A Fully Automated Approach to Segmentation of Irregularly Shaped Cellular Structures in EM Images

Aurélien Lucchi*, Kevin Smith, Radhakrishna Achanta,
Vincent Lepetit, and Pascal Fua

Computer Vision Lab, Ecole Polytechnique Fédérale de Lausanne, Switzerland

Abstract. While there has been substantial progress in segmenting natural images, state-of-the-art methods that perform well in such tasks unfortunately tend to underperform when confronted with the different challenges posed by electron microscope (EM) data. For example, in EM imagery of neural tissue, numerous cells and subcellular structures appear within a single image, they exhibit irregular shapes that cannot be easily modeled by standard techniques, and confusing textures clutter the background. We propose a fully automated approach that handles these challenges by using sophisticated cues that capture global shape and texture information, and by learning the specific appearance of object boundaries. We demonstrate that our approach significantly outperforms state-of-the-art techniques and closely matches the performance of human annotators.

1 Introduction

State-of-the-art segmentation algorithms which perform well on standard natural image benchmarks such as the Pascal VOC dataset [7] tend to perform poorly when applied to EM imagery. This is because the image cues they rely upon tend not to be discriminative enough for segmenting structures such as mitochondria. As shown in Fig. 1(a), they exhibit irregular shapes not easily captured using standard shape modeling methods. Their texture can easily be confused with that of groups of vesicles or endoplasmic reticula. Mitochondrial boundaries are difficult to distinguish from other membranes that share a similar appearance. Overcoming these difficulties requires taking all visible image cues into account simultaneously. However, most state-of-the-art techniques are limited in this respect. For example, TextonBoost uses sophisticated texture and boundary cues, but simple haar-like rectangular features capture shape [14]. In [5], SIFT descriptors capture local texture and gradient information, but shape is ignored.

Previous attempts at segmenting neural EM imagery include a normalized cuts based approach in [8]. More recently, [3] used a level set approach which is sensitive to initialization and limited to one object. [15] is an active contour approach designed to detect elliptical blobs but fails to segment mitochondria which often take non-ellipsoid shapes. In [4], a convolutional neural network considers only local information using a watershed-based supervoxel segmentation. Finally, [12] uses a classifier on texton features to learn mitochondrial texture, but ignores shape information.

* This work was supported in part by the MicroNano ERC project and by the Swiss National Science Foundation Sinergia Project CRSII3-127456.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 463–471, 2010.
© Springer-Verlag Berlin Heidelberg 2010

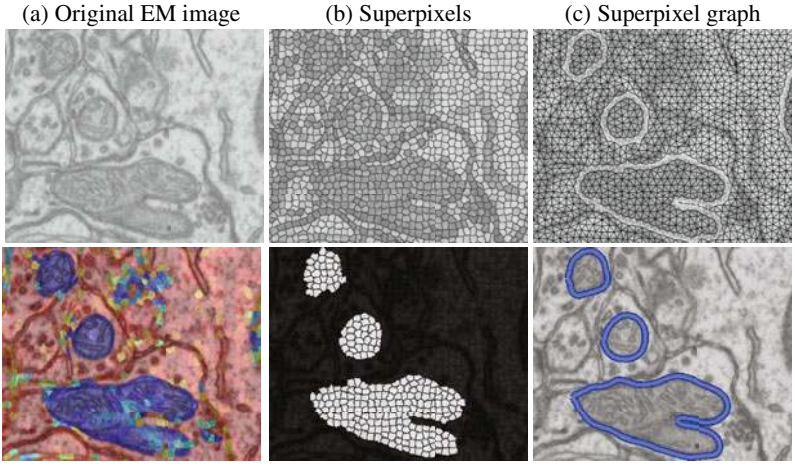

Fig. 1. Overview. **(a)** A detail of the original EM image. **(b)** Superpixel over-segmentation. **(c)** Graph defined over superpixels. White edges indicate pairs of superpixels used to train an SVM that predicts mitochondrial boundaries. **(d)** SVM prediction where blue indicates a probable mitochondrion. **(e)** Graph cut segmentation. **(f)** Final results after automated post-processing. Note: the same image is used in this figure for clarity; images in the training & testing sets are disjoint.

In this paper, we propose to overcome these limitations by:

1. **Using all available image cues simultaneously:** We consider powerful shape cues that do not require an explicit shape model in addition to texture and boundary cues.
2. **Learning the appearance of boundaries on a superpixel graph:** We train a classifier to predict where mitochondrial boundaries occur using these cues.

An overview of our approach appears in Fig. 1. We first produce a superpixel over-segmentation of the image to reduce computational cost and enforce local consistency. The superpixels define nodes in a graph used for segmentation. We then extract sophisticated shape, texture, and boundary cues captured by Ray [10] and Rotational [9] features for each superpixel. Support vector machine (SVM) classifiers are trained on these features to recognize the appearance of superpixels belonging to mitochondria, as well as pairs of superpixels containing a mitochondrial membrane. Classification results are converted to probabilities, which are used in the *unary* and *pairwise* terms of a graph-cut algorithm that segments the superpixel graph. Finally, an automated post-processing step smooths the segmentation. We show qualitatively and quantitatively that our approach yields substantial improvements over existing methods. Furthermore, whatever mistakes remain can be interactively corrected using well known methods [6].

2 Our Approach

2.1 Superpixel Over-Segmentation

Our first step is to apply a novel *k-means* based algorithm [13] to aggregate nearby pixels into *superpixels* of nearly uniform size whose boundaries closely match true image

boundaries, as seen in Fig. 1(b). It has been shown that using superpixels can be advantageous because they preserve natural image boundaries while capturing redundancy in the data [5]. Furthermore, superpixels provide a convenient primitive from which to compute local image features while reducing the complexity of the optimization by reducing the number of nodes in the graph.

2.2 Segmentation by Graph Partitioning

Graph-cuts is a popular approach to segmentation that splits an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ into partitions by minimizing an objective function [16]. As shown in Fig. 1(c), the graph nodes $\mathcal{V}$ correspond to superpixels x_i. Edges $\mathcal{E}$ connect neighboring superpixels. The objective function takes the form

$$E(c|x, w) = \sum_i \underbrace{\psi(c_i|x_i)}_{\text{unary term}} + w \sum_{(i,j)\in\mathcal{E}} \underbrace{\phi(c_i, c_j|x_i, x_j)}_{\text{pairwise term}} , \tag{1}$$

where $c_i \in \{foreground, background\}$ is a class label assigned to superpixel x_i. The so-called *unary* term ψ assigns to each superpixel its potential to be foreground or background based on a probability $P(c_i|\mathbf{f}(x_i))$ computed from the output of an SVM

$$\psi(c_i|x_i) = \frac{1}{1 + P(c_i|\mathbf{f}(x_i))}. \tag{2}$$

The *pairwise* term ϕ assigns to each pair of superpixels a potential to have similar or differing labels (indicating boundaries), based on a second SVM output

$$\phi(c_i, c_j|x_i, x_j) = \begin{cases} \frac{1}{1+P(c_i,c_j|\mathbf{f}(x_i),\mathbf{f}(x_j))} & \text{if } c_i \neq c_j, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The weight w in Eq. 1 controls the relative importance of the two terms. Our segmentation is achieved by minimizing Eq. 1 using a mincut-maxflow algorithm.

2.3 Superpixel-Based Shape and Local Features

The SVMs in Eqs. 2 and 3 predict which superpixels contain mitochondria and which neighboring superpixels contain a mitochondrial boundary. As discussed in Section 1, shape, texture, and boundary cues are all essential to this process. Features $\mathbf{f}(x_i)$ extracted from the image at superpixel x_i combine these essential cues

$$\mathbf{f}(x_i) = [\mathbf{f}^{\text{Ray}}(x_i)^\top,\ \mathbf{f}^{\text{Rot}}(x_i)^\top,\ \mathbf{f}^{\text{Hist}}(x_i)^\top]^\top , \tag{4}$$

where $\mathbf{f}^{\text{Ray}}$ represents Ray descriptors that capture object shape, $\mathbf{f}^{\text{Rot}}$ are rotational features describing texture and boundaries [9], and $\mathbf{f}^{\text{Hist}}$ are histograms describing the local intensity. These features, shown in Fig. 2, are detailed below.

Ray Descriptors describe the shape of local objects for each point in the image in a way that standard shape modeling techniques can not. Typically, other methods represent object shape using contour templates [1] or fragment codebooks [2]. While these

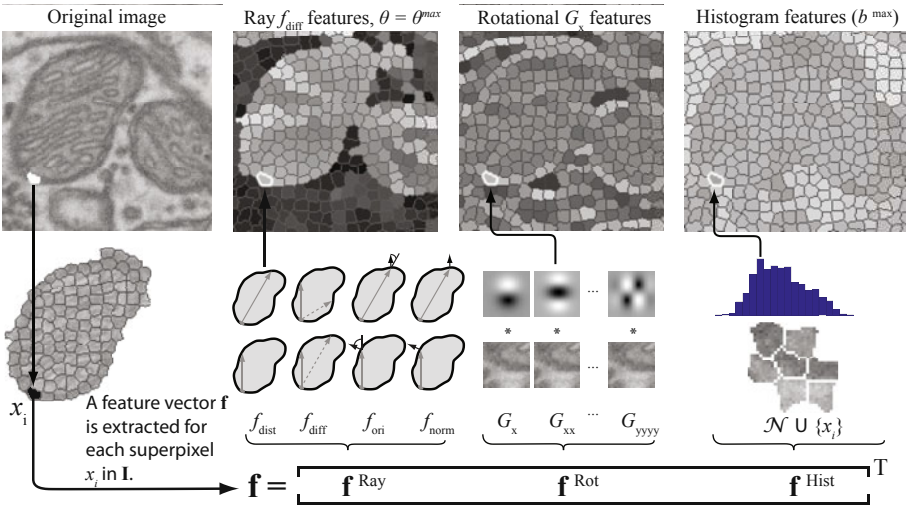

Fig. 2. For each superpixel, the SVM classifiers in Eqs. 2 and 3 predict the presence of mitochondria based on a feature vector $\mathbf{f}$ we extract. $\mathbf{f}$ captures shape cues with a Ray descriptor $\mathbf{f}^{\text{Ray}}$, texture and boundary cues with rotational features $\mathbf{f}^{\text{Rot}}$, and intensity cues in $\mathbf{f}^{\text{Hist}}$.

approaches can successfully segment a single object with known shape, they tend to fail when the shape is highly variable or when many objects appear in the image.

For a given point x_i in the image, four types of Ray features are extracted by projecting rays from x_i at regular angles $\mathbf{\Theta} = \{\theta_1, \ldots, \theta_N\}$ and stopping when they intersect a detected edge (r) [10]. The distance from x_i to r form the first type of feature f_{dist}. The other three types of features compare the relative distance from x_i to r for rays in two different directions (f_{diff}), measure the gradient strength at r (f_{norm}), and measure the gradient orientation at r relative to the ray (f_{ori}). While [10] uses individual Ray features as AdaBoost learners, we aggregate all features extracted for a single point into a *Ray descriptor* $\mathbf{f}^{\text{Ray}} = [f_{\text{dist}}\ f_{\text{diff}}\ f_{\text{norm}}\ f_{\text{ori}}]^\top$. We make it rotation invariant by shifting the descriptor elements so that the first element corresponds to the longest ray. Fig. 3 demonstrates the Ray descriptor's ability to compactly represent object shape.

Rotational Features capture texture and image cues indicating boundaries such as edges, ridges, crossings and junctions [9]. They are projections of image patches around a superpixel center x_i into the space of Gaussian derivatives at various scales, rotated to a local orientation estimation for rotational invariance.

Histograms complement $\mathbf{f}^{\text{Ray}}$ and $\mathbf{f}^{\text{Rot}}$ with simple intensity cues from superpixel x_i's neighborhood $\mathcal{N}$. $\mathbf{f}^{\text{Hist}}$ is written $\mathbf{f}^{\text{Hist}}(\mathbf{I}, x_i) = \sum_{j \in \mathcal{N} \cup \{i\}} h(\mathbf{I}, x_j, b)$ where $h(\mathbf{I}, x_j, b)$ is a b-bin histogram extracted from $\mathbf{I}$ over the pixels contained in superpixel x_j.

2.4 Learning Object Boundaries

Most graph-cut approaches model object boundaries using a simple pairwise term

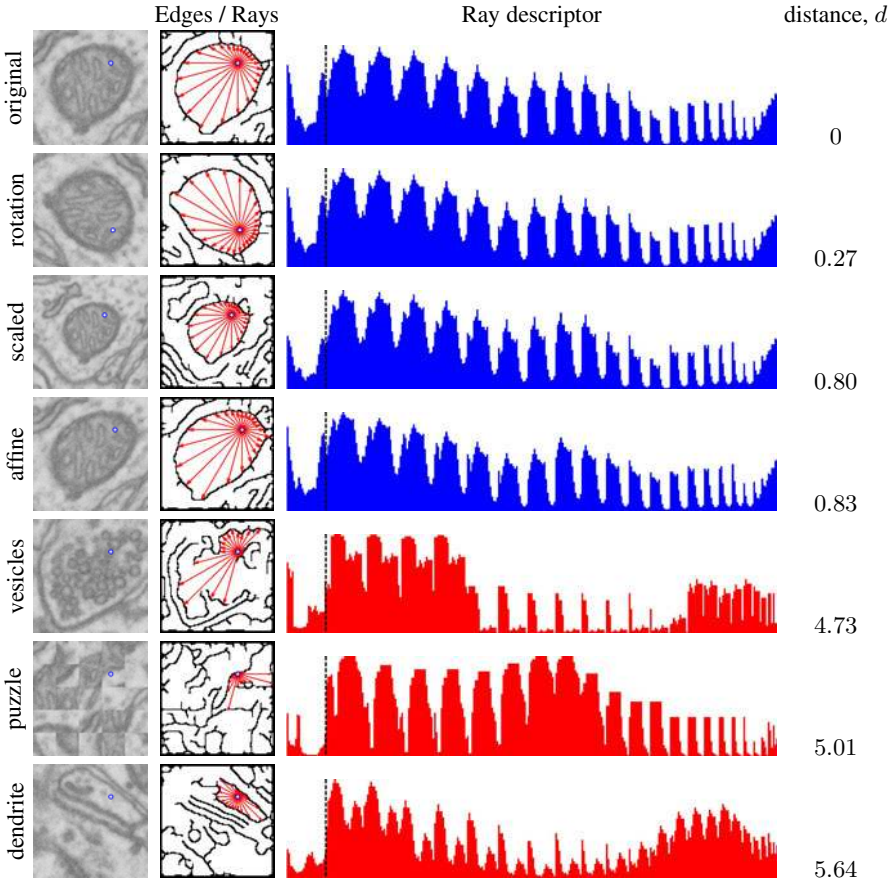

Fig. 3. Ray descriptors built from features in [10] provide a compact representation of local shape for each point in an image. The descriptors are stable when subjected to rotation, scale, and affine transformations, but change dramatically for other shapes including vesicles, dendrites, and randomly rearranged tiles from the original image (puzzle). d is the Euclidean distance between the descriptor extracted from the original image and descriptors extracted from other images.

$$\phi(c_i, c_j | x_i, x_j) = \begin{cases} \exp\left(-\frac{||I(x_i)-I(x_j)||^2}{2\sigma^2}\right) & \text{, if } c_i \neq c_j \\ 0 & \text{, otherwise,} \end{cases} \tag{5}$$

which favors cuts at locations where color or intensity changes abruptly, as in [16]. While similar expressions based on Laplacian zero-crossings and gradient orientations exist [16], very few works go beyond this standard definition. As illustrated in Fig. 4 *(left)*, this approach results in a poor prediction of where mitochondrial boundaries actually occur, as strong gradients from other membranes cause confusion. By learning what image characteristics indicate a true object boundary, we can improve the segmentation [11]. We train an SVM using features extracted from pairs of

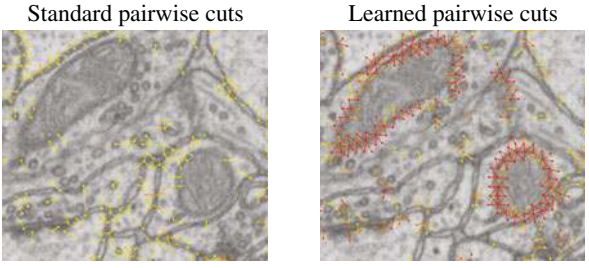

Fig. 4. *(left)* Boundaries predicted by a standard pairwise term (Eq. 5) correspond to strong gradients, but not necessarily to mitochondrial boundaries. *(right)* A learned pairwise term (Eq. 3) using more sophisticated cues $[\mathbf{f}_i^\top,\ \mathbf{f}_j^\top]^\top$ results in better boundary predictions. Red lines indicate strong probable boundaries, yellow lines indicate weaker boundaries.

superpixels containing true object boundaries, indicated by white graph edges in Fig. 1. The pairwise feature vector $\mathbf{f}_{i,j}$ is a concatenation of $\mathbf{f}_i$ and $\mathbf{f}_j$ extracted from each superpixel $\mathbf{f}_{i,j} = [\mathbf{f}_i^\top,\ \mathbf{f}_j^\top]^\top$, providing rich image cues for the SVM to consider.

3 Results

We tested our approach on a data set consisting of 23 annotated high resolution EM images. Each image is 2048×1536 pixels, and the entire data set contains 1023 total mitochondria. We used $k = 5$ k-fold cross validation for training and testing. Our evaluation compares segmentation results for the following methods:

TextonBoost A boosted texton-based segmentation algorithm [14],
Fulkerson09 A superpixel-based algorithm using SIFT features [5],
Standard-f* Our algorithm trained with the *standard* pairwise term of Eq. 5 and histogram and rotational features $[\mathbf{f}^{\text{Hist}\top}\mathbf{f}^{\text{Rot}\top}]^\top$,
Standard-f Our algorithm trained with the *standard* pairwise term of Eq. 5 and feature vector $\mathbf{f}$ incorporating shape and texture cues given in Eq. 4,
Learned-f *Our complete algorithm* trained with the *learned* pairwise term of Eq. 3 and feature vector $\mathbf{f}$ incorporating shape and texture cues given in Eq. 4.

Parameter settings for [14] used 50 textons and 2000 rounds of boosting. For [5], Quickshift superpixels were used, and SIFT descriptors were extracted over 9 scales at a fixed orientation and quantized into 50 clusters. For our approach, we used superpixels containing approximately 100 pixels, extracted Rays at 30° angles, computed rotational

Table 1. Segmentation Results

	TextonBoost [14]	Fulkerson09 [5]	Standard-$\mathbf{f}^*$	Standard-$\mathbf{f}$	Learned-$\mathbf{f}$
Accuracy	95%	96%	94%	96%	**98%**
VOC score [7]	61%	69%	60%	68%	**82%**

features using first to fifth Gaussian derivatives with $\sigma = \{3, 6, 9, 12\}$, and built histograms with $b = 20$ bins. A post-processing step depicted in Fig. 1(f) was used to smooth the results produced by all the algorithms.

Discussion. Table 1 summarizes results for the entire data set. Our approach achieved a pixel-wise accuracy of 98%. By the same metric, TextonBoost and Fulkerson09 also performed well, but visually the results are inferior, as seen in Fig. 5. This is because mitochondria account for very few pixels in the image. The VOC score $= \frac{TP}{TP+FP+FN}$, introduced in [7][1], is designed to be more informative in such cases, and reflects the superior quality of our segmentations. Because it is pixel-based and lacks shape cues, TextonBoost poorly estimates mitochondrial membranes. The use of superpixels in Fulkerson09 seems to improve results slightly over [14], but the lack of shape cues or learned boundaries still degrades its performance. Comparing the Standard-$\mathbf{f}^*$ and Standard-$\mathbf{f}$ variations of our approach, we see that adding shape cues boosts performance, and learning boundaries in the pairwise term leads to a further increase in Learned-$\mathbf{f}$.

4 Conclusion

We proposed a fully automated approach to segment irregularly shaped cellular structures that outperforms state-of-the-art algorithms on EM imagery. We also demonstrated that Ray descriptors increase performance by capturing shape cues without having to define an explicit model. Finally, we showed that a learning approach to the pairwise term of the energy function further helps find true object boundaries.

Acknowledgements. We wish to thank Graham Knott and Marco Cantoni for providing us with high-resolution imagery and invaluable advice. We also thank German Gonzalez for providing code for Rotational Features.

References

1. Ali, A., Farag, A., El-Baz, A.: Graph Cuts Framework for Kidney Segmentation With Prior Shape Constraints. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 384–392. Springer, Heidelberg (2007)
2. Levin, A., Weiss, Y.: Learning to Combine Bottom-Up and Top-Down Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 581–594. Springer, Heidelberg (2006)
3. Vazquez-Reina, A., Miller, E., Pfister, H.: Multiphase Geometric Couplings for the Segmentation of Neural Processes. In: CVPR (2009)
4. Andres, B., Koethe, U., Helmstaedter, M., Denk, W., Hamprecht, F.: Segmentation of Sbfsem Volume Data of Neural Tissue by Hierarchical Classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
5. Fulkerson, B., Vedaldi, A., Soatto, S.: Class Segmentation and Object Localization With Superpixel Neighborhoods. In: ICCV (2009)
6. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" - Interactive Foreground Extraction Using Iterated Graph Cuts. In: SIGGRAPH (2004)

[1] TP=true positives, FP=false positives, FN=false negatives.

7. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC 2010) Results (2010)
8. Frangakis, A., Hegerl, R.: Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis. Journal of Structural Biology (2002)
9. Gonzalez, G., Fleuret, F., Fua, P.: Learning Rotational Features for Filament Detection. In: Conference on Computer Vision and Pattern Recognition (June 2009)
10. Smith, K., Carleton, A., Lepetit, V.: Fast Ray Features for Learning Irregular Shapes. In: ICCV (2009)
11. Prosad, M., Zisserman, A., Fitzgibbon, A., Kumar, M., Torr, P.: Learning Class-Specific Edges for Object Detection and Segmentation. In: Kalra, P.K., Peleg, S. (eds.) ICVGIP 2006. LNCS, vol. 4338, pp. 94–105. Springer, Heidelberg (2006)
12. Narashimha, R., Ouyang, H., Gray, A., McLaughlin, S., Subramaniam, S.: Automatic Joint Classification and Segmentation of Whole Cell 3D Images. Pattern Recognition 42(2007), 1067–1079 (2009)
13. Radhakrishna, A., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels. Technical Report 149300, EPFL (June 2010)
14. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954. Springer, Heidelberg (2006)
15. Thévenaz, P., Delgado-Gonzalo, R., Unser, M.: The Ovuscule. PAMI (to appear, 2010)
16. Boykov, Y., Jolly, M.: Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. In: ICCV (2001)

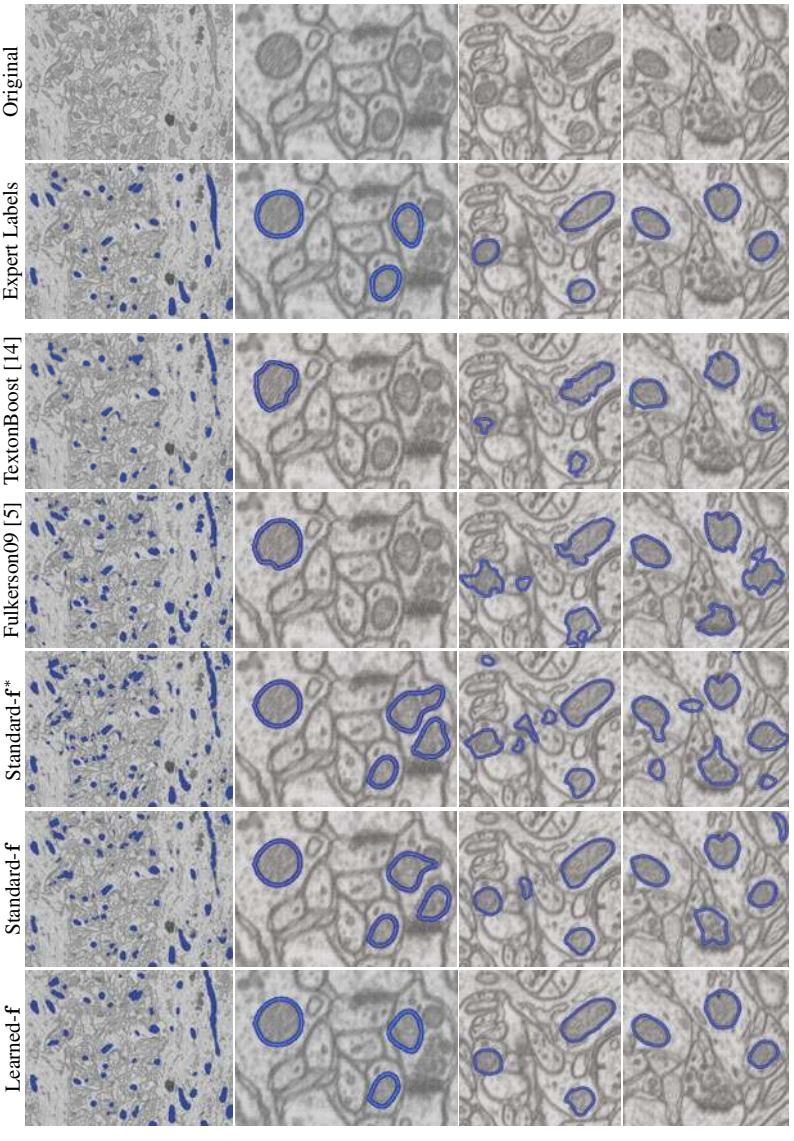

Fig. 5. Segmentation results on EM images. Column 1 contains the 2048×1536 micrograph at reduced resolution. Columns 2-4 contain details from column 1. Row 1 contains the original EM image. Row2 contains the expert annotations. Further rows contain results of the various methods. The lack of shape cues and learned boundaries result in inaccurate segmentations for TextonBoost and Fulkerson09, especially near distracting textures and membranes. Our method without shape or learned boundaries, Standard-$\mathbf{f}^*$, performs similarly. By injecting shape cues in Standard-$\mathbf{f}$, we see a significant improvement as more mitochondria-like shapes appear in the segmentation. However, some mistakes in the boundary persist. In Learned-$\mathbf{f}$ we add the learned pairwise term, eliminating the remaining errors and producing a segmentation that very closely resembles the human annotation.