Automatic Detection and Segmentation of Mitochondria from SEM Images using Deep Neural Network

Jing Liu[1], Weifu Li[1], Chi Xiao[1], Bei Hong[1], Qiwei Xie[1,2] and Hua Han[1,2,3]

Abstract— **Investigating the link between mitochondrial function and its physical structure is a hot topic in neurobiology research. With the rapid development of Scanning Electron Microscope (SEM), we can look closely into the fine mitochondrial structure with high resolution. Consequently, many meaningful researches have focused on how to detect and segment the mitochondria from EM images. Due to the complex background, hand-crafted features designed by traditional algorithms cannot provide satisfying results. In this paper, we propose an effective deep neural network improved from Mask R-CNN to produce the detection and segmentation results. On this base, we use the morphological processing and mitochondrial context information to rectify the local misleading results. The valuation was performed on two widely used datasets (FIB-SEM and ATUM-SEM), and the results demonstrate that the proposed method has comparable performance than state-of-the-art methods.**

Index Terms— **Mask R-CNN; Deep learning; Mitochondria; Electron Microscopy**

I. INTRODUCTION

It is generally known that mitochondria are the most essential and versatile organelles in eukaryotic cells that produce the overwhelming majority of adenosine triphosphate (ATP), and take substantial responsibility in the regulation of cellular life and death. Increasing evidence suggests that the regulation of mitochondrial shape are crucial for cellular physiology, as changes in mitochondrial shape have been linked to neurodegeneration, calcium signalling, lifespan and cell death [1]. Besides, the dysfunctional mitochondria which are the putative mediators of cell death are tightly related to cancers and several diseases, including Alzheimer's disease, Parkinson's disease, and Huntington's disease [2]. For example, previous researches [3], [4] indicated that the mitochondria in cancer cell can alter the function of resisting apoptosis, which naturally leads the research [5] for cancer therapy to target on mitochondria by stimulating mitochondrial membrane permeabilization or changing mitochondrial metabolism. Hence, automatic detection and segmentation of mitochondria from Electron Microscope (EM) images with high resolution are of great significance to the research on cellular physiology and cancer.

Owing to the existence of various subcellular structures, noises and imaging artifacts, the detection and segmentation of mitochondria remains a challenging problem. In recent years, many researchers have developed specialized algorithms for automatic detection and segmentation of mitochondria. Typically, Vitaladevuni *et al.* [6] used a GentleBoost classifier to detect mitochondria based on textural features. In [7], Narasimha *et al.* performed various classifiers for automatic texton-based mitochondria classification and segmentation in melanoma cells. Considering the important shape cues, Lucchi *et al.* [8] proposed an automated graph partitioning scheme incorporated with shape features, which operates on supervoxels instead of voxels to reduce the computational complexity. And they made further improvement by explicitly modeling membranes and introducing context-based features [9]. Note the fact that mitochondria have thick dark membranes, Jorstad *et al.* proposed an explicit active surface scheme for refining the boundary surfaces of mitochondrial segmentation [10]. Besides, some researchers achieved reasonable results by focusing on the graphical models commonly used in image segmentation, such as MRFs and CRFs [11], [12], [13]. However, all the works mentioned above need to design hand-crafted features of mitochondria. Recent years have witnessed the great success of CNNs in the field of computer vision since the representation capability of learned CNN features is more powerful than the traditional hand-crafted features. Based on this fact, we propose to design an improved Mask R-CNN for mitochondrial detection and segmentation, which can be trained end to end producing detection and segmentation results in parallel [14]. On this base, we use the morphological processing and mitochondrial context information to rectify the local misleading results.

The remainder of this study is outlined as follows: Section II presents a specific method for mitochondrial detection and segmentation. Then, the experimental results are shown to verify the effectiveness of proposed method in Section III. Finally, in Section IV, this study's conclusions are made, and some future research issues are discussed.

II. METHODS

In this section, we present our proposed algorithm that comprises image preprocessing, mitochondrial detection and segmentation, and post processing.

This research study is supported by the National Science Foundation of China (NO. 61673381, NO. 61201050, NO. 61306070, NO. 61701497, NO. 11771130), Special Program of Beijing Municipal Science & Technology Commission (NO. Z161100000216146), Strategic Priority Research Program of the CAS (NO. XDB02060001), and Scientific research instrument and equipment development project of Chinese Academy of Sciences (YZ201671).

[1]Jing Liu, Weifu Li, Chi Xiao, Bei Hong, Qiwei Xie and Hua Han are with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China.

[2]Qiwei Xie and Hua Han are with the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 20000, P. R. China.

[3]Hua Han is with the School of Future Technology, University of Chinese Academy of Sciences, Beijing 100190, P. R. China. (e-mail: hua.han@ia.ac.cn).

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

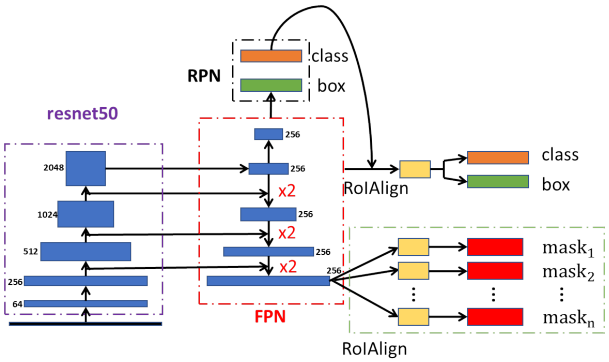

Fig. 1. The architecture of proposed network. The black block, blue blocks, orange blocks, green blocks, red block, and yellows blocks represent the input gray image, convolution layers, classifier layers, regression layers, mask output of original image, and fixed size feature maps obtained from RoIAlign, respectively. In addition, the purple dotted box contains the layers of resnet50, the black dotted box includes the two output layers of RPN, and the red dotted box denotes the top-down pathway of FPN. The green dotted box shows the multiple mask outputs, which is also the main difference between Mask R-CNN and our method.

A. Image Preprocessing

Deep learning needs a great deal of training data to improve the robustness of neural network. So we enlarge the training dataset by means of data augmentation, including rotation, flipping, adding random noise.

B. Mitochondrial Detection and Segmentation

In this subsection, we present an improved Mask R-CNN for mitochondrial detection and segmentation. The architecture of proposed network is illustrated in Fig. 1. Since the characteristics of mitochondria are relatively simple, the very deep network may cause over-fitting. In this paper, resnet50 [15] is adopted as the backbone network, and Feature Pyramid Network (FPN) [16] is used for detecting objects at different scales. The main task of Region Proposal Network (RPN) is producing the candidate object regions. Then the R-CNN makes further classification and regression based on the results from RPN. To get segmentation results, a mask branch is added to predict the object mask. Predicting pixel-accurate masks needs more detailed information, so only finest layer of FPN is connected to the mask branch. It is worth noting that the mask output is limited by the bounding box of detection output in original Mask R-CNN. That is to say, if the detection results cannot provide the complete location information of mitochondria, then the segmentation results will be incomplete. Inspired by previous researches, we move the FoV (Field of View) of the mask branch to get multiple mask outputs at a fixed step in four directions (up, down, left and right). The moving will be terminated if the percentage of background in the bounding box reaches a given threshold. The final segmentation is obtained by combining the multiple masks. Under this strategy, the error caused by detection branch can be weakened relatively. In addition, RoIPool is replaced by RoIAlign in order to overcome the misalignment problem.

C. Post Processing

In this subsection, we focus on optimizing the segmentation results obtained by proposed network. Since we fuse many masks to obtain relatively complete results, many trivial false segmentations will appear. Therefore, suitable post processing procedures will greatly improve the segmentation performance. The post processing procedures are mainly divided into three parts. First, we use morphological opening operation by a disk with radius 10 to eliminate the small regions and smooth the big regions. Second, note the fact that the mitochondrial sizes are far more than the resolution in $z-$direction. We use the multi-layer information fusion algorithm in [20] to obtain the mitochondria in 3D and eliminate these mitochondria with the "length" (times arising in $z-$direction) less than L (such as 15). Third, note that the segmentation results should keep consistent in adjacent layers, especially in FIB-SEM dataset. We compute the IoUs of segmentation results in adjacent layers, where a larger IoU means a better segmentation. Based on this assumption, we find the indexes of IoUs less than a given threshold T_1 (such as 0.7), where the segmentation results are considered as suspectable. Then the following criteria are formulated. If the ratio of suspectable number is greater than a given threshold T_2 (such as 0.3), this mitochondrion will be considered as false. In addition, if a small number of suspectable segmentations arise in both end-points, we replace them by the nearest segmentation by morphological erosion operation with radius 2.

III. Experimental and Results

In this section, we first provide some details related to the experimental dataset and experimental setup. Then, the results of proposed method including the mitochondrial detection results, segmentation results, and 3D visualization results are presented.

A. Datasets

In this subsection, the experiments are performed on two datasets (FIB-SEM dataset and ATUM-SEM dataset) that are widely used in this topic. Fig. 2 shows specific examples in each dataset and the related details are depicted as follows.

FIB-SEM dataset: the FIB-SEM dataset is publicly available[1] from mouse hippocampus and composed by a training volume and a testing volume. Here, each volume with a resolution of $5\times5\times5$ $nm^3/voxel$ consists of 165 slices and the size of each slice is 1024×768.

ATUM-SEM dataset: the ATUM-SEM dataset is acquired by the Institute of Automation, Chinese Academy of Sciences from mouse cortex [20]. The ground truth of mitochondria are annotated by experienced students using the software TrakEm2. The dataset with a resolution of $2\times2\times50$ $nm^3/voxel$ consists of 31 slices, each of which has a size of 4096×4096.

[1] http://cvlab.epfl.ch/data/em

TABLE I

AVERAGE *precision* AND *recall* GRAPH VERSUS ϕ

Dataset	ϕ	0.5	0.6	0.7	0.8	0.9
FIB-SEM	*precision*	0.872	0.882	0.901	0.911	0.930
	recall	0.894	0.886	0.878	0.865	0.842
ATUM-SEM	*precision*	0.812	0.842	0.863	0.891	0.922
	recall	0.952	0.948	0.945	0.942	0.925

B. Experimental Setup

We implement the proposed network using the Keras open-source deep learning library [17]. During the experiments, the original images are directly used as input for FIB-SEM dataset while the original images are cropped into small images (1024×1024 used in the experiments) as input for ATUM-SEM dataset. The network is trained using the stochastic gradient descent. The related parameters are as follows: momentum is 0.9, weight decay is 0.0001, and learning rate is initially set as 0.001 and decreases by a factor 10 when learning stagnates. The training and testing tasks are conducted on a server equipped with an Intel i7 CPU of 512 GB main memory and a Tesla K40 GPU.

C. Experimental Results

In this subsection, we present some detection and segmentation results in Fig. 2. It is clearly that the proposed method can detect and segment the mitochondria at different scales and sizes. And we evaluate the detection results in terms of two fundamental performance indicators, *precision* and *recall*. Specifically, *precision* is the ratio of detection outcome being correct, and *recall* is the ratio of the true elements being successfully detected, i.e.

$$precision = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP}), \tag{1}$$

$$recall = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}). \tag{2}$$

Here a detection result is considered as positive if the overlap between the detection region and corresponding ground truth occupies at least 70% of the area of the ground truth. Tab. I shows the *precision* and *recall* at different confidence threshold ϕ for the two datasets, respectively. We can see that the *recall* decreases and the *precision* increases with the increase of confidence threshold ϕ. Different from the detection, the segmentation accuracy is measured by the *Jaccard index*, which is a common criteria operated on pixels in image segmentation. It is computed as

$$Jaccard\ index = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP}+\mathrm{FN})}. \tag{3}$$

We present quantitative comparison with state-of-the-art methods in Tab. II (for FIB-SEM dataset) and Tab. III (for ATUM-SEM dataset) in terms of *Jaccard index*, where the highest values are marked in bold for distinction. It can be seen that the proposed method have obviously better segmentation performance than previous methods. To clearly see the mitochondrial structure in 3D, we import our segmentation results into software AMIRA for 3D visualization. Fig. 3 displays a specific example.

TABLE II

SEGMENTATION RESULTS ON FIB-SEM DATASET

Methods	*Jaccard index*
Non-parametric Higher-order Random Fields [13]	0.762
Improved KernelBoost[19]	0.776
Kernelized SSVM/CRF [11]	0.840
Our method	**0.849**

TABLE III

SEGMENTATION RESULTS ON ATUM-SEM DATASET

Methods	*Jaccard index*
Proposed method in [20]	0.747
U-Net[18]	0.837
Our method	**0.864**

IV. CONCLUSIONS

In this paper, we have presented an improved Mask R-CNN to obtain the mitochondrial locations and morphology. The application of FPN achieves more accurate detection, especially for the smaller and larger mitochondria, and the mask branch can get segmentation results directly based on the detection results. This mechanism can avoid the interrupt of noise and artifacts comparing with semantic segmentation network. And the application of moving FoV can refine the boundary of mitochondria which reduces the influence of detection results. Lastly, some post processing methods are used to improve the segmentation accuracy. Along the line of present research, we plan to expand this network to 3D in the future to explore more effective architecture. In addition, the mask output can be directly used as another input of mask branch, thereby forming a recurrent network which is more elegant.

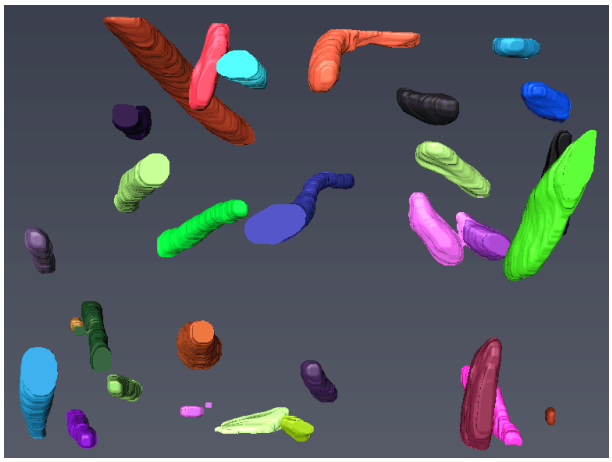

Fig. 3. Computerized reconstruction results on FIB-SEM.

REFERENCES

[1] S. Campello and L. Scorrano, "Mitochondrial shape changes: orchestrating cell pathophysiology," EMBO Rep., vol. 11, no. 9, pp. 678-684, Sep. 2010.

[2] MB. de Moura, LS. dos Santos, and B. Van Houten. "Mitochondrial dysfunction in neurodegenerative diseases and cancer," *Environ. Mol. Mutagen.*, vol. 51, no. 5, pp. 391-405, 2010.

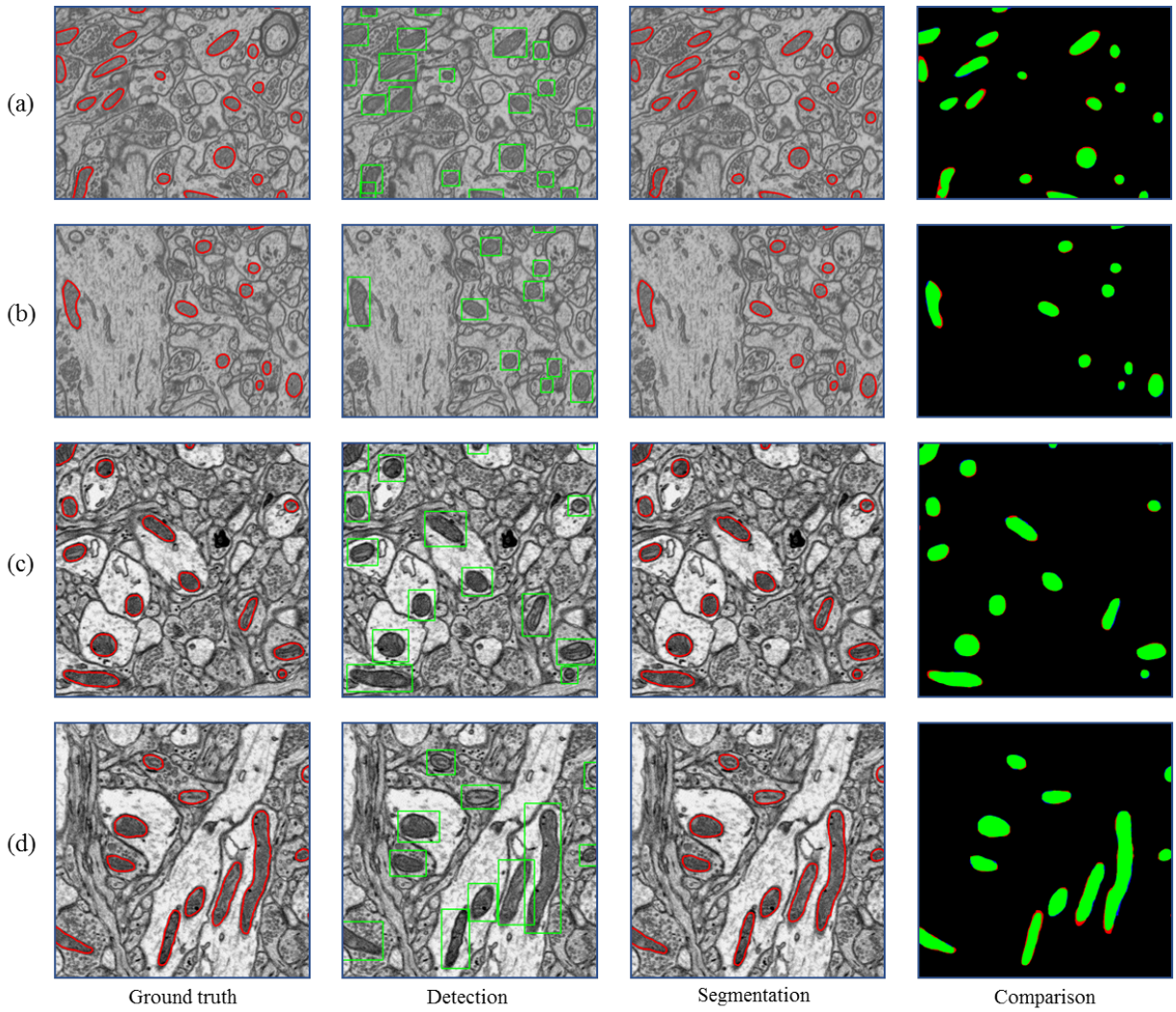

Fig. 2. (a)(b): FIB-SEM data; (c)(d): ATUM-SEM data. From left to right: ground truth of mitochondria, detection results indicated by green rectangles, segmentation results indicated by red curves, and the comparison to ground truth, where the green pixels denote true positives (TP), the blue pixels denote false positives (FP), the red pixels represent false negatives (FN), and the black pixels represent true negatives (TN).

[3] V. Gogvadze, S. Orrenius and B. Zhivotovsky, "Mitochondria in cancer cells: what is so special about them?" *Trends Cell Bio.*, vol. 18, pp. 165-173, 2008.

[4] D. C. Wallace, "Mitochondria and cancer," *Nat. Rev. Cancer*, vol. 12, pp. 685-698, 2012.

[5] F. Simone, G. Lorenzo and K. Guido, "Targeting mitochondria for cancer therapy," *Nat. Rev. Drug Discov.*, vol. 9, no. 6, pp. 447-464, 2010.

[6] S. Vitaladevuni, Y. Mishchenko, A. Genkin, D. Chklovskii and K. Harris, "Mitochondria detection in electron microscopy images," In Conf. Workshop on Microscopic Image Analysis with Applications in Biology. 2008.

[7] R. Narasimha, H. Ouyang, A. Gray, S. McLaughlin and S. Subramaniam, "Automatic joint classification and segmentation of whole cell 3D images," *Pattern Recogn.*, vol. 42, no. 6, pp. 1067-1079, 2009.

[8] A. Lucchi, K. Smith, R. Achanta, G. Knott and P. Fua, "Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features," *IEEE Trans Med. Imaging*, vol. 31, no. 2, pp. 474-486, 2012.

[9] A. Lucchi, C. Becker, P. M. Neila and P. Fua, "Exploiting enclosing membranes and contextual cues for mitochondria segmentation," in *Conf.* MICCAI, 65-72, 2014.

[10] A. Jorstad and P. Fua, "Refining mitochondria segmentation in electron microscopy imagery with active surfaces," in *Conf.* ECCV Workshops 4, 367-379, 2014.

[11] A. Lucchi, Y. Li, K. Smith and P. Fua, "Structured image segmentation using kernelized features," in *Conf.* ECCV, 2012.

[12] A. Lucchi, Y. Li and P. Fua, "Learning for structured prediction using approximate subgradient descent with working sets," in *Conf.* CVPR, 2013.

[13] P. Márquez-Neila, P. Kohli, C. Rother and L. Baumela, "Non-parametric higher-order random fields for image segmentation," in *Conf.* ECCV, 2014.

[14] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017.

[15] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Conf.* CVPR, 2016.

[16] T. Y. Lin, P. Dollar, R. Girshick, K. M. He and B. Hariharan, "Feature pyramid networks for object detection," in *Conf.* CVPR, 2017.

[17] Keras: Deep learning library for theano and tensorflow, 2015. http://keras.io/.

[18] O. Ronneberger, P. Fischer and T. Brox. "U-Net: Convolutional networks for biomedical image segmentation," in *Conf.* MICCAI, 2015.

[19] R. Rigamonti, V. Lepetit and P. Fua, "Beyond kernelboost," CoRR, vol.abs/1407.8518, 2014.

[20] W. F. Li, H. Deng, Q. Rao, Q. W. Xie, X. Chen and H. Han, "An automated pipeline for mitochondrial segmentation on ATUM-SEM Stacks," *J. Bioinf. Comput. Biol.* vol. 15, no. 3, pp. 1750015 (26 pages), 2017.