Automatic Segmentation of Mitochondria from EM Images via Hierarchical Context Forest

Jiajin Yi, Zhimin Yuan, and Jialin Peng[✉]

College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China
2004pjl@163.com

Abstract. To solve the problem of automatic mitochondria segmentation from electron microscope (EM) images, a hierarchical context forest (HCF) model using multi-level context features was proposed. Exploring effective contextual information is crucial to address the challenges caused by the varied appearances and shapes of mitochondria, and the complicated image content. To this end, a novel class of features named local patch pattern (LPP) are designed to characterize local contextual information, which are used to resolve the ambiguity caused by similar appearances and intensities of different organelles. Furthermore, to capture long-range contextual information, we also extract LPP features on intermediate probability predictions of the HCF model. Moreover, a multiscale strategy is used to capture different sizes of mitochondria. Solid validations of our method conducted on public dataset demonstrated the effectiveness of both of the proposed LPP features and the proposed model. The result of comparison showed that, the proposed method achieved distinct improvement of results in terms of Precision, Recall and F1-value score.

Keywords: Electron microscopy image · Mitochondria segmentation · Random forest · Contextual information · LPP feature

1 Introduction

An automatic segmentation of mitochondria from electron microscope (EM) images can greatly facilitate the analysis of mitochondria by precisely quantifying their volume, morphology and distribution, which have been directly linked to aging, cancer, and neurodegenerative diseases [1, 2]. High resolution EM is one of the state-of-the-art imaging devices for investigating the ultrastructure of cell, where mitochondria are subcellular organelle. However, with images in the scale of nm^3 and so in huge size, manual segmentation, even with assisted tools [3], is labor-intensive, time-consuming and also subjective. In fact, the very precise annotation could be a challenging task, even for experts, due to the intrinsic ambiguity in the image delineation process (see Fig. 1).

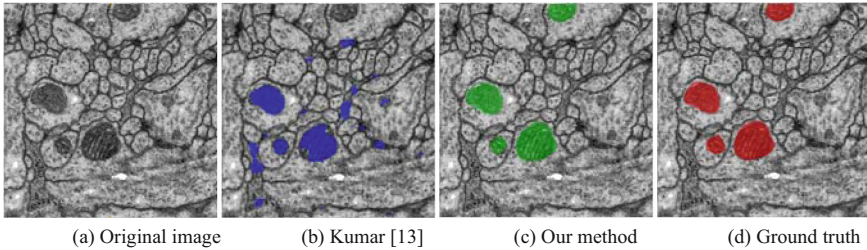|(a) Original image|(b) Kumar [13]|(c) Our method|(d) Ground truth|

Fig. 1. Segmentation of mitochondria from EM images, which is a challenge task.

Therefore, automatic methods are highly desirable to process the large volumes of EM images and provide more reproducible results [4]. However, the large variability of mitochondria in density, location, size and shape makes it a challenging task. Moreover, the appearance and content of EM image are rather complex as shown in Fig. 1. For example, image intensities representing mitochondria in EM images show large overlap with intensities of other structures; strong gradients do not necessarily correspond to the semantic boundaries of the target mitochondria. Thus, in order to determine the presence of mitochondria at a given position and delineate their boundaries accurately, high-level contextual information has to be explored, which is the focus of this study.

Recently, the detection and segmentation of mitochondria from EM images have attracted a variety of studies [4–9, 12–15]. For example, Macke *et al.* [5] introduced a semi-automated method based on level set for tracing of cell membranes, in which segmentation from adjacent slice was utilized as prior constraint. Although the method is much faster than manual tracing, the performance is limited by its simple intensity based features. For mitochondria segmentation, Narasimha *et al.* [6] explored a filter-bank and textons as texture feature encoders, which were combined with several kinds of popular classifiers. Neila *et al.* [7] extracted local visual features with both 2D and 3D image filters (e.g., first and second order derivatives) in multiple scales to take into account the varied shape of mitochondria and anisotropic image voxels; conditional random field with an anisotropy-aware regularization is then employed for image segmentation. However, the discriminative ability of these general texture features is still limited.

Rather than texture features, Smith *et al.* [8] devised a new class of shape features, that considered image characteristics at distant contour points to capture the irregular shape information of mitochondria. In [9], Lucchi *et al.* combined Ray features with intensity histogram features on supervoxels, and the segmentation was obtained using graph cut with learned potentials. However, this graph learning and partitioning method is costly in memory, which was improved in [10] with max-margin learning and a subgradient descent using working sets. In [11], elliptical descriptor at different scales combined with local Gaussian filters as [7] were investigated. To identify precise boundaries of mitochondria, Giuly *et al.* [12] performed random-forest-based classification on candidate boundary contours using geometric and intensity based features, while candidate contours were obtained by thresholding prediction results of another random forest classifier; the final fine segmentation is obtained with a geodesic active surface method. To alleviate the negative impact of the complex image content of EM image, Kumar *et al.* [13] proposed a class of more powerful features, Radon-like features, that allowed

for aggregation of spatially distributed image statistics into compact feature descriptors in the EM image. Final segmentation was extracted with thresholdings. Seyedhosseini *et al.* [14] extracted shape and textural features of mitochondria by algebraic curve, using the random forest as the classifier. The mitochondrial boundary curves can be obtained effectively by the method, however, it is easily attracted by the ridge structure of the mitochondria.

In this work, we propose a fully automatic method for mitochondria segmentation by integrating multilevel contextual and appearance features. More specifically, with a class of newly designed features, we propose to extract multilevel context and appearance features with a classifier in cascaded architecture that is similar to the auto-context model in [16, 25]. One contribution of this study is that, we design a novel class of features, named local patch pattern (LPP) to encode local context and appearance from raw images and middle-level context from intermediate predictions. To be specific, under the framework of hierarchical classifier, we extract middle-level features using LPP on prediction results from classifiers in lower layer. In this way, the features extracted include information from a much larger receptive field, the size of which plays a crucial role for the segmentation performance [17]. In this way, the amount of surrounding contextual information is enhanced. To further capture mitochondria of different sizes, we utilize a multi-scale strategy, which improves the efficiency and robustness of the model.

2   Method

In this section, we provide a detailed description of our method. Taking into account anisotropic voxels of our EM images, we perform segmentation in a slice-by-slice strategy using image patches. Specifically, we split each slice of the image into overlapped patches of size $l \times l$, and assign the most probable label to each pixel x of a target image according to the information from the patch I_x, centered on the pixel x. To achieve fine segmentation, we train a hierarchical set of classifier, in which random forest model is employed as the basic classifier due to its efficiency and scalability. The workflow of the proposed framework is shown in Fig. 2.

2.1   Feature Extraction

Local Appearance and Texture Fatures. Gray-scale intensity, gradient magnitude, and local binary pattern (LBP) are employed to capture appearance, edge and texture information. In each patch, we compute 6 statistics (including mean, variance, median, entropy, kurtosis and skewness) of intensities, gradients and local binary pattern (LBP) [18] features in local image patches as local appearance and texture descriptors. Gray-scale intensities and its statistics are the most prominent features of EM images; gradients are the basic features indicating edges in images. LBP is one of the most widely used texture features; it is computed by comparing each pixel in the image with each of its P (such as 8) neighbors, giving an P-digit binary number, which is usually converted to decimal (see Fig. 3 (a)). Specifically, LBP operator is defined as:
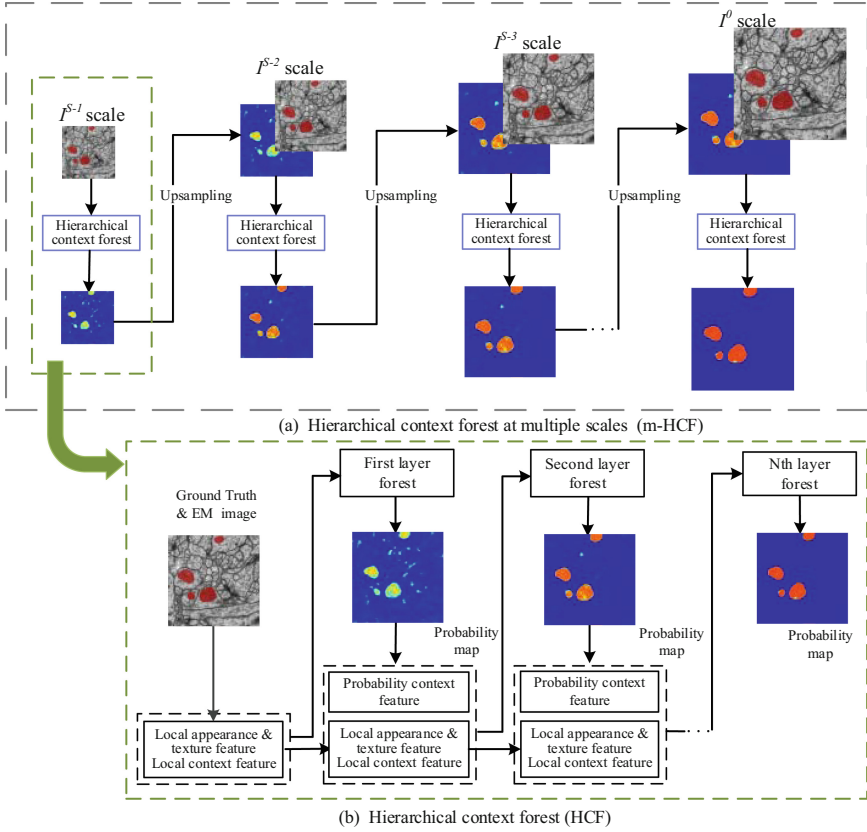
(a) Hierarchical context forest at multiple scales (m-HCF)

(b) Hierarchical context forest (HCF)

Fig. 2. Workflow of the multiscale hierarchical context forest (m-HCF) we proposed. The heat map in the figure is a graphical representation of probability map, in which the colder the color is, the smaller the magnitude of estimated probability is. (Color figure online)
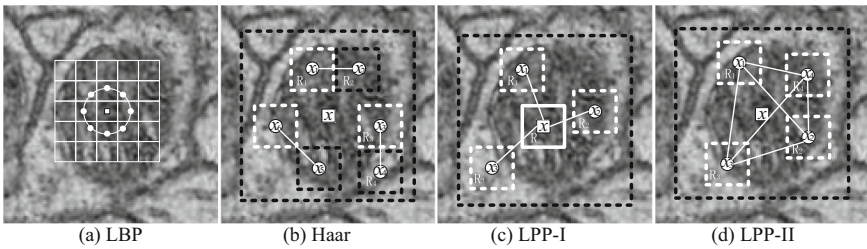

(a) LBP   (b) Haar   (c) LPP-I   (d) LPP-II

Fig. 3. Illustration of LBP feature, Haar feature and the proposed LPP features.

$$f_{\text{LBP}}(x, I) = \sum_{p=1}^{P} 2^{p-1} \delta\left(x_p - x_c\right), \qquad (1)$$

where

$$\delta(z) = \begin{cases} 1 & z \geq 0, \\ 0 & z < 0. \end{cases}$$

LBP is widely used due to its powerful ability for texture discrimination and simple calculation. However, it only captures the information in a smaller neighborhood, and unable to encode the more discriminative information of context in longer range. Since there are a large number of cell tissues with similar appearances and textures in the EM image, the local features mentioned above is still limited in discrimination.

Local Patch Pattern (LPP) Features. The association between each pixel of the image and its extended neighborhoods, called context information, that provides a way of eliminating the ambiguity of the local appearance and texture features [19, 20]. Inspired by the classical LBP feature and the Haar features [21], we propose a novel class of features (LPP) based on the local patch, which are used to capture local context information effectively. The classical Haar features consider sub-regions of adjacent positions and then calculate the difference of the sum of pixel intensities in these sub-regions, as shown in Fig. 3 (b), it can quickly generate a rich set of features that are robust against the noise. In this paper, we extend the LBP and Haar features into LPP features that capture the context information effectively. The LPP features contain two subtypes, i.e. LPP-I and LPP-II, which are calculated by extracting sparsely distributed sub-regions $\{R_i\}_{i=1}^{L}$ on the image patch I_x.

LPP-I Feature. This feature is a kind of LBP-like Haar feature. For the LPP-I feature extraction, we denote the central sub-region of the patch I_x as R_0, then record the differences between R_0 and other sub-regions R_i, and also the binary code formed by all positive, negative differences. The former is a real type LPP feature (LPP-Ir), the latter is a binary type LPP-Ib feature (actually convert the binary code to decimal).

$$f_{\text{LPP-Ir}}(I_x, \ R_i) = \frac{1}{|R_i|} \sum_{u \in R_i} I_x(u) - \frac{1}{|R_0|} \sum_{u \in R_0} I_x(v), \tag{2}$$

$$f_{\text{LPP-Ib}}(I_x) = \sum_{i=1}^{L} 2^{i-1} \delta(\sum_{u \in R_i} I_x(u) - \sum_{v \in R_0} I_x(v)). \tag{3}$$

Specifically, the real valued LPP features $f_{\text{LPP-Ir}}$ for R_i $(i = 1, \cdots, L)$ is obtained by calculating the difference of the average pixel intensities of sub-regions $\{R_i\}_{i=1}^{L}$ (as shown in Fig. 3 (c)) and the central one R_0 in I_x. We also record L binarized summaries of $f_{\text{LPP-Ir}}$ features through using different permutations of the $f_{\text{LPP-Ir}}$ values. For each permutation, a $f_{\text{LPP-Ib}}$ feature is calculated by recording the sign of the $f_{\text{LPP-Ir}}$ values, which is converted to decimal for convenience. From these steps, we have achieved the extension of classical LBP feature to the context feature LPP-I.

LPP-II Features. We extend the calculation of the traditional Haar feature to randomly select multiple pairs of sub-region of the same size R_i and R_j from a local image patch (such as R_1 and R_2 that are shown in Fig. 3 (d)). For each pair of sub-regions R_i and

R_j, $f_{\text{LPP-II}}$ is the difference between their average pixel intensities, that is defined as follows,

$$f_{\text{LPP-II}}\big(I_x, R_i, R_j\big) = \frac{1}{|R_i|} \sum_{u \in R_i} I_x(u) - \frac{1}{|R_j|} \sum_{v \in R_j} I_x(v). \tag{4}$$

LPP-II features encode the differences between different sub-regions in each local image patch. It is a variant of Haar feature with randomly-extracted non-adjacent sub-regions.

Compared with LBP and Haar features, the LPP-I and LPP-II features reflect contextual information in relatively long-range that the category of a pixel depends on. The contextual information in longer range is extracted by computing LPP features on probabilistic predictions (see Sect. 2.2), which will be used as additional information to refine the segmentation.

2.2 Hierarchical Context Forest (HCF)

We use the classical random forest [22] as the classifier, as it can effectively address the relative large-scale and high-dimensional data. Moreover, it is easily parallelized. Random forest obtains the final classification result by integrating the classification results of an ensemble of separately trained binary decision trees; each tree in the forest is trained only on the subset of the data and feature. The introduction of randomness makes the random forest model relatively robust to noise and mitigate the over-fitting problem. Each random forest can be described by the number of the trees used, the maximum depth and the weak learner used.

Due to the complexity of the EM image content, the context features extracted from raw image may be limited in discrimination. Besides, segmentation by independently inferring the label of each pixel is sub-optimal, as images with semantic content show obvious label dependencies between adjacent pixels. Therefore, it is preferred that the ultimate prediction can be influenced by the model's beliefs about nearby positions.

To solve these limitations, 1) *we use a hierarchical classification model, which cascades multiple random forests to improve the segmentation results iteratively*; 2) *we extract LPP features on the output probabilities of random forest in lower layers as additional features to the random forest in the current layer*. Importantly, the LPP features extracted from the intermediate segmentation probability map of the hierarchical model have a much larger receptive field, since the probability map of each position is influenced by information in the $l \times l$ local patch. Different from the original image, features on the probability map that is produced by the previous classifier can reflect the defect of the previous classifier, which help the latter to pay more attention to the poorly segmentation region in the previous probability map. Therefore, the LPP features on probability maps are expected to be useful to enhance the original context information. The above method follows the philosophy of the auto-context framework proposed in [16]. However, our model differs in that we use the LPP features proposed above to characterize the contextual information in long range.

Specifically, in the training step, a sequence of classifiers $F = \{F_t, t = 1, \cdots, T\}$ are iteratively trained, which with a total of T random forests, as shown in Fig. 2 (b). The

first forest is trained by extracting the local image features described in Sect. 2.1 from the training patches; subsequent random forests $F_t(t > 1)$ not only use the original image as training data, but also use the previously-estimated probability map as augmented training data. Importantly, we also extract the LPP features from the probability map as the probability context feature, which used to enhance the discrimination of features extracted on the grayscale images.

In the testing step, each trained classifier F_t is used sequentially to predict the test EM images I_t. In the first iteration, we extract the local image features on I_t described in Sect. 2.1, then predict it with F_1 to obtain initial mitochondrial probability map. In the second iteration, we not only use the features extracted in I_t, but also input the probability context features extracted from the previously-estimated mitochondria probability map as augmented features into F_2 for prediction. Repeat the above steps until all trained classifiers are applied. The output of the final cascaded forest is the segmentation result.

2.3   Hierarchical Context Forest in Multi-scale (M-HCF)

To accurately capture mitochondria of different sizes and improve the validity, robustness of the model, we use the multi-scale strategy to further enhance the performance of the model.

Specifically, at the first, we obtain S images of different resolutions $\{I^0, I^1, \ldots, I^{S-1}\}$ by sequentially downsampling each training image I, where I^0 is the original scale image I and I^j is obtained by downsampling I^{j-1} with 1/2 times. Then, we apply the HCF model on multiscale images (m-HCF), see Fig. 2 (a), in which the input of the model at high resolution is the corresponding scale image and the classification prediction probability map at low resolution. The introduction of multi-scale strategy makes the model utilize context information of different spatial extents. Actually, the multi-scale method has been proved to be effective on both natural image [20] and medical image segmentation [23, 26].

In the model prediction step, the S images of different resolutions are obtained by downsampling the test image I_t at first, and then input the image patches extracted at different resolutions into the m-HCF model, which is already obtained by training. Finally, we can get the result of segmentation from the m-HCF model.

3   Experiments

3.1   Experimental Setting

We evaluated the performance of our method on the Drosophila first instar larva ventral nerve cord (VNC) dataset [3, 24]. DVNC dataset contains 30 Drosophila abdominal nerve images of size 512×512 with a resolution of $5 \times 5 \times 40$ nm^3/voxel, which was acquired using continuous slice transmission electron microscopy (ssTEM). The first 15 images were selected for training and the rest were used for testing. Also we evaluated the robustness of the proposed method with 5 different splits of the dataset.

To evaluate the performance of segmentation, we selected Precision, Recall, F1-value as the measurements, which are defined as follows,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{6}$$

$$F1\text{ - value } = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{7}$$

where TP are the true positives, FP are the false positives, and FN are the false negatives. F1-value is a metric that combines precision and recall, equivalent to the Dice similarity coefficient, and it is one of the most commonly used comprehensive metric for evaluating segmentation results. The larger the scores of the three metrics are, the better the segmentation result is.

The model parameters were set as follows: for the multi-scale hierarchical context forest (m-HCF), the scale number S was set as $S = 2$, which included 1/4 original image size (I^1) and the original image size (I^0); for the hierarchical context forest (HCF), we used 2 and 5 layers of random forests at the two scales, respectively; the patch size on the size of 1/4 original image was set as 11×11, and on the original image size, set as 19×19. The discussion about patch size at I^0 scale in Sect. 3.3. For parameters in the random forest, the number of decision trees was 15, the maximum depth of each tree was 50, and the minimum number of samples at the leaf node in each tree was 10. In the experiment, we found that increasing the number of the trees, the depth of the trees and the number of features selected by the non-leaf nodes, can improve the performance of the model. However, this will result in much longer training time for the model and also much larger memory requirement. The parameter setting in our experiments is a balance of efficiency and accuracy.

3.2 Results and Comparisons

We evaluated the segmentation performance of our method and compared it with recent mitochondria segmentation methods on DVNC, which are summarized in Table 1. We compared our method with 1) the method of Kumar *et al.* [13], which devised a class of powerful feature named Radon-like features for mitochondria segmentation; 2) the novel pipeline of Giuly *et al.* [12], which combined patch classification, contour pair classification, and automatically seeded level sets; 3) the method of Seyedhosseini *et al.* [14], which originally extracted the shape and textural features of mitochondria by algebraic curve, using the random forest as the classifier. We also tested the random forest with the proposed features and the proposed HCF method using different scales of images.

As shown in Table 1, our HCF method using 2 scales of images, with a balance of computational efficiency and accuracy, achieved an average F1-value of 80.3%, 84.9% Precision, and 78.5% Recall, indicating that it outperforms other methods in terms of all metrics. In fact, the scores of the HCF model with only 1 scale ($S = 1$) is also higher than most of other methods. Furthermore, using the proposed features, the simple random forest method already achieved promising results, which show the discriminative ability of the proposed features.

Table 1. Performance comparison of different methods with multiple metrics on DVNC.

Methods	F1-value (%)	Precision (%)	Recall (%)
Kumar [13]	56.6	59.3	54.2
Random forest (with proposed features)	62.6	71.4	55.7
Giuly [12]	60.4	64.2	57.0
Seyedhosseini [14]	72.9	78.5	68.0
Proposed method ($S = 1$)	77.8	72.0	**87.5**
Proposed method ($S = 2$)	**80.3**	**84.9**	78.5

3.3 Model Analysis

The Effectiveness of the Proposed LPP Features (see Fig. 4). In order to verify the validity of the LPP features designed in this paper, we tested the multiscale HCF method using different features and different feature combinations. To be specific, we firstly set the baseline features (B in Fig. 4) as the 6 statistics (i.e., mean, variance, median entropy, kurtosis, skewness) and central point pixel intensity from the image patch in the grayscale image, gradient image, and probability map. Then, the baseline features, Radon-like feature [13], LBP feature [18], Haar-like feature, the proposed feature LPP including LPP-I and LPP-II and their combinations are compared.
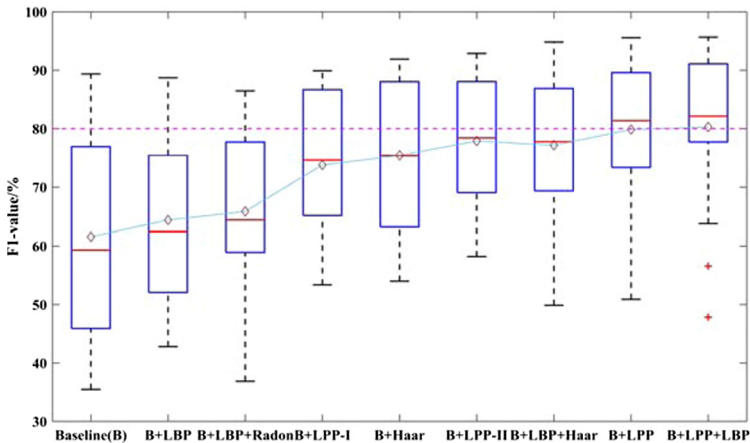
Fig. 4. Comparison of segmentation results with different choices of features.

The results of comparison F1-value metric are demonstrated with a box-plot shown in Fig. 4. It can be observed that LPP features are superior to the classical LBP, Haar features and also the Radon features. The F1-value using baseline features is 61.5%; compared with the LBP feature, the LPP-I feature (i.e., LBP-like feature) was better, and the F1-value was improved from 64.4% to 73.8%; in comparison of LPP-II features

with the Haar features, and the Haar **+** LBP features, we can see that the F1-value increased from 75.5%, 77.2% to 77.9%, so the LPP-II features are not only better than the Haar features, but also better than the combination of Haar and LBP features. In addition, the full LPP features are significantly better than the combination of LBP and Haar features, and the F1-value was increased from 77.2% to 79.8%.

These results proved the advantages of the LPP features designed in this paper. To further increase the robustness of our model, we combined the LPP features with the LBP features in the final model, and the F1-value of the final segmentation was 80.3%. Three reasons may explain the improvements using LPP features: 1) the LPP features contain sufficient image contextual information to resolve the ambiguity of local appearance and texture features; 2) the sub-patch based LPP features are more robust to noise; 3) a large number of features provides guarantee for tree depth, which improves the performance of the model.

The Influence of Patch Size (see Table 2). To verify the influence of different patch sizes at I^0 scale, we compared the segmentation results using different patch sizes at I^0 scale. As shown in Table 1, our model is relatively robust to different patch sizes. Moreover, our model performed best when the patch size was set as 19×19.

Table 2. Segmentation results of different patch sizes at I^0 scale.

Metric	Patch size (pixel)							
	13×13	15×15	17×17	19×19	21×21	23×23	25×25	27×27
F1-value (%)	77.4	79.5	79.8	**80.3**	80.2	79.6	79.5	78.6

The Influence of Scale Parameter (see Fig. 5). To verify the validity of multi-scale strategy, we compare the segmentation results of single-scale ($S = 1$) and multi-scale ($S = 2$). Figure 5 shows the intermediate results in the model iteration process, and the comparative results of performance are shown in Table 1. Comparing the probability maps of the second column of Fig. 5, it can be seen that the left lower non-mitochondrial region detected error at I^0 scale can be easily detected correctly at I^1 scale, and the probability map at the I^1 scale is much more "cleaner" than the probability map at the I^0 scale, which is more conducive to the subsequent classifier to refine the mitochondrial classification probability map.

The Bias of Training Data Selection (see Fig. 6). To verify the robustness of our model, we employed the 5 different splits of the dataset to train model. For each split of the data, we randomly selected 15 images from the dataset as the training set, and the rest as the testing set. Figure 6 shows the results of segmentation accuracy with the 5 different splits. The relatively low standard deviation of the performance show that our method is less affected by the bias of training data selection, and the average value of the results is close to our segmentation result shown in Table 1.
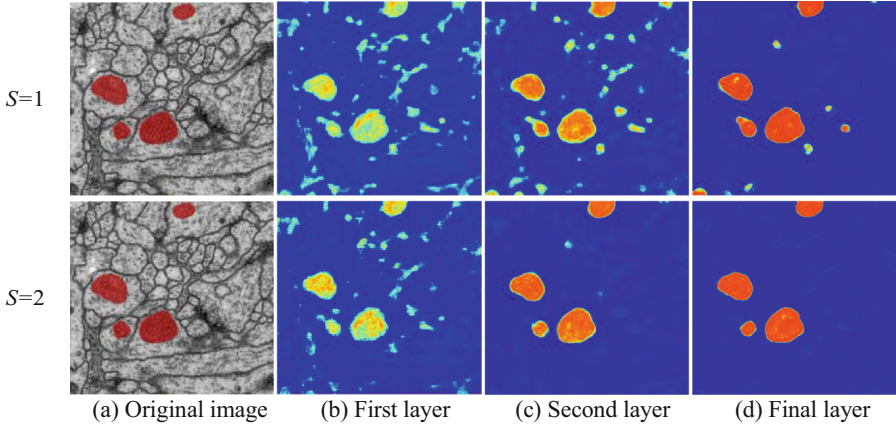
| | (a) Original image | (b) First layer | (c) Second layer | (d) Final layer |

Fig. 5. Probability maps at different layers of each scale. (a) Original image & ground truth; (b) probability map of first layer; (c) probability map of second layer; (d) probability map of final layer. The first row (b)–(d) at I^0 scale; the second row (b) at I^1 scale, (c)–(d) at I^0 scale.
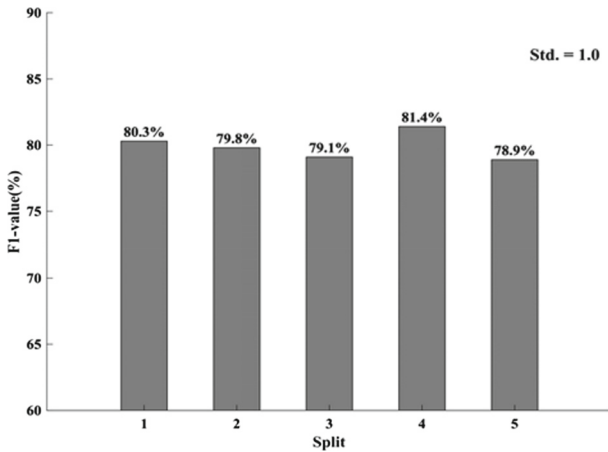

Fig. 6. Segmentation results of our method with 5 different splits of dataset for training and testing. The standard deviation of the 5 segmentation results is 1.0.

4 Conclusion

This paper presented a multiscale HCF model for the automatic segmentation of mitochondria from EM images. To address the challenges of complex image contents, we proposed a novel set of discriminative contextual features, named LPP, to encode multi-level contextual information. Specifically, under the framework of iterative refinement, we extracted features both on raw images and intermediate predictions, which can effectively alleviate the negative impact of complex content of EM images on the mitochondrial segmentation task. Our approach is generic and can be used for similar segmentation tasks. Validation and comparisons on the challenging DNVC dataset indicated that the

proposed method is robust and effective for the automatic segmentation of mitochondria form EM images. In the future work, we plan to extend our work to 3D analysis with exploiting the 3D information of the EM image stack. In addition, we intend to apply our model to other objects segmentation task of EM images.

References

1. Campello, S., Scorrano, L.: Mitochondrial shape changes: orchestrating cell pathophysiology. EMBO Rep. **11**(9), 678–684 (2010)
2. Alston, C.L., Rocha, M.C., Lax, N.Z., Turnbull, D.M., Taylor, R.W.: The genetics and pathology of mitochondrial disease. J. Pathol. **241**(2), 236–250 (2017)
3. Cardona, A., et al.: TrakEM2 software for neural circuit reconstruction. PLoS ONE **7**(6), e38011 (2012)
4. Uzunbas, M.G., Chen, C., Metaxas, D.: An efficient conditional random field approach for automatic and interactive neuron segmentation. Med. Image Anal. **27**, 31–44 (2016)
5. Macke, J.H., Maack, N., Gupta, R., Denk, W., Schölkopf, B., Borst, A.: Contour-propagation algorithms for semi-automated reconstruction of neural processes. J. Neurosci. Methods **167**(2), 349–357 (2008)
6. Narasimha, R., Ouyang, H., Gray, A., McLaughlin, S.W., Subramaniam, S.: Automatic joint classification and segmentation of whole cell 3D images. Pattern Recogn. **42**(6), 1067–1079 (2009)
7. Márquez Neila, P., Baumela, L., González-Soriano, J., Rodríguez, J., DeFelipe, J., Merchán-Pérez, Á.: A fast method for the segmentation of synaptic junctions and mitochondria in serial electron microscopic images of the brain. Neuroinformatics **14**(2), 235–250 (2016). https://doi.org/10.1007/s12021-015-9288-z
8. Smith, K., Carleton, A., Lepetit, V.: Fast ray features for learning irregular shapes. In: 12th International Conference on Computer Vision, pp. 397–404. IEEE, Kyoto (2009)
9. Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P.: Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features. IEEE Trans. Med. Imaging **31**(2), 474–486 (2012)
10. Wei, L., et al.: Learning-based deformable registration for infant MRI by integrating random forest with auto-context model. Med. Phys. **44**(12), 6289–6303 (2017)
11. Cetina, K., Buenaposada, J.M., Baumela, L.: Multi-class segmentation of neuronal structures in electron microscopy images. BMC Bioinform. **19**(1), 298 (2018)
12. Giuly, R.J., Martone, M.E., Ellisman, M.H.: Method: automatic segmentation of mitochondria utilizing patch classification, contour pair classification, and automatically seeded level sets. BMC Bioinform. **13**(1), 29 (2012)
13. Kumar, R., Vázquez-Reina, A., Pfister, H.: Radon-like features and their application to connectomics. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 186–193. IEEE, San Francisco (2010)
14. Seyedhosseini, M., Ellisman, M.H., Tasdizen, T.: Segmentation of mitochondria in electron microscopy images using algebraic curves. In: 2013 IEEE 10th International Symposium on Biomedical Imaging, pp. 860–863. IEEE, San Francisco (2013)
15. Li, W., Liu, J., Xiao, C., Deng, H., Xie, Q., Han, H.: A fast forward 3D connection algorithm for mitochondria and synapse segmentations from serial EM images. BioData Mining **11**(1), 24 (2018). https://doi.org/10.1186/s13040-018-0183-7
16. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **32**(10), 1744–1757 (2010)

17. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018)
18. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
19. Zhang, H., et al.: Context encoding for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 7151–7160. IEEE, Salt Lake City (2018)
20. Ding, H., Jiang, X., Shuai, B., Qun Liu, A., Wang, G.: Context contrasted feature and gated multi-scale aggregation for scene segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2393–2402. IEEE, Salt Lake City (2018)
21. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition, p. I. IEEE, Kauai (2001)
22. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
23. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
24. Cardona, A., et al.: An integrated micro- and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol. **8**(10), e1000502 (2010)
25. Gadde, R., Jampani, V., Marlet, R., Gehler, P.V.: Efficient 2D and 3D facade segmentation using auto-context. IEEE Trans. Pattern Anal. Mach. Intell. **40**(5), 1273–1280 (2017)
26. Roth, H.R., et al.: A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 417–425. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_48