Connectomics Machine Learning Competition

Andrew Warrington
&
Frank Wood

February 9, 2016

Contents

Introduction - I

- Last term : Chemical, Cutting, Recording
- This term : Data processing

Introduction - II

- Massive computational effort.
- Synapse detection elemental.
- Ultimately need connectivity.

Competition

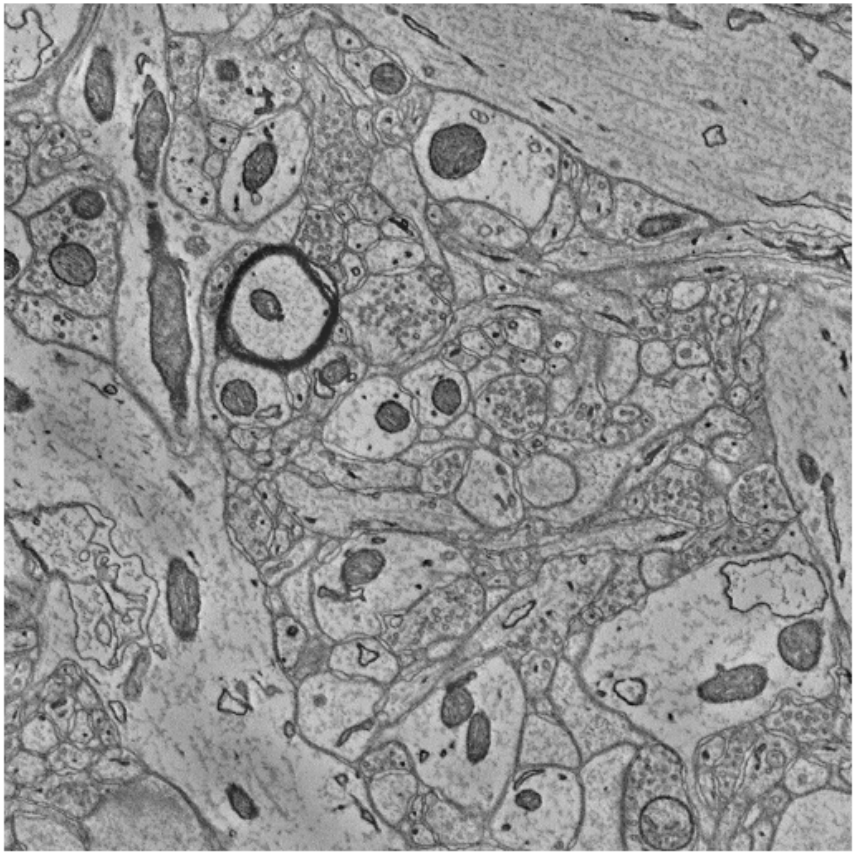
Synapse detection.

Contents

Data - I

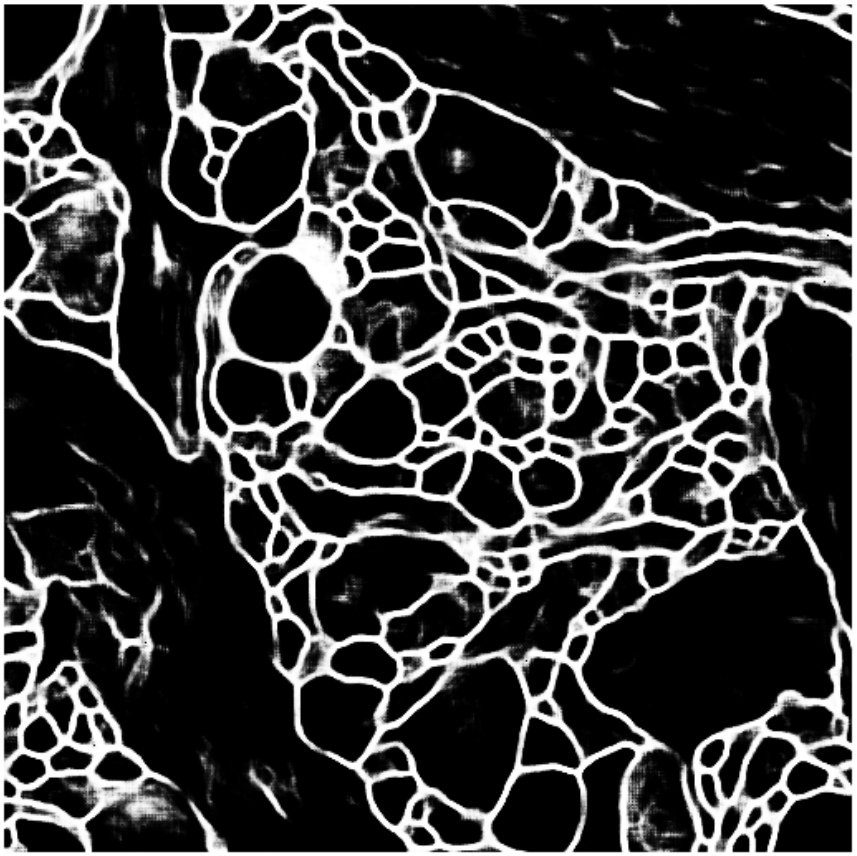
- High resolution electron microscopy data.
- Ground truth annotations have been made for small volumes [1, 2].

Data - II



- Raw volumetric data [1, 2]
- `submit:/scratch/3yp/train/image.tif`

Data - III



- Membrane detector output [1, 2]
- `submit:/scratch/3yp/train/membrane.tif`

Data - IV

- Vesicle detector output [1, 2]
- `submit:/scratch/3yp/train/vesicle.tiff`

Data - V

- Synapse detector output [1, 2]
- `submit:/scratch/3yp/train/synapse.tiff`

Data - VI

Provided

- Training volume $1024 \times 1024 \times 125$ voxels.
- Validation $1024 \times 1024 \times 50$.
- Test $1024 \times 1024 \times 75$ (final Competition).

`submit:/scratch/3yp/{train,validate,test}`

Contents

The Competition - I

- To create your own methodology and software to ingest raw images and predict which voxels are parts of synapses.
- Raw image, synapse, vesicle, and membrane data will be provided for the training set.
- Raw image and synapse data will be provided for the validation set. (use for development evaluation only).
- *Raw image data only* will be provided for the final test set (Last week of Hilary).

The Competition - II

- Training volume contains 148 synapses (synapses whose centroid is inside the volume).
- Validation volume contains 13 synapses.
- Test ?

Contents

Scoring I

A (voxel) classifier's results will include (given truth on across columns and classifier output down rows)

Table: Classifier output possibilities

	T	F
T	TP	FP
F	FN	TN

Scoring II

F1-Score [3]

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

with

$$precision = \frac{TP}{TP + FP}$$

and

$$recall = \frac{TP}{TP + FN}.$$

F1 can be computed as

$$F1 = \frac{2 \times TP}{2 \times TP + FN + FP}$$

which lies between 0 and 1.

Scoring III

Score

- F1-Score of synapse voxel classification
- matlab : `sos_evaluate_F1_tiff(ground,test)`

Can compute individually. Submitted result scores will be computed weekly and reported in class.

Contents

Setup - I

- Make sure you have
  - `scp`
  - `ssh`
- `ssh login.robots.ox.ac.uk`
- CHANGE PASSWORD!
- ssh submit
- Familiarize yourself with Torque and Maui, particularly submitting matlab jobs through a script.
  - `qsub`
  - `qstat -Q -f`

Setup - II

- Create a work directory in `/scratch`
- `cd` into it and clone `https://github.com/openconnectome/CAJAL`
- Clone `https://bitbucket.org/andrewwarrington/saving-oneself-algorithms`
- In matlab
  - `addpath` *saving-oneself-algorithms-dir*

Contents

Tools - I

- Lectures will be given introducing you to a range of techniques.
- Week 2 - Decision Forests - simple tool for classification.
- Week 3 - Neural Networks - supervised function approximation.
- Week 4/6 - Segmentation - separating whole image into distinct regions.
- Your solution may use any or all of the lectured material (or more!)

Tools - II

- `https://bitbucket.org/andrewwarrington/saving-oneself-algorithms` contains stock algorithms for data import and results processing.

Tools - III

- `sos_overlay_logical_labels(image, labels)` overlays non-zero labels over the image.
- `sos_stack_viewer(stack)` allows you to scroll through a stack using the mouse wheel when stack is a 3D grayscale matrix.
- `sos_cell_viewer(stack)` allows you to scroll through a stack using the mouse wheel when stack is a vector of cells containing RGB data.

Tools - IV

`sos_evaluate_F1_tiff(ground_path,predicted_path)` returns the F1 score of predicted synapse voxels compared to the ground truth synapse voxels

Tools - V

- Tools provided are implemented in Matlab.
- Feel free to work in other languages.
- Other tools are available however, look at Fiji, Espina, Ilastik etc for inspiration.

Submission

- Results must be submitted as `tiff` stack
- Results must be placed in your *home directory*, exactly `/3yp/validate/{author,validate.tif}`
- Then run `/scratch/3yp/bin/upload validate` to submit validate (and eventually test) results.
- Results released each week in class

References

[1] N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. A. Conchello, S. Knowles-Barley, D. Lee, A. Vazquez-Reina, V. Kaynig, T. R. Jones, M. Roberts, J. L. Morgan, J. C. Tapia, H. S. Seung, W. G. Roncal, J. T. Vogelstein, R. Burns, D. L. Sussman, C. E. Priebe, H. Pfister, and J. W. Lichtman. Saturated Reconstruction of a Volume of Neocortex. *Cell*, 162(3):648–661, Jul 2015.

[2] Openconnectome project. `http://www.openconnectomeproject.org/`. Accessed: 2016-01-18.

[3] Wikipedia f1 score page. `https://en.wikipedia.org/wiki/F1_score`. Accessed: 2016-01-18.