

Visualizing and Interacting with Large Imaging Data

Eric Perlman¹

¹ Independent Researcher, Baltimore, MD, USA.

Increased access to high-throughput microscopy methods like light-sheet and high-throughput electron microscopy (EM) are leading to a data deluge. The large volume is often cumbersome to interact with due to the combination of ad-hoc storage methods and tools for visualization which often require high-end workstations with direct access to the data. With minimal modification to an existing systems, one can enable new ways to interact with this data through the combination of new open source visualization tools, intelligent file formats, and increasingly affordable network and cloud storage.

Neuroglancer [1] is one such visualization tool to enable new methods of data exploration. Neuroglancer is web-based tool for visualizing volumetric data and has gained significant use within the connectomics community [Figure 1a]. It facilitates interaction with remote data-sources of arbitrary sizes, and enables easy collaboration: any view can be shared by copying the URL from the web browser. (Other open-source tools can fit in this role of visualization front-end, notably BigDataViewer [2], a Java-based viewer for arbitrarily large image volumes within the ImageJ ecosystem, and the basis of Paintera [3], an annotation environment for neuron reconstruction in EM data.)

As an open-source project, Neuroglancer is highly extensible. Multiple groups have contributed new features to support additional scientific domains and modalities. For example, NeuroData [4,5] makes regular use of several extensions, including improved support for multichannel light microscopy through simple color selection and blending modes, the use of anatomical reference atlases [Figure 2a] to identify regions, and loading stored configuration data from a remote URL.

Neuroglancer provides a powerful python interface, enabling python code to control the entire interface, including the creation of virtual image and annotation layers. Geometric primitives can be defined programmatically and drawn onto imaging data to identify biological features [Figure 1b]. Combining this with the result of an automated analysis algorithm can provide an environment to qualitatively evaluate the results [Figure 2b]. Beyond visualization, extendibility through key bindings enables the construction of simple proof-reading environments with minimal amounts of code.

File formats play an important role in being able to access data efficiently by incorporating down-sampled image hierarchy and spatial partitioning techniques. The resulting files can be stored on local or remote cloud infrastructure. A number of tools exist to convert from common formats (e.g. TIFF stacks or hdf5) into an accessible format. Two examples include cloud-volume [6] for generating Neuroglancer-compatible data and N5utils [7] for data compatible with BigDataViewer-based tools. For larger environments and distributed applications a block based storage service such as BossDB [8] or DVID [9] can be an alternative to direct file access. Both services are supported by Neuroglancer and BigDataViewer and are actively used in research environments [12].

References:

- [1] Neuroglancer, <https://github.com/google/neuroglancer> (accessed Feb 9, 2019).
 [2] T Pietzsch et al., *Nature Methods*, **12** (2015), p. 481.
 [3] paintera, <https://github.com/saalfeldlab/paintera> (accessed Feb 19, 2019).
 [4] Vogelstein et al., *Nature Methods*, **15** (2018), p. 846.
 [5] NeuroData, <https://neurodata.io/> (accessed Feb 9, 2019).
 [6] cloud-volume, <https://github.com/seung-lab/cloud-volume> (accessed Feb 9, 2019).
 [7] n5-utils, <https://github.com/saalfeldlab/n5-utils> (accessed Feb 9, 2019).
 [8] BossDB, <https://bosddb.org> (accessed Feb 9, 2019).
 [9] DVID, WT Katz and SM Plaza, *Frontiers in Neural Circuits* **13** (2019), p. 5.
 [10] Zheng et al., *Cell*, **174-3** (2018), p. 730.
 [11] CREMI, <https://cremi.org> (accessed Feb 9, 2019).
 [12] The author acknowledges Vikram Chandrashekhar, Ailey Crow, Ben Falk and Brian Hsueh for their data, and the developers associated with the referenced software projects.

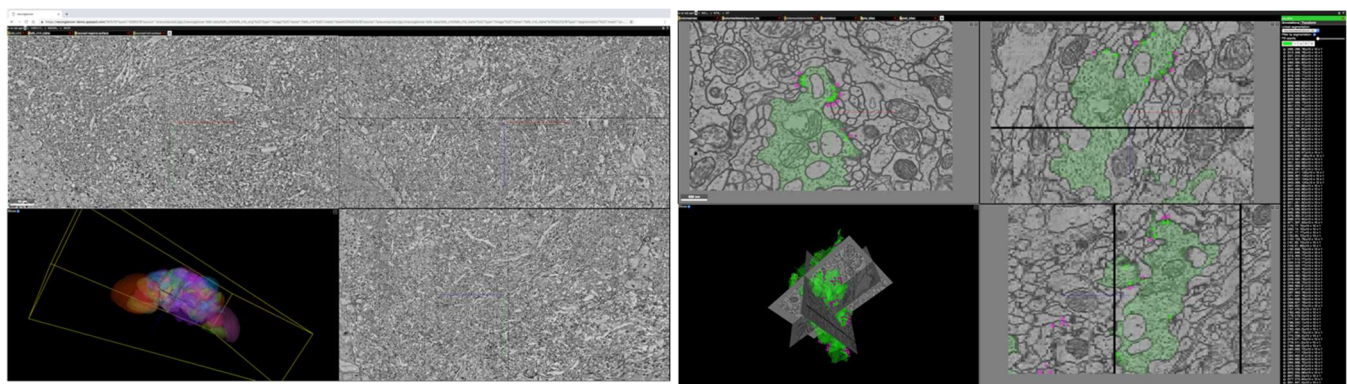


Figure 1. (a) Viewing the publicly-hosted FAFB dataset [10] (b) Visualization of one a CREMI [11] test volumes, backed by neuroglancer-python to load pre- and post-synaptic sites associated with a chosen neuron.

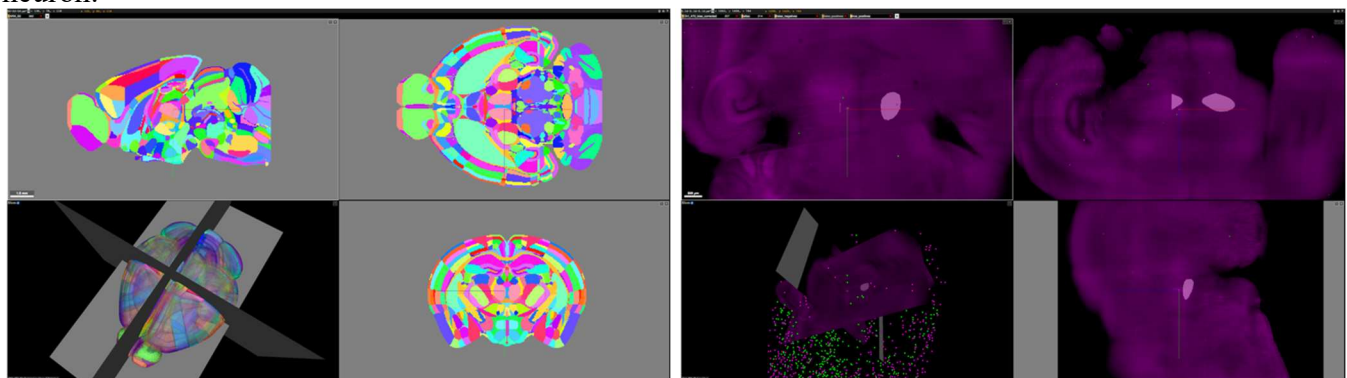


Figure 2. (a) Visualization of the Allen Brain Reference Adult Mouse Atlas (2011) (b) Visual evaluation of the results of a cell detection algorithm run on light sheet data, along with an aligned reference atlas to identify approximate brain regions. [Data courtesy B. Hsueh, A. Crow and K. Deisseroth with analysis by V. Chandrashekhar]