When the Gold Standard goes Gray

An essay in response to the question: At what point does the human diagnostician's eye no longer remain the "gold standard"?

Center for Bio-Image Informatics
University of California, Santa Barbara
Manuel Arturo Deza Figueroa
arturodeza@gmail.com

Even though a diagnostician is usually portrayed as a lab-coated doctor who may receive multiple semi-automated tools such as interactive segmentation for heart imaging - medicine and biology, are still moving forward side by side in a field that has now become very important : neuroscience. In this essay, we will focus on how computer vision has fused with neuroinformatics and bioimaging to help and aid biologists, physiologists, and cognitive scientists evaluate diseases such as alzheimer's and autism, by producing accurate 3D neuron maps called traces [11-13,16].

Where is the state of the art in neuroscience then? Recently, there have been a succession of projects, challenges and initiatives from different research labs and companies worldwide to try to achieve what is the "Holy Grail" of neuroscience : To find a complete map of the brain [4]. Thus the ultimate goal of neuroscience is to understand how the nervous system works. This cannot be achieved without obtaining the structure of a real neuronal network, which reveals how neurons are connected [7]. Unfortunately, to train a computer to perform this task automatically there must be a gold standard. This mens biologists must perform manual traces of neurons, analyzing dendrite morphology, spending hours and hours in this time consuming task. Once multiple biologists have traced a particular neuron or collections of neurons, there is a 'Gold Standard' for the given set to which other biologists can later compare their automated traces of the same data [15].

Manual traces can take months, and today's computers are powerful enough to perform the same operations biologists can in hours or minutes, if they are trained and programmed correctly.

Initiatives such as the DIADEM Challenge, established by the Allen Institute for Brain Science and the Janelia Farm Research Campus of the Howard Hughes Medical Institute (HHMI), have identified this problem: **The lack of powerful – and effective – computational tools to automatically reconstruct neuronal arbors has emerged as a major technical bottleneck in neuroscience research** [1]. This in consequence, prevents major breakthroughs in treating neurodegenerative & non-neurodegenerative diseases, that need better understanding of intrinsic dendritic morphology and how they evolve over time.

In the DIADEM Challenge (short for Digital Reconstruction of Axonal and Dendritic Morphology) the goal is to create algorithmic methods for automated neuronal tracing. Five finalist teams presented multiple automated algorithms that integrate multiple computer vision and artificial intelligence techniques : open contour snakes, Frangi - vesselness measures, seed point detection / regions of interest, minimum spanning trees and Dijkstra's shortest path algorithm are only some of the algorithms used. Even though most automated scores range above 80% of the proposed DIADEM metric on the Neuromuscular Projection Fibers dataset, semi-automated score's: automated reconstructions that are later edited by a biologist, score around 95% [5]. We can thus infer that, the time consuming task of manual tracing is being saved by running

automated traces that are later refined, by our ‘Gold Standard’ expert: the biologist.

However datasets in the DIADEM challenge are only of a small number of neuron types, and their morphology, all though quite complex, is still easy to process given the multiple confocal / electron microscopy images taken - stack by stack. The biologist’s eye runs into trouble when it is confronted by a mess of chaotic morphology and overlapping connections that although detectable are confusing. Dealing with terabytes of data is painful for them, luckily supercomputers don’t face the same problem [8,14]. They will ideally produce a near perfect output trace once the algorithm is accurately designed; or if machine learning approaches are used, once it has learned the correct parameters, previously tuned by biologists.

Although we are getting close, diagnosticians eyes still remain the gold standard to manually produce a trace of the brain, but what is really important is to visualize the connections and how multiple neurons fire. Thus the problem and task of biologists should be focused in the area of exploring the connections and discovering meaning, instead of a somewhat exhausting yet relatively trivial task of manual neuron tracing. As Hebbian theory suggests, neurons that fire together, wire together, and here is where efforts should be shifted towards : Let the biologist understand the firing and computer vision take care of the wiring, as opposed to the biologist doing both.

At Ecole Polytechnique Federale de Lausanne in Switzerland, this approach is being developed by a team of researchers leaded by Henry Markram in a project called : “The Blue Brain Project” [4]. The Blue Brain Project (BBP) is an attempt to create a synthetic brain by reverse engineering the mammalian brain down to the molecular level. In essence, as they make progress towards their final goal they are getting a better understanding of the brain’s functionality and architecture. For example, in one of their recent papers they found out experimental evidence regarding Edelman’s “Neuronal Group Selection Theory”, where unlike Hebb’s theory, it suggests that functional neural circuitry arises by selection among neuronal groups that already emerged during embryonic development independent of experience [3].

Nevertheless, experimental results like these took around 10 years (1995-2005) for the manual tracing from a rat neocortical column, which is composed of around 10, 000 neurons and 100 million synapses [4,9] , while a human neocortical column contains approximately 60, 000 neurons. If we were to estimate the time to manually trace a human neocortical column, it would take 60 years, and these manual tracing techniques can not be used if we want to achieve the long-term goal of mapping the complete human neocortex. With the computational tools currently being developed, we should be able to achieve this in a shorter time frame [1,2,8,15].

The software that we have today can fortunately produce both automated and semi-automated tracing, yet the problem faced in the computer vision realm is not the tracing per se, but the processing of electron microscopy images and extrapolation of data that has not been detected correctly. Real images exhibit variable image contrast, poor axial resolution compared to the lateral resolution, and frequent loss of continuity in structure, especially near branching points or crossovers [17]. Nevertheless, these limitations can be overcome by modeling the underlying phenomena. Examples of software, that can cope with these limitations are Neuromorpho, NeuronTracing, Neurolucida, FIJI (an Image-J plug-in), Neuromantic, FARSIGHT [5].

Besides analyzing the correctness of reconstructions, correct metrics must be created that can represent a dendritic morphology semantically, so that automated reconstructions may be evaluated accurately. New metrics have progressed over time, starting with only morphological considerations. The TED : Tree Edit Distance metric is a graphical representation of branching

points in .swc format [19], that models the dendritic morphology as a free deformable body. Newer metrics such as the DIADEM metric incorporate exact 3D space position, lengths and distances between these branching points and nodes along the dendritic body [7]. With this approach, precision is awarded for taking in account multiple branching details that might be taken for granted in a free deformable body modeling of a dendrite.

Diagnosticians must strive continually on trying to improve this metric, so precision of traces can be scored accurately. This task is probably as important as designing an automated tracing algorithm, because with a correct metric the diagnostician's eye can be removed from the system and the evaluation process can be automated, producing a numerical output that gives instant information of how accurate the dendrite is reconstructed and which areas need to be revised.

In conclusion, it is not feasible for diagnosticians eyes to act as the Gold Standard when the data is of astronomical size. From a neuroinformatic point of view, the effort must be oriented towards developing an algorithm that can perform as well as a human, and that can 'see' as clearly as one, to transfer the responsibility for Gold Standard reconstructions to supercomputers. As a result, this will enable doctors and scientists from different fields to contribute to the treatment and understanding of neuro-related diseases, by giving them a better understanding of the dendritic morphology that lies behind it.

References

[1] **DIADEM Challenge official website** www.diademchallenge.org
[2] **The human connectome: a structural description of the human brain.** O. Sporns, G. Tononi, R. Kotter *PLoS Computational Biology*
[3] **A synaptic organizing principle for cortical neuronal groups.** R. Perin, T.K. Berger, H. Markram. *PNAS 2011.*
[4] **Henry Markram TED talk : Supercomputing the Brain's Secrets** http://blog.ted.com/2009/10/15/supercomputing/
[5] **A Broadly Applicable 3-D Neuron Tracing Method Based on Open-Curve Snake.** Y. Wang, A. Narayanaswamy, C. Tsai, B. Roysam *Neuroinformatics 2011*.
[6] **Automated Reconstruction of Neuronal Morphology Based on Local Geometrical and Global Structural Models.** T. Zhao , J. Xie, F. Amat, N. Clack, P. Ahammad, H. Peng, F. Long, E. Myers. *Neuroinformatics 2011*
[7] **The DIADEM Metric: Comparing Multiple Reconstructions of the Same Neuron.** T.A. Gillette, K.M. Brown, G.A. Ascoli. *Neuroinformatics 2011*
[8] **The FARSIGHT Trace Editor: an Open source Tool for 3-D Inspection and Efficient Pattern Analysis Aided Editing for 3-D Inspection and Efficient Pattern Analysis Aided Editing of Automated Neuronal Reconstructions.** J. Luisi, A. Narayanaswamy, Z. Galbreath, B. Roysam *Neuroinformatics 2011*
[9] **The Cat is Out of The Bag: Cortical Simulations with 10^9 neurons and 10^13 synapses.** R. Ananthanarayanan, S.K. Esser, H.D. Simon, D.S. Modha *SuperComputing 2009*
[10] **Machines that learn to segment images: a crucial technology for connectomics**. V. Jainm, H.S. Seung, S.C. Turaga *Current Opinion in NeuroBiology 2010*
[11] **About Alzheimer's Disease: Alzheimer's Basics** - National Institute of Health, National Institute of Aging 2012. http://www.nia.nih.gov/alzheimers/topics/alzheimers-basics
[12] **Parkinson's disease: clinical features and diagnosis** J. Jankovic *Journal of Neurology, Neurosurgery & Psychiatry 2007*
[13] **Autism** S.E. Levy, D.S. Mandell, R.T. Schultz *The Lancet 2009*

[14] **Automated Tracing of Neurites from Light Microscopy Stacks of Images** P. Chotani, V. Mehta, A. Stepanyants *Neuroinformatics 2011*
[15] **Racing down the highway** - News at Northeastern University 2010 http://www.northeastern.edu/news/stories/2010/08/Diadem.html
[16] **Principal Curves as Skeletons of Tubular Objects** E. Bas, D. Erdogmus *Neuroinformatics 2011*
[17] **3-D Image Pre-processing Algorithms for Improved Automated Tracing of Neuronal** Arbors A. Narayanaswamy, Y. Wang, B. Roysam *Neuroinformatics 2011*
[18] **The Organization of behaviour** D.O. Hebb Wiley & Sons 1949
[19] **What is .swc format?** DIADEM Challenge FAQ http://diademchallenge.org/faq.html