Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2020 December 03.

Published in final edited form as:

Med Image Comput Comput Assist Interv. 2020 October ; 12265: 66–76.

doi:10.1007/978-3-030-59722-1_7.

MitoEM Dataset: Large-scale 3D Mitochondria Instance Segmentation from EM Images

Donglai Wei[1], **Zudi Lin**[1], **Daniel Franco-Barranco**[2,3], **Nils Wendt**[4,*], **Xingyu Liu**[5,*], **Wenjie Yin**[1,*], **Xin Huang**[6,*], **Aarush Gupta**[7,*], **Won-Dong Jang**[1], **Xueying Wang**[1], **Ignacio Arganda-Carreras**[2,3,8], **Jeff W. Lichtman**[1], **Hanspeter Pfister**[1]

[1]Harvard University

[2]Donostia International Physics Center

[3]University of the Basque Country

[4]Technical University of Munich

[5]Shanghai Jiao Tong University

[6]Northeastern University

[7]Indian Institute of Technology Roorkee

[8]Ikerbasque, Basque Foundation for Science

Abstract

Electron microscopy (EM) allows the identification of intracellular organelles such as mitochondria, providing insights for clinical and scientific studies. However, public mitochondria segmentation datasets only contain hundreds of instances with simple shapes. It is unclear if existing methods achieving human-level accuracy on these small datasets are robust in practice. To this end, we introduce the *MitoEM* dataset, a 3D mitochondria instance segmentation dataset with two $(30\mu m)^3$ volumes from human and rat cortices respectively, 3, 600× larger than previous benchmarks. With around 40K instances, we find a great diversity of mitochondria in terms of shape and density. For evaluation, we tailor the implementation of the average precision (AP) metric for 3D data with a 45× speedup. On MitoEM, we find existing instance segmentation methods often fail to correctly segment mitochondria with complex shapes or close contacts with other instances. Thus, our MitoEM dataset poses new challenges to the field. We release our code and data: https://donglaiw.github.io/page/mitoEM/index.html.

Keywords

Mitochondria; EM Dataset; 3D Instance Segmentation

donglai@seas.harvard.edu.

*Works are done during internship at Harvard University.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1. Introduction

Mitochondria are the primary energy providers for cell activities, thus essential for metabolism. Quantification of the size and geometry of mitochondria is not only crucial to basic neuroscience research, *e.g.*, neuron type identification [26], but also informative to clinical studies, *e.g.*, bipolar disorder [13] and diabetes [35]. Electron microscopy (EM) images have been used to reveal their detailed 3D geometry at the nanometer level with the terabyte scale [22]. Consequently, to enable an in-depth biological analysis, we need high-throughput and robust 3D mitochondria instance segmentation methods.

Despite the advances in the large-scale instance segmentation for neurons from EM images [12], such effort for mitochondria has been overlooked in the field. Due to the lack of a large-scale public dataset, most recent mitochondria segmentation methods were benchmarked on the EPFL Hippocampus dataset [20] (referred to as *Lucchi* later on), where mitochondria instances are small in number and simple in morphology (Fig. 1). Even for the non-public dataset [1,8], mitochondria instances do not have complex shapes due to the limited dataset size and the non-mammalian tissue. However, in mammal cortices, the complete shape of mitochondria can be sophisticated, where even state-of-the-art neuron instance segmentation methods may fail. In Fig. 2a, we show a mitochondria-on-a-string (MOAS) instance [36], prone to the false split error due to the voxel-level thin connection. We also show multiple instances entangling with each other with unclear boundaries, prone to the false merge error in Fig. 2b. Therefore, we need a large-scale mammalian mitochondria dataset to evaluate current methods and foster new researches to address the complex morphology challenge.

To this end, we have curated a large-scale 3D mitochondria instance segmentation benchmark, **MitoEM**, which is 3,600× larger than the previous benchmark [20] (Fig. 1). Our dataset consists of two 30 μm^3 3D EM image stacks, one from an adult rat and one from an adult human brain tissue, facilitating large-scale cross-tissue comparison. For evaluation, we adopt the average precision (AP) evaluation metric and design an efficient implementation for 3D volumes to benchmark state-of-the-art methods. Our analysis of model performance sheds light limitations of current automatic instance segmentation methods.

1.1 Related Works

Mitochondria Segmentation.—Most previous segmentation methods are benchmarked on the aforementioned Lucchi dataset [20]. For mitochondria semantic segmentation, earlier works leverage traditional image processing and machine learning techniques [27,29,18,19], while recent methods utilize 2D or 3D deep learning architectures for mitochondria segmentation [24,4]. More recently, Liu *et al.* [17] showed the first instance segmentation approach on the Lucchi dataset with a modified Mask R-CNN [10], and Xiao *et al.* [30] obtained the instance segmentation through an IoU tracking approach. However, it is hard to evaluate their robustness in a large-scale setting due to the lack of a proper dataset.

Instance Segmentation for Biomedical Images.—Instance segmentation methods in the biomedical domain have been used for the segmenting glands from histology images and neurons from EM images. For gland, state-of-the-art methods [3] train deep learning models

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to predict both the semantic segmentation mask and the boundary map in a multi-task setting. Additional targets [32] and shape-preserving loss functions [33] are proposed for further improvement.

For neurons, there are two main methodologies. The first one trains 2D or 3D CNNs to predict an intermediate representation such as boundary [6,25,34] or affinity maps [28,15]. Then, clustering techniques such as watershed [7,37] or graph partition [14] transform these intermediate output into a segmentation. Adjacent segments are further agglomerated by a similarity measure using either the intermediate output [9] or a new classifier [11,23,37]. In the other methodology, CNNs are trained recursively to grow the current estimate of a single segmentation mask [12], which is extended to handle multiple objects [21]. Compared to neuron instances, the sparsity of mitochondria instances and the close appearance to other organelles make it hard to directly apply those segmentation methods tuned for neuron segmentation.

2. MitoEM Dataset

Dataset Acquisition.

Two tissue blocks were imaged using a multi-beam scanning electron microscope: *MitoEM-H*, from Layer II in the frontal lobe of an adult human and *MitoEM-R*, from Layer II/III in the primary visual cortex of an adult rat. Both samples are imaged at a resolution of $8 \times 8 \times 30\ nm^3$. After stitching and aligning the images, we cropped a $(30\ \mu m)^3$ sub-volume, avoiding large blood vessels where mitochondria are absent. To focus on the mitochondria morphology challenge, We made the specific design choice of the dataset size and region, which contains complex mitochondria without introducing much of the domain adaptation problem due to the diverse image appearance.

Dataset Annotation.

We facilitated a semi-automatic approach to annotate this large-scale dataset. We first manually annotated a $5\mu m^3$ volume for each tissue, then trained a state-of-the-art 3D U-Net (U3D) model [5] to predict binary masks for unlabeled regions, which are transformed into instance masks with connected-component labeling. Then expert annotator proofread and modify the prediction. With this pipeline, we iteratively accumulated ground truth instance segmentation for the 5,10,20,30 μm^3 sub-volumes for each tissue. Considering the complex geometry of large mitochondria, we ordered the labeled instances by volume size and conducted a second round of proofreading with 3D mesh visualization. Finally, we asked three neuroscience experts to go through the dataset to proofread until no disagreement.

Dataset Analysis.

The physical size of our two EM volumes is more than *3,600*× larger than the previous Lucchi benchmark [20]. MitoEM-H and MitoEM-R have around 24.5k and 14.4k mitochondria instances, respectively, over *500*× more than that of Lucchi [20]. We show the distribution of instance sizes for both volumes in Fig. 1. Both MitoEM-H and MitoEM-R follow the exponential distribution with different rate parameters. MitoEM-H has more small mitochondria instances, while MitoEM-R has more big ones. To illustrate the diverse

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

morphology of mitochondria, we show all 3D meshes of small objects (<5k voxels) and large objects (>30k voxels) from both tissues (Fig. 3, Top). Despite their differences in species and cortical regions, the mitochondria-on-a-string (MOAS) are common in both volumes, where round balls are connected by ultra-thin tubes. Furthermore, we plot the length versus volume of mitochondria instances for both volumes, where the length of the mitochondria is approximated by the number of voxels in its 3D skeleton (Fig. 3, Bottom left). There is a strong linear correlation between the volume and length mitochondria in both volumes, which is the average thickness of the instance. While the MitoEM-H has more small instances, the MitoEM-R has more large instances with complex morphologies. We sample mitochondria of different length along the regression line and find instances share similar shapes to MOAS in both volumes (Fig. 3, Bottom right).

3. Method

For the 3D mitochondria instance segmentation task, we first introduce the evaluation metric and provide an efficient implementation. Then, we categorize state-of-the-art instance segmentation methods for later benchmarking (Section 4).

3.1 Task and Evaluation Metric

Inspired by the video instance segmentation challenge [31], we adapt the COCO evaluation API [16] designed for 2D instance segmentation to our 3D volumetric segmentation. Out of COCO evaluation metrics, we choose AP-75 requiring at least 75% intersection over union (IoU) with the ground truth for a detection to be a true positive. In comparison, AP-95 is too strict even for human annotators and AP-50 is too loose for the high-precision biological analysis.

Efficient Implementation.—The original AP implementation for natural image and video datasets is suboptimal for the 3D volume. Two main bottlenecks are the saving/loading of individual masks from an intermediate JSON file, and the IoU computation. For our case, it is storage-efficient to directly input the whole volume, thus removing the overhead for data conversion. For an efficient IoU computation, we first compute the 3D bounding boxes of all the instance segmentation by iterating through each 2D slice in all three dimensions. It reduces the complexity to $3N + \mathcal{O}(1)$ compared to $KN + \mathcal{O}(1)$ by naively iterating through all instances, where N is the number of voxels and K is the number of instances. To compute the intersection region with ground truth instances, we only need to do local calculation within the precomputed bounding box. Compared to the previous version on the MitoEM-H dataset, our implementation achieves a **45**× speed-up for 4k instances within a 0.4 Gigavoxel volume.

3.2 State-of-the-Art Methods

We categorize state-of-the-art instance segmentation methods not only from mitochondria literature but also from neuron and gland segmentation (Fig. 4).

Bottom-up Approach.—Bottom-up approaches often use 3D U-Net to predict the binary segmentation mask [25] (U3D-B), affinity map [15] (U3D-A), or binary mask with instance

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contour [3] (U3D-BC). However, since those predictions are not the instance masks, several post-processing algorithms have been utilized for object decoding. Those algorithms include connected component labeling (CC), graph-based watershed, and marker-controlled watershed (MW). For rigorous evaluation of the state-of-the-art methods, we examine different combinations of model predictions and decode algorithms on our MitoEM dataset.

Top-down Approach.—Methods like Mask-RCNN [10] are not applicable due to the undefined scale of bounding boxes in the EM volume. Previously FFN [12] has shown promising results on neuron segmentation by gradually growing precomputed seeds. We therefore test FFN in the experiments.

4. Experiments

4.1 Implementation Details

For a fair comparison of bottom-up approaches, we use the same residual 3D U-Net [15] for all representations. For training, we use the same data augmentation and learning schedule as in [15]. The input data size is 112×112×112 for Lucchi and 32×256×256 for MitoEM due to its anisotropicity. We use weighted BCE loss for the prediction. For the FFN model [12], we only train it on the small Lucchi dataset, which already took 4 hours for label pre-processing. We use the official implementation online and train it until convergence.

4.2 Benchmark Results on Lucchi Dataset

We first show previous semantic segmentation results in Table 1a. To evaluate the metric sensitivity to the annotation, we perturb ground truth labels with 1-voxel dilation or erosion, which has similar performance to those from the previous methods. As the annotation is not pixel-level accurate, previous methods have already achieved human-level performance for semantic segmentation.

For the top-down approaches, we tried our best to tune the FFN method without obtaining desirable results (Tab. 1b). In particular, FFN achieves around 0.7 AP-50 but 0.2 AP-75, showing its weakness in capture object geometry.

For the bottom up approaches (Tab. 1c), U-Net models with standard training practice achieves on-par results with specifically designed methods [4]. However, the AP-75 instance metric can still reveal the false split and false merge errors in the prediction. All four representations provide similar semantic results and the U3D-BC+MW achieves the best instance decoding result with the help of the additional instance contour information.

4.3 Benchmark Results on MitoEM Dataset

We evaluate previous state-of-the-art methods on our MitoEM dataset. Specifically, both human (MitoEM-H) and rat (MitoEM-R) datasets are partitioned into consecutive train, val and test splits with 40%, 10% and 50% of the total amount of data. We select the hyper-parameters on the val split and report the final results on the test split. As mitochondria has diverse sizes, we also report the AP-75 results for small, medium and large instances separately with the volume threshold of 5K and 15K voxels.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As shown in Table 2, all methods perform consistently better on the human tissue (MitoEM-H) than the rat tissue. Besides, marker-controlled watershed (MW) is significantly better than connected-component (CC) and IoU-based tracking (IoU) for processing both binary mask (U3D-B) and binary mask + instance contour (BC). Furthermore, U3D-BC+MW achieves the best performance considering the mean AP-75 scores for both tissues. Our MitoEM posts new challenges for methods which are nearly perfect on the Lucchi dataset.

We show qualitative results of U3D-BC+MW (Fig. 5). Such method successfully captures many mitochondria with non-trivial shapes, but it is still not robust to the ambiguous boundary and overlapping surface. Further improvement can be achieved by considering 3D shape prior of mitochondria.

4.4 Cross-Tissue Evaluation

In this experiment, we examine the cross-tissue performance of the U3D-BC model. That is, we run inference on the MitoEM-Human dataset using the model trained on the MitoEM-Rat dataset, and vice versa. We observe that the MitoEM-R model achieves better performance on the human dataset than the MitoEM-H model, while the MitoEM-H model performs worse than MitoEM-R on the rat dataset (Table 3). Since the rat dataset contains more large objects with complex morphologies, it is reasonable that the models trained on rat datasets generalize better and can handle more challenging instances.

5. Conclusion

In this paper, we introduce a large-scale mitochondria instance segmentation dataset that reveals the limitation of state-of-the-art methods in the field to deal with mitochondria with complex shape or close contacts with others. Similar to ImageNet for natural images, our densely annotated MitoEM can have various applications beyond its original task, *e.g.*, feature pre-training, 3D shape analysis, and testing approaches on active learning and domain adaptation.

Acknowledgments.

This work has been partially supported by NSF award IIS-1835231 and NIH award 5U54CA225088-03.

References

1. Ariadne.ai: Automated segmentation of mitochondria and ER in cortical cells (2018 (accessed July 7, 2020)), https://ariadne.ai/case/segmentation/organelles/CorticalCells/
2. Beier T, Pape C, Rahaman N, Prange T, Berg S, Bock DD, Cardona A, Knott GW, Plaza SM, Scheffer LK, et al.: Multicut brings automated neurite segmentation closer to human performance. Nature methods 14(2) (2017)
3. Chen H, Qi X, Yu L, Heng PA: DCAN: deep contour-aware networks for accurate gland segmentation In: CVPR. pp. 2487–2496. IEEE (2016)
4. Cheng HC, Varshney A: Volume segmentation using convolutional neural networks with limited training data In: ICIP. pp. 590–594. IEEE (2017)
5. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O: 3d u-net: learning dense volumetric segmentation from sparse annotation In: MICCAI. pp. 424–432. Springer (2016)
6. Ciresan D, Giusti A, Gambardella LM, Schmidhuber J: Deep neural networks segment neuronal membranes in electron microscopy images. In: NeurIPS. pp. 2843–2851 (2012)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

7. Cousty J, Bertrand G, Najman L, Couprie M: Watershed cuts: Minimum spanning forests and the drop of water principle. TPAMI 31, 1362–1374 (2008)
8. Dorkenwald S, Schubert PJ, Killinger MF, Urban G, Mikula S, Svara F, Kornfeld J: Automated synaptic connectivity inference for volume electron microscopy. Nature methods 14(4), 435–442 (2017) [PubMed: 28250467]
9. Funke J, Tschopp F, Grisaitis W, Sheridan A, Singh C, Saalfeld S, Turaga SC: Large scale image segmentation with structured loss based deep learning for connectome reconstruction. TPAMI 41(7), 1669–1680 (2018)
10. He K, Gkioxari G, Dollár P, Girshick R: Mask r-cnn In: ICCV. pp. 2961–2969. IEEE (2017)
11. Jain V, Turaga SC, Briggman K, Helmstaedter MN, Denk W, Seung HS: Learning to agglomerate superpixel hierarchies In: NeurIPS. pp. 648–656 (2011)
12. Januszewski M, Kornfeld J, Li PH, Pope A, Blakely T, Lindsey L, Maitin-Shepard J, Tyka M, Denk W, Jain V: High-precision automated reconstruction of neurons with flood-filling networks. Nature Methods (2018)
13. Kasahara T, Takata A, Kato T, Kubota-Sakashita M, Sawada T, Kakita A, Mizukami H, Kaneda D, Ozawa K, Kato T: Depression-like episodes in mice harboring mtdna deletions in paraventricular thalamus. Molecular psychiatry (2016)
14. Krasowski N, Beier T, Knott G, Köthe U, Hamprecht FA, Kreshuk A: Neuron segmentation with high-level biological priors. TMI 37(4) (2017)
15. Lee K, Zung J, Li P, Jain V, Seung HS: Superhuman accuracy on the snemi3d connectomics challenge. arXiv:1706.00120 (2017)
16. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL: Microsoft coco: Common objects in context In: ECCV. pp. 740–755. Springer (2014)
17. Liu J, Li W, Xiao C, Hong B, Xie Q, Han H: Automatic detection and segmentation of mitochondria from sem images using deep neural network In: EMBC. IEEE (2018)
18. Lucchi A, Li Y, Smith K, Fua P: Structured image segmentation using kernelized features In: ECCV. Springer (2012)
19. Lucchi A, Márquez-Neila P, Becker C, Li Y, Smith K, Knott G, Fua P: Learning structured models for segmentation of 2-d and 3-d imagery. TMI 34(5), 1096–1110 (2014)
20. Lucchi A, Smith K, Achanta R, Knott G, Fua P: Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. TMI 31(2), 474–486 (2011)
21. Meirovitch Y, Mi L, Saribekyan H, Matveev A, Rolnick D, Shavit N: Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics In: CVPR. IEEE (2019)
22. Motta A, Berning M, Boergens KM, Staffler B, Beining M, Loomba S, Hennig P, Wissler H, Helmstaedter M: Dense connectomic reconstruction in layer 4 of the somatosensory cortex. Science 366(6469) (2019)
23. Nunez-Iglesias J, Kennedy R, Parag T, Shi J, Chklovskii DB: Machine learning of hierarchical clustering to segment 2d and 3d images. PloS one (2013)
24. Oztel I, Yolcu G, Ersoy I, White T, Bunyak F: Mitochondria segmentation in electron microscopy volumes using deep convolutional neural network. In: Bioinformatics and Biomedicine (2017)
25. Ronneberger O, Fischer P, Brox T: U-net: Convolutional networks for biomedical image segmentation In: MICCAI. pp. 234–241. Springer (2015)
26. Schubert PJ, Dorkenwald S, Januszewski M, Jain V, Kornfeld J: Learning cellular morphology with neural networks. Nature communications (2019)
27. Smith K, Carleton A, Lepetit V: Fast ray features for learning irregular shapes In: ICCV. IEEE (2009)
28. Turaga SC, Briggman KL, Helmstaedter M, Denk W, Seung HS: Maximin affinity learning of image segmentation. In: NeurIPS. pp. 1865–1873 (2009)
29. Vazquez-Reina A, Gelbart M, Huang D, Lichtman J, Miller E, Pfister H: Segmentation fusion for connectomics In: ICCV. IEEE (2011)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

30. Xiao C, Chen X, Li W, Li L, Wang L, Xie Q, Han H: Automatic mitochondria segmentation for em data using a 3d supervised convolutional network. Frontiers in neuroanatomy 12, 92 (2018) [PubMed: 30450040]
31. Xu N, Yang L, Fan Y, Yue D, Liang Y, Yang J, Huang T: Youtube-vos: A large-scale video object segmentation benchmark In: ECCV. Springer (2018)
32. Xu Y, Li Y, Wang Y, Liu M, Fan Y, Lai M, Eric I, Chang C: Gland instance segmentation using deep multichannel neural networks. Transactions on Biomedical Engineering 64(12), 2901–2912 (2017)
33. Yan Z, Yang X, Cheng KTT: A deep model with shape-preserving loss for gland instance segmentation In: MICCAI. pp. 138–146. Springer (2018)
34. Zeng T, Wu B, Ji S: Deepem3d: approaching human-level performance on 3d anisotropic em image segmentation. Bioinformatics 33(16), 2555–2562 (2017) [PubMed: 28379412]
35. Zeviani M, Di Donato S: Mitochondrial disorders. Brain 127(10) (2004)
36. Zhang L, Trushin S, Christensen TA, Bachmeier BV, Gateno B, Schroeder A, Yao J, Itoh K, Sesaki H, Poon WW, Gylys K: Altered brain energetics induces mitochondrial fission arrest in alzheimer's disease. Scientific reports 6, 18725 (2016) [PubMed: 26729583]
37. Zlateski A, Seung HS: Image segmentation by size-dependent single linkage clustering of a watershed basin graph. arXiv:1505.00249 (2015)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

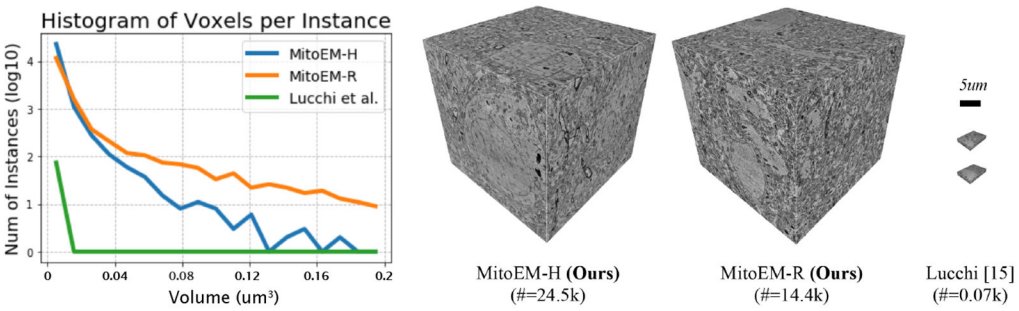

Fig. 1:

Comparison of mitochondria segmentation datasets. (Left) Distribution of instance sizes. (Right) 3D image volumes of our MitoEM and Lucchi [20]. Our MitoEM dataset has greater diversity in image appearance and instance sizes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

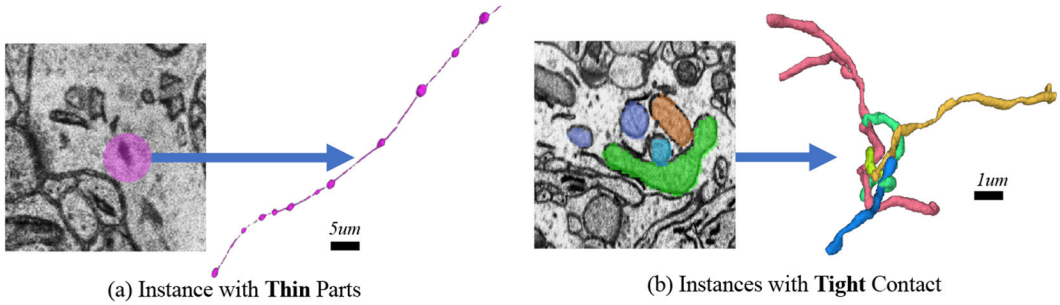

Fig. 2:

Complex mitochondria in our MitoEM dataset: **(a)** mitochondria-on-a-string (MOAS) [36], and **(b)** dense tangle of touching mitochondria. Those challenging cases are prevalent but not covered by existing labeled datasets.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

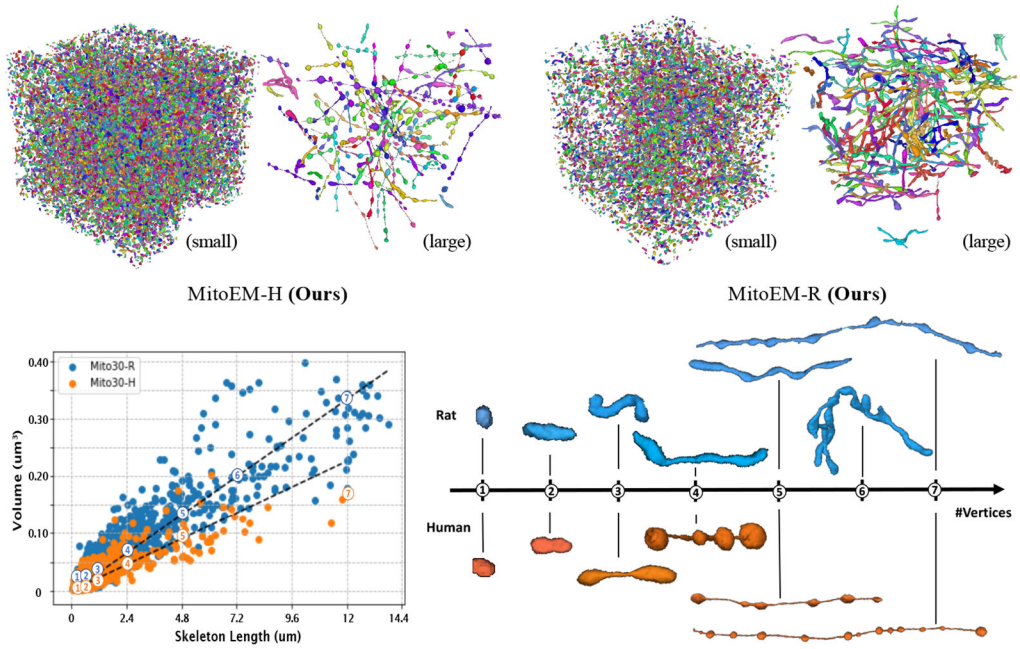

Fig. 3:

Visualization of MitoEM-H and MitoEM-R datasets. (Top) 3D meshes of small and large mitochondria, where MitoEM-R has a higher presence of large mitochondria; (Bottom left) scatter plot of mitochondria by their skeleton length and volume; (Bottom right) 3D meshes of the mitochondria at the sampled positions.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

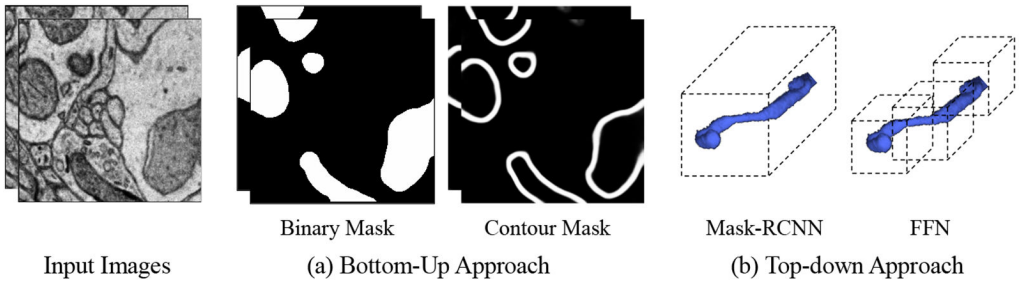

Fig. 4:
Instance segmentation methods in two types: bottom-up and top-down.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

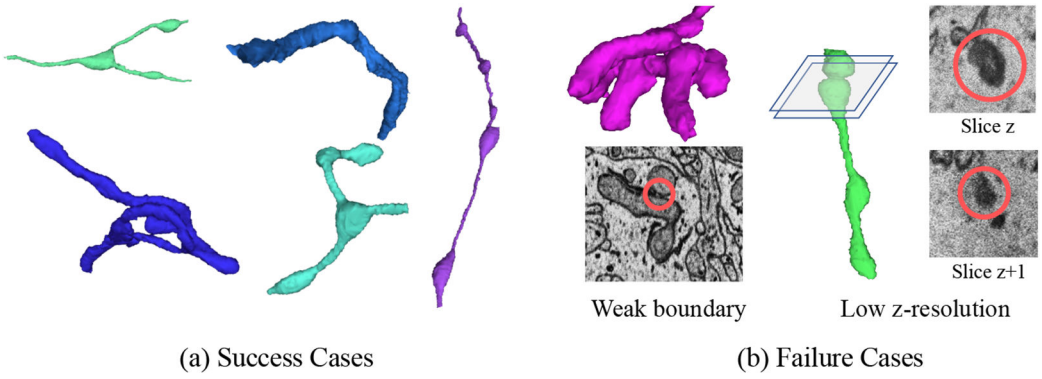

Fig. 5:

Qualitative results on MitoEM. (a) The U3D-BC+MW method can capture complex mitochondria morphology. (b) Failure cases are resulted from ambiguous touching boundaries and highly overlapping cross sections.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2020 December 03.

Table 1:

Mitochondria Segmentation Results on Lucchi Dataset.

We show results for (a) previous semantic segmentation methods, (b) a top-down, and (c) bottom-up approaches with different instance decoding methods.

(a) Previous approaches

Method	Jaccard↑	AP-75↑
CNN+post [24]	0.907	N/A
Working Set [19]	0.895	N/A
U3D-B [4]	0.889	N/A
GT+dilation-1	0.885	0.881
GT+erosion-1	0.904	0.894

(b) Top-down approaches

Method	Jaccard↑	AP-75↑
FFN[12]	0.554	0.230

(c) Bottom-up approaches

	Method	Jaccard↑	AP-75↑
U3D-A	+waterz [9]	0.877	0.802
	+zwatershed [15]		0.801
U2D-B	+CC [25]	0.882	0.760
	+MC [2]		0.521
U3D-B	+CC [5]	0.881	0.769
	+IoU [30]		0.770
	+MW		0.770
U3D-BC	+CC [3]	0.887	0.770
	+IoU		0.771
	+MW		**0.812**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2:

Main benchmark results on the MitoEM dataset.

We compare state-of-the-art methods on the MitoEM dataset using AP-75. Following MS-COCO evaluation [16], we report the results for instances of different sizes.

Method		MitoEM-H				MitoEM-R			
		Small	**Med**	**Large**	**All**	**Small**	**Med**	**Large**	**All**
U3D-A	+zwatershed [37]	**0.564**	0.774	0.615	**0.617**	**0.408**	0.235	**0.653**	0.328
	+waterz [9]	0.454	0.763	0.628	0.572	0.324	0.149	0.539	0.294
U2D-B	+CC [25]	0.408	0.814	**0.711**	0.597	0.104	0.628	0.481	0.355
U3D-B	+CC [5]	0.109	0.497	0.437	0.271	0.017	0.390	0.275	0.208
	+MW	0.439	0.794	0.567	0.561	0.254	0.692	0.397	0.447
U3D-BC	+CC [3]	0.480	0.801	0.611	0.594	0.187	0.551	0.402	0.397
	+MW	0.489	**0.820**	0.618	0.605	0.290	**0.751**	0.490	**0.521**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3:

Cross-tissue evaluation on MitoEM.

The U3D-BC model trained on rat (R model) is tested on human (MitoEM-H), and vice versa. R model generalizes better as the MitoEM-R dataset has higher diversity and complexity.

Method		MitoEM-H (R model)				MitoEM-R (H model)			
		Small	Med	Large	All	Small	Med	Large	All
U3D-BC	+CC [3]	0.533	0.833	0.664	0.650	0.218	0.640	0.354	0.407
	+MW	**0.587**	**0.862**	**0.669**	**0.690**	**0.224**	**0.674**	**0.359**	**0.411**

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2020 December 03.