
Semi-Supervised Segmentation of Neurons From Brainbow Images

Uygar Sümbül

Grossman Center for the Statistics of Mind and Department of Statistics
Columbia University
uygar@stat.columbia.edu

Suraj Keshri

Department of Industrial Engineering and Operations Research
Columbia University

Min-hwan Oh

Department of Industrial Engineering and Operations Research
Columbia University

Dawen Cai

Department of Cell and Developmental Biology
University of Michigan Medical School

John P. Cunningham

Grossman Center for the Statistics of Mind and Department of Statistics
Columbia University

Liam Paninski

Grossman Center for the Statistics of Mind and Department of Statistics
Columbia University

1 Abstract

Understanding the organization of brain circuits is a basic goal in neuroscience. However, reconstructing neuroanatomy at the level of individual cells and on a large scale remains a challenge. Imaging neural tissues with electron microscopy and the ensuing reconstruction of the circuits have been used successfully on small tissues. Yet, this resource-intensive method faces scaling challenges in the imaging, storage and reconstruction steps.

Stochastic expression of multiple fluorescent proteins in individual cells of transgenic animals can provide similar information using light microscopy, where the role of the fluorescent proteins is to provide a color code for each neuron. This extra color information can resolve neighboring cells that would otherwise remain unresolved due to the diffraction limit. Ideally, individual cells express consistent and random combinations of colors throughout their anatomy. This “Brainbow” approach to neural circuit mapping offers high-throughput and the potential to monitor the circuits over time. While the approach has been successful in body parts where cellular anatomy is simpler, signal quality considerations have limited its use in the nervous system.

We propose a method to automate the segmentation of individual neurons from Brainbow tissues. Brainbow neurons express consistent colors and they are connected components in the morpholog-

ical sense. We formalize these notions using a graphical representation of the color-coded, three-dimensional image stack. However, the sheer size of the image stacks precludes a direct application where nodes represent individual voxels. Moreover, colors of individual voxels are unreliable due to low signal-to-noise ratio.

We first apply collaborative filtering to denoise the image stack. This approach creates an extra dimension in space by stacking similar volume patches, and applies low-pass filtering in this extra dimension to preserve the neuron-neuron and neuron-background interfaces. Next, we apply the watershed transform on the image of the absolute values of the spatial gradients to identify connected sets of voxels whose colors are homogeneous. These “supervoxels” represent the nodes of the graphical representation. We summarize a supervoxel by a single color and its spatial distances to other supervoxels. The edges of the graph represent the affinity of the supervoxels (i.e., the belief that the two nodes belong to the same neuron). We estimate these edge strengths using a simple exponential decay model in the spatio-color space: $A(i, j) \sim \max(1, \alpha e^{-\beta d_s(i, j)}) e^{-\theta d_c(i, j)}$, where $d_s(i, j)$ and $d_c(i, j)$ denote the spatial and color distances between the supervoxels i and j , respectively, (α, β, θ) are nonnegative parameters, and $A(i, j)$ denotes the affinity between i and j . We retain only a fraction of the entries of the resulting affinity matrix for storage and computational efficiency.

We apply spectral clustering to identify clusters of supervoxels that have homogeneous colors and form connected components (i.e., individual neurons). Neurons with brighter colors comprise supervoxels with higher signal-to-noise ratios and, consequently, are easier to segment. We also observe that supervoxels at the neuron-background boundary may appear dimmer than the rest of the supervoxels of the same neuron. Therefore, we propose a semi-supervised scheme where an initial oversegmentation of brighter supervoxels is used to generate *must-connect* and *can't-connect* sets to guide further segmentation.

We obtain the top few eigenvectors of the normalized graph Laplacian matrix once (e.g., using a power method). Then, we adapt the semi-supervised k-means clustering literature for our problem. Given the top few eigenvectors and an initial segmentation that does not violate the constraints, the algorithm iterates between a k-means step, a constraint satisfaction step, and a merging step, which merges potentially oversegmented neurons into single clusters, until convergence or the maximum number of iterations is reached. Note that merging two clusters decreases the cluster count by 1, and the corresponding feature vectors are simply shortened by 1 without the need to recalculate the eigenvectors.