Updated Baselines on MitoEM Dataset for Mitochondria Instance Segmentation

Abstract. Electron microscopy (EM) allows identifying intracellular organelles such as mitochondria, providing insights for clinical and scientific studies. We introduced the *MitoEM* dataset, a 3D mitochondria instance segmentation dataset with two $(30\mu\text{m})^3$ volumes from human and rat cortices, respectively, $3,600\times$ larger than previous benchmarks. This note shows the updated baseline results on the MitoEM dataset using improved data augmentation and model optimization techniques. Code publicly available at `https://github.com/zudi-lin/pytorch_connectomics/tree/master/configs/MitoEM`.

Keywords: Mitochondria · EM Dataset · 3D Instance Segmentation.

1 Introduction

In this note, we show the updated baseline results demonstrated in the MitoEM challenge (ISBI 2021) [1], where the U3D-BC model achieves an AP-75 score of 0.81 on the two large-scale ($(30\ \mu m)^3$) test volumes. There are two major changes in the experiment settings compared to the MICCAI 2020 paper [7]:

1. Annotation Update. We conducted another round of proofreading and finalized the annotation for the public challenge. We removed many small-sized mitochondria-like segments and corrected some mitochondria-on-a-string (MOAS) segments.
2. Model Update. For the best-performing baseline method, U3D-BC, we used our *Pytorch Conncetomics* package [5]. After the paper submission, we enhanced the package with more advanced data augmentation methods and better learning rate schedulers, which significantly improved the MitoEM dataset results.

2 Details on the Model Update

Additional Augmentations. Besides the brightness, flip, elastic transform, and missing parts augmentations used in the original MitoEM paper [7], we added misalignment, CutBlur [8], CutNoise, and motion-blur augmentations to simulate more kinds of data artifacts. We also increased the applying probability and intensity of all augmentations to improve the robustness of the model.

Optimization. Despite the faster training speed of ADAM optimizer [4], recent work [9] has shown that ADAM-alike adaptive optimization algorithms do not

Table 1: **Effects of the Annotation Update.** We report the results for U2D-B+CC on both the original [7] and the updated challenge dataset [1] using the same experiment settings. With the improvement of the annotation, the method has around 5% absolute performance drop in terms of AP-75.

Method		MitoEM-H				MitoEM-R				Overall
		Small	Med	Large	All	Small	Med	Large	All	
U2D-B+CC	v1 [7]	0.408	0.814	0.711	0.597	0.104	0.628	0.481	0.355	0.476
	v2 [1]	0.272	0.636	0.493	0.534	0.133	0.523	0.289	0.330	0.432

Table 2: **Effects of the annotation update and model update.** We report the U3D-BC+MW results on both the original dataset [7] with the old implementation and the updated challenge dataset [1] with the updated implementation [5]. With the improved data augmentations and model optimization techniques, the U3D-BC method has around 25% absolute performance gain in terms of AP-75.

Method		MitoEM-H				MitoEM-R				Overall
		Small	Med	Large	All	Small	Med	Large	All	
U3D-BC+MW	v1 [7]	0.489	0.820	0.618	0.605	0.290	0.751	0.490	0.521	0.563
	v2 [1]	0.426	0.838	0.798	0.804	0.311	0.845	0.803	0.816	**0.810**

generalize as well as SGD [6,2]. Therefore in the improved model, we switch the optimizer to SGD and use the *cosine learning rate decay* policy [3] to update the learning rate. The number of training iterations and the initial learning rate are not adjusted.

3 Results

The results indicate that the predicted binary mask with a 2D UNet model can hardly handle the 3D mitochondria segmentation task (Table 1). On the other hand, with more advanced data augmentation techniques and better optimization strategies, the 3D UNet model, which predicts both the binary foreground mask and instance contour map, can be significantly improved and achieves an AP-75 score of 0.81[1] (Table 2).

4 Future Directions

Despite the improved baseline results, the following two challenges remain for the community to tackle.

[1] `https://mitoem.grand-challenge.org/evaluation/challenge/leaderboard/`

1. Model Challenge: In the current full-supervised setting, *i.e.*, 40-10-50 data split, our updated baseline method U3D-BC serves as a strong baseline, achieving a 0.81 AP-75 score. However, for practical large-scale deployment, the instance segmentation methods need to achieve 0.9 (AP-75) to make the saturated proofreading feasible. In addition to the challenges mentioned in the paper, *e.g.*complex geometries and crowded instances, we find current methods do not perform well on the "small" segments as there are many objects that look like mitochondria.
2. Low data-regime Challenge: In practice, the annotation budget is often around 5-10% of the whole volume. Therefore, it is critical to develop data-efficient methods, *e.g.*, unsupervised, semi-supervised, active learning methods, that can achieve 0.9 (AP-75) with a limited amount of annotation. Our MitoEM dataset can be used to simulate those learning settings.

References

1. Mitoem challenge: Large-scale 3d mitochondria instance segmentation, `https://mitoem.grand-challenge.org/` 1, 2
2. Bottou, L.: Stochastic gradient learning in neural networks. Proceedings of Neuro-Nımes **91**(8), 12 (1991) 2
3. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 558–567 (2019) 2
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 1
5. Lin, Z., Wei, D.: Pytorch connectomics. `https://github.com/zudi-lin/pytorch_connectomics` (2019) 1, 2
6. Robbins, H., Monro, S.: A stochastic approximation method. The annals of mathematical statistics pp. 400–407 (1951) 2
7. Wei, D., Lin, Z., Franco-Barranco, D., Wendt, N., Liu, X., Yin, W., Huang, X., Gupta, A., Jang, W.D., Wang, X., et al.: Mitoem dataset: Large-scale 3d mitochondria instance segmentation from em images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 66–76. Springer (2020) 1, 2
8. Yoo, J., Ahn, N., Sohn, K.A.: Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8375–8384 (2020) 1
9. Zhou, P., Feng, J., Ma, C., Xiong, C., HOI, S., et al.: Towards theoretically understanding why sgd generalizes better than adam in deep learning. arXiv preprint arXiv:2010.05627 (2020) 1