Automated Lesson Synthesis: Using Misconceptions to Improve Neuron Tracing

Anant Mittal
University of Washington
Seattle, Washington, USA
anmittal@cs.washington.edu

ABSTRACT

Mozak is a scientific discovery game for neuroscience where players collaboratively reconstruct complex 3D representations of neurons by tracing their volumetric image. This paper introduces an automated system which uses incorrect traces (misconceptions) to generate meaningful lessons for Mozak players. The hypothesis is that these targeted lessons can assist novice players become expert tracers such that their traces closely resemble those created by neuroscience experts. We discuss the architecture of our system and efficacy of lessons generated.

Author Keywords

Neuroscience; Neuronal Reconstruction; Misconceptions; Lesson Generation; Citizen Science

INTRODUCTION

There are billions of neurons in our brain, yet their precise count is undecided and keeps varying [4]. The study of the structure of neuronal cells and how they connect with each other is a crucial to understanding how human brain functions. The advancement in precision imaging technology has allowed neuroscience researchers to capture volumetric 3D images of neurons on an individual level. Neuron tracing or neuron reconstruction of 3D images is an important step in uncovering how brain diseases develop and could be treated [8]. Although a wide variety of automated techniques have been developed to generate 3D reconstructions from microscopic images of neurons [10, 11], human edited neuron traces still remain the benchmark for evaluation [3].

Mozak (Figure 1) [12] is an online citizen science game which aims to create a new community equipped with the skills required to accurately and rapidly trace neurons. Since its launch in 2017, Mozak has helped trace around 544 neurons till date. Neurons published on Mozak are received from neuroscience labs as volumetric images. Once players start tracing neurons, the trace data is stored as graphs where each node is mapped to its *x, y, z* coordinates in the pixel space. As they keep tracing, in the background a *consensus* graph gets created continuously. This consensus graph is generated if two or more players trace near each other, analogous to an inter-annotator agreement. It then gets sent back to the neuroscience lab for proof-editing and completing any untraced or poorly traced parts of the neuron. Final output of this process is a weakly connected directed graph in *SWC* format [9] referred to as the *gold standard*.

ISBN 978-1-4503-2138-9...$15.00
DOI: 10.1145/1235

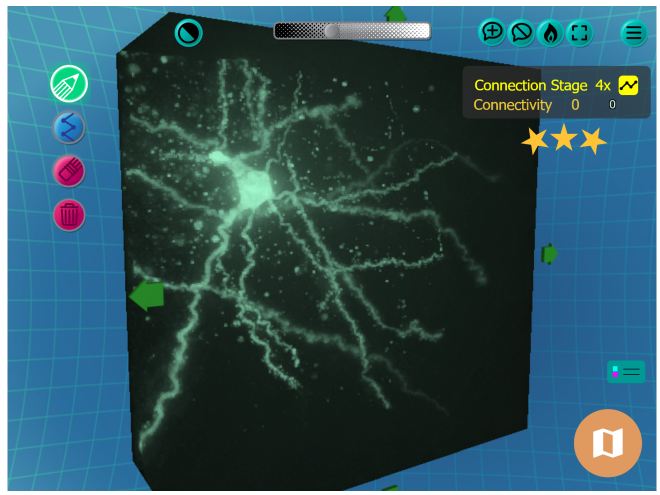


Figure 1: Mozak is a citizen science game for 3D reconstruction of neurons.

The uniqueness and complexity of neuron structures make it hard to quantify which citizen behaviors lead to tracing inaccuracies. In this ongoing research, we look at this ambiguity through a learning sciences lens. We developed a data-driven approach to automatically generate visually interactive lessons by identifying tracing misconceptions on Mozak. The goal of these lessons is to help Mozak players become expert neuron tracers.

In the coming sections, we will discuss the system architecture and look at a few lessons generated through it.

Related Work

There have been prior studies to understand player motivations in citizen science games [6, 2, 5]. The closest citizen science

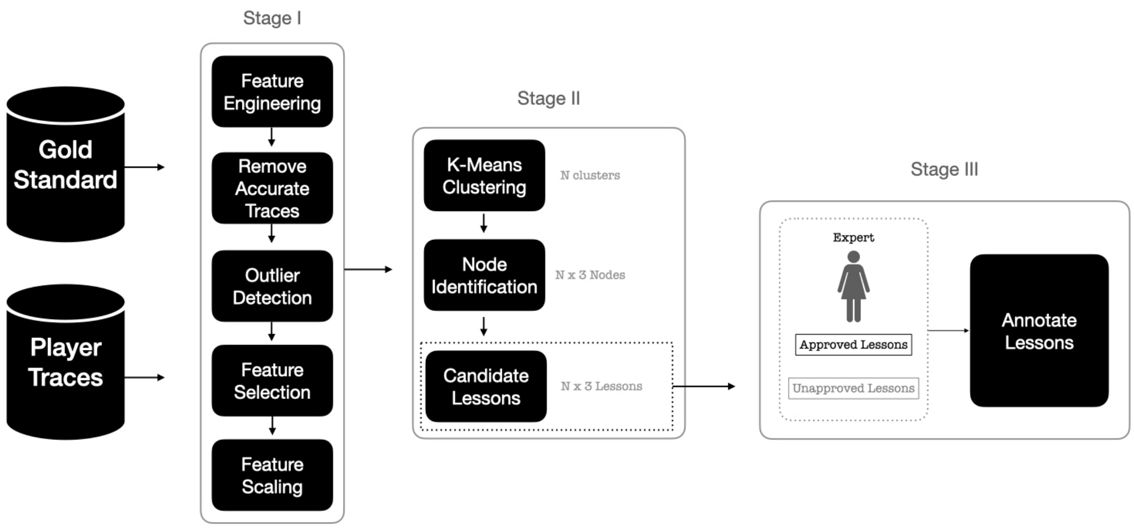


Figure 2: In Stage I, raw gold standard and player traces data are processed. In Stage II, processed data is clustered and from each cluster 3 candidate lessons are generated. Finally, in Stage III approved lessons are annotated which are then deployed on Mozak.

game to Mozak is Eyewire [7], where survey-based methods have been used to understand player motivations [13] but a learning sciences perspective has not been applied to study misconceptions neuron tracers face.

SYSTEM ARCHITECTURE

Figure 2 shows the entire architecture of our lesson generation system. The system uses *Gold Standards* and *Player Traces* from a previously traced neuron on Mozak as raw data. Computationally this data consists of graphs represented as nodes and edges in a 3-dimensional space.

The entire pipeline of the lesson generation system is divided in to three stages.

Stage I

We take nodes from player traces and gold standard as raw data and run a typical data science pipeline on them. In the *Remove Accurate Traces* sub-stage we remove the nodes which meet the following conditions:

- In Player Traces, nodes which are (1) close to gold standard and close to another user (consensus), and (2) away from gold standard and away from any other user (potentially random user tracing).
- In Gold Standard, nodes which were close to 2 or more unique players.

Because of this sub-stage, we end up removing high-quality traces and keeping the inaccuracies. The assumption is that low quality or inaccurate traces are a result of players' misconceptions.

Stage II

We cluster the results from *Stage I* using k-means++ [1] with the elbow method to find the optimum number of N clusters. Since k-means is a centroid based clustering method, we then find the 3 closest nodes (using euclidean distance) from each centroid and create a bounding box of a certain threshold surrounding each of them. Using the (x,y,z) dimension of the bounding box, we plot the player traces, consensus, and the neurite in pixel space to visually understand the misconceptions, given they exist within these coordinates. These visually marked and annotated bounding boxes are nothing but *Candidate Lessons*. The rationality behind clustering is that we want to target the most common misconceptions faced by the players.

Stage III

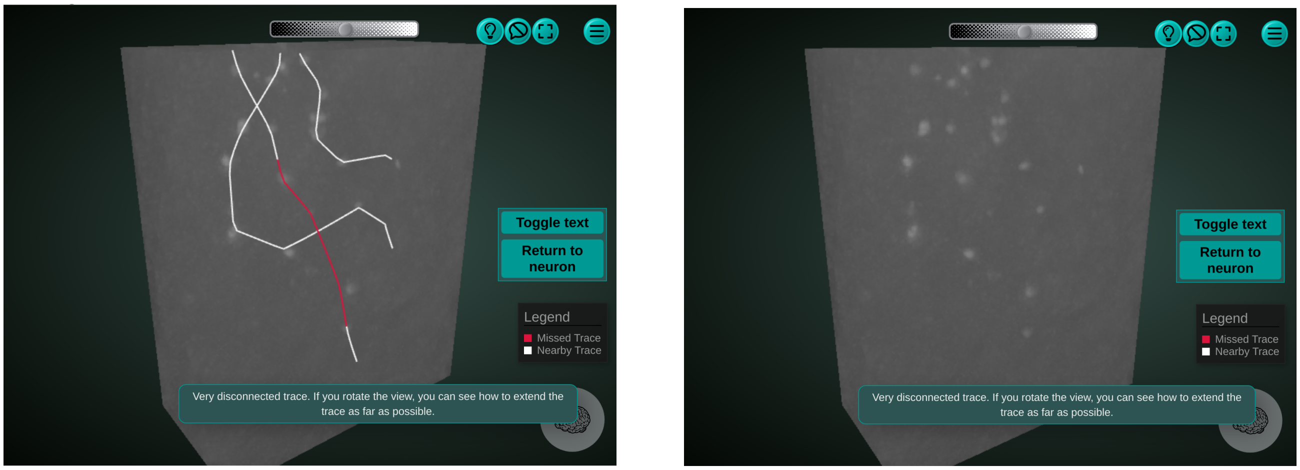
In this final stage, neuroscience experts look at candidate lessons to approve or reject them. The approved lessons are annotated and edited to remove extraneous details. It is important to note that except in the final annotation phase, no manual intervention is required in generating these lessons for Mozak. These lessons are interactive allowing players to rotate and zoom in a 3D space to investigate misconceptions.

GENERATED LESSONS

In this section, we discuss three lessons that were generated by our system. A video slide of these lessons have been posted here: https://youtu.be/9Wc2o06Cnu8.

The approved and annotated lessons on Mozak have pulsating visualizations. In Figures 3, 4, and 5, sub-figure *(a)* shows the lesson when the pulse is ON and sub-figure *(b)* shows it with the pulse OFF.

In Figure 3, we can see that players are probably missing this trace because they are misinterpreting the disconnected neuron signal as noise. For the lesson shown in Figure 4, we can see the neuron signal is very faint (Figure 4a). On adjusting the brightness, light blobs seem to appear which should have ideally been traced (Figure 4b). In Figure 5, we notice that players accurately trace a running signal (yellow



(a) Lesson Visualization with Pulse ON.

(b) Lesson Visualization with Pulse OFF.

Figure 3: Lesson 1 – Disconnected trace misconception

(a) Lesson Visualization with Pulse ON.

(b) Lesson Visualization with Pulse OFF and increased brightness.

Figure 4: Lesson 2 – Faint blob misconception

line) but only one player traces the parallel running neuron signal (blue line). In this case, the proximity of parallel neuron signals probably made players misinterpret them as a single signal and miss the accurate trace.

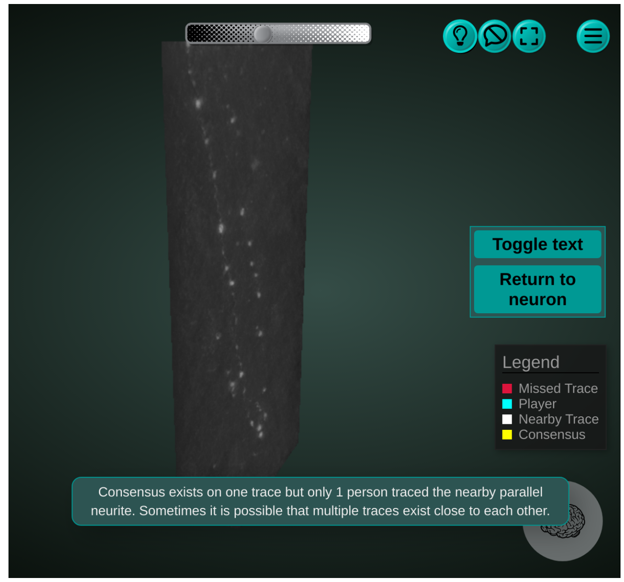
DISCUSSION AND ONGOING WORK

To evaluate the effectiveness of lessons and quantify if they helped improve player tracing skills, we ran a randomized controlled trial in the beginning of March 2021. We took volumetric images of some neurons from Allen Institute for Brain Science (AIBS) for this experiment. These neurons were chosen because AIBS already had their gold standards and they had never been traced on Mozak. In treatment condition players were presented with lessons generated by our automated system on Mozak. Due to a low turnout of active players on Mozak, it was not possible to observe statistically significant behavioral changes and the study was aborted. An elaborate in-lab study is required to understand if beginners/novice players are able to pick up tracing skills much faster than players who don't see such lessons.

From our preliminary experiments, we also found that players were merely tracing 30% of neurons on an average, something which is difficult to quantify without gold standards. This was probably because players usually start tracing from *soma* where the neuron signals are brightest and most clear. As they get to the tips of the neuron the signal becomes faint and extremely hard to trace. The Mozak development team is working on building features which could help players focus more on such tough areas.

In this paper, we looked at neuron tracing improvements from a different perspective of learning through misconceptions. The lessons our system is able to generate have been vetted by neuroscience experts but to critically evaluate it we would need to conduct a robust experiment with enough player data.

(a) Lesson Visualization with Pulse ON.



(b) Lesson Visualization with Pulse OFF.

Figure 5: Lesson 3 – Missing the parallel neurite trace misconception

REFERENCES

[1] David Arthur and Sergei Vassilvitskii. 2006. *k-means++: The advantages of careful seeding*. Technical Report. Stanford.

[2] Vickie Curtis. 2015. Motivation to participate in an online citizen science game: A study of Foldit. *Science Communication* 37, 6 (2015), 723–746.

[3] Duncan E Donohue and Giorgio A Ascoli. 2011. Automated reconstruction of neuronal morphology: an overview. *Brain research reviews* 67, 1-2 (2011), 94–102.

[4] Suzana Herculano-Houzel. 2012. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences* 109, Supplement 1 (2012), 10661–10668. DOI: `http://dx.doi.org/10.1073/pnas.1201895109`

[5] Ioanna Iacovides, Charlene Jennett, Cassandra Cornish-Trestrail, and Anna L Cox. 2013. Do games attract or sustain engagement in citizen science? A study of volunteer motivations. In *CHI'13 extended abstracts on human factors in computing systems*. 1101–1106.

[6] Rebecca C Jordan, Steven A Gray, David V Howe, Wesley R Brooks, and Joan G Ehrenfeld. 2011. Knowledge gain and behavioral change in citizen-science programs. *Conservation biology* 25, 6 (2011), 1148–1154.

[7] Vivien Marx. 2013. Neuroscience waves to the crowd. *Nature methods* 10, 11 (2013), 1069–1074.

[8] Erik Meijering. 2010. Neuron tracing in perspective. *Cytometry Part A* 77, 7 (2010), 693–704.

[9] n/a. 2021. SWC Format Specificaiton. `http://www.neuronland.org/NLMorphologyConverter/MorphologyFormats/SWC/Spec.html`. (2021). [Online; accessed 30-March-2021].

[10] Hanchuan Peng, Fuhui Long, and Gene Myers. 2011. Automatic 3D neuron tracing using all-path pruning. *Bioinformatics* 27, 13 (2011), i239–i247.

[11] Miroslav Radojević and Erik Meijering. 2017. Automated neuron tracing using probability hypothesis density filtering. *Bioinformatics* 33, 7 (2017), 1073–1080.

[12] Jane Roskams and Zoran Popović. 2016. Power to the people: addressing big data challenges in neuroscience by creating a new cadre of citizen neuroscientists. *Neuron* 92, 3 (2016), 658–664.

[13] Ramine Tinati, Markus Luczak-Roesch, Elena Simperl, and Wendy Hall. 2017. An investigation of player motivations in Eyewire, a gamified citizen science project. *Computers in Human Behavior* 73 (2017), 527–540. DOI: `http://dx.doi.org/https://doi.org/10.1016/j.chb.2016.12.074`