

# Contrastive Learning for Mitochondria Segmentation

Zhili Li, Xuejin Chen, Jie Zhao and Zhiwei Xiong

**Abstract**—Mitochondria segmentation in electron microscopy images is essential in neuroscience. However, due to the image degradation during the imaging process, the large variety of mitochondrial structures, as well as the presence of noise, artifacts and other sub-cellular structures, mitochondria segmentation is very challenging. In this paper, we propose a novel and effective contrastive learning framework to learn a better feature representation from hard examples to improve segmentation. Specifically, we adopt a point sampling strategy to pick out representative pixels from hard examples in the training phase. Based on these sampled pixels, we introduce a pixel-wise label-based contrastive loss which consists of a similarity loss term and a consistency loss term. The similarity term can increase the similarity of pixels from the same class and the separability of pixels from different classes in feature space, while the consistency term is able to enhance the sensitivity of the 3D model to changes in image content from frame to frame. We demonstrate the effectiveness of our method on MitoEM dataset as well as FIB-SEM dataset and show better or on par with state-of-the-art results.

**Index Terms**—Electron Microscopy; Mitochondria; Image Segmentation; Contrastive Learning

## I. INTRODUCTION

Known as the powerhouses of cells, mitochondria play a crucial role in the regulation of cellular life and death for they carry out all types of important cellular functions by producing the overwhelming majority of cellular adenosine triphosphate (ATP). Many clinical studies have revealed the correlation between the biological function of mitochondria and their size and geometry [1], [2]. In the last few years, with the development of imaging technology, electron microscopy (EM) images have been widely used for the analysis of nanometer-level structures including mitochondria. Knott *et al.* [3] deliver a dataset of an adult rodent brain tissue using focused ion beam scanning electron microscopy (FIB-SEM) to see axons, dendrites and their synaptic contacts. More recently, Wei *et al.* [4] publish a large-scale EM dataset (MitoEM) for 3D mitochondria instance segmentation. MitoEM dataset is  $3,600\times$  larger than FIB-SEM dataset [3] and consists of more mitochondria with sophisticated shapes and ultra-structures that are similar in appearance to mitochondria. Therefore, this dataset poses new challenges to the segmentation of mitochondria.

Many earlier approaches combine traditional image processing and machine learning techniques for mitochondria

segmentation. Jorstad and Fua [5] propose an active surface-based method to refine the boundary surfaces by exploiting the thick and dark membranes of mitochondria. Lucchi *et al.* [6] utilize an approximate subgradient descent algorithm to minimize the hinge loss in the margin-sensitive frameworks. However, these methods rely on hand-crafted features to build the classifiers, limited in efficiency and generalizability. Recently, deep learning-based approaches [7], [8], [9], [10], [11], [12] show promising results in mitochondria segmentation. Liu *et al.* [7] adopt a modified Mask R-CNN for mitochondria instance segmentation on the FIB-SEM dataset [3]. Oztel *et al.* [8] propose a fully convolutional network (FCN) to segment mitochondria and then use various post-processing, such as 2D spurious detection filtering, boundary refinement and 3D filtering, to improve the segmentation. Xiao *et al.* [9] propose an effective 3D residual fully convolutional network (FCN) to avert the vanishing gradient problem for mitochondria segmentation. Most of the methods listed above improve the segmentation by dedicatedly designing the network architecture, or adding various post-processing algorithms. However, these methods rarely pay extra attention to hard examples, such as noises, imaging artifacts and sub-cellular ultra-structures, which widely exist in EM images. As a result, these approaches can scarcely segment the hard examples correctly, preventing further improvement in mitochondria segmentation accuracy.

In the last few years, contrastive learning has been widely applied in computer vision tasks [13], [14], [15], [16], [17]. Contrastive learning is an approach to formulate the task of finding similar and dissimilar things for a deep learning model. [13], [14] adopt contrastive learning for unsupervised visual representation learning. For image classification, [15] extend the self-supervised batch contrastive approach to the fully-supervised setting, which effectively leverage label information. More recently, [16] propose to first pretrain the CNN feature extractor using a label-based contrastive loss for semantic segmentation task. [17] adopt point-wise contrastive learning to improve boundary estimation for ultrasound image segmentation. Inspired by [17], we extend contrastive learning to solve hard examples in mitochondria segmentation, based on the intra-class similarity of mitochondria textures and the inter-class discrepancy of textures between mitochondria and hard examples.

In this work, we propose a contrastive learning based framework for mitochondria segmentation task. In the training phase, we first sample representative points (i.e. pixels) from hard examples. Then we adopt a contrastive loss, which is computed on the feature space of these points, to facilitate model training. Our contrastive loss consists of two terms,

Corresponding author: Xuejin Chen, xjchen99@ustc.edu.cn

Zhili Li, Xuejin Chen, Jie Zhao and Zhiwei Xiong are with National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application, University of Science and Technology of China and Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China.

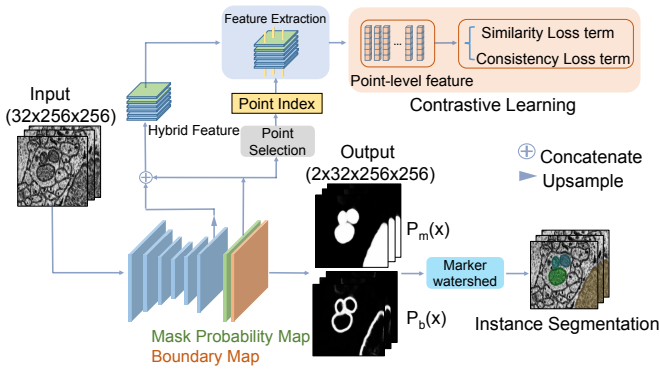


Fig. 1. The pipeline of our proposed framework.

similarity term and consistency term. The similarity loss term is designed to improve the feature similarity of the points from the same class and decrease that of different classes. The consistency loss term focuses on the feature similarity of points in adjacent frames to enhance model robustness of image content changes. We conduct many experiments on both MitoEM dataset and FIB-SEM dataset to verify the effectiveness of proposed method.

## II. OUR METHOD

Our method consists of three components: a backbone network in a symmetric encoder-decoder structure, a point selection block to sample representative points, and a contrastive learning head to compute the contrastive loss during the training stage, as shown in Fig. 1.

### A. Network Architecture

We build a semantic segmentation backbone network based on the 3D U-Net [18]. Following [19], we generate a mask probability map and a boundary map from the backbone segmentation network. At the end of our framework, a marker-controlled watershed post-processing algorithm [20] is applied to isolate each instance from the semantic segmentation map. In this paper, we denote our backbone network by U3D-BC following [4].

Training the segmentation network greatly relies on the labeled data which is expected to be diverse and representative. For each input image  $X \in R^{W \times H}$ , we have its dense mask label  $M \in \{0, 1\}^{W \times H}$  as well as the boundary label  $B \in \{0, 1\}^{W \times H}$ , which can be generated from the mask label, same as [4]. In previous methods, a common way to train the segmentation network is to minimize the overall cross-entropy loss defined as

$$\begin{aligned}
 L_{CE} = & \sum_{x \in X} -\log P_m(x)M(x) - \log(1 - P_m(x))(1 - M(x)) \\
 & + \sum_{x \in X} -\log P_b(x)B(x) - \log(1 - P_b(x))(1 - B(x))
 \end{aligned} \quad (1)$$

where  $x$  is a pixel of the image  $X$ ,  $P_m$  denotes the predicted probability map of mitochondria masks, and similarly  $P_b$  is the predicted probability of boundaries.

The cross-entropy loss penalizes the predicted probability of all pixels without distinction. Moreover, there are usually far more easy examples than hard samples in EM images. Therefore, it's challenge for existing methods to learn good representations of hard examples, making the predictions of them inaccurate. To solve this problem, we propose to first select representative points from both hard examples and easy examples, and then penalize the predictions of hard points additionally in a contrastive manner. In this way, hard examples are paid extra attention and consequently their segmentation results can be possibly aided.

### B. Point Selection Block

The point selection block is designed to sample representative points during the training phase. We sample two types of points from the input image. One is uncertain points, which are sampled from hard examples, and the other one is certain points, which have high prediction confidence in the mask probability map. Let  $N$  denote the number of sampled points, then the specific process of point selection is as follows. Based on the mask probability map, we randomly sample a representative point set  $\mathcal{S}$ , which consists of  $\beta N$  uncertain points and  $(1 - \beta)N$  certain points. The uncertain points are sampled based on the prediction error, which is the absolute value of the difference between the predicted probability and the ground truth label. Top  $\beta N$  points with the highest prediction error make up uncertain point set  $\mathcal{S}_u$ . Likewise, we choose the top  $(1 - \beta)N$  points with the highest mask probability to constitute the certain point set  $\mathcal{S}_c$ .

### C. Contrastive Learning

With  $N$  representative points, we adopt contrastive learning to mine hard examples in the training phase. Contrastive learning is mostly utilized in feature space to facilitate model training. The point features are derived from the backbone network. In the backbone segmentation network, we extract a mask probability map  $M$  and a feature map  $F \in R^{\frac{W}{2} \times \frac{H}{2} \times C}$  from the last two layers. We use trilinear interpolation to upsample the feature map to size of  $W \times H$  and concatenate it with the probability map  $M$  into a hybrid feature map. The feature vectors of representative points in the hybrid feature map can be extracted according to their point index.

With the feature vectors of  $N$  representative points, we compute a contrastive loss to “weight” hard points. Our contrastive loss consists of two terms: similarity loss term and consistency loss term. The similarity term can improve the feature similarity of points from the same class, while reducing that of points from different classes. Thus the segmentation of hard examples can be possibly aided. Complementary to the similarity term, the consistency term is designed to enhance the feature similarity between points of the same class at the same position in adjacent frames, and contrastively decrease the similarity of those from different classes. As a result, the inter-frame continuity of segmentation predictions will be improved. Both of the two loss terms contribute to a more discriminative and robust model.

The steps of similarity loss term computation are as follows. First, we split point set  $\mathcal{S}$  into three sets, certain foreground set  $\mathcal{S}_{cf}$ , uncertain foreground set  $\mathcal{S}_{uf}$  and background set  $\mathcal{S}_b$  according to their predicted probability and class label. Every point labeled as foreground in the certain point set  $\mathcal{S}_c$  will be clustered to  $\mathcal{S}_{cf}$ . Similar clustering rule is also applied to the uncertain point set  $\mathcal{S}_u$ : for each point in  $\mathcal{S}_u$ , if it belongs to the foreground then it will be assigned to  $\mathcal{S}_{uf}$ . The remaining points in  $\mathcal{S}$  are clustered to  $\mathcal{S}_b$ . Afterwards, the certain foreground set and the uncertain foreground set form the positive pair while the certain foreground set and the background set build the negative pair. Let  $\mathbf{p}_r$  represent the feature vector of point  $r$  and  $r_z$  be the normalized  $z$  coordinate of point  $r$ . Then the similarity loss  $L_{sim}$  is computed in the following formula:

$$Sim(\mathbf{p}, \mathbf{q}) = \frac{\mathbf{p} \cdot \mathbf{q}}{\|\mathbf{p}\| \cdot \|\mathbf{q}\|} \quad (2)$$

$$w(r, s) = \exp(\alpha \cdot (r_z - s_z)^2) \quad (3)$$

$$pos(r) = \frac{1}{|\mathcal{S}_{uf}|} \sum_{s \in \mathcal{S}_{uf}} w(r, s) \cdot \exp(Sim(\mathbf{p}_r, \mathbf{p}_s)) \quad (4)$$

$$neg(r) = \frac{1}{|\mathcal{S}_b|} \sum_{t \in \mathcal{S}_b} w(r, t) \cdot \exp(Sim(\mathbf{p}_r, \mathbf{p}_t)) \quad (5)$$

$$L_{sim} = \gamma - \frac{1}{|\mathcal{S}_{cf}|} \sum_{r \in \mathcal{S}_{cf}} \log\left[1 + \frac{pos(r)}{neg(r)}\right] \quad (6)$$

We adopt  $Sim(\mathbf{u}, \mathbf{v})$  to compute the cosine similarity between two vectors  $\mathbf{u}$  and  $\mathbf{v}$ .  $|\cdot|$  denotes the number of elements in a set.  $w(r, s)$  is the similarity weight based on the distance along  $z$  axis between point  $r$  and point  $s$ .  $\alpha$  is a constant term that adjusts the correlation of features between adjacent frames.  $\gamma$  is a constant used to ensure that  $L_{sim}$  is always greater than zero.

The consistency loss term is computed as follows. For each point  $s$  with coordinates  $(x, y, z)$  in the sampled point set  $\mathcal{S}$ , we sample another point  $s'$  on  $(x, y, z - 1)$  from adjacent frame. If point  $s$  has the same class label with point  $s'$ , they will be viewed as a positive pair and otherwise be regarded as a negative pair. We denote the positive pair set as  $\mathcal{P}$ , and the negative pair set as  $\mathcal{N}$ . The formulation of loss term  $L_{con}$  can be written down as follows:

$$pos = \frac{1}{|\mathcal{P}|} \sum_{(s, s') \in \mathcal{P}} \exp(Sim(\mathbf{p}_s, \mathbf{p}'_s)) \quad (7)$$

$$neg = \frac{1}{|\mathcal{N}|} \sum_{(s, s') \in \mathcal{N}} \exp(Sim(\mathbf{p}_s, \mathbf{p}'_s)) \quad (8)$$

$$L_{con} = \log\left(1 + \frac{pos}{neg}\right) \quad (9)$$

The overall loss function of our framework is:

$$L_{total} = L_{CE} + \lambda_1 L_{sim} + \lambda_2 L_{con} \quad (10)$$

$L_{CE}$  denotes the cross entropy loss on the output mask probability map and boundary map.  $\lambda_1$  and  $\lambda_2$  are two constant terms that control the weight of  $L_{sim}$  and  $L_{con}$ , separately.

TABLE I  
SEGMENTATION RESULTS ON MITOEM DATASET.

Dataset	Method	AP-75			
		Small	Medium	Large	All
MitoEM-R	U3D-BC	0.139	0.724	0.895	0.844
	Our method	<b>0.203</b>	<b>0.743</b>	<b>0.913</b>	<b>0.870</b>
MitoEM-H	U3D-BC	0.232	0.747	0.825	0.773
	Our method	<b>0.296</b>	<b>0.778</b>	<b>0.830</b>	<b>0.787</b>

### III. EXPERIMENTS AND RESULTS

In this section, we first introduce some details related to the experimental datasets and implementation. Then the results of our method and the quantitative and qualitative comparisons with other approaches are presented.

#### A. Datasets

We evaluate our method on two datasets (MitoEM dataset and FIB-SEM dataset). The related details are depicted as follows:

**MitoEM dataset [4]:** This dataset consists of two  $30 \mu m^3$  volumes,  $1000 \times 4096 \times 4096$  in voxels at  $30 \times 8 \times 8 nm^3$  resolution. The image volumes are acquired from a rat (MitoEM-R) and a human (MitoEM-H) tissue, respectively. For each image stack, only the labels of the first 500 slices have been published, and the labels of the remaining 500 slices are unreleased. In our experiments, we use the first 400 slices of published data for training and evaluate the segmentation performance on the last 100 slices.

**FIB-SEM dataset [3]:** This dataset is obtained from mouse hippocampus and composed by a training volume and a testing volume. Each volume with a resolution of  $5 \times 5 \times 5 nm^3/voxel$  consists of 165 slices and the size of each slice is  $1024 \times 768$ .

#### B. Implementation Details

We implement the proposed method using the Pytorch open-source deep learning library [21]. In our experiments, the input data size for training is  $32 \times 256 \times 256$  for both MitoEM and FIB-SEM dataset. The network is trained using the stochastic gradient descent [22]. We follow the same data augmentation and learning schedule setting as [23]. We set the hyper-parameters in this work as follows:  $N=1024$ ,  $\beta=0.75$ ,  $\alpha=-4.0$ ,  $\lambda_1=\lambda_2=0.2$ ,  $\gamma=\log(1 + e^2)$ .

#### C. Evaluation Metrics and Results

For segmentation evaluation, AP-75 and Jaccard index are two main metrics to measure the results. AP-75 evaluates the segmentation performance in instance level. In AP-75 evaluation metric, a predicted instance is considered as a TP only if the voxel-wise overlap between the prediction and corresponding ground truth reaches at least 75%. Jaccard index is a common segmentation criteria to measure the accuracy in pixel level, and its formula is as follows:

$$Jaccard\ index = \frac{TP}{(TP + FP + FN)} \quad (11)$$

TABLE II  
EFFECT OF CONTRASTIVE LOSS TERMS ON MITOEM DATASET.

Dataset	$L_{CE}$	$L_{sim}$	$L_{con}$	AP-75			
				Small	Medium	Large	All
MitoEM-R	✓	✗	✗	0.139	0.724	0.895	0.844
	✓	✗	✓	0.166	0.679	0.909	0.858
	✓	✓	✗	0.185	<b>0.747</b>	0.909	0.866
	✓	✓	✓	<b>0.203</b>	0.743	<b>0.913</b>	<b>0.870</b>
MitoEM-H	✓	✗	✗	0.232	0.747	0.825	0.773
	✓	✗	✓	0.216	0.755	0.829	0.778
	✓	✓	✗	0.292	0.777	0.825	0.781
	✓	✓	✓	<b>0.296</b>	<b>0.778</b>	<b>0.830</b>	<b>0.787</b>

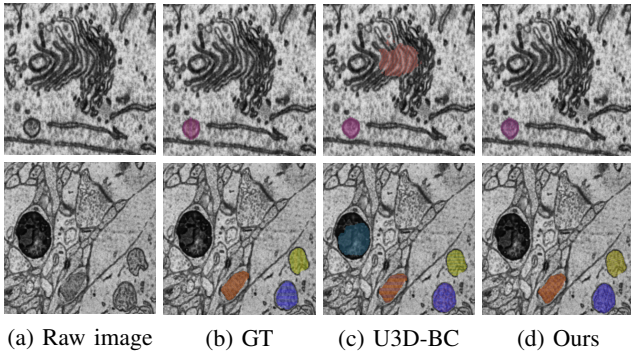


Fig. 2. Visualization results on MitoEM dataset. From left to right: raw images, ground truth, results of U3D-BC, and results of our method.

Following [4], we report the AP-75 results for small, medium, large and all instances separately to evaluate the segmentation results on MitoEM dataset. Table I shows the performance comparison of U3D-BC and our method on MitoEM-H and MitoEM-R validation datasets. It’s obvious to see that our method brings continuous improvement over baseline on all indicators for both two datasets, which demonstrates the effectiveness of our approach. We also conduct ablation study about the effect of two contrastive loss terms on MitoEM dataset as shown in Table II. When utilizing similarity loss term ( $L_{sim}$ ) or consistency loss term ( $L_{con}$ ), our method can always achieve better results over the baseline. After combining them together, the segmentation performance can be further improved. This suggests that our two contrastive loss terms are essential and indispensable for training a good segmentor.

Fig. 2 illustrates the qualitative comparison of segmentation results between the baseline and our method on MitoEM dataset. Our method can correctly distinguish mitochondria-like organelles from real mitochondria. This indicates that our method can solve hard examples better. To clearly show the 3D mitochondria structure, we import our segmentation results into VAST [24] for visualization, as shown in Fig. 3.

On FIB-SEM dataset, we utilize Jaccard index (foreground IoU) to measure segmentation accuracy following [8], [9], [10], [11]. The quantitative comparisons with previous methods are presented in Table III, where the top three values

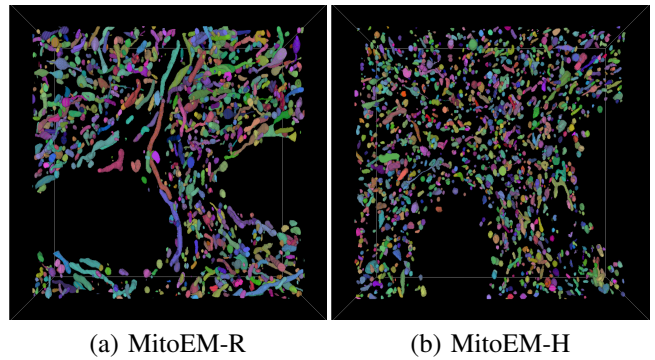


Fig. 3. Instance segmentation results on MitoEM-R and MitoEM-H.

TABLE III  
SEGMENTATION RESULTS ON FIB-SEM DATASET

Method	Jaccard index
Lucchi <i>et al.</i> 2013 [6]	0.867
3D U-Net [18]	0.874
Oztel <i>et al.</i> 2017 [8]	<b>0.907</b>
Cheng <i>et al.</i> 2017 [10]	0.889
Improved Mask R-CNN [7]	0.849
Xiao <i>et al.</i> 2018 [9]	<b>0.900</b>
Casser <i>et al.</i> 2020 [11]	0.890
Liu <i>et al.</i> 2020 [12]	0.864
Ours	<b>0.895</b>

are masked in bold for distinction. It can be seen that the Jaccard index of our approach is higher than most previous algorithms, demonstrating that our method achieves state-of-the-art results. To explicitly compare the qualitative segmentation differences between our method and others, we displayed 3 examples in Fig. 4. Note that our approach segments most of the mitochondria and effectively reduces FPs and FNs, achieving better qualitative results than [8], [11] and comparable results to [9].

#### IV. CONCLUSION

In electron microscopy images, there are many hard examples which limit the improvement of mitochondria segmentation accuracy. To address this issue, we proposed an effective contrastive learning-based framework in which we first pick out representative points, then build positive pair and negative pair based on their labels and prediction probability, and finally compute a contrastive loss to facilitate model training. Our contrastive loss can increase intra-class compactness, inter-class separability and feature consistency between adjacent frames, thereby enabling a better feature representation. The experimental results on both two EM datasets demonstrate the effectiveness of our method in mitochondria segmentation. In the future, we plan to explore the application of contrastive learning in semi-supervised mitochondria segmentation task.

#### V. ACKNOWLEDGMENT

This work was supported in part by National Key R&D Program of China under Grant 2020AAA0105702, in part by



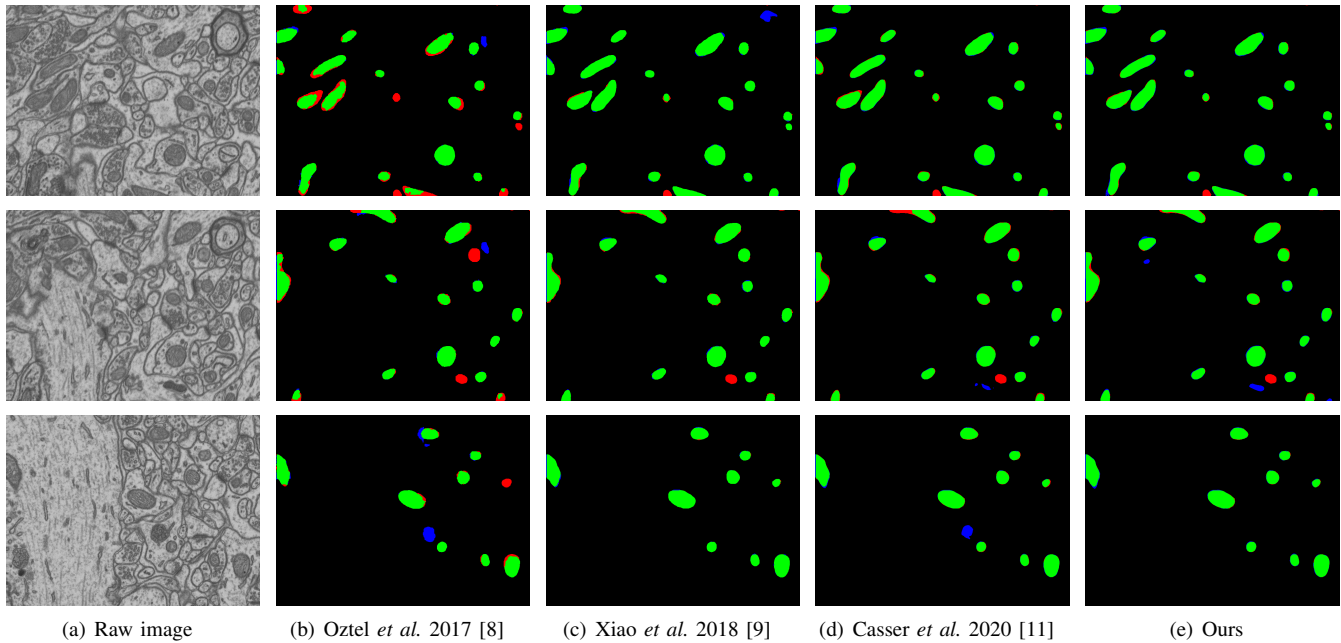


Fig. 4. The qualitative comparisons between different methods on FIB-SEM dataset. Green pixels denote true positives (TP), red pixels denote false negatives (FN), blue pixels denote false positives (FP), and black pixels denote true negatives (TN).

the National Natural Science Foundation of China (NSFC) under Grants 62076230, in part by the University Synergy Innovation Program of Anhui Province GXXT-2019-025.

#### REFERENCES

- [1] R. Roychaudhuri, M. Yang, Minako M. Hoshi, and David B. Teplow. Amyloid  $\beta$ -protein assembly and alzheimer disease. *Journal of Biological Chemistry*, 284:4749–4753, 2009.
- [2] S. Campello and L. Scorrano. Mitochondrial shape changes: orchestrating cell pathophysiology. *EMBO Reports*, 11:678–684, 2010.
- [3] G. Knott, H. Marchman, D. Wall, and B. Lich. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *Journal of Neuroscience*, 28(12):2959–2964, 2008.
- [4] D. Wei, Z. Lin, D. Barranco, N. Wendt, X. Liu, W. Yin, X. Huang, A. Gupta, W. Jang, X. Wang, I. Arganda-Carreras, J. Lichtman, and H. Pfister. MitoEM dataset: Large-scale 3D mitochondria instance segmentation from EM images. In *MICCAI*, 2020.
- [5] A. Jorstad and P. Fua. Refining mitochondria segmentation in electron microscopy imagery with active surfaces. In *Computer Vision - ECCV 2014 Workshops*, pages 367–379, 2015.
- [6] A. Lucchi, Y. Li, and P. Fua. Learning for structured prediction using approximate subgradient descent with working sets. In *CVPR*, pages 1987–1994, 2013.
- [7] J. Liu, W. Li, C. Xiao, B. Hong, Q. Xie, and H. Han. Automatic detection and segmentation of mitochondria from SEM images using deep neural network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 628–631, 2018.
- [8] I. Oztel, G. Yolcu, I. Ersoy, T. White, and F. Bunyak. Mitochondria segmentation in electron microscopy volumes using deep convolutional neural network. In *BIBM*, pages 1195–1200, 2017.
- [9] C. Xiao, X. Chen, W. Li, L. Li, L. Wang, Q. Xie, and H. Han. Automatic mitochondria segmentation for EM data using a 3D supervised convolutional network. *Frontiers in Neuroanatomy*, 12:92, 11 2018.
- [10] H. Cheng and A. Varshney. Volume segmentation using convolutional neural networks with limited training data. *ICIP*, pages 590–594, 2017.
- [11] V. Casser, K. Kang, H. Pfister and D. Haehn. Volume segmentation using convolutional neural networks with limited training data. *MIDL*, pages 111–120, 2020.
- [12] J. Liu, L. Li, Y. Yang, B. Hong, X. Chen, Q. Xie, and H. Han. Automatic reconstruction of mitochondria and endoplasmic reticulum in electron microscopy volumes by deep learning. *Frontiers in neuroscience*, 14, 2020.
- [13] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9726–9735, 2020.
- [14] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, volume 119, pages 1597–1607, 2020.
- [15] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.
- [16] X. Zhao, R. Vemulapalli, P. Mansfield, B. Gong, B. Green, L. Shapira, and Y. Wu. Contrastive learning for label-efficient semantic segmentation. *ArXiv*, abs/2012.06985, 2020.
- [17] H. Li, X. Yang, J. Liang, W. Shi, C. Chen, H. Dou, R. Li, R. Gao, G. Zhou, J. Fang, X. Liang, R. Huang, A. Frangi, Z. Chen, and D. Ni. Contrastive rendering for ultrasound image segmentation. In *MICCAI*, pages 563–572, 2020.
- [18] Ç. Özgün, A. Ahmed, L. Soeren S., B. Thomas, and R. Olaf. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *MICCAI*, pages 424–432, 2016.
- [19] H. Chen, X. Qi, L. Yu, and P. Heng. DCAN: Deep contour-aware networks for accurate gland segmentation. In *CVPR*, pages 2487–2496, 2016.
- [20] F. Meyer and S. Beucher. Morphological segmentation. In *Journal of Visual Communication and Image Representation*, pages 21–46, 1990.
- [21] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *CoRR*, 2019.
- [22] B. Léon. Stochastic gradient descent tricks. *Neural networks: Tricks of the trade*, 421–436, 2012.
- [23] K. Lee, J. Zung, P. Li, V. Jain, and H. Seung. Superhuman accuracy on the SNEMI3D connectomics challenge. *CoRR*, 2017.
- [24] Daniel R. Berger, H. Sebastian Seung, and Jeff W. Lichtman. Vast (volume annotation and segmentation tool): Efficient manual and semi-automatic labeling of large 3D image stacks. *Frontiers in Neural Circuits*, 12:88, 2018.