Maintenance and enhancement of Funke lab's automated reconstruction pipeline for volumetric Electron Microscopy

Samia Mohinta, Albert Cardona

Jan Funke's lab at HHMI Janelia has developed a set of algorithms and software tools that, together, form a deep-learning based ecosystem for electron microscopy (EM) connectomics. Funke lab's tools achieve state-of-the-art results even with constrained computational resources, which make these tools extremely attractive for application at scale to ever-growing EM volumes. We are using the Funke lab's tool set to automate a connectomic neural circuit mapping workflow (segment neurons and microtubules, detect synapses, and infer neurotransmitters) for our multiple terabyte to petabyte scale nanometer-resolution EM volumes at the University of Cambridge and at the MRC LMB. However, the tools of this fabulous ecosystem require intimate knowledge of its constituent software frameworks and libraries, which makes harder than it should be to reproduce the setup of the pipelines, some of which are not actively maintained by the authors, and therefore have not yet achieved the desirable 'off-the-shelf' status that these magnificent and efficient software tools deserve.

We recognise the extraordinary value of the multi-year effort that the Funke lab has put into creating these important models that overcome the critical Machine Learning (ML) challenge of robust instance segmentation and detection with *Drosophila* EM datasets. Therefore, we here request support for maintaining, improving, integrating, documenting and distributing all relevant software packages from the Funke lab.

While we have started bringing some of these 4 software tools up to date with modern libraries and reproducible environments such as docker for our internal use, we aim to (1) complete the update of all of them, for their biggest impact is in using them together; and (2) share their updated off-the-shelf versions with the global connectomics and image processing community, so that the outstanding work from the Funke lab becomes both a standard toolkit in biology labs and a foundation for others to develop upon.

1. Local Shape Descriptors for Neuron Segmentation (Sheridan et al. 2022)

Local shape descriptors (LSD) is an affinity-based fully-convolutional (UNet) segmentation method, which leverages the prediction of local features in relation to the shape of the neuron to be segmented. The proposed algorithm achieves segmentation accuracy en par with Google's state-of-the-art Flood Filling Networks, while being two orders of magnitude faster in processing large-scale volumetric datasets, a key requirement to enable segmentation of future peta-scale volumes. This makes it suitable for use in small academic labs, who are generating a lot of data but lack access to the lavish computational resources of large technology companies to run complex and large models.

Volumetric segmentation models with LSD are presently implemented using a prior version of TensorFlow (1.x). However, the documentation has been expanded to include working tutorials of 2D networks (and one vanilla 3D UNet) in PyTorch, to be referenced for PyTorch-based 3D reimplementation. Additionally, the software has been updated for extended applicability on other types of datasets (e.g., non-EM). Still, both legacy TensorFlow and PyTorch LSD packages may run into compatibility issues if they are not continuously upgraded or frozen in time, due to the future evolution of the libraries.

We propose to (A) offer containerized versions of the 'development' and 'production' -level software versions to both future-proof and facilitate installation and execution without running into compatibility issues, and (B) further expand the documentation to cover their deployment across operating systems and hardwares.

2. Synful: Automatic Detection of Synaptic Partners in a Whole-Brain *Drosophila* EM Dataset (Buhmann et al. 2020)

Synful automatically detects pre- and postsynaptic sites with a full-convolutional neural network (UNet), no longer requiring neuron segmentations, clefts and vesicles as input but only synaptic point annotations. This model has been used to detect 244 million putative synaptic sites in the Female Adult Fly Brain (FAFB; Buhmann et al. 2021), while achieving 92% and 96% accuracy in identifying weakly-connected and strongly-connected neurons respectively.

Synful's present implementation targets synapse detection in anisotropic volumes and is also implemented using library versions that aren't easily available and have fallen out of date.

We propose to both (A) generalize the implementation to isotropic volumes and (B) modernize its library dependencies. We will then offer (C) a containerized version of the software for trivial deployment and which guarantees its execution independently of evolving software libraries or operating systems.

3. Synister: Neurotransmitter Classification from Electron Microscopy Images at Synaptic Sites in *Drosophila* (Eckstein, Bates, et al. 2020)

Synister uses a feedforward VGG network to predict neurotransmitters from the phenotype of synapses in *Drosophila* EM volumes. It overcomes the challenge of mapping Light Microscopy gene expression to EM neuron reconstructions, with large savings in time and resources. Instead, VGG can ascertain neurotransmitter identity on EM-imaged synaptic sites with 87% accuracy within a matter of seconds, and 99% when considering whole neurons, therefore serving as an extremely important tool to easily add further information to the reconstructed connectomes, especially of invertebrates like *Drosophila melanogaster*.

Presently, library dependencies of Synister are rapidly aging.

We propose to (A) update the library dependencies; (B) ensure the code is usable with datasets other than the anisotropic FAFB volume, which includes generalizing to isotropic volumes such as imaged with FIB-SEM; and (C) offer a containerized version for easy deployment with guaranteed execution in the future.

4. Micron: Microtubule Tracking in Electron Microscopy Volumes (Eckstein, Buhmann, et al. 2020)

Micron couples a novel integer linear programming (ILP) method with a 3D UNet to solve a constrained optimization problem for microtubule tracking in EM volumes. It achieves a 3X speed-up and an increase in 53% accuracy over prior efforts. Efficient tracking of microtubules can both identify neurons and improve proof-reading times, with the labeling of neural backbones and twigs, essential to construct an enriched wiring diagram (Schneider-Mizell et al. 2016).

Micron depends on a legacy TensorFlow library (1.x).

We propose to (A) update the library dependencies and (B) offer a containerized version for easy deployment.

Scope of the project

Goal 1: Maintenance of Jan Funke's automated volumetric EM reconstruction software ecosystem

1. LSD, Synful and Micron were developed in TensorFlow 1.x, which lacks eager execution and has a steep learning curve. Most of the Machine Learning (ML) community have moved on to newer frameworks like PyTorch, which offer eager execution, better programming flexibility and debugging,

optimal data parallelism for scaling models in production environments. As a result, we propose to completely port these ML models into PyTorch to make it suitable for future use by a wider audience. For LSD, we propose to extend the existing PyTorch-based network examples to include training and validation cycles in the same script for 3D volumes.

2. Synister uses PyTorch, however, several dependencies of the code on external libraries have not been clearly recorded, which makes it difficult to reproduce the network performance even on the suggested public datasets (CREMI). We shall be resolving these compatibility issues, fixing issues with edge cases and maintaining it for both backwards and forward compatibility.

3. Additionally, we shall expand existing documentation to not only explain code snippets but also design/logical choices (unclear or unavailable for the current source codes) for any user.

4. We aim to package all up-to-date software into shareable containers using platforms such as Docker and Apptainer (previously Singularity). This will also ensure training models at scale using Kubernetes to orchestrate dockers, without having to worry about hardware and operating systems.

Goal 2: Enhancements and new features

1. Funkelab's ML ecosystem heavily relies on Gunpowder, a data loading and augmentation pipeline. Currently, we cannot integrate any other augmentation library (e.g Kornia; supports GPU-based augmentation) into Gunpowder, without explicitly writing a wrapper 'node' for it. It would be beneficial to add this feature to Gunpowder to overcome redundancy and ensure faster development.

2. All of the above ML models were generally trained on a single GPU. Given the amount of the data being generated across labs on a daily basis, it is now pertinent to add multi-GPU training (distributed training) and inference capabilities to achieve quicker gains.

3. We shall also endeavor to refactor these softwares to integrate with LabKit in Fiji and BioImage Model Zoo in Ilastik as pretrained models so as to make them accessible to a broader image processing community.

Milestones

At the end of the first year, we plan to have a working and shareable version of Synful, which include their complete transition to PyTorch, fixed corner cases and compatibility issues, and necessary improvements to the Gunpowder software framework for faster training and inference. We shall also containerize the LSD software to ensure future runnability and expand the documentation to cover their deployment. By the end of second year, we plan to have Synister and Micron up and running, while simultaneously delivering on the integration of these models into popular image processing frameworks such as Fiji and Ilastik.

References

Buhmann, Julia, Arlo Sheridan, Stephan Gerhard, Renate Krause, Tri Nguyen, Larissa Heinrich, Philipp Schlegel, et al. 2020. "Automatic Detection of Synaptic Partners in a Whole-Brain Drosophila EM Dataset." *bioRxiv*. https://doi.org/10.1101/2019.12.12.874172.

Eckstein, Nils, Alexander S. Bates, Michelle Du, Volker Hartenstein, Gregory S. X. Jefferis, and Jan Funke. 2020. "Neurotransmitter Classification from Electron Microscopy Images at Synaptic Sites in Drosophila." *bioRxiv*. https://doi.org/10.1101/2020.06.12.148775.

Eckstein, Nils, Julia Buhmann, Matthew Cook, and Jan Funke. 2020. "Microtubule Tracking in Electron Microscopy Volumes." In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 99–108. Springer International Publishing.

Schneider-Mizell, Casey M., Stephan Gerhard, Mark Longair, Tom Kazimiers, Feng Li, Maarten F. Zwart,

Andrew Champion, et al. 2016. “Quantitative Neuroanatomy for Connectomics in Drosophila.” *eLife* 5 (March). https://doi.org/10.7554/eLife.12059.

Sheridan, Arlo, Tri Nguyen, Diptodip Deb, Wei-Chung Allen Lee, Stephan Saalfeld, Srini Turaga, Uri Manor, and Jan Funke. 2022. “Local Shape Descriptors for Neuron Segmentation.” *bioRxiv*. https://doi.org/10.1101/2021.01.18.427039.