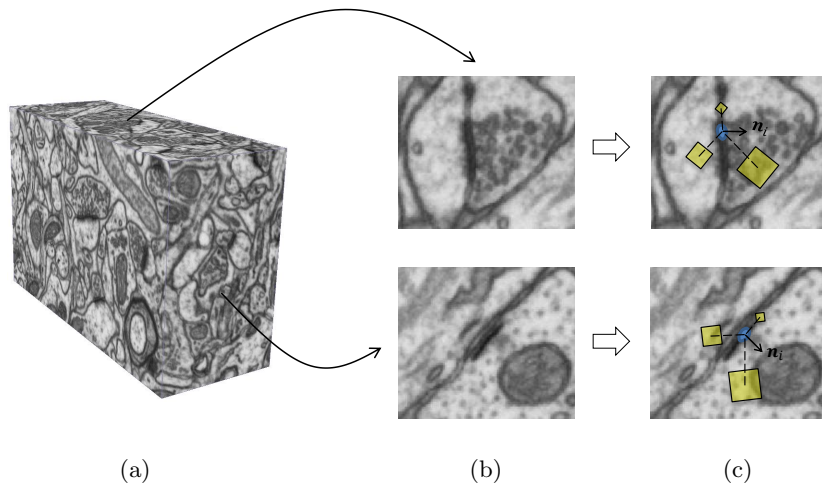


Fig. 1: Importance of context in synapse segmentation. The dark structures in the middle of both images look similar locally. However, only the ones in the top image constitute a synapse: the lack of vesicles in the bottom one is a indicative sign. The features we use are designed to capture this fact. To classify a voxel (blue), we consider sums over image cubes (shown as yellow squares) whose respective positions are defined relative to an estimated normal vector  $\mathbf{n}_i$ .

## 2 Related Work

Several fully automated approaches to reliable segmentation of organelles, such as mitochondria [6, 7] or neuronal membranes [8, 9], from 3D EM stacks have recently been proposed. However none of these methods exploit context in a meaningful way. Though features are extracted in a neighborhood around the voxel of interest, they are either pooled into global histograms [6, 7] or computed at predetermined spatial locations [8, 9]. The resulting classifier is therefore unable to hone in on arbitrary localized context cues. Along similar lines, a method for automated 3D segmentation of synapse in EM volumes using a Random Forest classifier has recently been proposed [4]. While this technique produces interesting results, it does not account for context as the features it uses simply measure various filter responses at the voxel of interest. The method is therefore unable to distinguish synaptic voxels from voxels exhibiting synapse-like textural properties. This is the limitation that our approach, specifically designed to utilize context cues, addresses. We run various filters over the EM stack but compute our features over arbitrarily sized cubes placed at arbitrary locations inside an extended neighborhood of the voxel to be classified. Next, we rely on a Boosting to select the relevant filter channels as well as the relevant cube locations and sizes. The resulting classifier is thus able to hone in on the presence of pre- and post-synaptic regions around the synapse. As a result, our method is shown to reduce the false alarm rate by a factor of 2 as compared to [4].

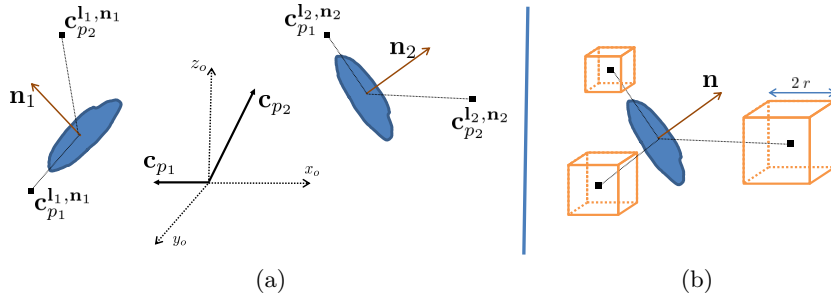


Fig. 2: (a)  $\mathbf{c}_p^{i, \mathbf{n}_i}$  define consistent locations relative to differently located and oriented voxels. (b) Cubes over which sums are computed at consistent locations.

### 3 Method

As shown in Fig. 1(b), it can be difficult to distinguish synapses from other structures based solely on local texture evidence. Human experts confirm the presence of a synapse by looking nearby for post-synaptic densities and vesicles. This protocol cannot be emulated by measuring filter responses at the target voxel [4], pooling features into a global histogram [6, 7] or relying on hand-determined locations for feature extraction [8, 9]. To emulate the human ability to identify synapses, we design features, termed context cues, that can be extracted in any cube contained within a large volume centered on the voxel to be classified, as depicted by Fig. 2(b). They are computed in several channels using a number of Gaussian kernels, as shown in Fig. 3. As this yields a total of 40,000 potential features, we rely on Boosting to select the most discriminative ones.

#### 3.1 Contextual Features

Given that synapses have arbitrary 3D orientations, we ensure that our context cues are computed at consistent locations across differently oriented synapses. We rely on the pose-indexing framework of [10] to enforce this consistency.

**Context Cue Location** Formally, let us consider voxel  $s_i$ , located at  $\mathbf{l}_i$  and an associated unit vector  $\mathbf{n}_i$ . In practice, we take  $\mathbf{n}_i$  to be the orientation of the eigenvector with largest eigenvalue of the Hessian operator, which can be expected to be normal to the synaptic cleft if there is one. Let

$$\mathbf{c}_p, \quad p = 1, \dots, P \quad (1)$$

denote a set of  $P$  locations expressed in the common  $x_0, y_0, z_0$  reference frame shown at the center of Fig. 2(a). These locations are translated and rotated to occur at consistent locations relative to a target voxel by defining,

$$\mathbf{c}_p^{i, \mathbf{n}_i} = \mathbf{l}_i + \mathbf{R}(\mathbf{n}_i)\mathbf{c}_p \quad (2)$$

where  $\mathbf{R}(\mathbf{n}_i)$  is a rotation matrix such that  $\mathbf{R}(\mathbf{n}_i)(0, 0, 1)^T = \mathbf{n}_i$ .

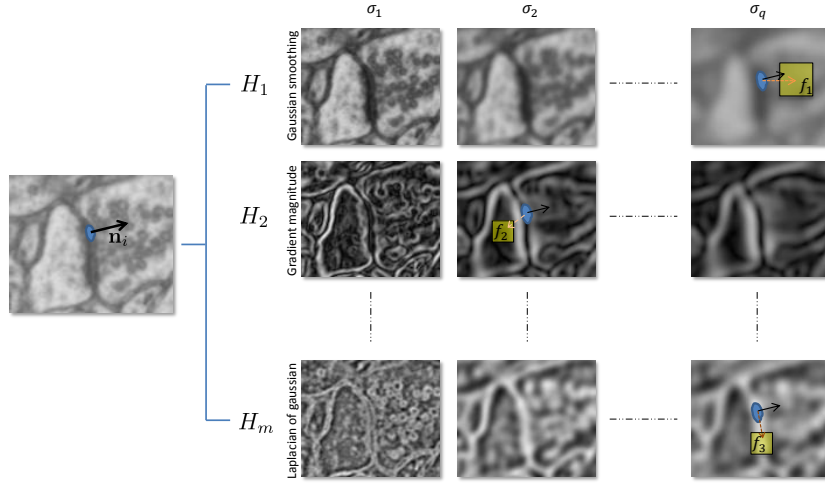


Fig. 3: Context cues. They are computed within the yellow rectangles whose coordinates are expressed with respect to the location of the voxel to be classified and the local orientation vector  $\mathbf{n}$ . Each  $H_i$  line depicts a specific channel designed to capture different statistical characteristics.

**Context Cue Features** Given the  $\mathbf{c}_p^{l_i, \mathbf{n}_i}$  locations of Eq. 2, our goal now is to compute image statistics inside cubic neighborhoods  $\mathcal{N}_r(\mathbf{c}_p^{l_i, \mathbf{n}_i})$  of edge length  $2r$  centered around these locations, such as those depicted in Fig. 2(b).

To this end, we precompute gradient magnitudes, Laplacians of Gaussian and eigenvalues of structure tensors and Hessians everywhere in the EM volume. Each of the resulting cubes of data, in addition to the original one, is treated as a data channel  $m$ ,  $1 \leq m \leq 5$ , and is smoothed using isotropic Gaussian kernels of increasing variance  $\sigma_n$  as in [4]. We denote the gray levels in the resulting data volumes as

$$H_{m, \sigma_n}(x, \mathbf{z}), \quad (3)$$

where  $x$  is the original EM volume and  $\mathbf{z}$  represents the 3D location. We take context cue features to be

$$f_{\mathbf{c}_p, m, \sigma_n, r}(x, \mathbf{l}_i, \mathbf{n}_i) = \sum_{\mathbf{z} \in \mathcal{N}_r(\mathbf{c}_p^{l_i, \mathbf{n}_i})} H_{m, \sigma_n}(x, \mathbf{z}). \quad (4)$$

In other words, we sum the smoothed channel output over the cubic boxes centered at all  $\mathbf{c}_p$  for all possible values of  $m$ ,  $\sigma_n$ , and  $r$ . This yields a set of  $K = 40000$  features, which we will denote for simplicity

$$f_k(x, \mathbf{l}_i, \mathbf{n}_i), \quad k = 1, \dots, K, \quad (5)$$

and which we use for classification purposes as discussed below.

### 3.2 Segmentation and Implementation Details

We create decision stumps by thresholding on the value of the  $f_k$  features of Eq. 5. These stumps are combined by a standard AdaBoost procedure [5] into a strong learner of the form

$$\varphi(x, \mathbf{l}_i, \mathbf{n}_i) = \sum_{t=1}^T \omega_t \mathbf{1}_{\{f_t(x, \mathbf{l}_i, \mathbf{n}_i) > \rho_t\}}. \quad (6)$$

Learning this classifier requires annotated training data. Since our contextual features are computed both for a given location and orientation, our training data must include both. As discussed above, we use the Hessian to compute the orientation for every voxel.

A potential difficulty arises from the fact that polarity also matters since pre- and post-synaptic regions look very different. We follow the pose-indexing methodology to exploit this structure. At training time, when dealing with synaptic voxels, we direct the orientation vector towards the pre-synaptic region in our positive examples and add the corresponding location with the flipped orientation vector to our list of negative examples. At run-time, we use the Hessian to compute  $\mathbf{n}_i$ , evaluate  $\varphi(\cdot)$  for both possible polarities, and retain the maximum response.

In practice, to speed-up the computation, we do not work on individual voxels of the EM volume. Instead, we group them into supervoxels that are regularly spaced small regions with relatively uniform gray level [11]<sup>1</sup>. We then run our classification scheme on their centers.

## 4 Experiments

We now demonstrate our approach and compare its performance to that of [4] on two different volumes from the adult rat brain, one from the somatosensory cortex, and the other from striatum. Their respective sizes are  $1500 \times 1125 \times 750$  and  $1423 \times 872 \times 318$ . The training set consists of 7 fully-labeled synapses in each volume plus negative samples labeled from non-synaptic voxels. For evaluation purposes, we labeled as synaptic or not each voxel in a somatosensory cortex subvolume of size  $655 \times 429 \times 250$ , which contains 24 synapses. It took approximately 40 minutes for each one, which highlights the need for automation. This voxel-wise ground truth allows for a robust quantitative evaluation and differs significantly from the protocol of [4], where synapses were annotated by an expert using a small sphere, making voxel-wise evaluation impossible.

Fig. 4 depicts the results of our quantitative evaluation on the ground-truth volume of the somatosensory cortex. We plot ROC curves, which show the True Positive Rate (TPR) as a function of the False Positive Rate (FPR). There are two parameters that we vary to obtain different sets of features  $f_k$ . The first is  $r_{\max}$ , the maximum possible size of the cubes over which sums are computed.

<sup>1</sup> Supervoxels source code available at <http://ivrg.epfl.ch/research/superpixels>

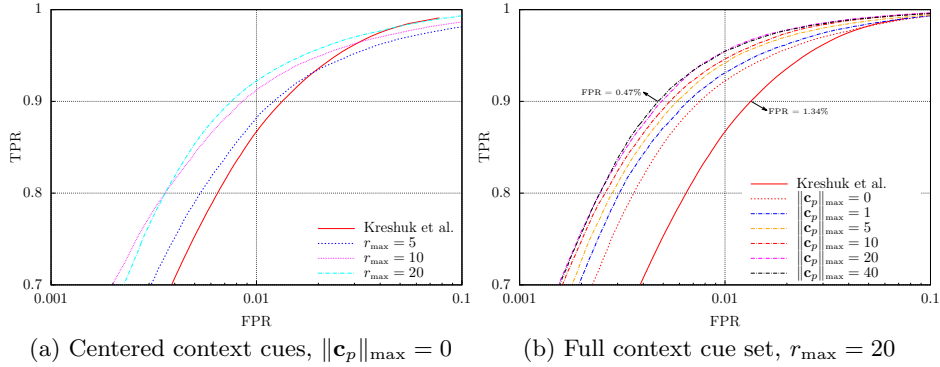


Fig. 4: Voxel-wise classification ROC curves for different radii of the sphere within which the contextual features are computed.

The second,  $\|\mathbf{c}_p\|_{\max}$ , is the radius of the sphere surrounding each voxel in which we extract our context features. Note that with  $\|\mathbf{c}_p\|_{\max} = 0$  and  $r_{\max} = 0$ , our feature set essentially reduces to that of [4] where responses are measured at the target voxel, without exploiting context.

In Fig. 4(a), we set  $\|\mathbf{c}_p\|_{\max} = 0$  and vary  $r_{\max}$ , meaning that the feature set only consists of cubes centered on the target voxel. As  $r_{\max}$  grows, more context is taken into account and performance improves. At 90% TPR our method outperforms [4] in all cases. In Fig. 4(b), we set  $r_{\max} = 20$  and vary  $\|\mathbf{c}_p\|_{\max}$ . In other words, we use our full feature set consisting of both centered and un-centered cubes. For  $\|\mathbf{c}_p\|_{\max} = 40$ , our approach yields a FPR almost three times smaller than the one of [4] at 90% TPR. Qualitatively, this difference in performance can be seen in Fig. 5 where the algorithm of [4] erroneously fires on non-synaptic tissue due to textural similarities.

Performing this computation on the whole somatosensory volume and thresholding at 90% TPR yields the result depicted by Fig. 6, where connected regions of less than 2000 voxels were removed and a gaussian filter of  $\sigma = 1$  was applied for visualization, as in [4]. The number of detected synapses is 358, which implies a density of 0.9 synapses per  $\mu m^3$  and is in agreement with the expected density for the somatosensory cortex region (layer II) [12]. Finally, Fig. 7 depicts similar qualitative results for the striatum dataset.

## 5 Conclusion

We presented a novel approach to synapse segmentation. It relies on a large set of image features, specifically designed to take spatial context into account, which are selected, weighed and combined using AdaBoost. We used two different EM datasets to demonstrate that our algorithm effectively distinguishes true synapses from other organelles that exhibit the same local texture.

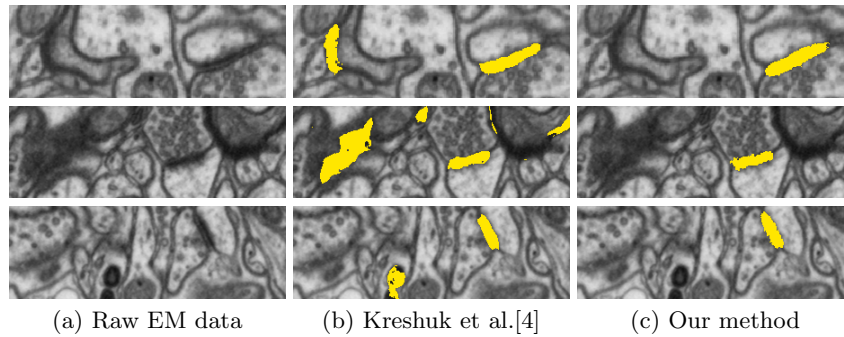


Fig. 5: Somatosensory cortex dataset results. Voxels labeled as synaptic by the method of [4] and ours overlaid in yellow. Threshold set to 90% TPR. Note the non-synaptic voxels found by [4] and correctly ignored by our method.

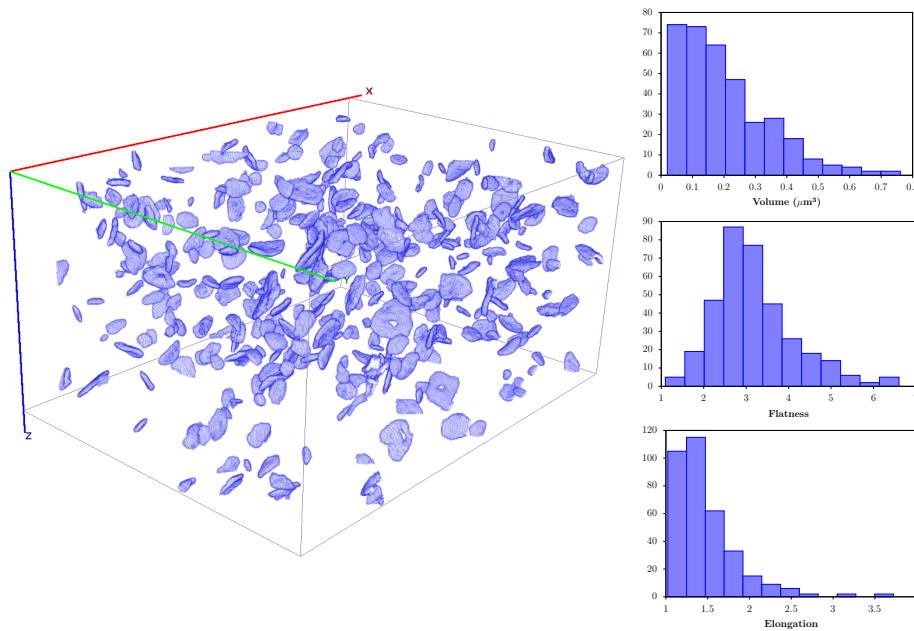


Fig. 6: 3D reconstruction of the detected synaptic voxels in the somatosensory cortex dataset. From our automated segmentation, various shape statistics can be computed for the purpose of analysis, such as the histograms on the right.

## References

1. Schmid, B., Schindelin, J., Cardona, A., Longair, M., Heisenberg, M.: A high-level 3D visualization API for Java and ImageJ. *BMC Bioinformatics* **11** (2010) 274

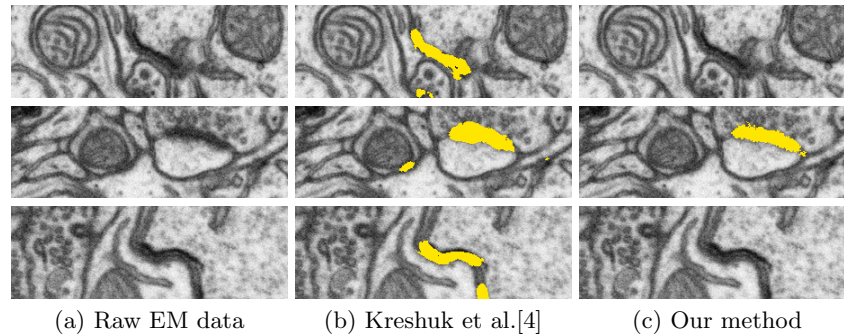


Fig. 7: Striatum dataset results. Voxels labeled as synaptic by the method of [4] and ours overlaid in yellow. Threshold set to 90% TPR.

2. Mishchenko, Y., Hu, T., Spacek, J., Mendenhall, J., Harris, K.M., Chklovskii, D.B.: Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. *Neuron* **67** (2010) 1009 – 1020
3. Herold, J., Schubert, W., Nattkemper, T.W.: Automated detection and quantification of fluorescently labeled synapses in murine brain tissue sections for high throughput applications. *Journal of Biotechnology* **149** (2010) 299 – 309
4. Kreshuk, A., Straehle, C., Sommer, C., Koethe, U., Knott, G., Hamprecht, F.: Automated segmentation of synapses in 3D EM data. In: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on.* (2011) 220 –223
5. Freund, Y., Schapire, R.: Experiments with a New Boosting Algorithm. In: *ICML.* (1996) 148–156
6. Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P.: Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features. *Medical Imaging, IEEE Transactions on* **31** (2012) 474 –486
7. Narasimha, R., Ouyang, H., Gray, A., McLaughlin, S.W., Subramaniam, S.: Automatic joint classification and segmentation of whole cell 3D images. *Pattern Recognition* **42** (2009) 1067 – 1079
8. Jurrus, E., Paiva, A.R., 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** (2010) 770 – 783
9. Venkataraju, K., Paiva, A., Jurrus, E., Tasdizen, T.: Automatic markup of neural cell membranes using boosted decision stumps. In: *Biomedical Imaging: From Nano to Macro, 2009. IEEE International Symposium on.* (2009) 1039 –1042
10. Fleuret, F., Geman, D.: Stationary Features and Cat Detection. *JMLR* **9** (2008) 2549–2578
11. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Suesstrunk, S.: SLIC Superpixels. Technical Report 149300, EPFL (2010)
12. DeFelipe, J., Marco, P., Busturia, I., Merchán-Prez, A.: Estimation of the number of synapses in the cerebral cortex: Methodological considerations. *Cerebral Cortex* **9** (1999) 722–732