BrainLine: An Open Pipeline for Connectivity Analysis of Heterogeneous Whole-Brain Fluorescence Volumes

Thomas L. Athey*[1,2], Matthew A. Wright[3], Marija Pavlovic[4], Vikram Chandrashekhar[1], Karl Deisseroth[4,5,6,7], Michael I. Miller[1,2,8,9], and Joshua T. Vogelstein[1,2,8,9]

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[2]Institute of Computational Medicine, Johns Hopkins University, Baltimore, MD, USA
[3]Department of Psychiatry, Weill Cornell Medicine, New York, NY, USA
[4]Department of Bioengineering, Stanford University, Stanford, CA, USA
[5]CNC Program, Stanford University, Stanford, CA, USA
[6]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA
[7]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA
[8]Center for Imaging Science, Johns Hopkins University, Baltimore, MD, USA
[9]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD, USA

Whole-brain fluorescence images require several stages of computational processing to fully reveal the neuron morphology and connectivity information they contain. However, these computational tools are rarely part of an integrated pipeline. Here we present BrainLine, an open-source pipeline that interfaces with existing software to provide registration, axon segmentation, soma detection, visualization and analysis of results. By implementing a feedback based training paradigm with BrainLine, we were able to use a single learning algorithm to accurately process a diverse set of brain images generated by light-sheet microscopy. BrainLine is available as part of our Python package brainlit: http://brainlit.neurodata.io/.

Main

Whole-brain image volumes at the micron scale are helping scientists characterize neuron-level morphology and connectivity, and discover new neuronal subtypes. These volumes require intense computational processing to uncover the rich neuronal information they contain. Currently, however, image acquisition is outstripping the availability and throughput of analysis pipelines. The steps in analyzing these images include registration, axon segmentation, soma detection, visualization and analysis of results. Several tools exist for these individual steps, but are rarely all part of an integrated pipeline and able to facilitate cloud-based collaboration [7, 9]. Further, many existing machine learning based tools are highly tuned to their training data and perform poorly when they encounter out-of-distribution artifacts or signal levels [6].

To address these challenges, we present BrainLine, an open-source, fully-integrated pipeline that performs registration, axon segmentation, soma detection, visualization, and analysis on whole-brain fluorescence volumes (Figure 1a). BrainLine combines state-of-the-art, already available open-source tools such as CloudReg [3] and ilastik [2] with brainlit, our Python package developed here. The BrainLine pipeline uses generalizable machine learning training schemes that adapt to out-of-distribution samples and facilitates cloud-based collaboration across institutions.

To share and interact with data across multiple institutions, BrainLine uses Amazon S3 to store data in precomputed format, so it can be viewed using Neuroglancer [1]. Specifically, we use CloudReg [3] for file conversion of the stitched image, and for image registration to the Allen atlas [10].

For axon segmentation and soma detection, we sought to leverage recent machine learning advances but experienced two major constraints. First, as generating ground truth image annotations is labor intensive, we wanted the approach to be effective on a small amount of training data. Second, images were provided to us in a sequential manner, and new samples would sometimes have unique artifacts or different levels of image quality (Figure 1b-c,e-f). We therefore sought a learning algorithm that could be quickly retrained on new data. Many learning algorithms assume that all training and testing data come from the same distribution and fail when this is not the case [8]. However, using our closed-loop training paradigm with ilastik [2], we were able to use a single ilastik project for all samples, only occasionally adding training data when difficult samples arose.

*tathey1@jhu.edu

We used an ilastik pixel classification workflow for both axon segmentation and soma detection, but in the latter case we applied a size threshold to the connected components following segmentation. In both cases, the training approach was the same. For each new whole-brain volume, we identified a set of subvolumes (99^3 voxels for axons, 49^3 for somas) across a variety of brain regions, and annotated only a few slices (three for axons, five for somas) in each subvolume for our validation set. This strategy is similar to that employed in Friedmann et al. [5]. If our model could not achieve a satisfactory f-score on this validation dataset, we would annotate more subvolumes from the sample and add them to the training set until satisfactory performance was achieved.

We observed that this heterogeneous training procedure (i.e. training on multiple brain samples) often improved performance on other samples as well. In an experiment where we controlled the number of subvolumes used for training, this approach was at least as good as a homogeneous approach, where all training subvolumes came from a single brain sample (Figure 1d,g).

The pipeline can display the axon segmentation and soma detection results in a variety of ways, including brain-region-based bar charts accompanied by statistical tests (Fig. 1a.i), 2D plots with the atlas borders (Fig. 1a.ii), and 3D visualizations using brainrender (Fig. 1a.iii) [4]. Since every experimental design is unique, we designed our pipeline in a modular way, so investigators can pick and choose which components they want to incorporate in their own analyses. We also leverage existing software and file formats to facilitate interoperability [9].

BrainLine enables accelerated analysis of brain-wide connectivity through parallel programming, the use of cloud-compliant file formats, and a machine learning training scheme that generalizes across brain samples. As a result, BrainLine has the potential to accelerate the characterization of morphology and connectivity profiles of neurons, adding in the discovery of new neuronal subtypes, and more. BrainLine is immediately applicable to future neuroscience investigations, enabling generalizability and cloud-based collaboration, rather than spending painstaking time and effort building custom analysis pipelines from scratch. BrainLine is available as a set of thoroughly documented notebooks and scripts in our Python package brainlit: `http://brainlit.neurodata.io/`.

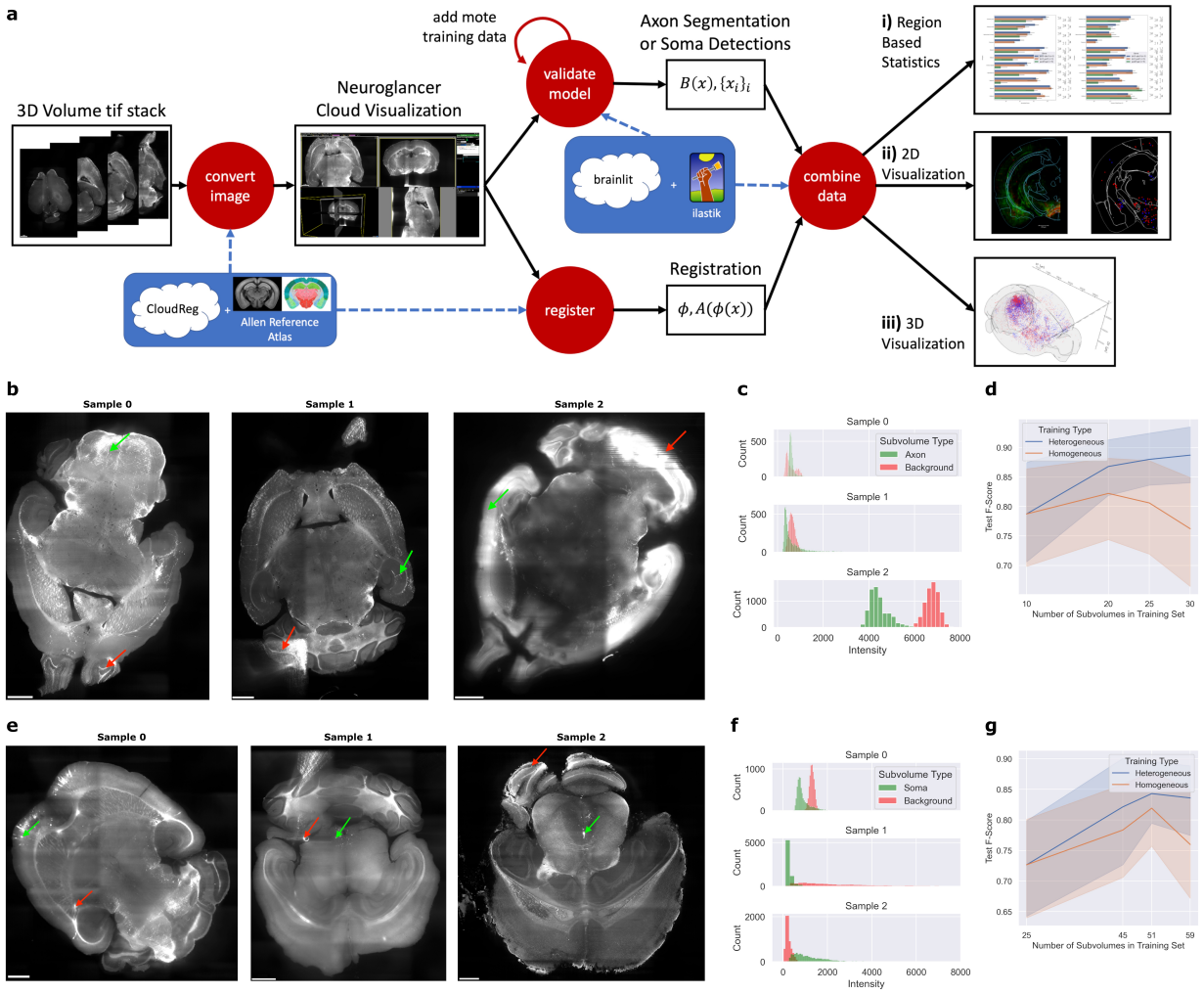


Figure 1: BrainLine allows for efficient processing of heterogeneous whole brain fluorescence volumes. **a** BrainLine combines CloudReg [3], ilastik [2] and our package, brainlit, to produce results in both quantitative (**a.i**) and visual (**a.ii-a.iii**) formats. **b** Example images with fluorescently labeled axon projections and arrows pointing to regions with (green) and without (red) labeled axons. **c** Intensity histograms of 20x20x20 voxel subvolumes located at the arrows in **b**. **d** Comparison between axon segmentation performance after training on subvolumes from different samples (heterogeneous) or the same sample (homogeneous). **e** Example images with fluorescently labeled cell bodies and arrows pointing to regions with (green) and without (red) labeled cell bodies. **f** Intensity histograms of 20x20x20 voxel subvolumes located at the arrows in **e**. **g** Comparison between soma detection performance after training on subvolumes from different brain samples (heterogeneous) or a single brain sample (homogeneous).

References and Notes

[1] Neuroglancer. URL https://github.com/google/neuroglancer. 1

[2] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N Straehle, Bernhard X Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, et al. Ilastik: interactive machine learning for (bio) image analysis. *Nature methods*, 16(12):1226–1232, 2019. 1, 3

[3] Vikram Chandrashekhar, Daniel J Tward, Devin Crowley, Ailey K Crow, Matthew A Wright, Brian Y Hsueh, Felicity Gore, Timothy A Machado, Audrey Branch, Jared S Rosenblum, et al. Cloudreg: automatic terabyte-scale cross-modal brain volume registration. *Nature methods*, 18(8):845–846, 2021. 1, 3

[4] Federico Claudi, Adam L Tyson, Luigi Petrucco, Troy W Margrie, Ruben Portugues, and Tiago Branco. Visualizing anatomically registered data with brainrender. *Elife*, 10:e65751, 2021. 2

[5] Drew Friedmann, Albert Pun, Eliza L Adams, Jan H Lui, Justus M Kebschull, Sophie M Grutzner, Caitlin Castagnola, Marc Tessier-Lavigne, and Liqun Luo. Mapping mesoscale axonal projections in the mouse brain using a 3d convolutional network. *Proceedings of the National Academy of Sciences*, 117(20):11068–11075, 2020. 2

[6] Ali Geisa, Ronak Mehta, Hayden S Helm, Jayanta Dey, Eric Eaton, Jeffery Dick, Carey E Priebe, and Joshua T Vogelstein. Towards a theory of out-of-distribution learning. *arXiv preprint arXiv:2109.14501*, 2021. 1

[7] Thomas J Pisano, Zahra M Dhanerawala, Mikhail Kislin, Dariya Bakshinskaya, Esteban A Engel, Ethan J Hansen, Austin T Hoag, Junuk Lee, Nina L de Oude, Kannan Umadevi Venkataraju, et al. Homologous organization of cerebellar pathways to sensory, motor, and associative forebrain. *Cell reports*, 36(12):109721, 2021. 1

[8] Joaquin Quinonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. *Dataset shift in machine learning*. MIT Press, Boston, 2008. 1

[9] Adam L Tyson and Troy W Margrie. Mesoscale microscopy and image analysis tools for understanding the brain. *Progress in Biophysics and Molecular Biology*, 168:81–93, 2022. 1, 2

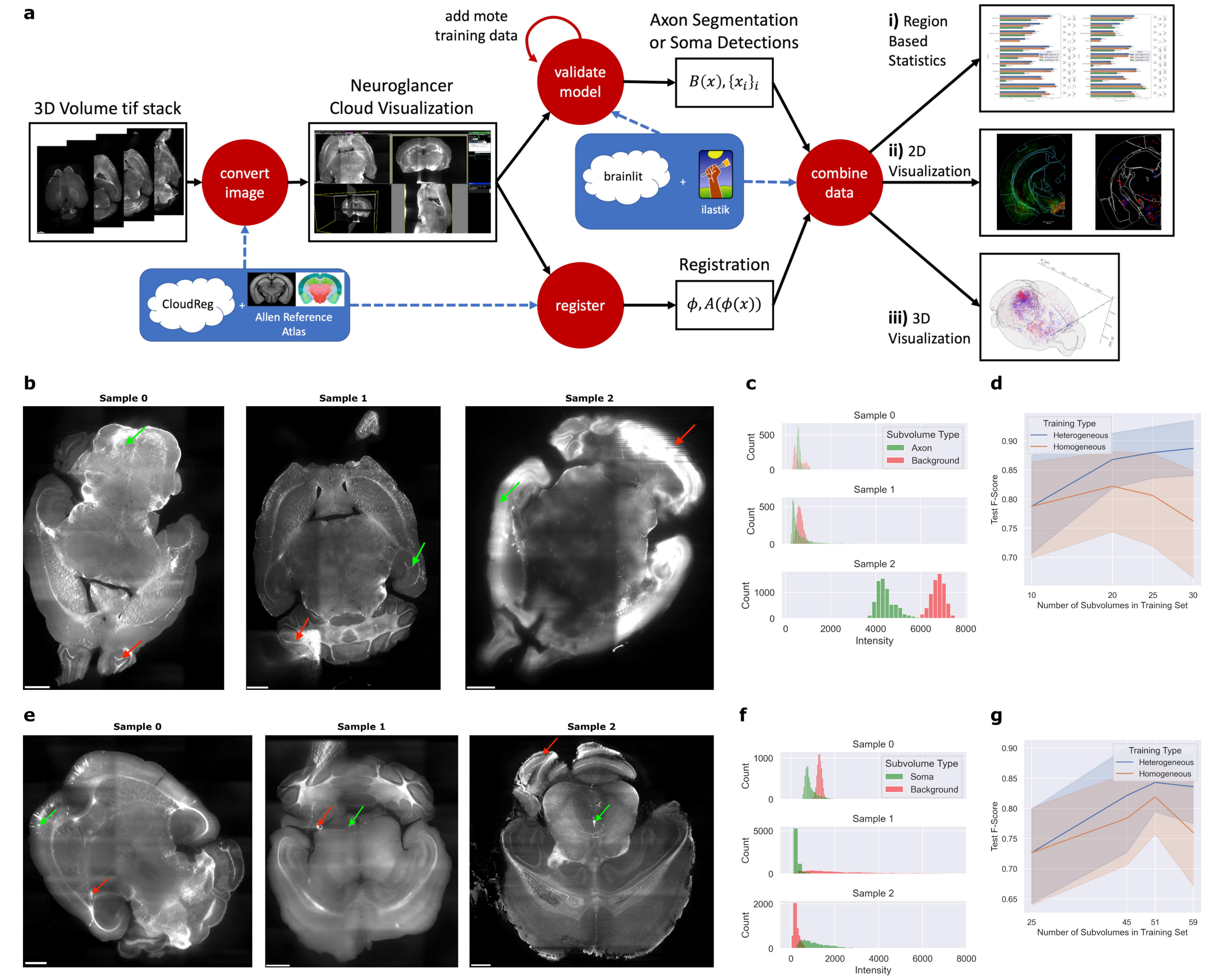
[10] Quanxin Wang, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, Maitham Naeemi, Benjamin Facer, Anh Ho, et al. The allen mouse brain common coordinate framework: a 3d reference atlas. *Cell*, 181(4):936–953, 2020. 1

Acknowledgements

This work is supported by NIH Grants RF1MH121539, U19AG033655, and RO1AG066184-01, NSF grants 2031985, 2014862 and the CAREER award. M.W. is supported by NIMH Grant K08MH113039. K.D. is supported by NIMH, NIDA, the NIH BRAIN Initiative, the Integrated Circuit Cracking NeuroNex Technology Hub funded by the National Science Foundation, the NOMIS Foundation, the Else Kröner Fresenius Foundation, the Gatsby Foundation and the AE Foundation.

Conflict of Interest Statement

M.I.M. owns a significant share of Anatomy Works with the arrangement being managed by Johns Hopkins University in accordance with its conflict of interest policies. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. V.C. owns a significant share of Neurosimplicity, LLC, which is a medical device and technology company focusing on medical image processing.


a
add mote training data
Axon Segmentation or Soma Detections
i) Region Based Statistics
3D Volume tif stack
Neuroglancer Cloud Visualization
validate model
$B(x), \{x_i\}_i$
convert image
brainlit
+
ilastik
combine data
ii) 2D Visualization
CloudReg
+
Allen Reference Atlas
register
Registration
$\phi, A(\phi(x))$
iii) 3D Visualization
b
Sample 0
Sample 1
Sample 2
c
Sample 0
Subvolume Type
Axon
Background
Count
Sample 1
Sample 2
Intensity
d
Training Type
Heterogeneous
Homogeneous
Test F-Score
Number of Subvolumes in Training Set
e
Sample 0
Sample 1
Sample 2
f
Sample 0
Subvolume Type
Soma
Background
Count
Sample 1
Sample 2
Intensity
g
Training Type
Heterogeneous
Homogeneous
Test F-Score
Number of Subvolumes in Training Set