Embracing Web Technology for more Open and Collaborative Science

Citizen Science and Crowdsourcing in Neuroscience

Anisha Keshavan

Why We Collect Big Data in Neuroimaging

Neuroimaging in the Age of Information

TO POWER
Multivariate Analyses

TO PREVENT
A Reproducibility Crisis

TO TREAT WITH
Precision Medicine

TO ENABLE
Generalizability

Why Not ?

The technology is there.

Storage is getting cheaper &

Computers are getting faster

The Problems We Face

The Repercussions of Using Big Data for Neuroimaging

PROBLEM #1

The Learning Curve

PROBLEM #2

Data Visualization

PROBLEM #3

The Manual Bottleneck

The Problems We Face

The Repercussions of Using Big Data for Neuroimaging

PROBLEM #1
The Learning Curve

PROBLEM #2
Data Visualization

PROBLEM #3
The Manual Bottleneck

The Learning Curve

Is too steep

Need to have domain knowledge & learn specialized software, have UNIX proficiency, coding experience. Difficult when academia and citizen scientists have a high turnover rate.

Solution: A platform that facilitates a lower learning curve, without being opaque. Should be easy to install.

The Problems We Face

The Repercussions of Using Big Data for Neuroimaging

PROBLEM #1

The Learning Curve

PROBLEM #2

Data Visualization

PROBLEM #3

The Manual Bottleneck

Data Visualization

Is more complicated

Multidimensional data is difficult to understand, and traditional plots don't work. If we find it difficult to see patterns in our data, how can we ask citizen scientists to do the same for us?

Solution: A platform that enables interaction, and that can reproduce a set of interactions.

The Problems We Face

The Repercussions of Using Big Data for Neuroimaging

PROBLEM #1

The Learning Curve

PROBLEM #2

Data Visualization

PROBLEM #3

The Manual Bottleneck

The Manual Bottleneck

Quality control, brain editing takes too long.

Manual intervention is error prone, and editing segmentations takes a long time. With thousands of scans to edit, distributing work between many researchers (& citizens) is needed, but hard to coordinate.

Solution: A platform that facilitates collaboration, and can manage feedback from many individuals.

Solution: The Web Browser

For more open and collaborative science

SOLVING PROBLEM #1

The Learning Curve

SOLVING PROBLEM #2

Data Visualization

SOLVING PROBLEM #3

The Manual Bottleneck

Solution: The Web Browser

For more open and collaborative science

SOLVING PROBLEM #1

The Learning Curve

SOLVING PROBLEM #2

Data Visualization

SOLVING PROBLEM #3

The Manual Bottleneck

The Learning Curve

is easier when we're familiar with the technology

Web browsers are ubiquitous, web apps have no install, and need a user interface. Easy access leads to a wider adoption by researchers and citizen scientists. Can "view source" of all websites, and access from any device.

There's an app for that: A web application written in HTML5 can be compiled to a mobile application.

Solution: The Web Browser

For more open and collaborative science

SOLVING PROBLEM #`

The Learning Curve

SOLVING PROBLEM #2

Data Visualization

SOLVING PROBLEM #3

The Manual Bottleneck

Data Visualization

is interactive

Web browsers run JavaScript, which can manipulate the content and view of the page. Many interactive data visualization libraries already exist, such as D3. Can interact with 3D models using WebGL.

Undo: State management libraries in JS keep track of all actions a user makes. Can go back (or forward) in time.

Solution: The Web Browser

For more open and collaborative science

SOLVING PROBLEM #`

The Learning Curve

SOLVING PROBLEM #2

Data Visualization

SOLVING PROBLEM #3

The Manual Bottleneck

The Manual Bottleneck

is manageable when we collaborate

Using the cloud, web servers can be scaled up to send and receive data from thousands of people at once. Cellular data enables us to be connected all the time. Collaborative tools, like Google Docs, are widely used every day.

Grab the sunscreen: In theory, we could be analyzing our data on the beach (if there is reception)!

The Web is Open

Why the web promotes sharing

5 Principles of the Web

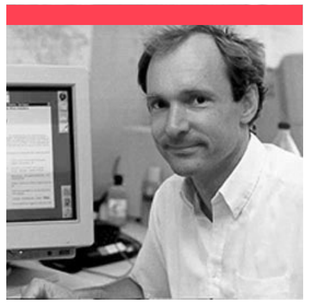
Non-Discrimination
All websites are treated equally

Bottom-Up Design
View source

Universality
Independent of hardware

Decentralization
No central authority, no permission needed

Consensus
Standards, agreed by everyone



Tim Berners Lee

Founder of the Internet

Started by Tim Berners Lee at CERN who wanted to share data between computers.

Source: https://webfoundation.org/about/vision/history-of-the-web/

Overview

What I'll be talking about

Web Neuroimaging

Neuroimaging tools in the browser

nbpapaya
Mindboggle & ROYGBIV
Mindcontrol
Brainbox
Brainspell
And more!

Crowdsourcing

In other domains

FoldIt
PlanetHunters
GlobalXplorer
Eyewire
Mozak

A Proposal

To crowdsource neuroimaging

Collaboration with Adam Thomas and Dylan Nielson to create a crowdsourcing service for manual image segmentation

nbpapaya

A Jupyter notebook plugin to view 3D nifti volumes

Jupyter Notebook

Browser tool for Python

Papaya Viewer

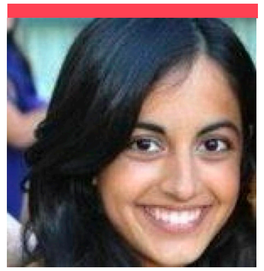
Javascript 3D/4D image viewer

nbpapaya:

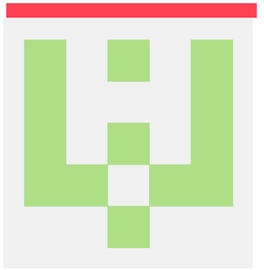
View nifti images with Papaya in the Jupyter notebook

View 3D files as you generate them in the notebook.

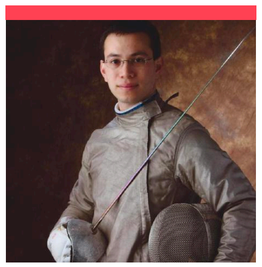
Contributors



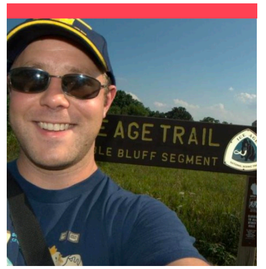
akeshavan



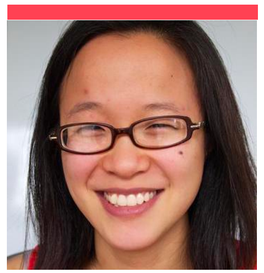
mrbago



alexlicohen



kastman



pandafulmanda

ROYGBIV

A browser-based visualization tool for Mindboggle

Mindboggle

Improved anatomical labelling

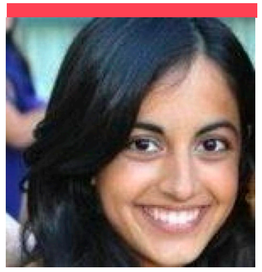
Improve the precision of anatomical labelling and compute additional shape features of the cortex.

ROYGBIV

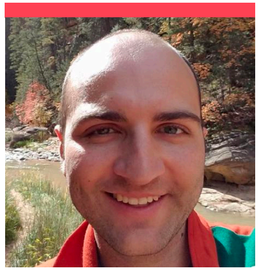
The Roy. G. Brain Imaging Viewer

After running Mindboggle, visually explore the results with an interactive 3D brain surface and accompanying box plots of shape measures.

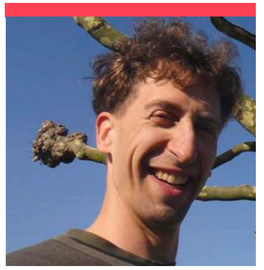
Contributors



akeshavan



curtpw



binarybottle

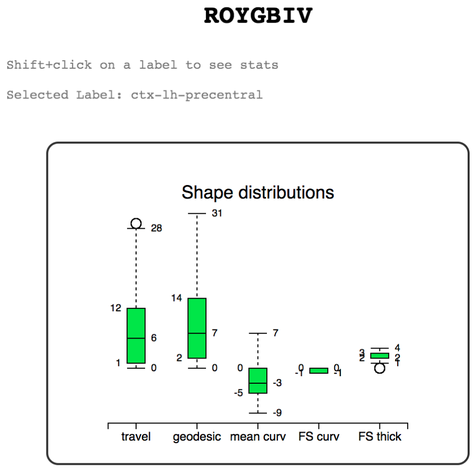
bcipolli

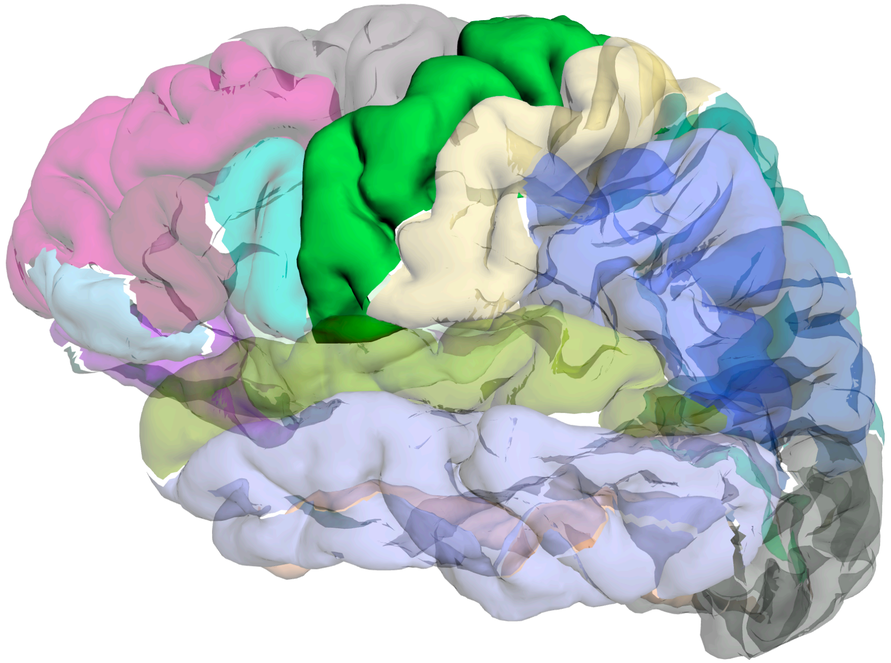
ROYGBIV

A browser-based visualization tool for Mindboggle

Live Demo

roygbiv.mindboggle.info





Mindcontrol

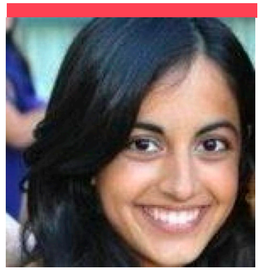
A web application for collaborative brain segmentation quality control

Motivation

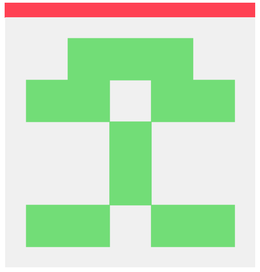
QC takes too long

The output of every pipeline needs to be visually inspected, and coordinating this effort is difficult. Current method: use imaging viewer to inspect, annotate, edit. Software to plot distribution and select outliers. Google docs to take notes/ collaborate. Is inefficient, error prone.

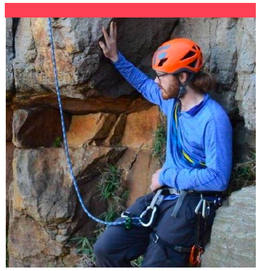
Contributors



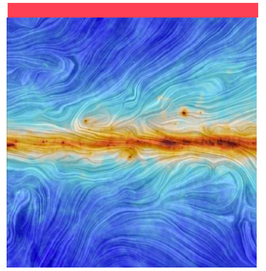
akeshavan



edatta



ltirrell



seldamat

immcdonough

Mindcontrol

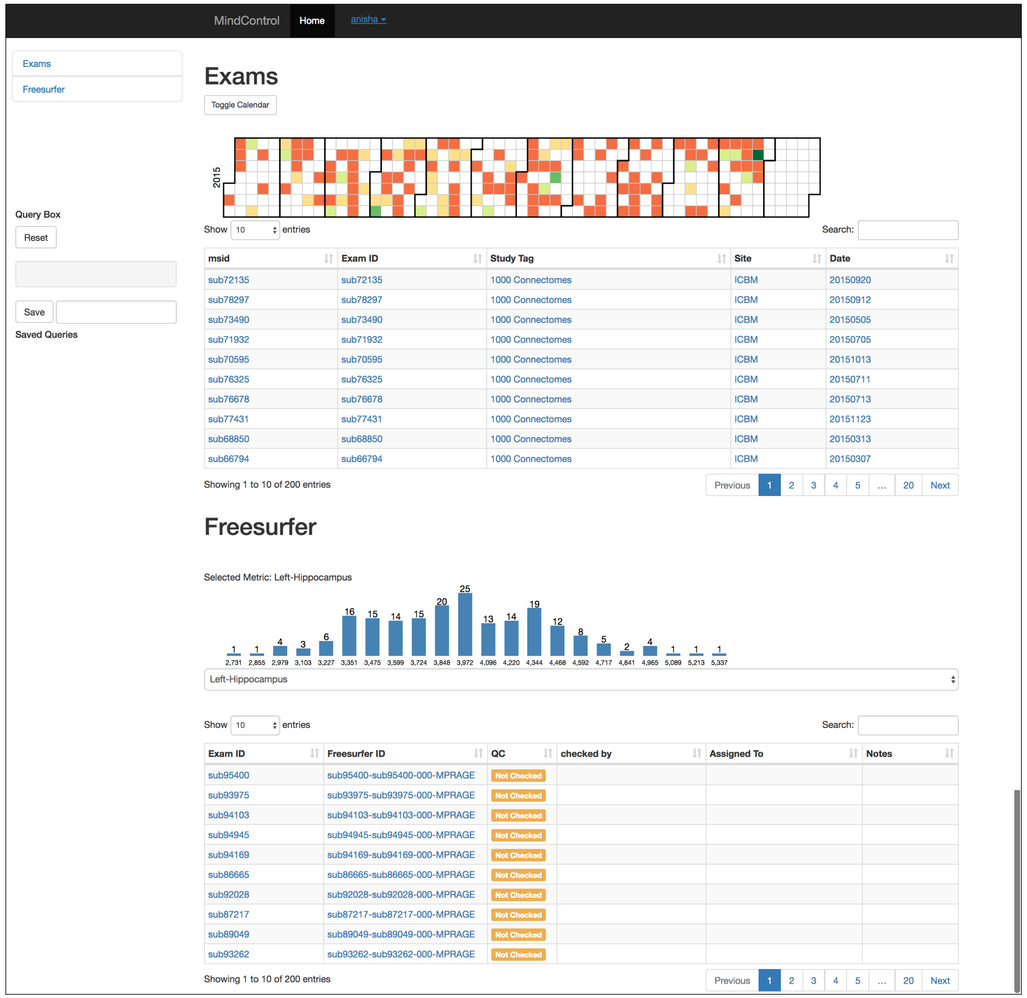
A web application for collaborative brain segmentation quality control

Solution

Configurable web application to QC

A web platform for QC, curation, annotation and editing of data. Task management and data exploration tool. Like a “triage” tool to prioritize images that need more TLC.

Mindcontrol can be configured for the output of any pipeline. Annotations and edits are stored in a database rather than the filesystem. Access from any device, including tablet.



Mindcontrol

A web application for collaborative brain segmentation quality control

Live Demo #1

mindcontrol.herokuapp.com

Explore Freesurfer segmentations of 200 scans from the 1000 Functional Connectomes project

Live Demo #2

mindcontrol-abide.herokuapp.com

Explore Quality Assurance Protocol (QAP) outputs from the ABIDE dataset, using an interactive parallel coordinates plot

Live Demo #3

Local UCSF pipeline outputs

Lesion segmentation annotations from neuroradiologists at UCSF

Future Directions

Combine Mindcontrol with Clowder for cloud computing with a click of the button.

Brainspell

A web application that lets you annotate and segment 3D brain imaging data in real time, collaboratively.

Motivation

Meta-analyses for more stable NI results

Brainspell

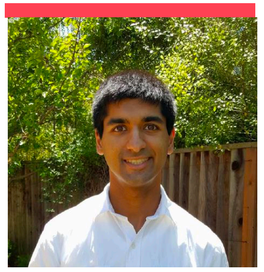
Collaborate on annotation for meta analyses

Neurosynth

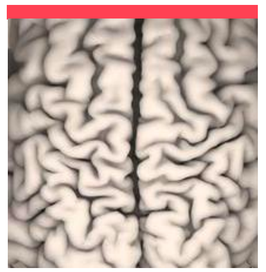
Automatically extract MNI coordinates from manuscripts

It is difficult to automatically extract coordinates from manuscripts, there is no standardization of results. Human eyes needed.

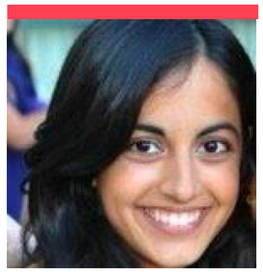
Contributors



neelsomani



r0b3rt0



akeshavan

sharabesh

jbpoline

Brainspell

A web application that lets you annotate and segment 3D brain imaging data in real time, collaboratively.

Live Demo

brainspell.herokuapp.com

Neuroimaging on the Browser

There are many more tools available, and more to come!

MicroDraw

Collaborate on editing BigBrain data

Neurovault

Upload and share your brain maps

BrainBox

Google docs for brain editing

AFQ-Browser

Browse and share your results from AFQ processing

bdpdb

Brain damage patient data browser

Cluster-viz

Browser tool to efficiently edit diffusion tractography output

We are using the Web to:

Share Data 

Share Information 

Share Work 

Overview

What I'll be talking about

Web Neuroimaging

Neuroimaging tools in the browser

nbpapaya
Mindboggle & ROYGBIV
Mindcontrol
Brainbox & Microdraw
Brainspell
And more!

Crowdsourcing

In other domains

FoldIt
PlanetHunters
GlobalXplorer
Eyewire
Mozak

A Proposal

To crowdsource neuroimaging

Collaboration with Adam Thomas and Dylan Nielson to create a crowdsourcing service for manual image segmentation

Crowdsourcing Principles

Science Communication
Effectively convey why the problem is important

Interaction Design
Easy to use, intuitive, and engaging

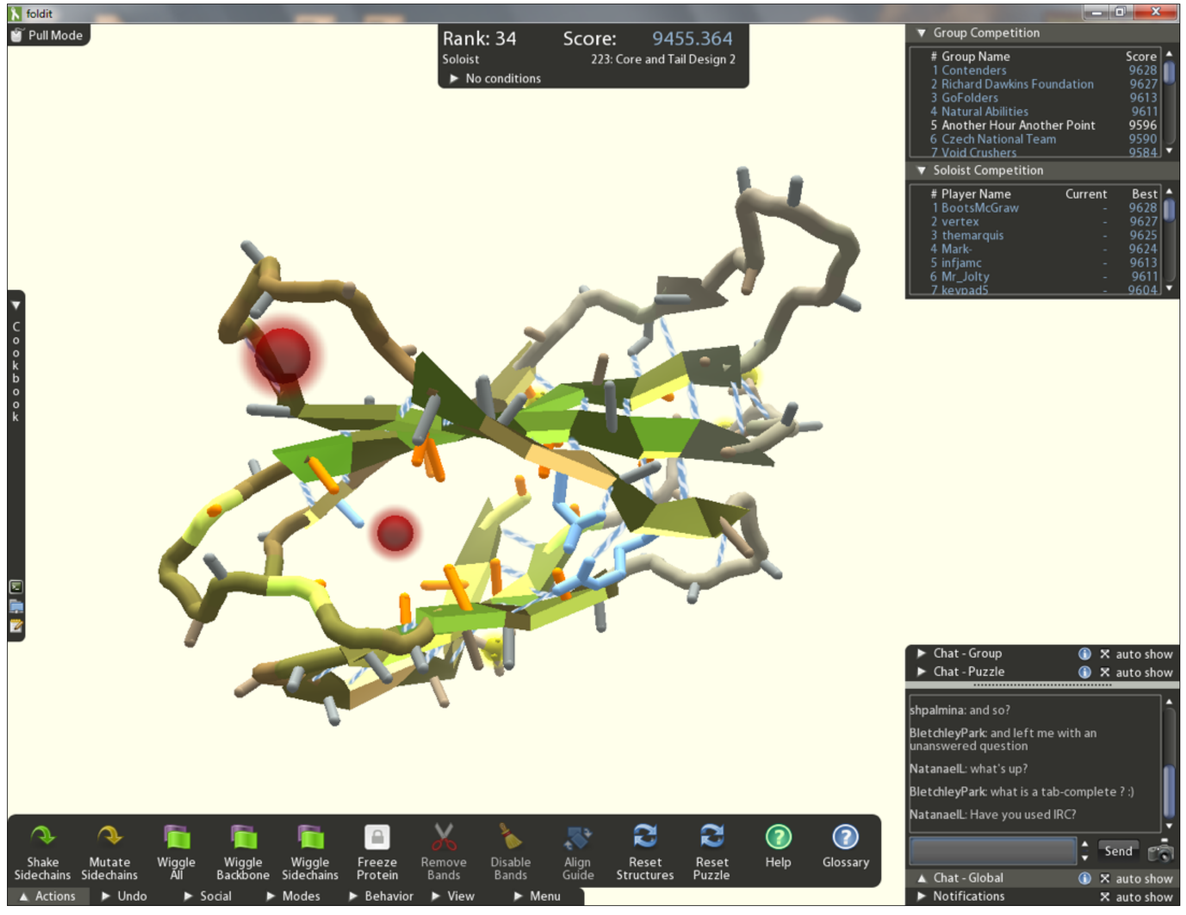
Microtasks
Convert large problem into small, manageable chunks

Community
Large communities work together, but also compete with each other!

FoldIt

A puzzle to solve for the structure of proteins

Released in 2008 and had over half a million players



Game:
Predict how a given protein will fold

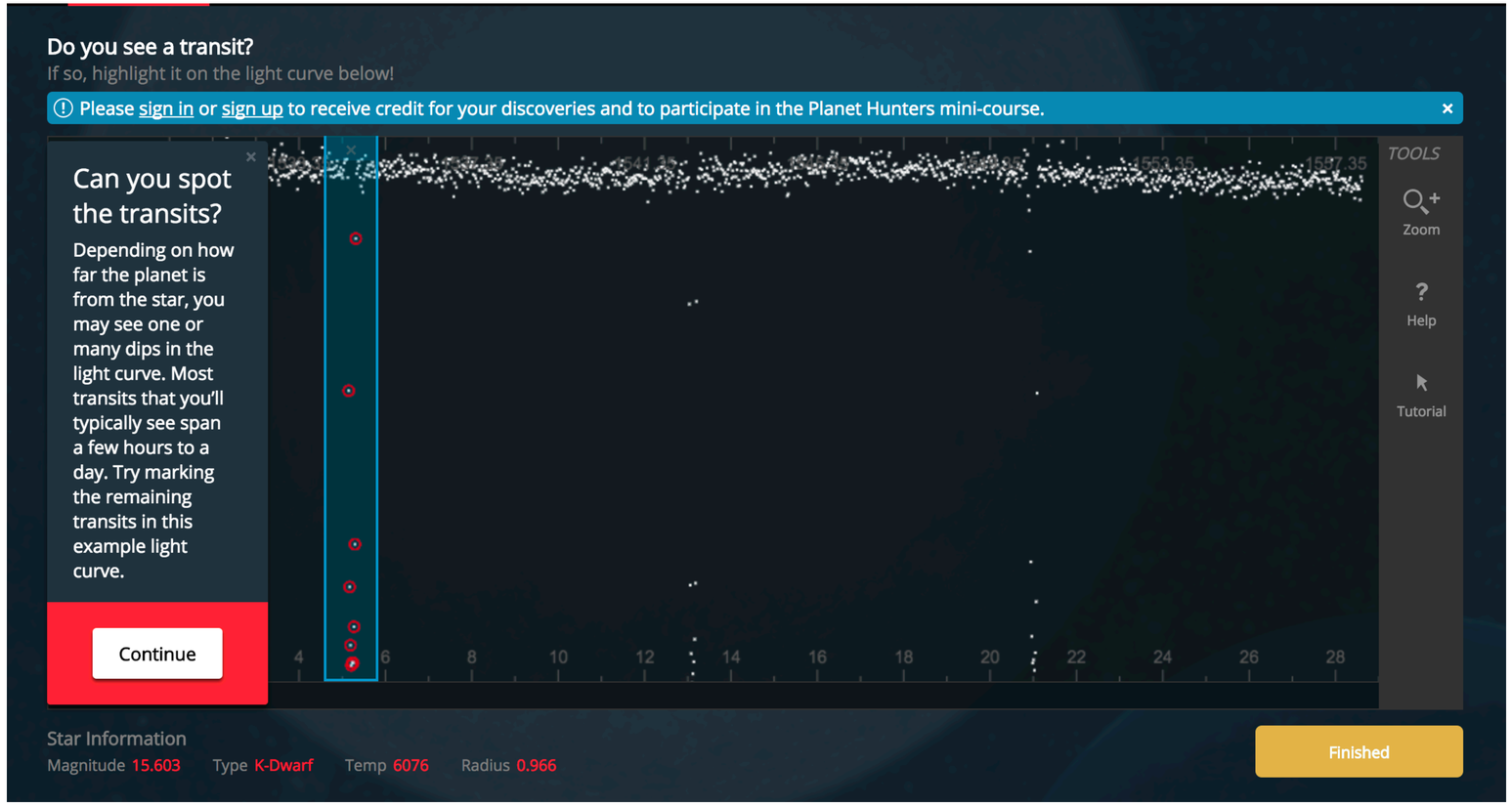
Unsolved structure of HIV protein was solved in 10 days

"This is about leveraging the things that computers can do well and the things that people do really well when they're trained, forming a symbiotic superpower capable of solving unsolved challenges in science."

- Popović

PlanetHunters

Find earth-like planets



Kepler satellite is looking for earth like planets

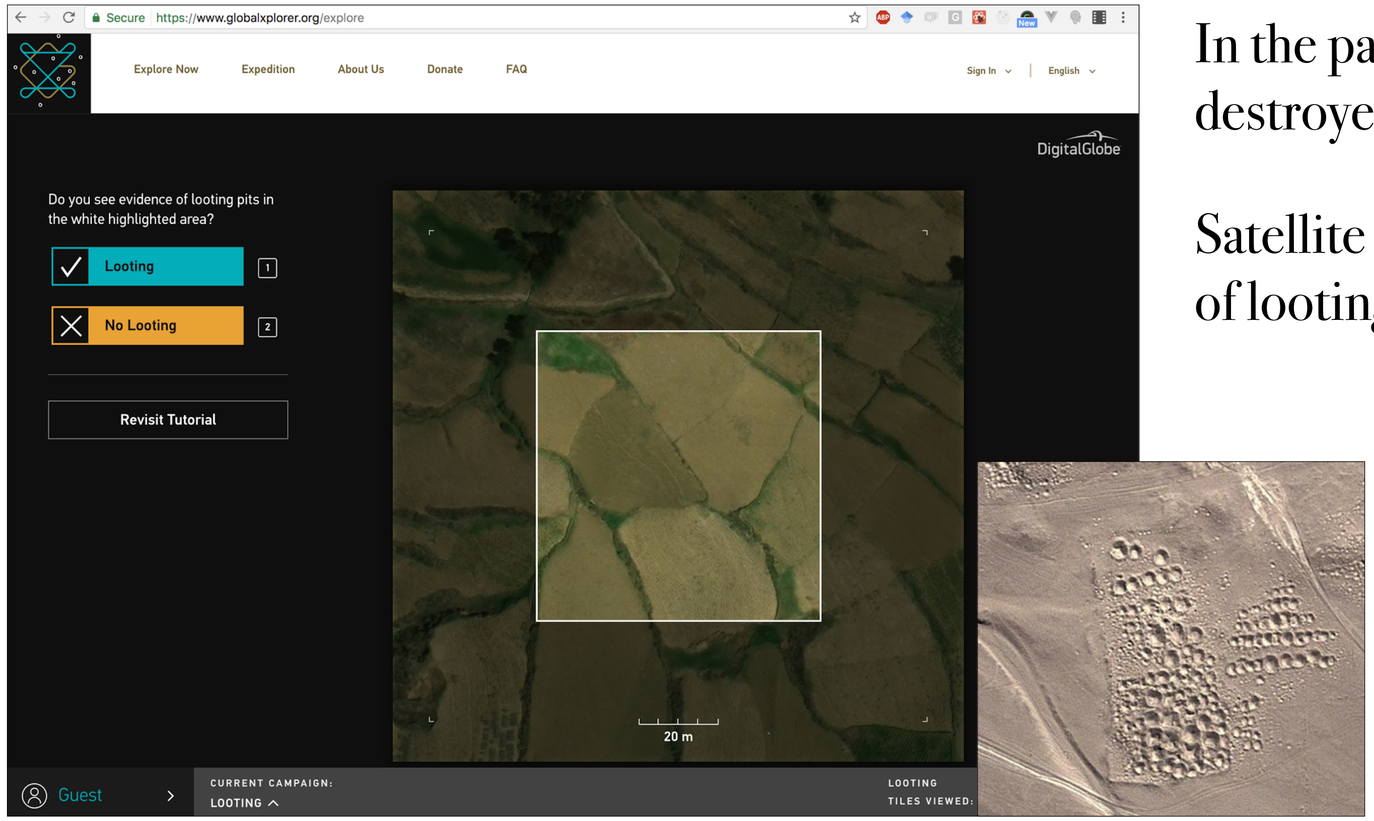
Computers couldn't do it

With citizen scientists, they've found dozens

And they've discovered a new mystery in the cosmos

GlobalXplorer

Become a space archaeologist and prevent looting of ancient ruins



In the past 5 years, looters have destroyed archeological sites

Satellite imagery shows distinct patterns of looting, and new undiscovered ruins

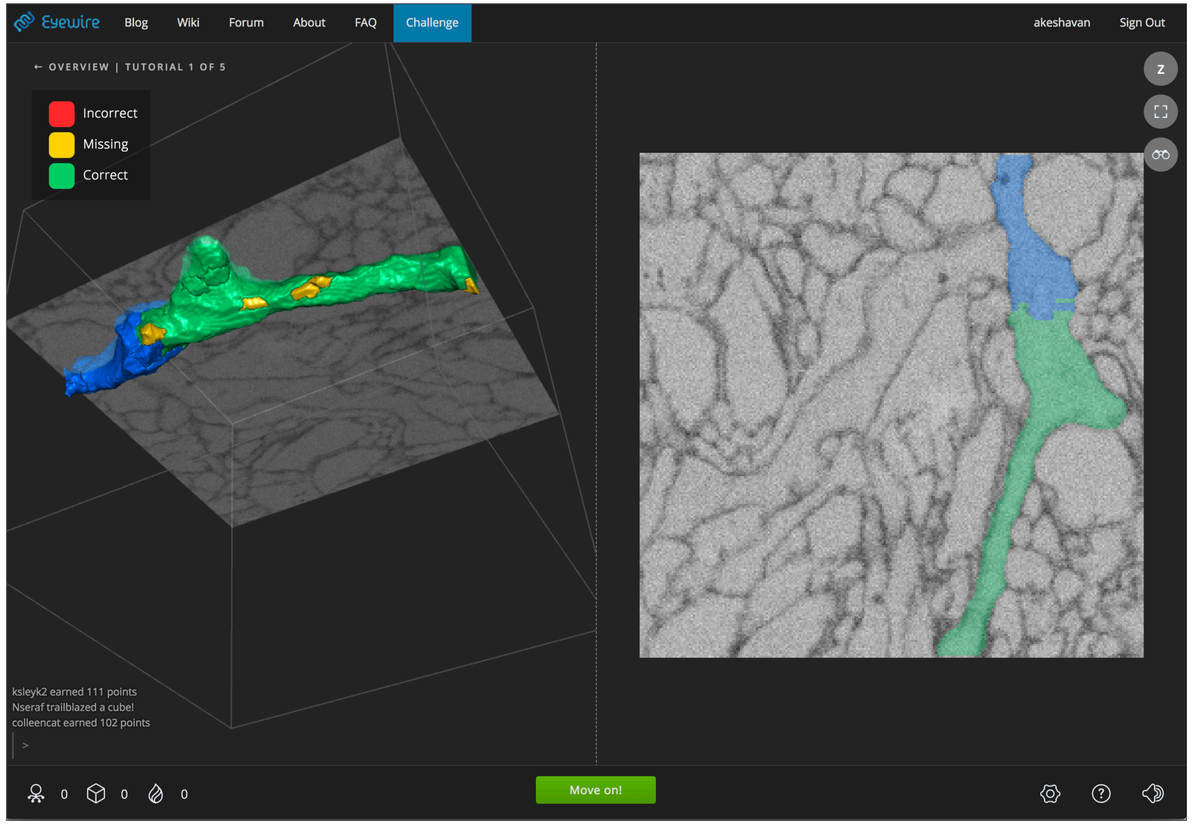
Citizen scientists are becoming “space archaeologists” to identify looting and discover new ruins

Works really well on mobile - play it while waiting for the bus!

Eyewire

Map the 3D structure of neurons

Recruited >100,000 people, over 130 countries to reconstruct neurons from microscopy data



Started with a "seed" segmentation, used CNNs to create "supervoxels": when you click once, it highlights a whole area

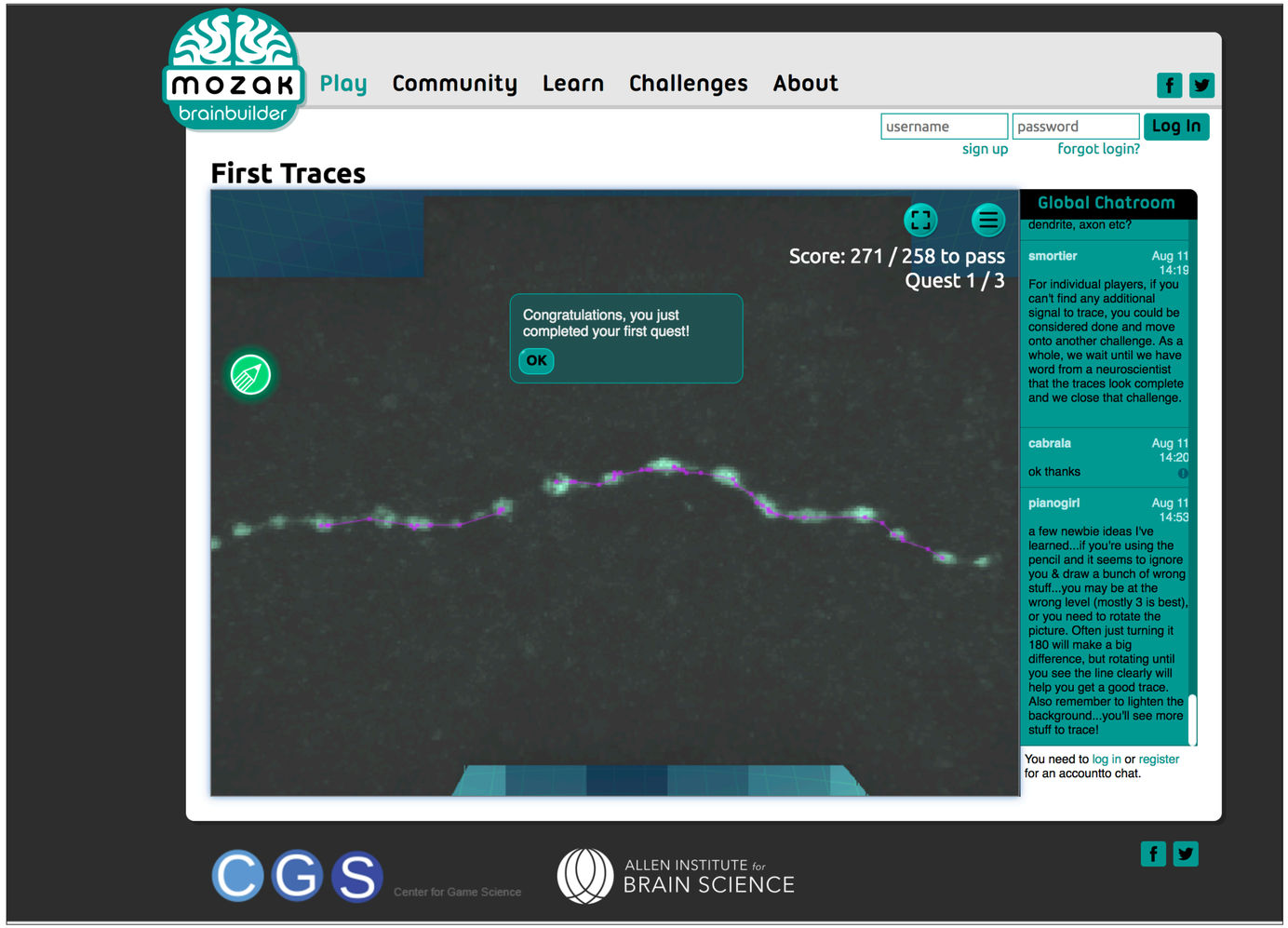
Finding new talent: The top 100 users did half the segmentations!

Looks beautiful, and has music

Two panels, left is interactive 3D w/ blue "seed" region, right shows slice of image, scroll to change slice

Mozak

Build models of brain cells, and help scientists learn more about the brain



Goal: Trace different neurons from microscopy images to study the diversity of cell types

Same group that made FoldIt

Around 200 players / day

Has cool music, leaderboard, chat

Some Difficulties

In making similar citizen science games

Making a game is really hard

Building a large community is really hard

I can’t reuse any of these games for my data.

We need a more general crowdsourcing platform.

General Crowdsourcing Platforms

Possible platforms to use for neuroimaging applications

mTurk

Amazon Mechanical Turk

PsiTurk: makes it easier to run mTurk experiments

ExpFactory: even easier to run mTurk experiments by standardization

Zooniverse

Open source!

Has a bunch of really cool projects. Also, has image annotation tools already built in. But, no voxel/ pixel editing tools yet. Has potential!

CrowdFlower

What industry uses

Too expensive. But a free version exists if you release data. Annotation tools, no pixel editing.

Overview

What I'll be talking about

Web Neuroimaging

Neuroimaging tools in the browser

nbpapaya
Mindboggle & ROYGBIV
Mindcontrol
Brainbox & Microdraw
Brainspell
And more!

Crowdsourcing

In other domains

FoldIt
PlanetHunters
GlobalXplorer
Eyewire
Mozak

A Proposal

To crowdsource neuroimaging

Collaboration with Adam Thomas and Dylan Nielson to create a crowdsourcing service for manual image segmentation and editing

The Plan

First pass at crowdsourcing image segmentation

Upload Data

To a central server

Include some training tiles (with “ground truth”) and testing tiles (without “ground truth”)

Train

Teach the citizen scientists

Serve training images & give feedback until the player has learned enough. Then start serving test images (no feedback)

Performance Drift

Make sure they aren’t slacking

Sporadically serve training images to measure drift, and refresh training. Also serve previous images to measure reliability.

Consensus

Reconstruct the final image through a weighted average. Then, send back to the researcher.

The Plan

First pass at crowdsourcing image segmentation

Upload Data

To a central server

Include some training tiles (with “ground truth”) and testing tiles (without “ground truth”)

Train

Teach the citizen scientists

Serve training images & give feedback until the player has learned enough. Then start serving test images (no feedback)

Performance Drift

Make sure they aren’t slacking

Sporadically serve training images to measure drift, and refresh training. Also serve previous images to measure reliability.

Consensus

Reconstruct the final image through a weighted average. Then, send back to the researcher.

The Plan

First pass at crowdsourcing image segmentation

Upload Data

To a central server

Include some training tiles (with “ground truth”) and testing tiles (without “ground truth”)

Train

Teach the citizen scientists

Serve training images & give feedback until the player has learned enough. Then start serving test images (no feedback)

Performance Drift

Make sure they aren’t slacking

Sporadically serve training images to measure drift, and refresh training. Also serve previous images to measure reliability.

Consensus

Reconstruct the final image through a weighted average. Then, send back to the researcher.

The Plan

First pass at crowdsourcing image segmentation

Upload Data

To a central server

Include some training tiles (with “ground truth”) and testing tiles (without “ground truth”)

Train

Teach the citizen scientists

Serve training images & give feedback until the player has learned enough. Then start serving test images (no feedback)

Performance Drift

Make sure they aren’t slacking

Sporadically serve training images to measure drift, and refresh training. Also serve previous images to measure reliability.

Consensus

Reconstruct the final image through a weighted average. Then, send back to the researcher.

The Plan

First pass at crowdsourcing image segmentation

Upload Data

To a central server

Include some training tiles (with “ground truth”) and testing tiles (without “ground truth”)

Train

Teach the citizen scientists

Serve training images & give feedback until the player has learned enough. Then start serving test images (no feedback)

Performance Drift

Make sure they aren’t slacking

Sporadically serve training images to measure drift, and refresh training. Also serve previous images to measure reliability.

Consensus

Reconstruct the final image through a weighted average. Then, send back to the researcher.

The Aims

Breaking down the project into 3 separate pieces

AIM #1

The Front End

AIM #2

The Back End

AIM #3

The Validation

The Aims

Breaking down the project into 3 separate pieces

AIM #1

The Front End

AIM #2

The Back End

AIM #3

The Validation

The Front End

The design and interaction of the application

Goals: 1) Implement draw, erase, fill, brightness/contrast, hide/show, undo features. 2) Design app to work on mobile, tablet, and laptop/desktop. 3) Create a tutorial. 4) Show feedback, scores and a leaderboard.

User experience is key! Without it, no one will play. Need to try many design iterations.

Front End Design Iteration #1

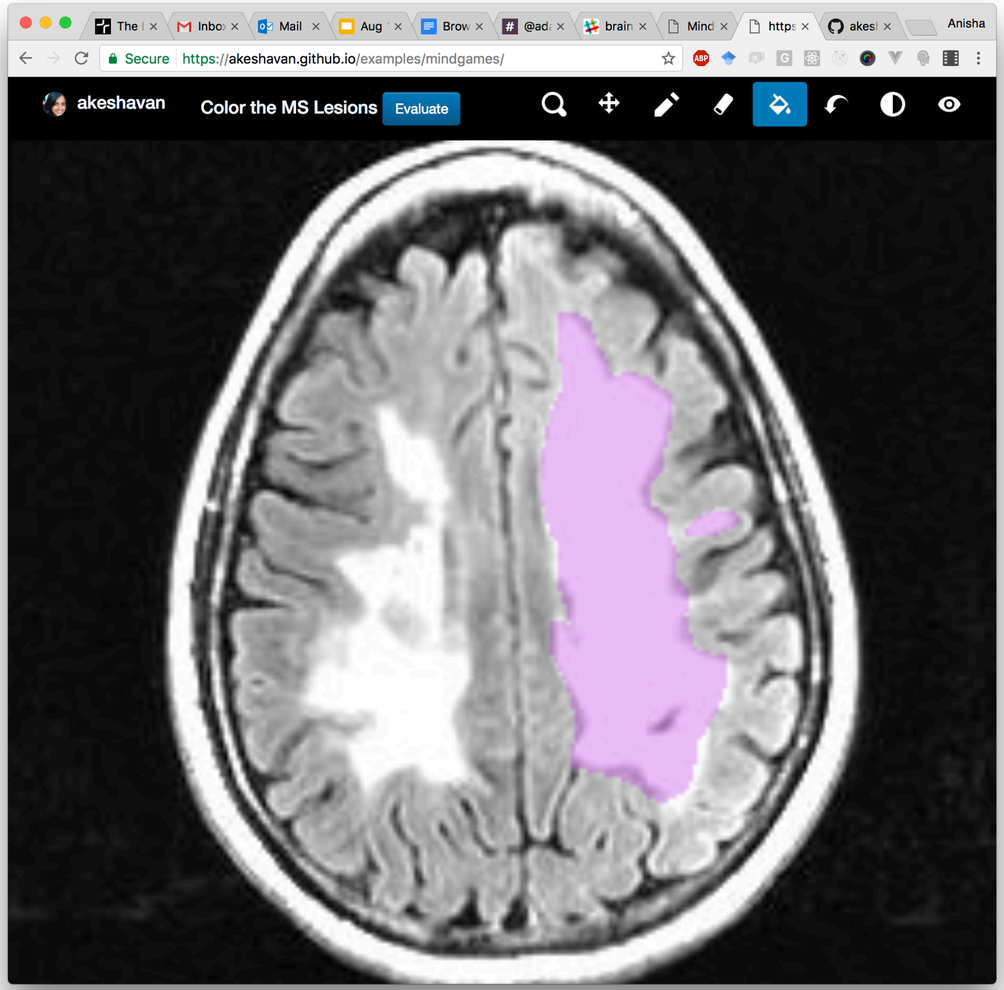
Just warming up

Basic Functionalities

zoom
pan
paint different colors
flood fill
undo
brightness/contrast

Front End Design Iteration #2

Adding more features



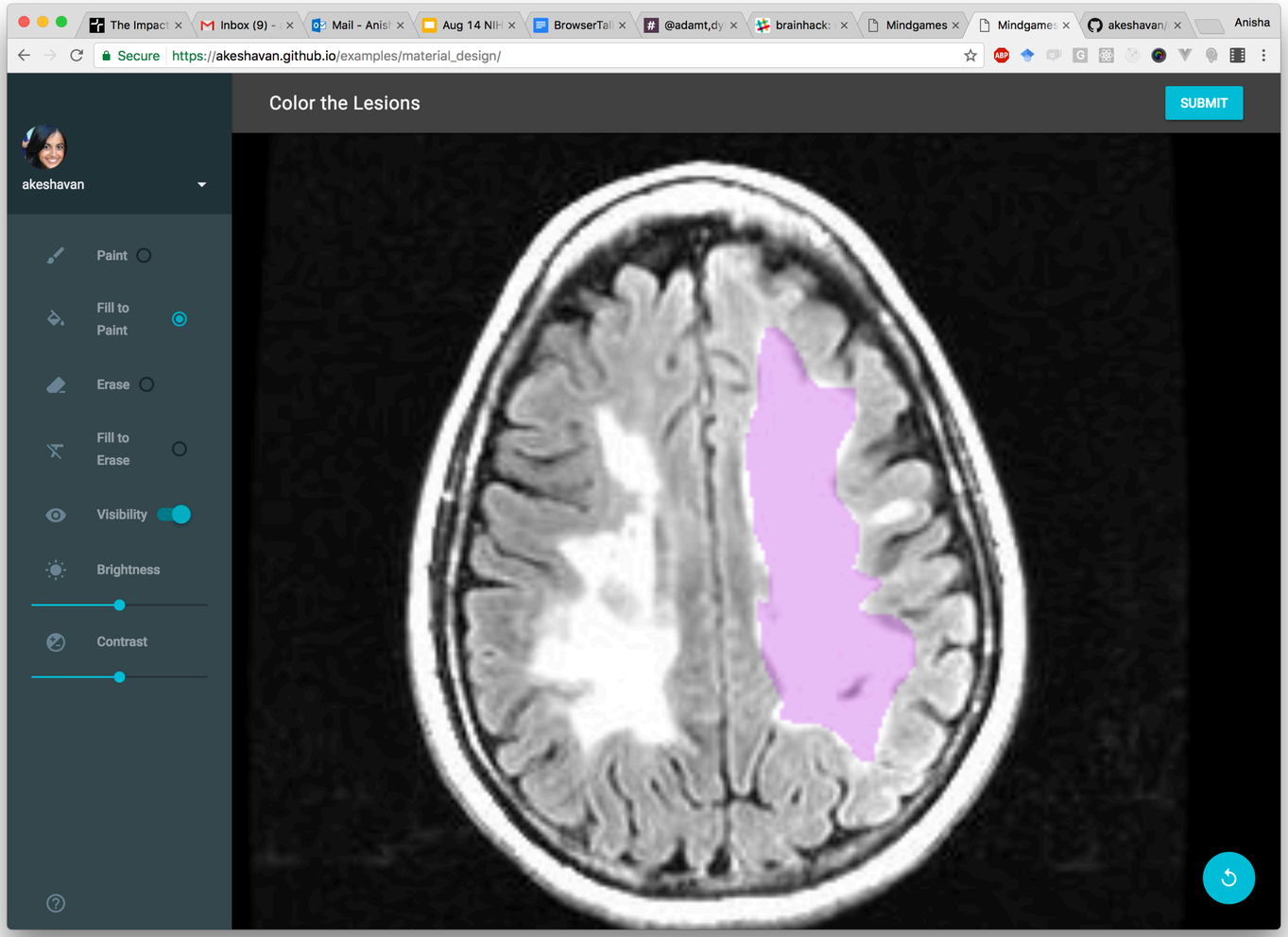
More Complexity

Changed to icons

Added authentication through GitHub

Front End Design Iteration #3

Trying a new layout



Responsive Design

Used Google's Material Design spec

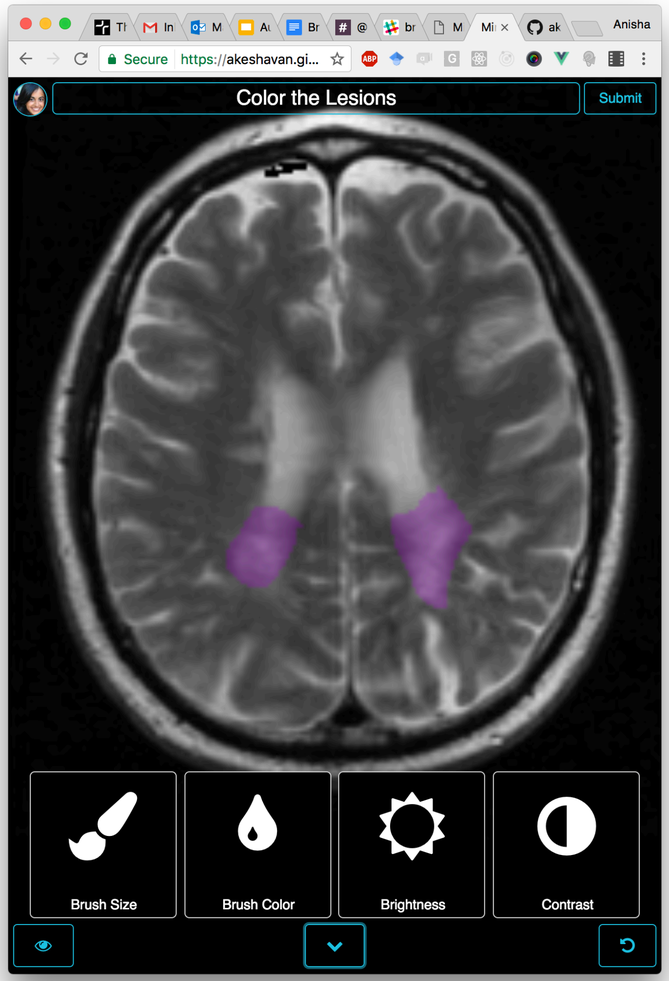
Zoom w/ scroll (desktop) or multitouch (on tablet/mobile)

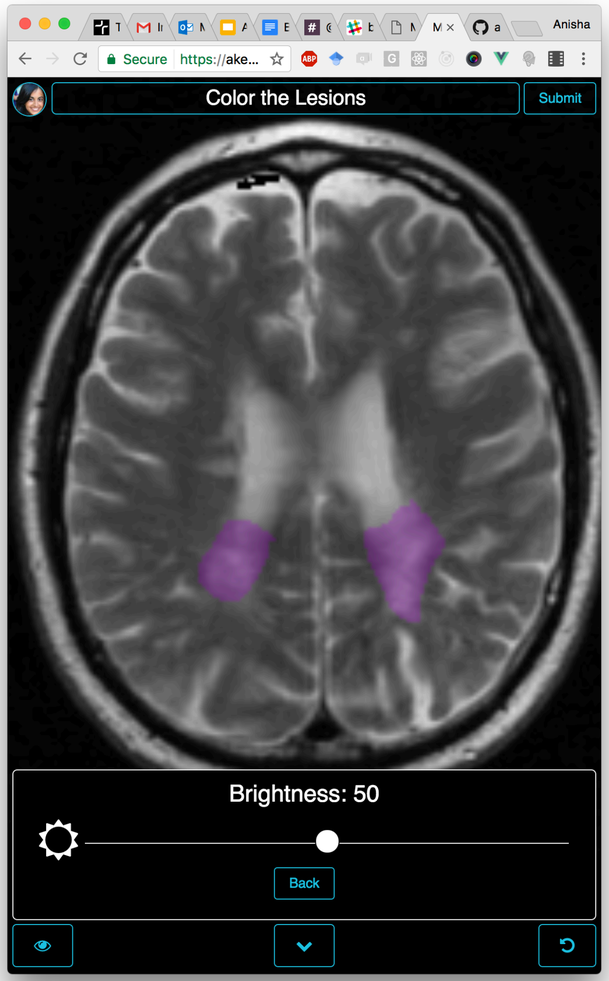
Pan w/ two fingers, or right click

Responsive design changes based on screen size

Front End Design Iteration #4

Focusing on mobile





More Complexity

Optimized for mobile layout

Instagram inspired design

Bottom navigation better than side nav

Front End Design Iteration #5

Refining the look

A More Defined Look

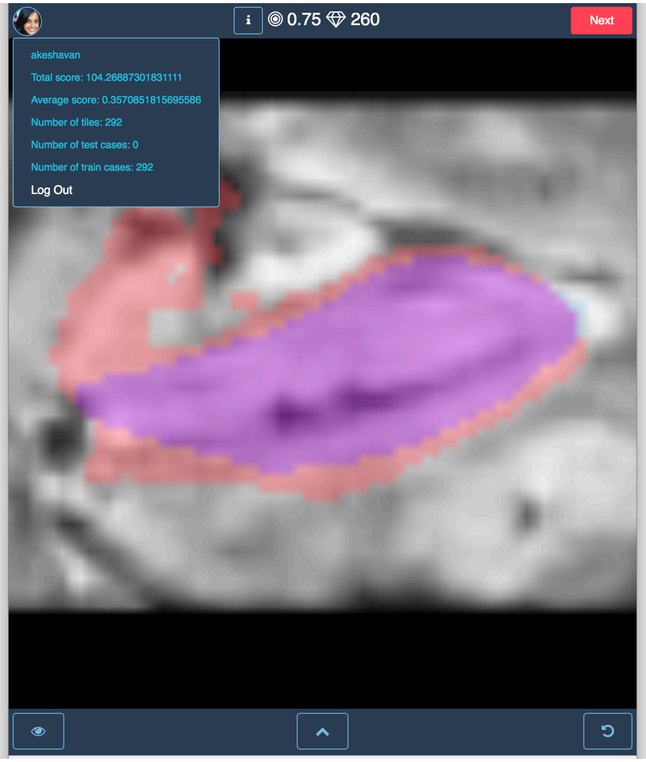
Color scheme

Smaller icons

Red “submit” button

Started tutorial (info button)

Red = false negative (you missed this part!)
Blue = false positive (you shouldn’t have drawn this part!)



Score = dice coefficient

The Aims

Breaking down the project into 3 separate pieces

AIM #1
The Front End

AIM #2
The Back End

AIM #3
The Validation

The Back End

Server-side data management

Goals: 1) Create methods, “endpoints”, to upload data, masks, new users, and their edits. 2) keep track of user’s scores 3) sample images randomly and uniformly 4) sample performance drift and reliability based on dice

This is (mostly) running on Amazon right now

Medulina

A game for segmenting the brain

Live Demo

medulina.com

Make sure you have a GitHub account

The Aims

Breaking down the project into 3 separate pieces

AIM #1

The Front End

AIM #2

The Back End

AIM #3

The Validation

The Validation

Is this feasible?

Measurements: 1) inter/intra-rater reliability 2) performance drift parameters 3) segmentation time vs performance 4) device type versus reliability 5) learning rate, and 6) calculate optimal aggregation

Future directions: Deep learning aids?

Conclusion

Embrace the web

As a Field

Browser tools becoming more common in neuroimaging

We need to continue using, creating and supporting web-based tools.

As a Developer

Learn web-dev, its worth the time investment

Your software will impact more people

As a PI

Support web-based tools and web developers

Your lab will get to the science more quickly with browser-based tools.

Contact Me

anishakeshavan@gmail.com www.github.com/akeshavan

Thanks!

Collaborators

Thanks for working on this with me!

Dylan Nielson
Adam Thomas

Future Funding

Postdoc Fellowship at UW

UW eScience Institute
UW Institute for Neuroengineering

Unofficial Mentors

(They may not know they are)

Satra Ghosh
Arno Klein
Roberto Toro
Pierre Bellec
Cameron Craddock
JB Poline

Official Mentors

Thanks for your support!

Roland Henry
Ariel Rokem
Jason Yeatman

Friends and Family

Thanks for your help!

Vidya Keshavan
Bago Amirbekian