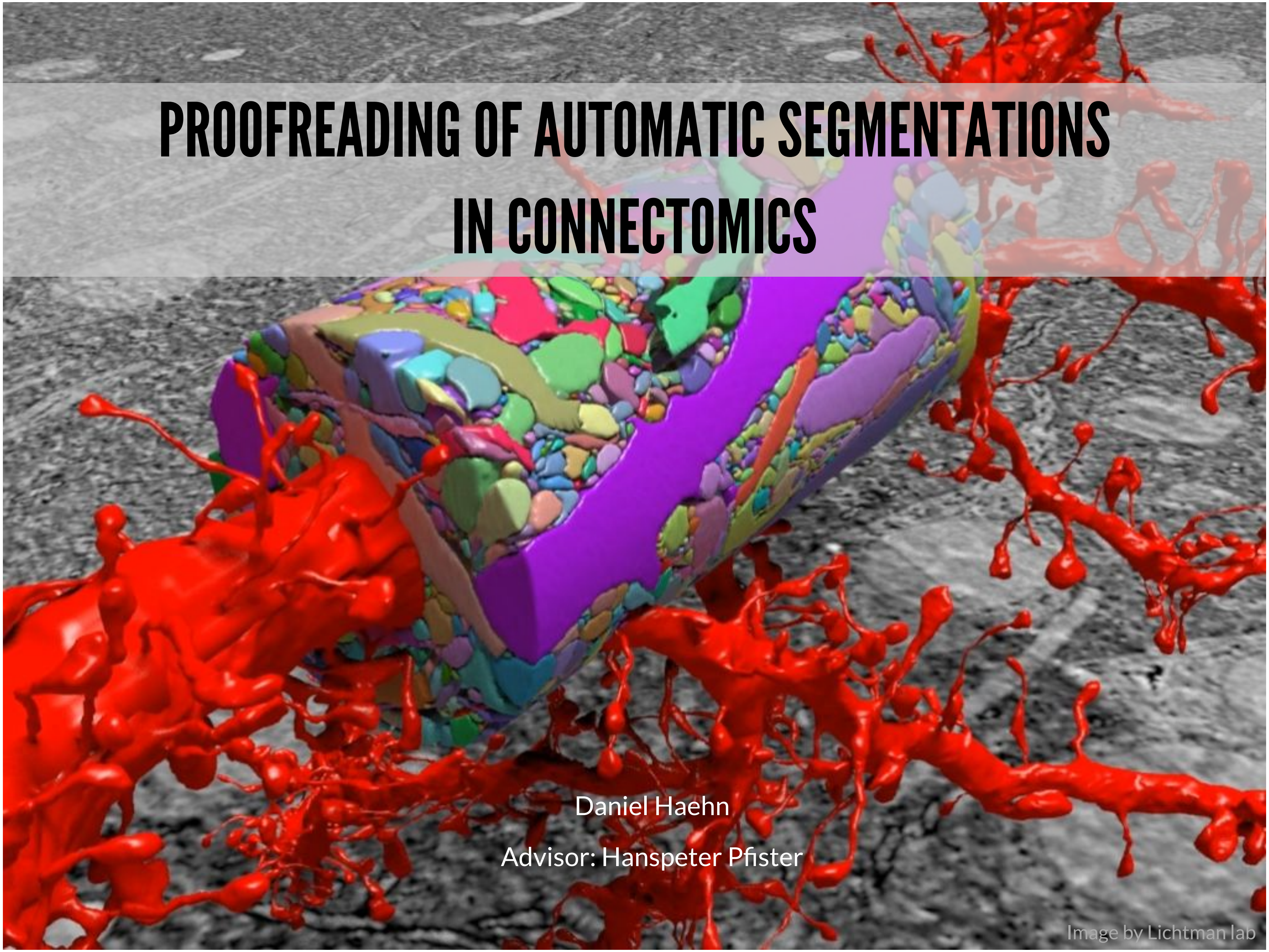
PROOFREADING OF AUTOMATIC SEGMENTATIONS IN CONNECTOMICS



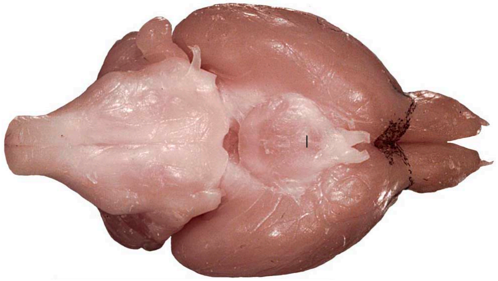
Daniel Haehn

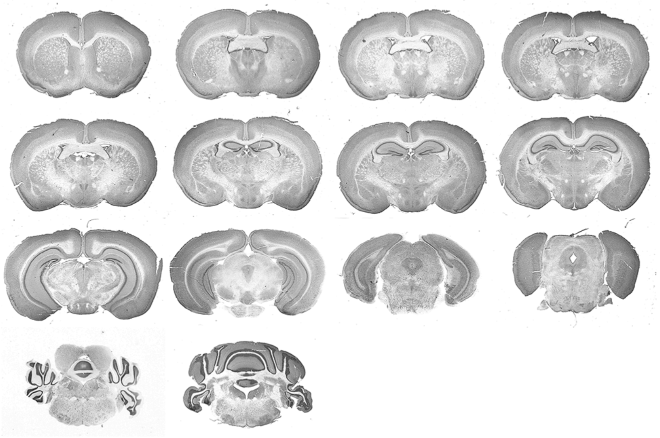
Advisor: Hanspeter Pfister

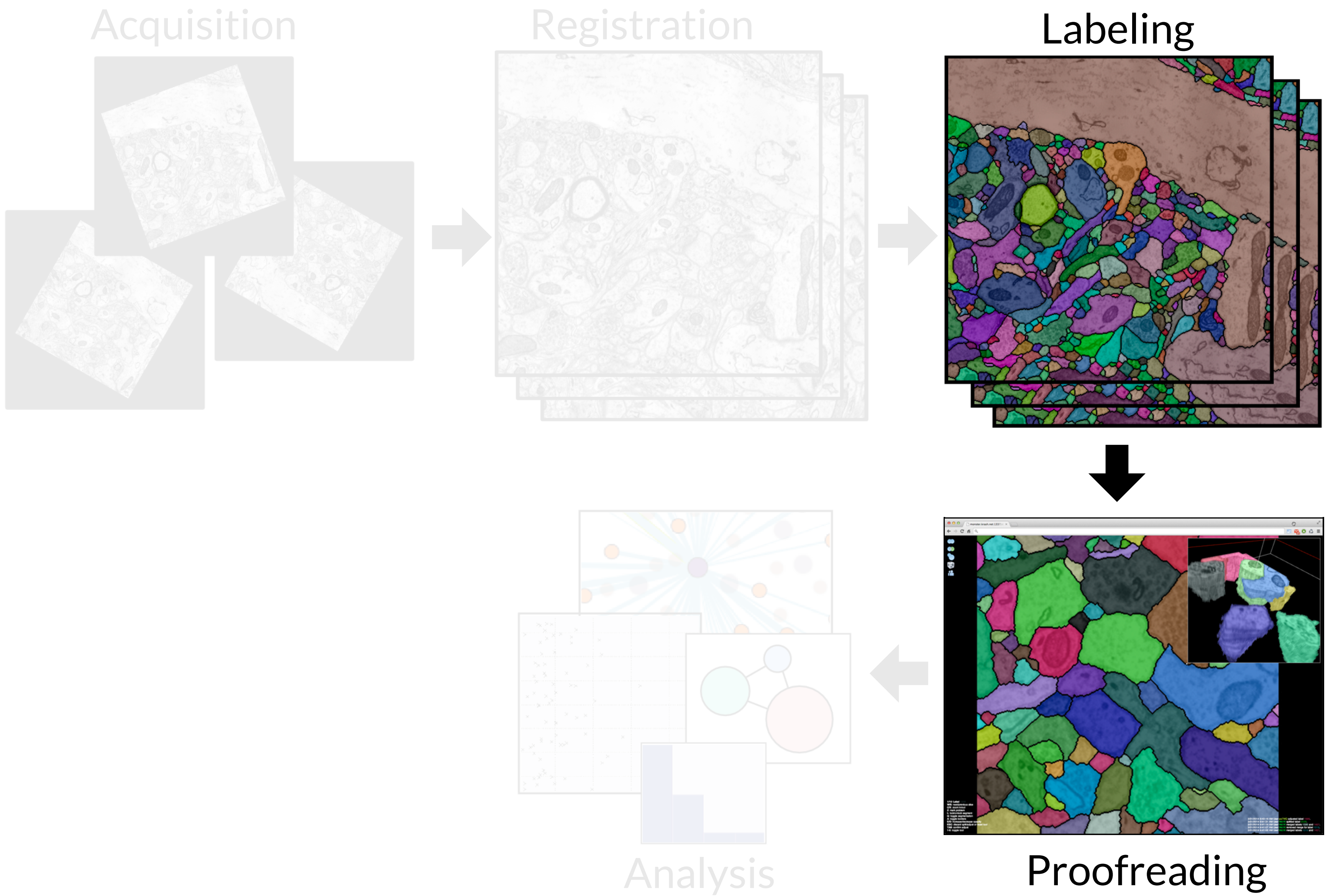
Image by Lichtman lab

CONNECTOMICS

Goal: Fully understand the wiring diagram of the brain

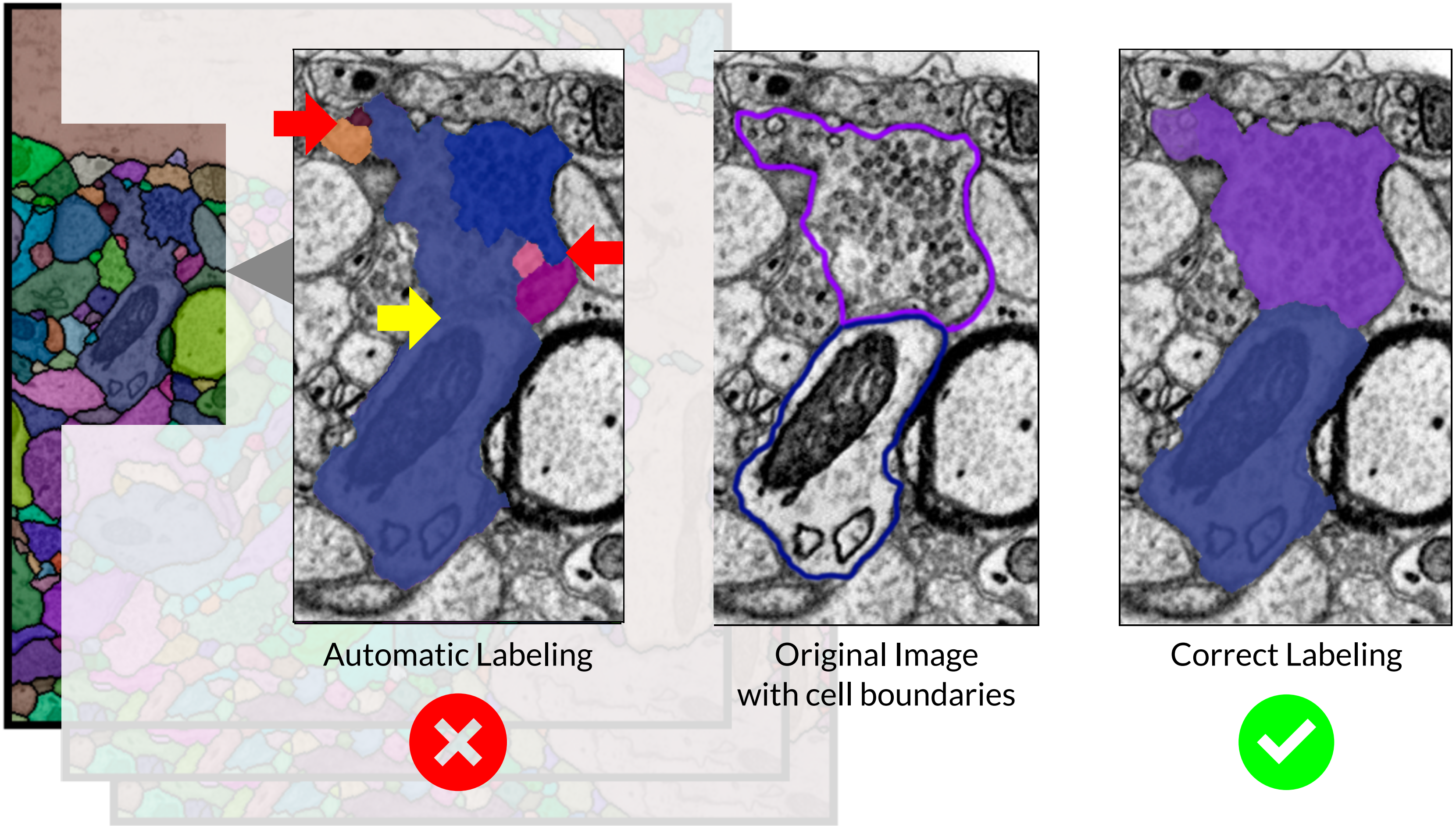






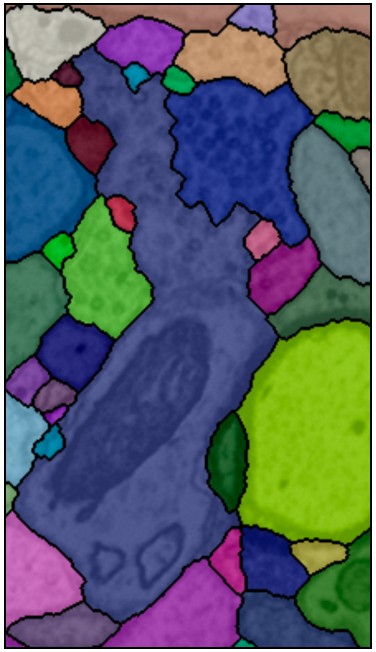
Acquisition
Registration
Labeling
Proofreading
Analysis

LABELING ERRORS

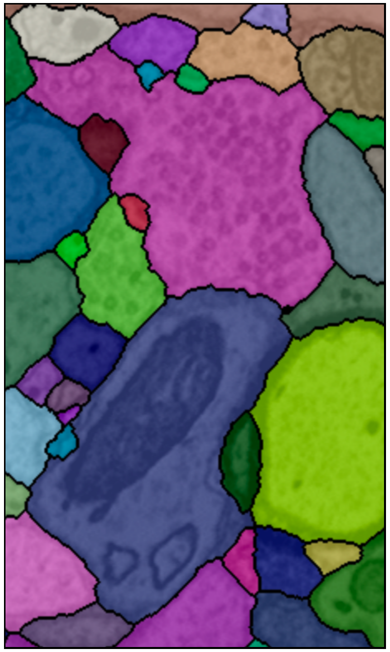


PROOFREADING

"Manual Correction of Automatic Labeling"

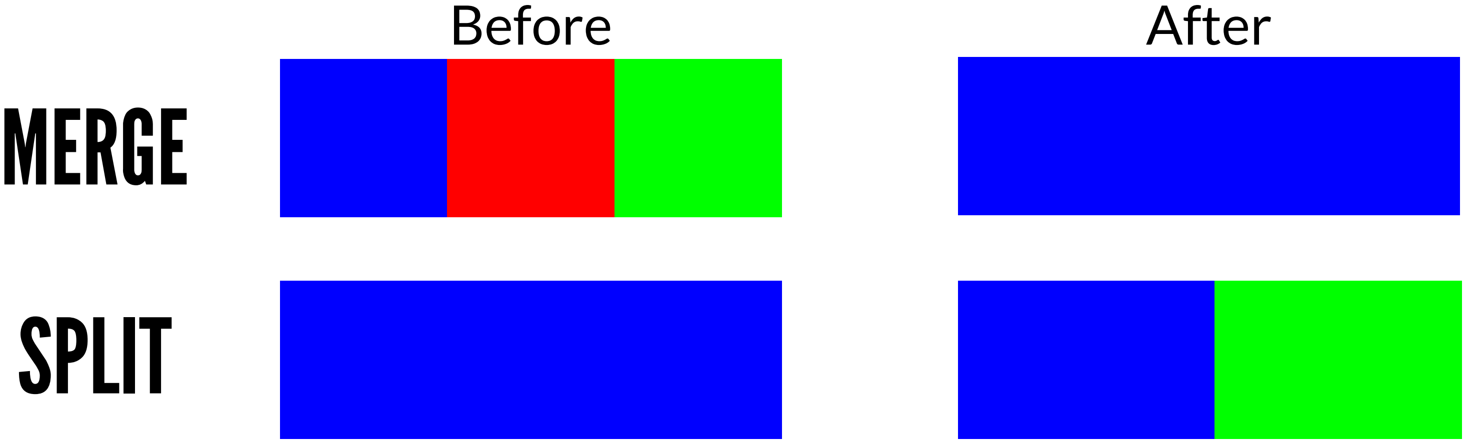


Before Proofreading



After Proofreading

PROOFREADING OPERATORS



Design and Evaluation of Interactive Proofreading Tools for Connectomics

Daniel Haehn, Seymour Knowles-Barley, Mike Roberts,
Johanna Beyer, Narayanan Kasthuri, Jeff W. Lichtman, and Hanspeter Pfister

Fig. 1: Proofreading with Dojo. *We present a web-based application for interactive proofreading of automatic segmentations of connectome data acquired via electron microscopy. Split, merge and adjust functionality enables multiple users to correct the labeling of neurons in a collaborative fashion. Color-coded structures can be explored in 2D and 3D.*

Abstract—Proofreading refers to the manual correction of automatic segmentations of image data. In connectomics, electron microscopy data is acquired at nanometer-scale resolution and results in very large image volumes of brain tissue that require fully automatic segmentation algorithms to identify cell boundaries. However, these algorithms require hundreds of corrections per cubic micron of tissue. Even though this task is time consuming, it is fairly easy for humans to perform corrections through splitting, merging, and adjusting segments during proofreading. In this paper we present the design and implementation of *Mojo*, a fully-featured single-user desktop application for proofreading, and *Dojo*, a multi-user web-based application for collaborative proofreading. We evaluate the accuracy and speed of *Mojo*, *Dojo*, and *Raveler*, a proofreading tool from Janelia Farm, through a quantitative user study. We designed a between-subjects experiment and asked non-experts to proofread neurons in a publicly available connectomics dataset. Our results show a significant improvement of corrections using web-based *Dojo*, when given the same amount of time. In addition, all participants using *Dojo* reported better usability. We discuss our findings and provide an analysis of requirements for designing visual proofreading software.

D. Haehn, et al., “Design and Evaluation of Interactive Proofreading Tools for Connectomics,” IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 2466-2475, 2014.

Guided Proofreading of Automatic Segmentations in Connectomics

Daniel Haehn[1] Verena Kaynig[1,2] James Tompkin[1] Jeff W. Lichtman[2] Hanspeter Pfister[1]

[1]Harvard Paulson School of Engineering and Applied Sciences
[2]Harvard Center for Brain Science
Cambridge, MA 02138, USA
haehn@seas.harvard.edu

Abstract

Automatic cell image segmentation methods in connectomics can lead to split *and* merge *errors, which require correction through proofreading. To aid error correction, we develop two classifiers that are able to recommend candidate errors and their corrections to the user. These classifiers are informed by training a convolutional neural network with known errors in automatic segmentations by comparison to expert-labeled ground truth. Our network architecture is able to detect potentially erroneous regions by considering a large context region around a segmentation boundary. With recommendations, proofreading of mouse cortex electron microscopy image segmentations can reduce variation of information scores on two different datasets from 0.4764 to 0.3996 and from 0.4847 to 0.3946, which we find is an improvement on pure manual and pure automatic cases.*

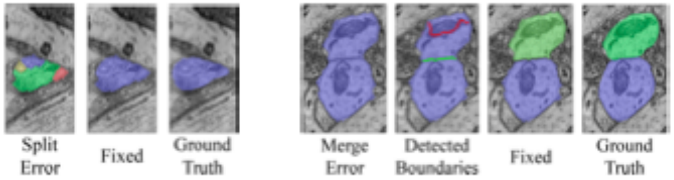


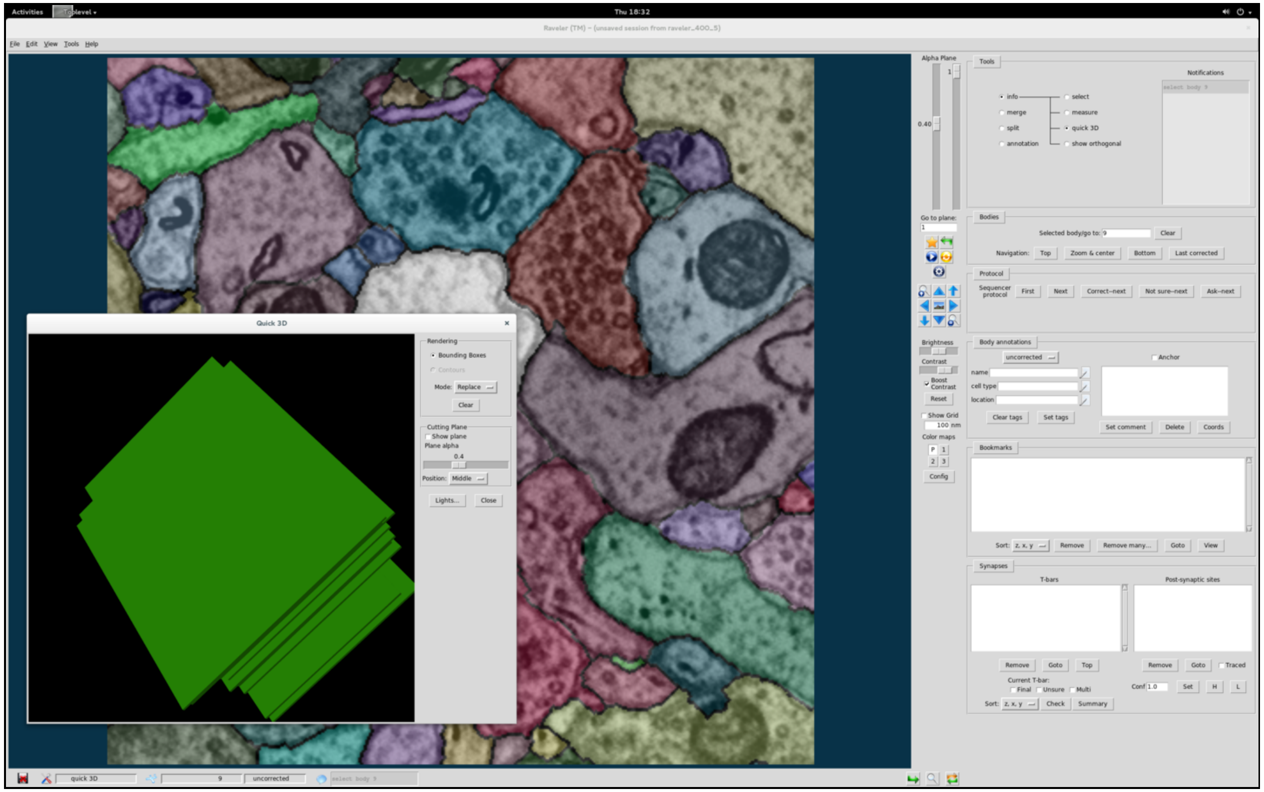
Figure 1: Split and merge error examples, their corrections, and their ground truths.

tion to be grouped into geometrically-consistent cells across registered sections, or cells are segmented across registered sections in 3D directly. Using dynamic programming techniques [22] and a GPU cluster, these classifiers can segment $\approx$ 1 terabyte of data per hour [16], which is sufficient to keep up with the 2D data capture process on state-of-the-art electron microscopes (though 3D registration is still an expensive offline operation).

All automatic methods make errors, and we are left with

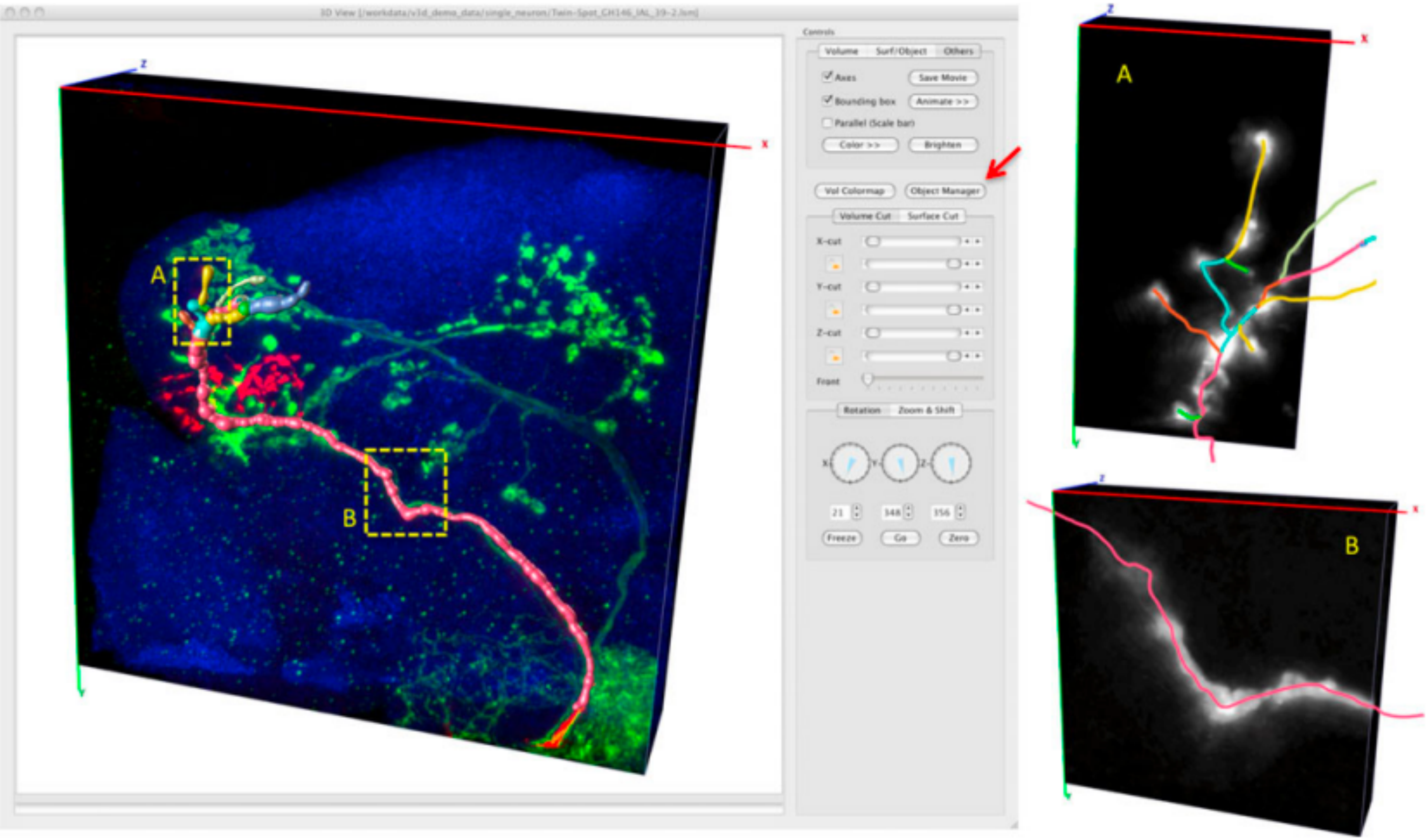
Rejected from MICCAI 2016, Submission to CVPR

RELATED WORK



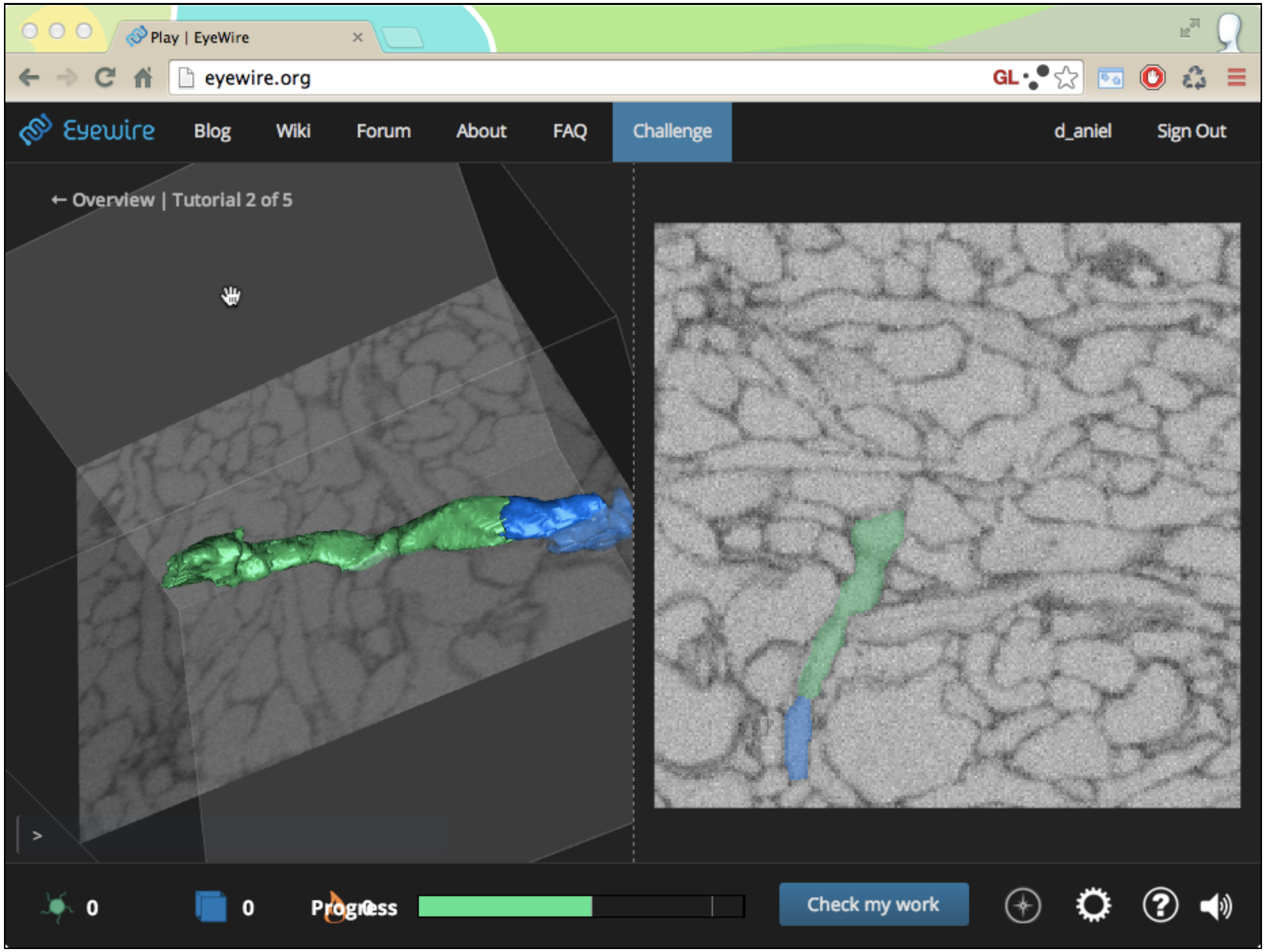
Chklovskii et al.: **Semi-automated reconstruction of neural circuits using electron microscopy**, Current Opinion in Neurobiology, 2010.

RELATED WORK (2)



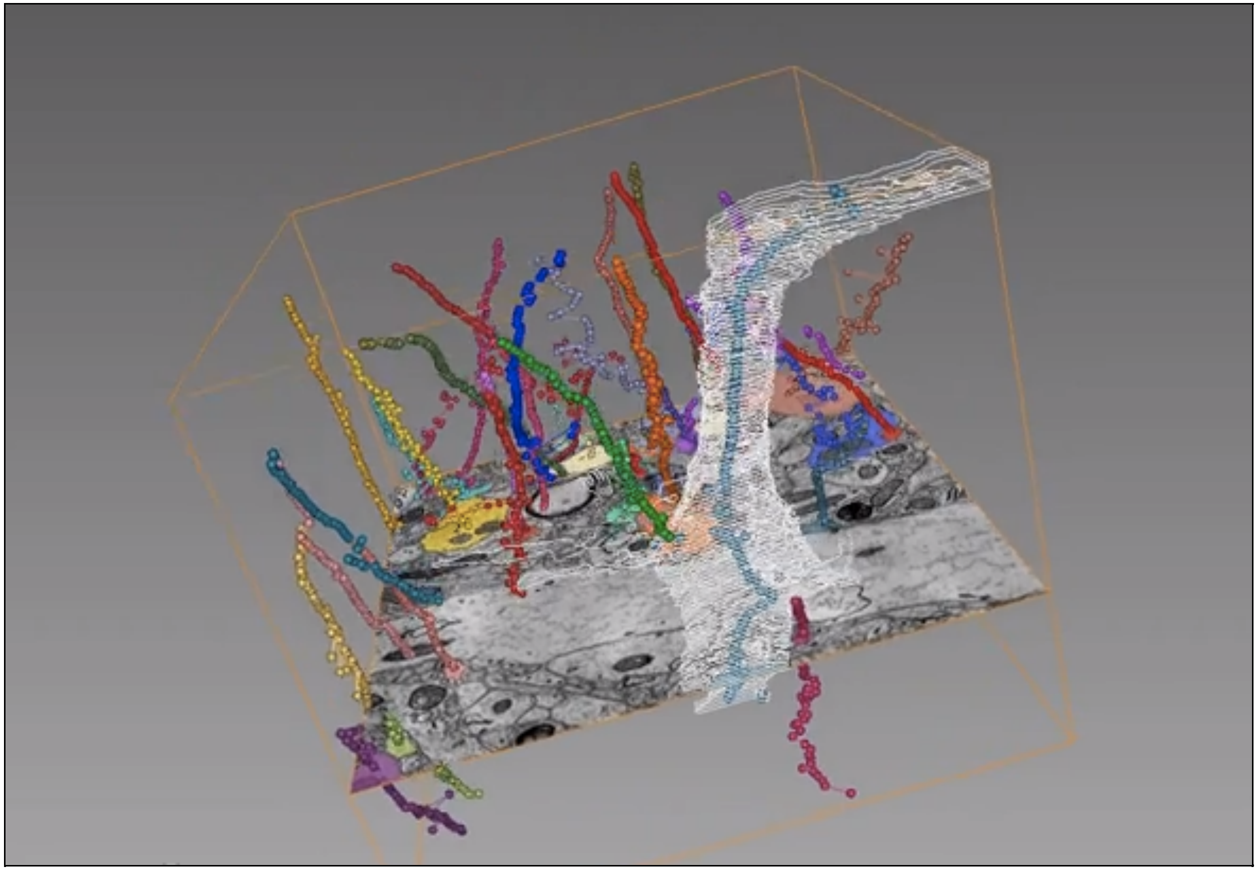
Peng et al.: **Proof-editing is the Bottleneck Of 3D Neuron Reconstruction: The Problem and Solutions**, Neuroinformatics, 2011.

RELATED WORK (3)



Seung et al.: **EyeWire,** http://eyewire.org, 2012.

RELATED WORK (4)



Sicat et al.: **Graph abstraction for simplified proofreading of slice-based volume segmentation**,

EUROGRAPHICS, 2013.

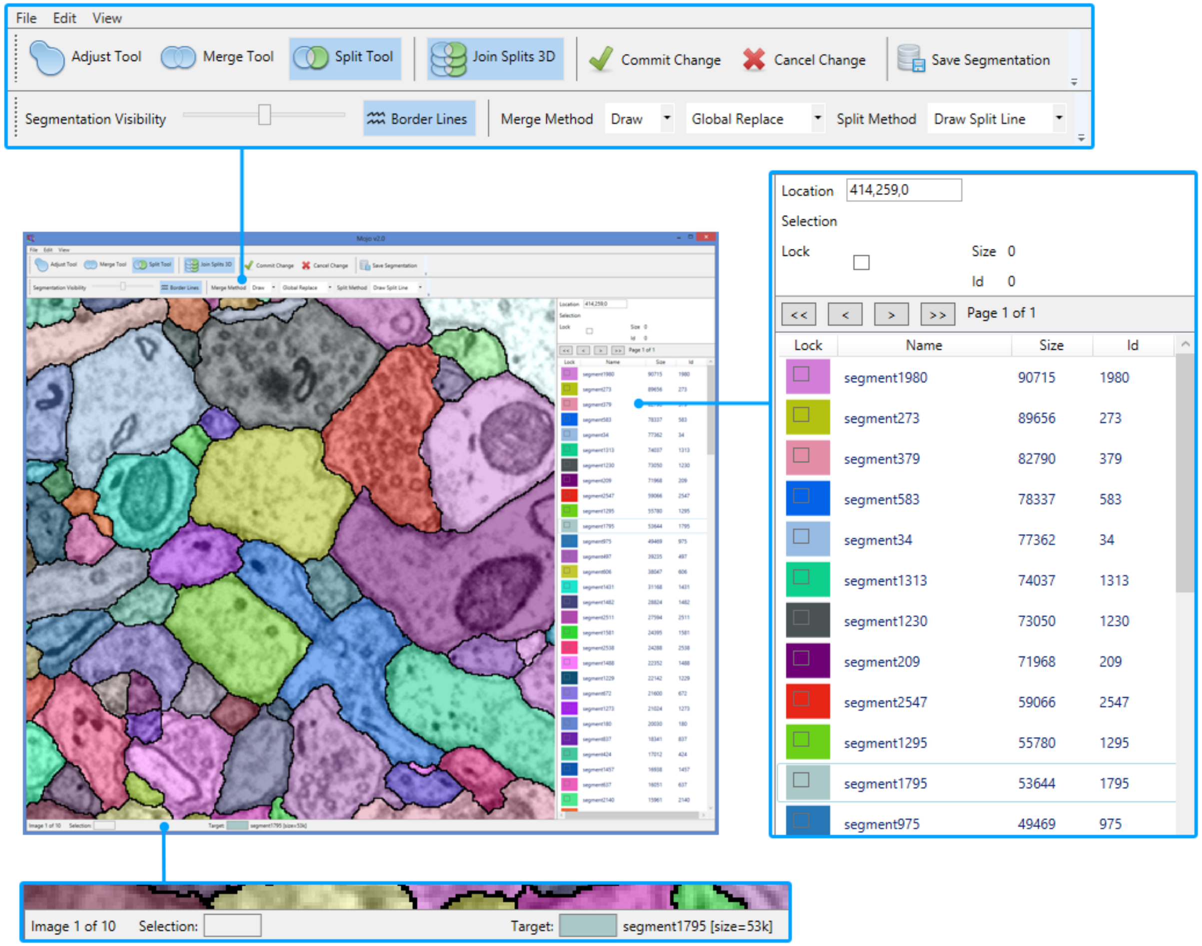
Design and Evaluation of Interactive Proofreading Tools for Connectomics

Daniel Haehn, Seymour Knowles-Barley, Mike Roberts,
Johanna Beyer, Narayanan Kasthuri, Jeff W. Lichtman, and Hanspeter Pfister

Fig. 1: Proofreading with Dojo. *We present a web-based application for interactive proofreading of automatic segmentations of connectome data acquired via electron microscopy. Split, merge and adjust functionality enables multiple users to correct the labeling of neurons in a collaborative fashion. Color-coded structures can be explored in 2D and 3D.*

Abstract—Proofreading refers to the manual correction of automatic segmentations of image data. In connectomics, electron microscopy data is acquired at nanometer-scale resolution and results in very large image volumes of brain tissue that require fully automatic segmentation algorithms to identify cell boundaries. However, these algorithms require hundreds of corrections per cubic micron of tissue. Even though this task is time consuming, it is fairly easy for humans to perform corrections through splitting, merging, and adjusting segments during proofreading. In this paper we present the design and implementation of *Mojo*, a fully-featured single-user desktop application for proofreading, and *Dojo*, a multi-user web-based application for collaborative proofreading. We evaluate the accuracy and speed of *Mojo*, *Dojo*, and *Raveler*, a proofreading tool from Janelia Farm, through a quantitative user study. We designed a between-subjects experiment and asked non-experts to proofread neurons in a publicly available connectomics dataset. Our results show a significant improvement of corrections using web-based *Dojo*, when given the same amount of time. In addition, all participants using *Dojo* reported better usability. We discuss our findings and provide an analysis of requirements for designing visual proofreading software.

MOJO



File
Edit
View
Adjust Tool
Merge Tool
Split Tool
Join Splits 3D
Commit Change
Cancel Change
Save Segmentation
Segmentation Visibility
Border Lines
Merge Method
Draw
Global Replace
Split Method
Draw Split Line
Location
414,259,0
Selection
Lock
Size 0
Id 0
<<
<
>
>>
Page 1 of 1
Lock
Name
Size
Id
segment1980 90715 1980
segment273 89656 273
segment379 82790 379
segment583 78337 583
segment34 77362 34
segment1313 74037 1313
segment1230 73050 1230
segment209 71968 209
segment2547 59066 2547
segment1295 55780 1295
segment1795 53644 1795
segment975 49469 975
Image 1 of 10
Selection:
Target:
segment1795 [size=53k]

USAGE OF MOJO

Summer interns, Researchers

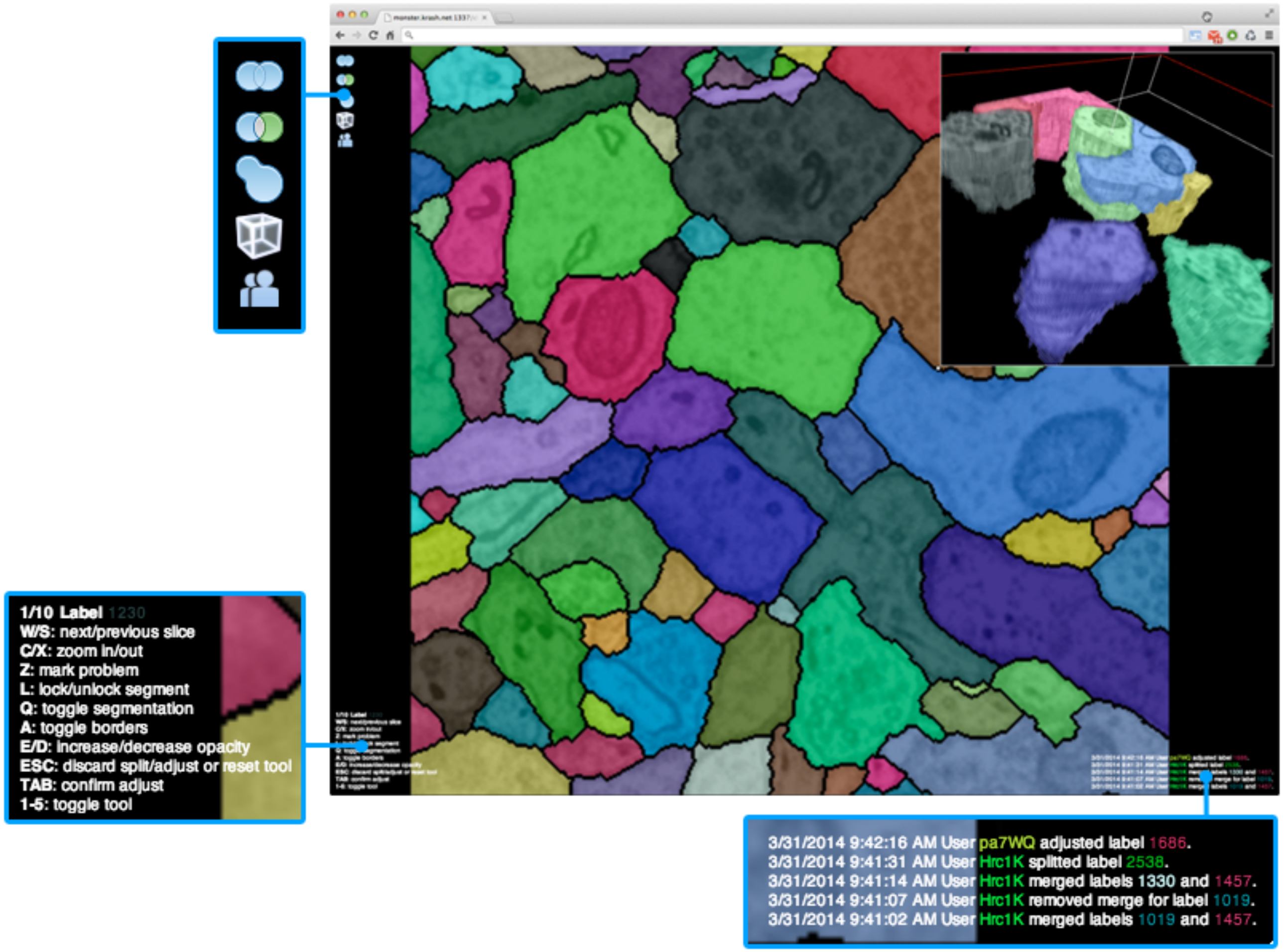
But so much and huge data..

➡ Highschool Students

Problem: very restrictive IT environment, Mojo too complicated and for single users

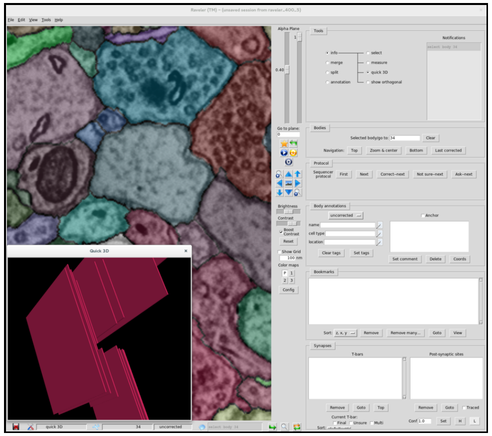
➡ Simple Web-based Solution

DOJO


1/10 Label 1230
W/S: next/previous slice
C/X: zoom in/out
Z: mark problem
L: lock/unlock segment
Q: toggle segmentation
A: toggle borders
E/D: increase/decrease opacity
ESC: discard split/adjust or reset tool
TAB: confirm adjust
1-5: toggle tool
3/31/2014 9:42:16 AM User pa7WQ adjusted label 1686.
3/31/2014 9:41:31 AM User Hrc1K splitted label 2538.
3/31/2014 9:41:14 AM User Hrc1K merged labels 1330 and 1457.
3/31/2014 9:41:07 AM User Hrc1K removed merge for label 1019.
3/31/2014 9:41:02 AM User Hrc1K merged labels 1019 and 1457.

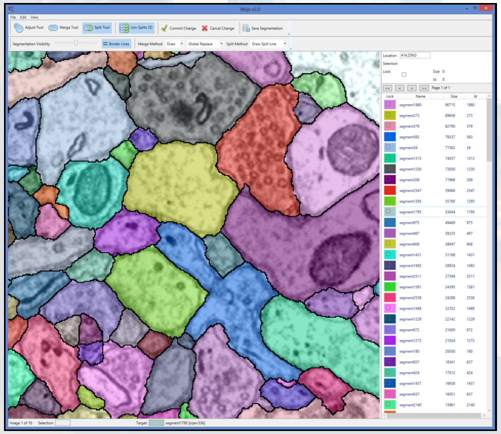
DOJO - LIVE DEMO

EXPERIMENT



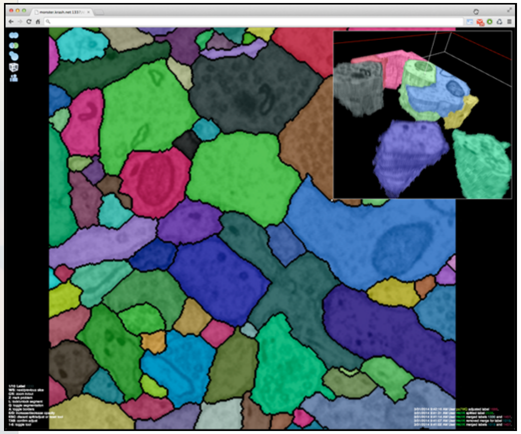
Raveler

vs



Mojo

vs



Dojo

EXPERIMENT (2)

between-subjects experiment with untrained participants (N=30)

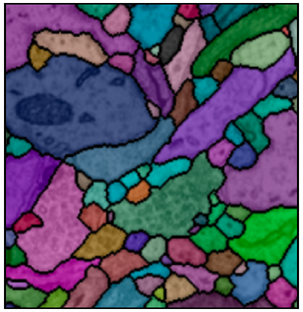
proofread a small dataset

in 30 minutes

also asked 2 experts to segment the dataset from scratch

EXPERIMENT (3)

DATASET



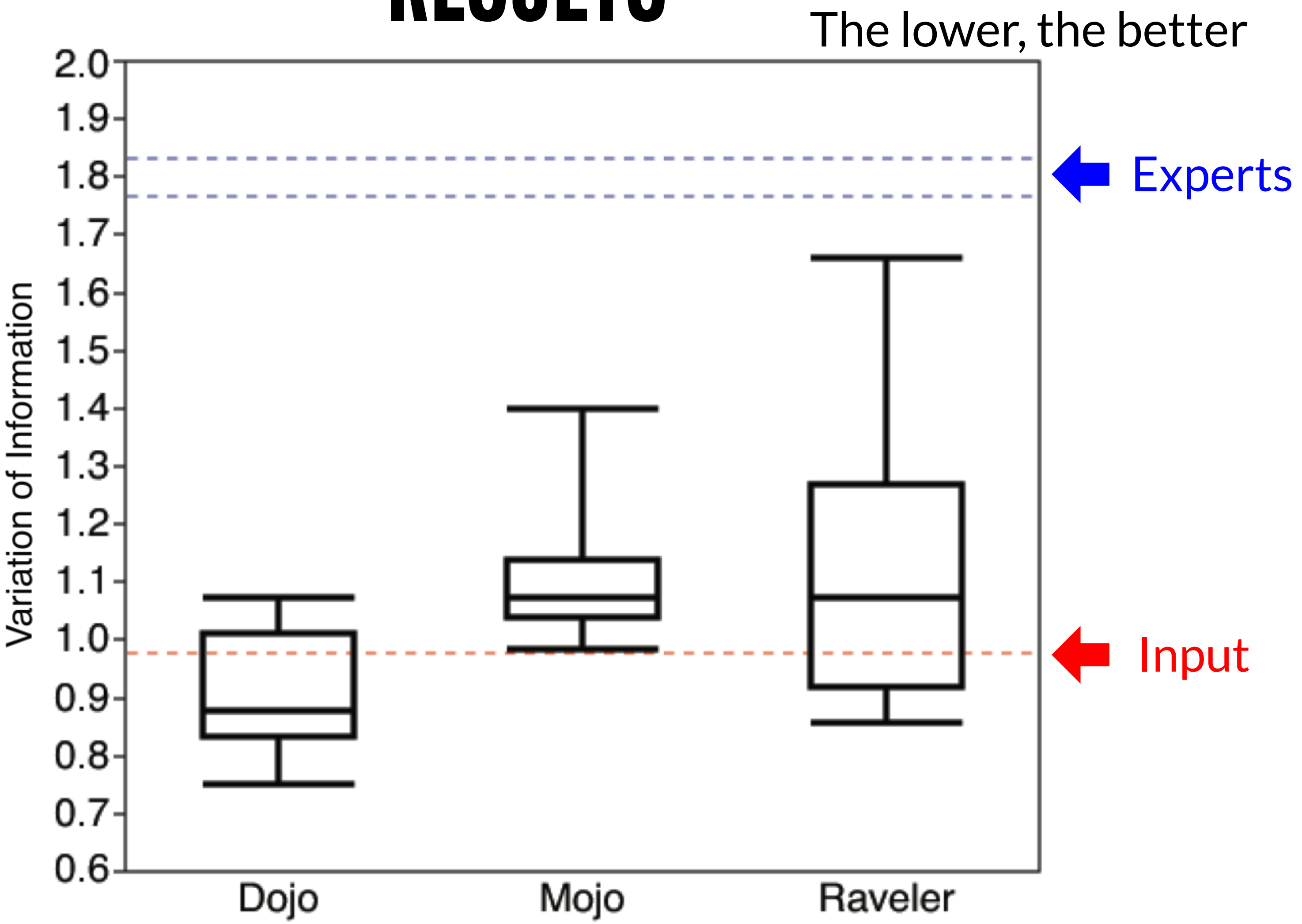
400x400x10 voxels, 6x6x30 nm/voxel

cut from publicly available dataset
(ISBI 2013 challenge)

with ground-truth available

automatically segmented

RESULTS



The lower, the better
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
Variation of Information
Dojo
Mojo
Raveler
Experts
Input

RESULTS (2)

Why was Dojo better?

- simple and minimalistic
- parameter-free operations
- 3D visualization

NeuroBlocks - Visual Tracking of Segmentation and Proofreading for Large Connectomics Projects

A. K. Al-Awami, J. Beyer, **D. Haehn**, N. Kasthuri, J. W. Lichtman, H. Pfister, M. Hadwiger
IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 738-746, 2016.

Automatic Neural Reconstruction from Petavoxel of Electron Microscopy Data

A. Suissa-Peleg, **D. Haehn**, S. Knowles-Barley, V. Kaynig, T. R. Jones, A. Wilson, R. Schalek, J. W. Lichtman, H. Pfister
Microscopy and Microanalysis, vol. 22, S3, pp. 536-537, 2016.

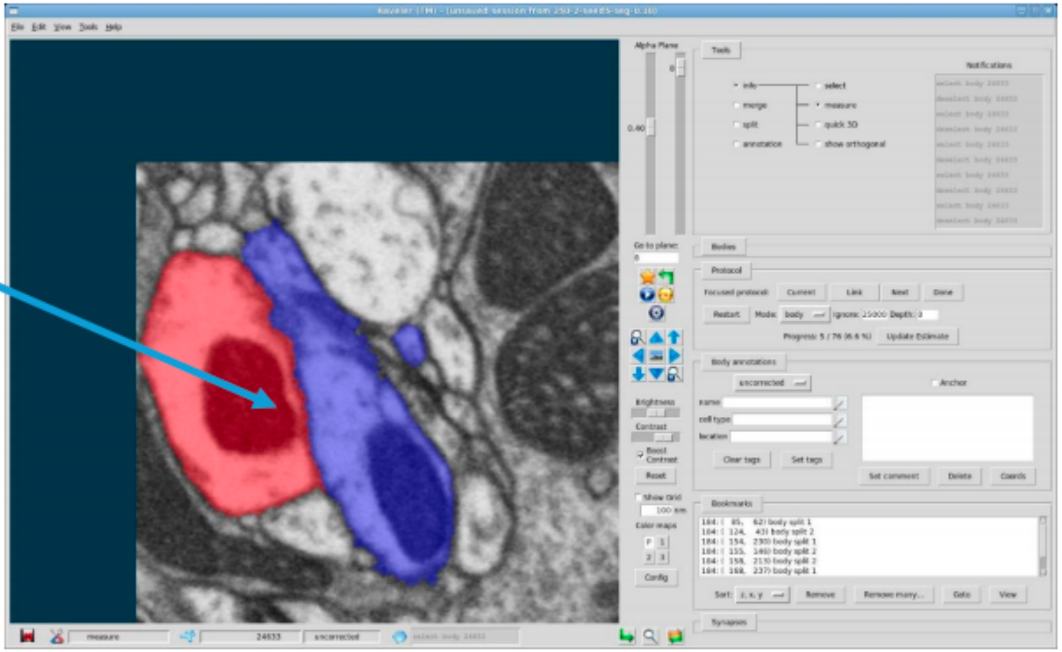
Imaging a 1 mm^3 Volume of Rat Cortex Using a MultiBeam SEM

R. Schalek, D. Lee, N. Kasthuri, A. Peleg, T. R. Jones, V. Kaynig, **D. Haehn**, H. Pfister, D. Cox, J. W. Lichtman
Microscopy and Microanalysis, vol. 22, S3, pp. 582-583, 2016.

The majority of time is spent finding errors.

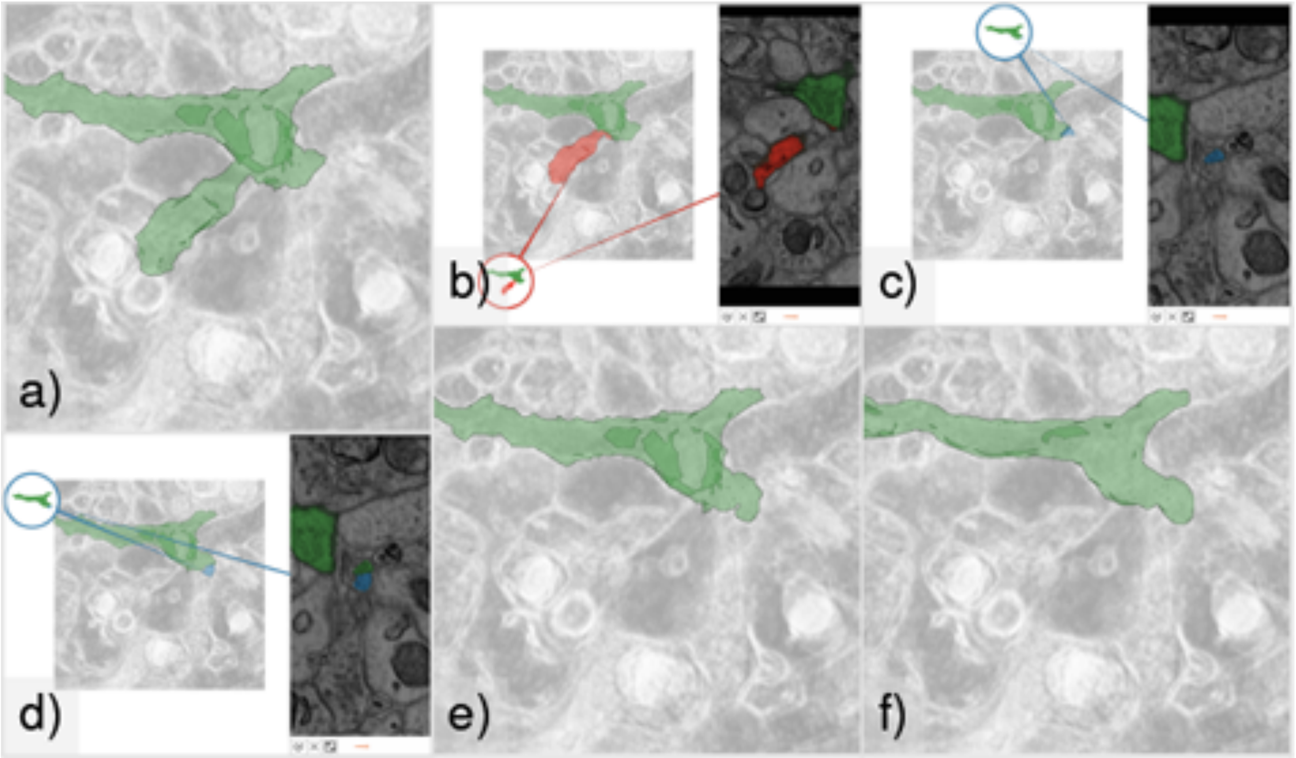
RELATED WORK (5)

focused protocol
(simple yes/no
decision)



Plaza et al.: **Focused Proofreading: Efficiently Extracting Connectomes from Segmented EM Images**, CoRR, 2014.

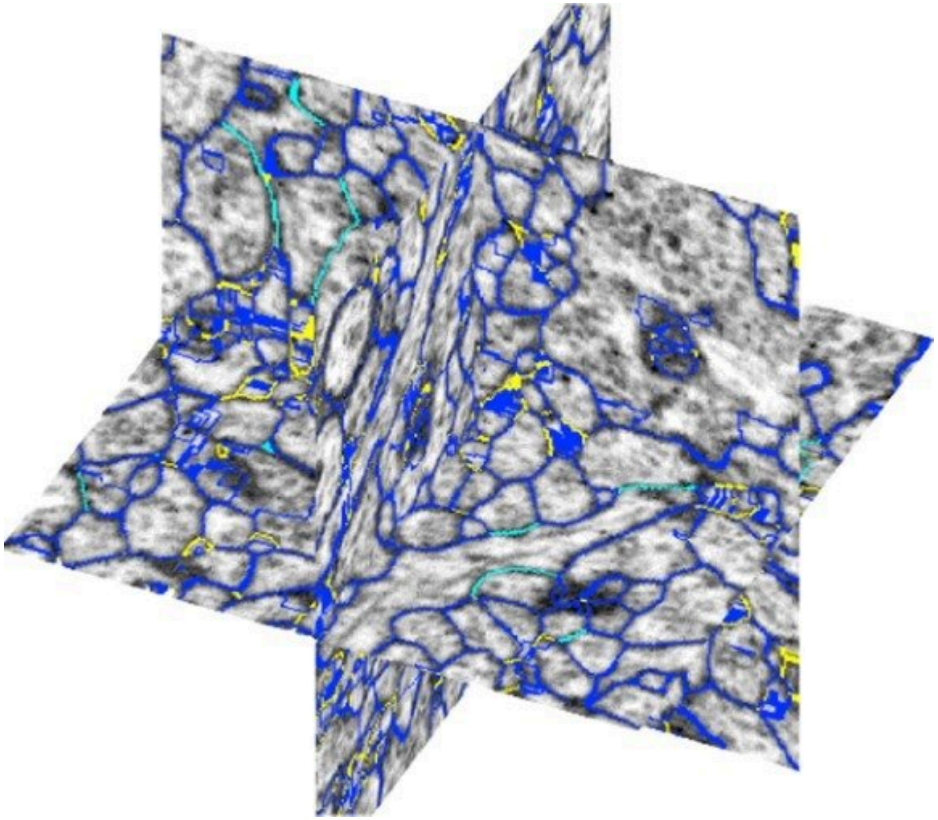
RELATED WORK (6)



Karimov et al.: **Guided Volume Editing based on Histogram Dissimilarity**,

Computer Graphics Forum, 2015.

RELATED WORK (7)



Uzunbas et al.: **An efficient conditional random field approach for automatic and interactive neuron segmentation**, Medical Image Analysis, 2016.

Guided Proofreading of Automatic Segmentations in Connectomics

Daniel Haehn[1] Verena Kaynig[1,2] James Tompkin[1] Jeff W. Lichtman[2] Hanspeter Pfister[1]

[1]Harvard Paulson School of Engineering and Applied Sciences
[2]Harvard Center for Brain Science
Cambridge, MA 02138, USA
haehn@seas.harvard.edu

Abstract

Automatic cell image segmentation methods in connectomics can lead to split *and* merge *errors, which require correction through proofreading. To aid error correction, we develop two classifiers that are able to recommend candidate errors and their corrections to the user. These classifiers are informed by training a convolutional neural network with known errors in automatic segmentations by comparison to expert-labeled ground truth. Our network architecture is able to detect potentially erroneous regions by considering a large context region around a segmentation boundary. With recommendations, proofreading of mouse cortex electron microscopy image segmentations can reduce variation of information scores on two different datasets from 0.4764 to 0.3996 and from 0.4847 to 0.3946, which we find is an improvement on pure manual and pure automatic cases.*

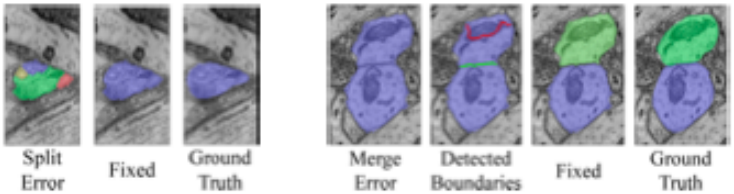


Figure 1: Split and merge error examples, their corrections, and their ground truths.

tion to be grouped into geometrically-consistent cells across registered sections, or cells are segmented across registered sections in 3D directly. Using dynamic programming techniques [22] and a GPU cluster, these classifiers can segment $\approx$ 1 terabyte of data per hour [16], which is sufficient to keep up with the 2D data capture process on state-of-the-art electron microscopes (though 3D registration is still an expensive offline operation).

All automatic methods make errors, and we are left with

GUIDED PROOFREADING - LIVE DEMO

SPLIT ERROR DETECTION

Ground Truth

Automatic Segmentation

SPLIT ERROR DETECTION (2)

Ground Truth

Automatic Segmentation

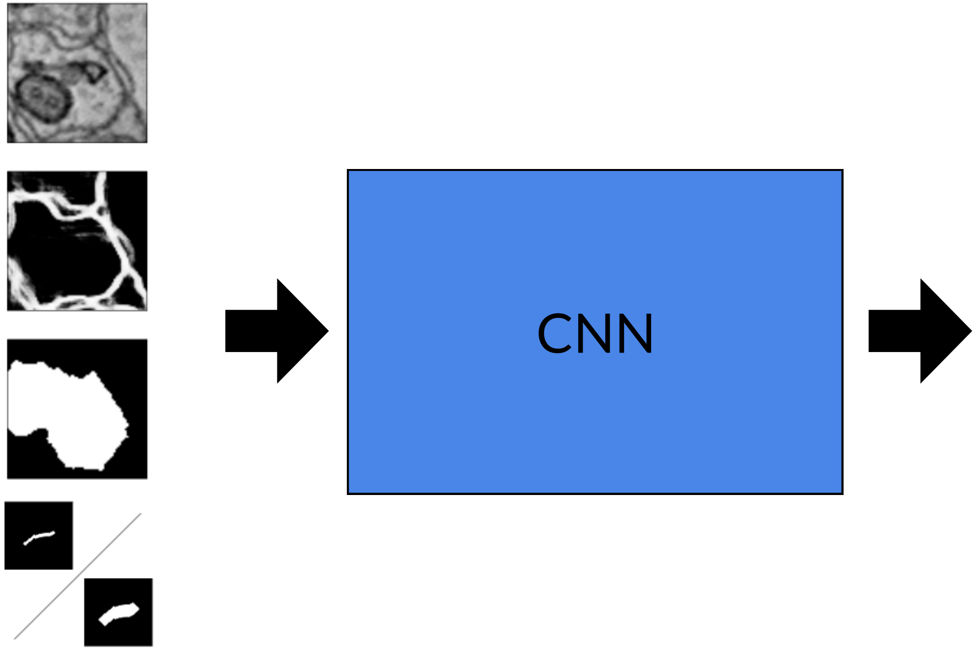
Inspired by Bogovic et al.: **Learned versus Hand-Designed Feature Representations for 3d Agglomeration**, CoRR, 2013.

Image

Membrane Prob.

Binary

Border

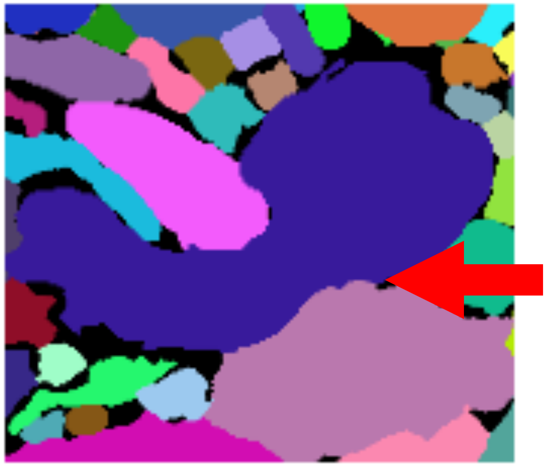


0: Correct

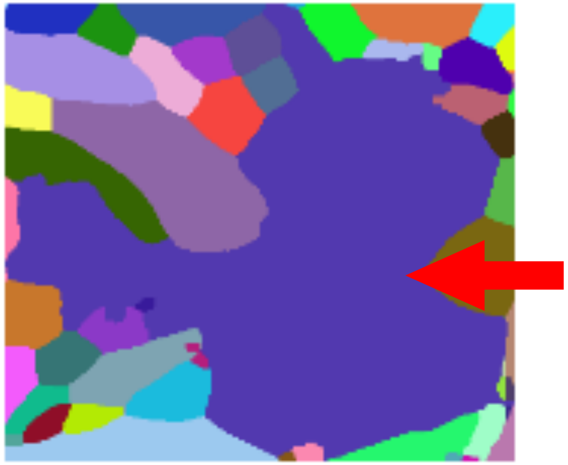
..

1: Split Error

MERGE ERROR DETECTION

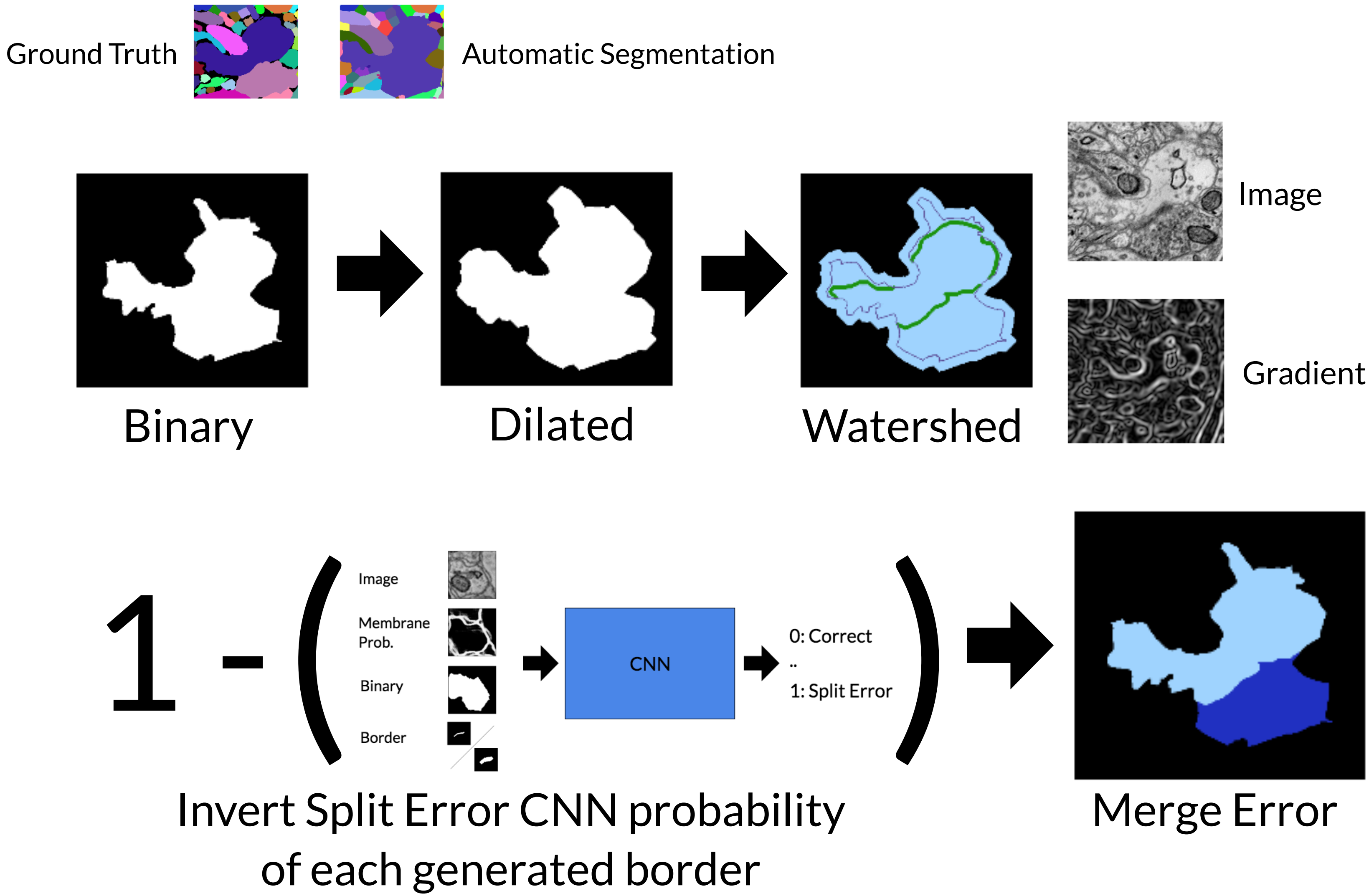


Ground Truth



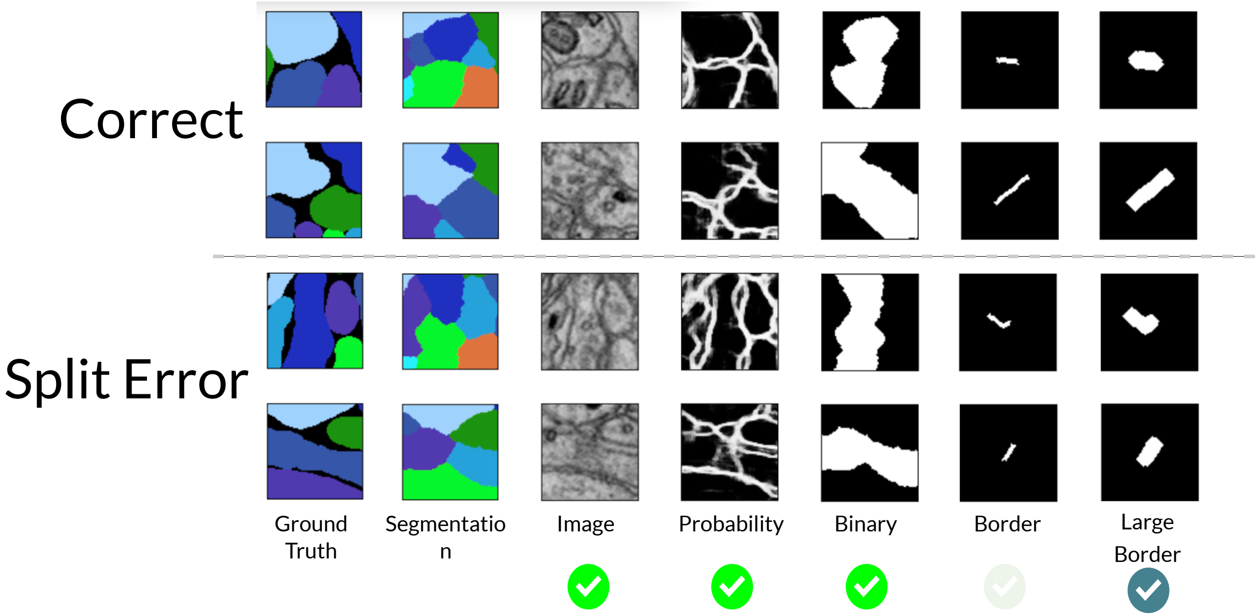
Automatic
Segmentation

MERGE ERROR DETECTION (2)

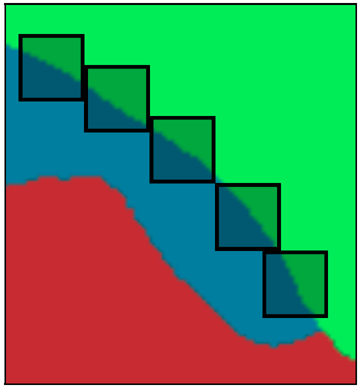


CNN INPUTS

Patch size 75x75 pixels, center pixel on membrane



Patch size covers ~80% of all segment boundaries



For larger boundaries: sample multiple times (non-overlapping) and combine (weighted arithmetic mean)

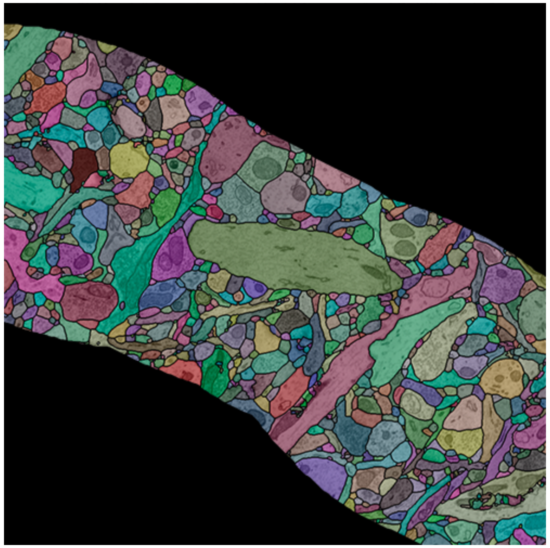
WHY 2D?

3D Alignment is very expensive

Fusion step after proofreading required (NeuroProof)

PATCH GENERATION

Left Cylinder Volume from Kasthuri et al.: **Saturated Reconstruction of a Volume of Neocortex**, Cell, 2015.



2048x2048x300 voxels, 3x3x30nm/voxel

ground truth 180k labels, automatic segmentation 500k labels

Training/Validation

2048x2048x250 voxels

266,080 correct patches

266,080 split error patches

normalized -0.5 .. 0.5

data augmentation: rotation

Test

2048x2048x50 voxels

26,816 correct patches

26,816 split error patches

normalized -0.5 .. 0.5

Patches from Segmentation rather than Ground Truth

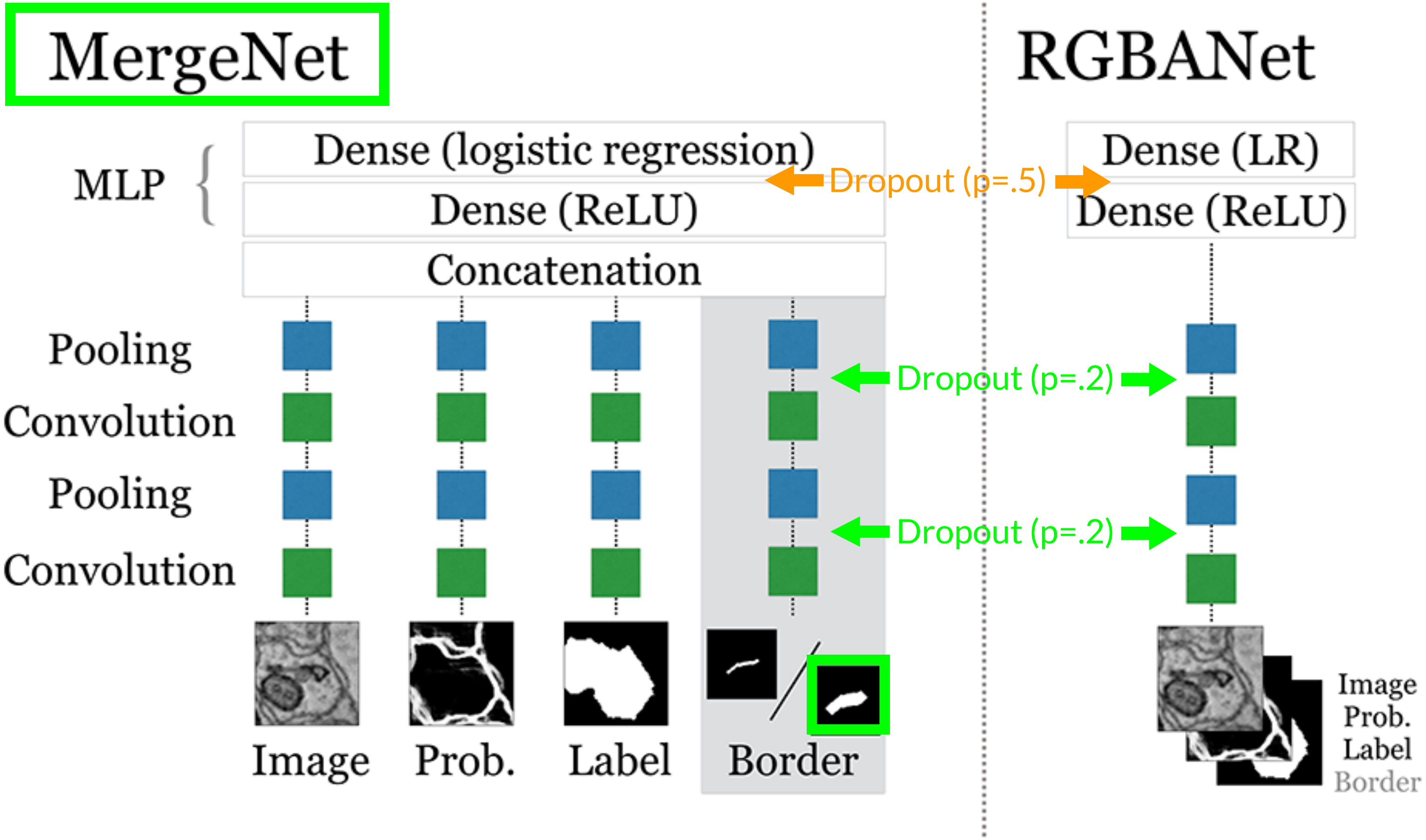
Ground Truth

Automatic Segmentation

Maximum Overlap

Wrong Boundaries

NETWORK ARCHITECTURE



Test Acc.: ~70% ~90% >91%

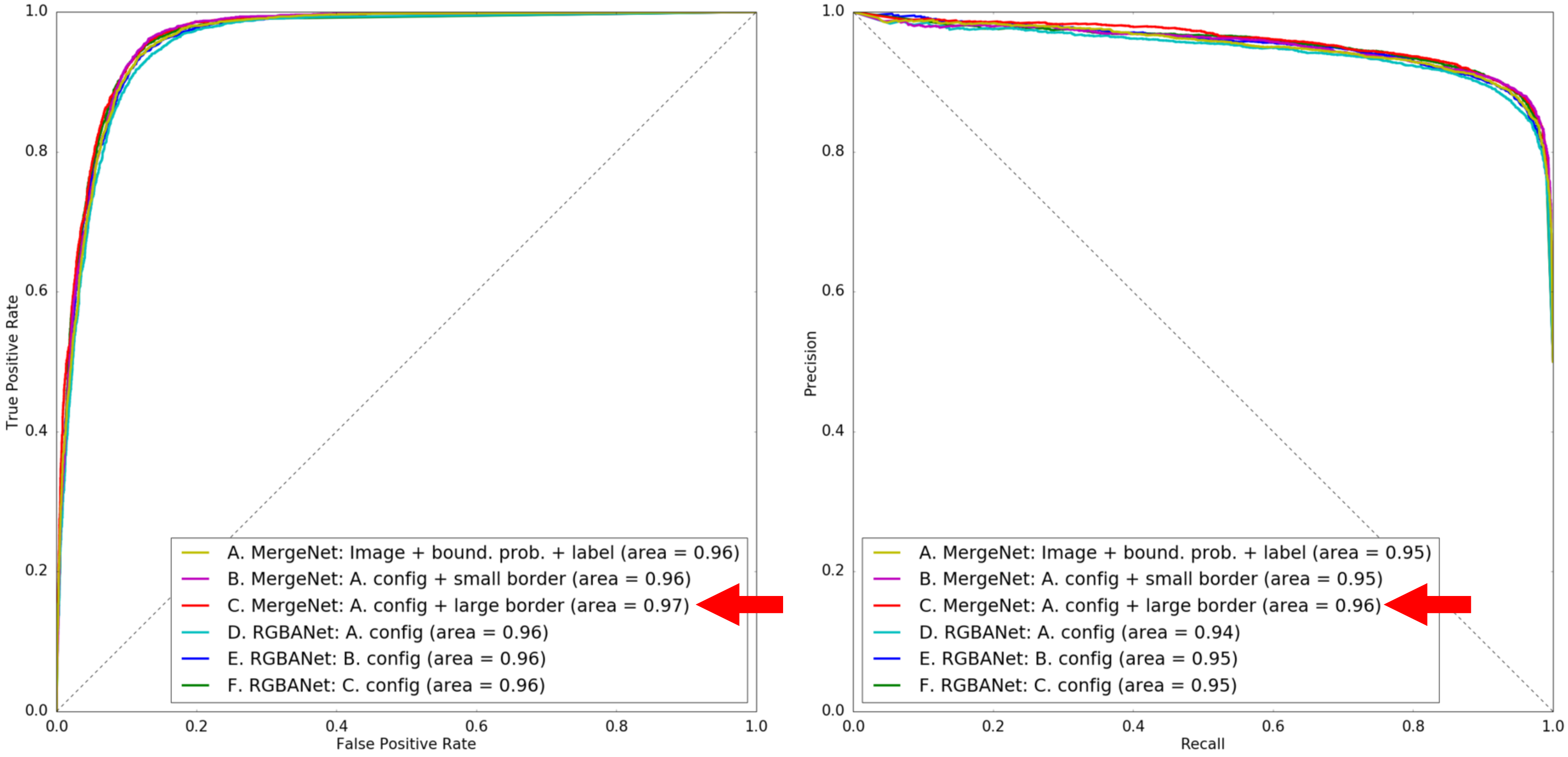
PARAMETERS

Brute force parameter permutation

3rd Conv2D layer	[true, false]	
Learning rate	[0.00001, 0.00001, 0.0001]	linear mapping between 0.03 .. 0.00001
Momentum	[0.9, 0.95, 0.99]	linear mapping between 0.9 .. 0.99
No. Filters	[16, 20, 32]	for RGBANet 64, 48
Filter size	[9x9, 13x13, 17x17]	
Mini-batch size	[10, 50, 100]	
Epochs	[500, 2000]	with patience counter 50

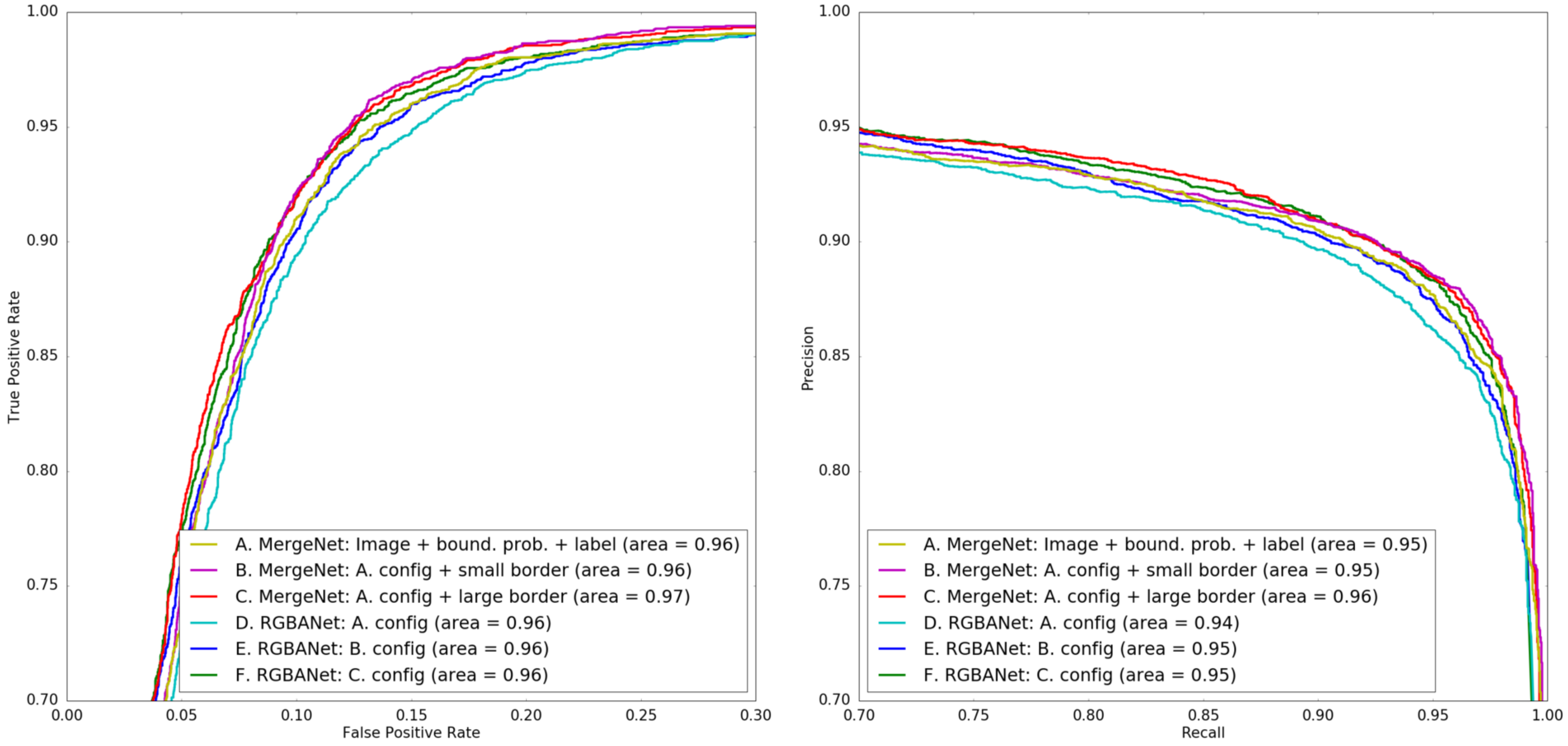
~1.5 million parameters for MergeNet and RGBANet

RESULTS



All configurations very similar, choosing **C**.

RESULTS (2)



True Positive Rate
False Positive Rate
A. MergeNet: Image + bound. prob. + label (area = 0.96)
B. MergeNet: A. config + small border (area = 0.96)
C. MergeNet: A. config + large border (area = 0.97)
D. RGBANet: A. config (area = 0.96)
E. RGBANet: B. config (area = 0.96)
F. RGBANet: C. config (area = 0.96)
Precision
Recall
A. MergeNet: Image + bound. prob. + label (area = 0.95)
B. MergeNet: A. config + small border (area = 0.95)
C. MergeNet: A. config + large border (area = 0.96)
D. RGBANet: A. config (area = 0.94)
E. RGBANet: B. config (area = 0.95)
F. RGBANet: C. config (area = 0.95)

AUTOMATIC CORRECTION

1. Find and correct merge errors

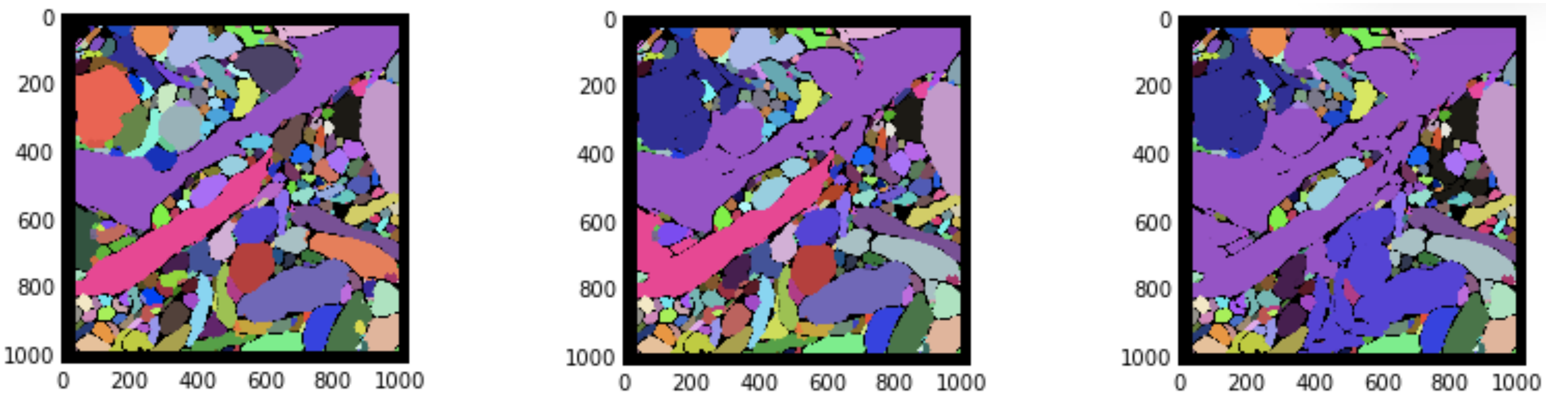
```
for all segments:
   generate N borders
   rank all generated borders
   choose lowest rated border and split segment
```

1. - p_t

2. Find and correct split errors

```
for all segments:
    rank borders and store ranking
    for next highest rated border:
       merge segments
       re-rank borders of merged segment
```

p_t = 0.95



USER IN THE LOOP

User double-checks suggestions of CNN

Tested our system with 1 novice and 2 experts

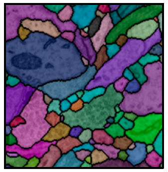
- ~3.2 seconds per correction with Guided Proofreading
- ~30 seconds per correction with Dojo

Simulated User

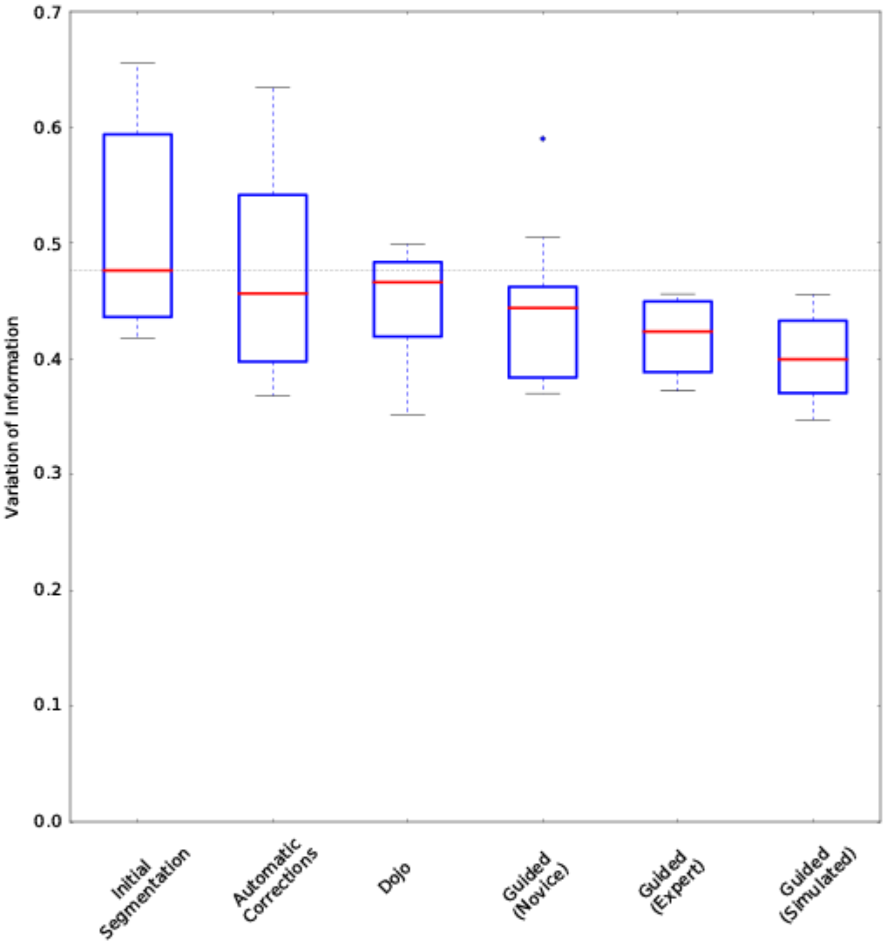
- Only accepts corrections which reduce VI
- Simulation budget: 5 seconds per correction
- Configurable error rate

EXPERIMENTS

Guided Proofreading vs. best Dojo user (30 minutes, same data)

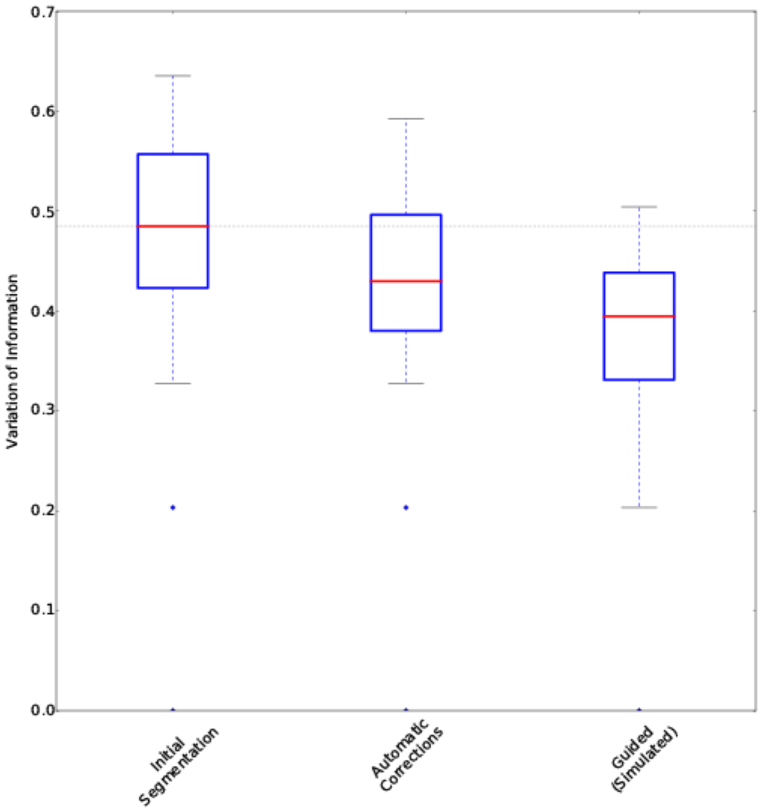


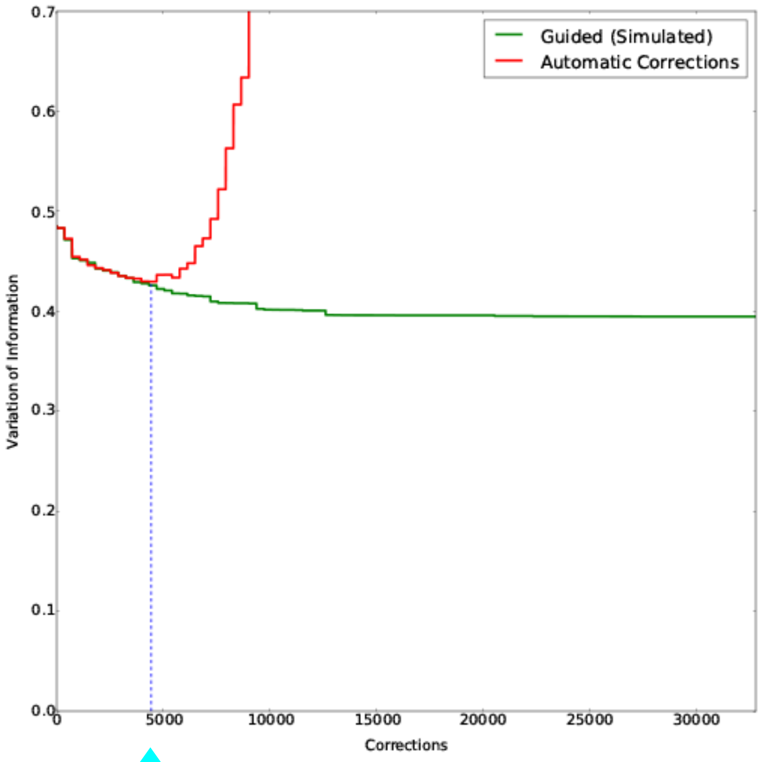
18 merge errors
842 split errors



EXPERIMENTS (2)

2k x 2k x 50 voxels, Automatic Corrections vs. Simulated User

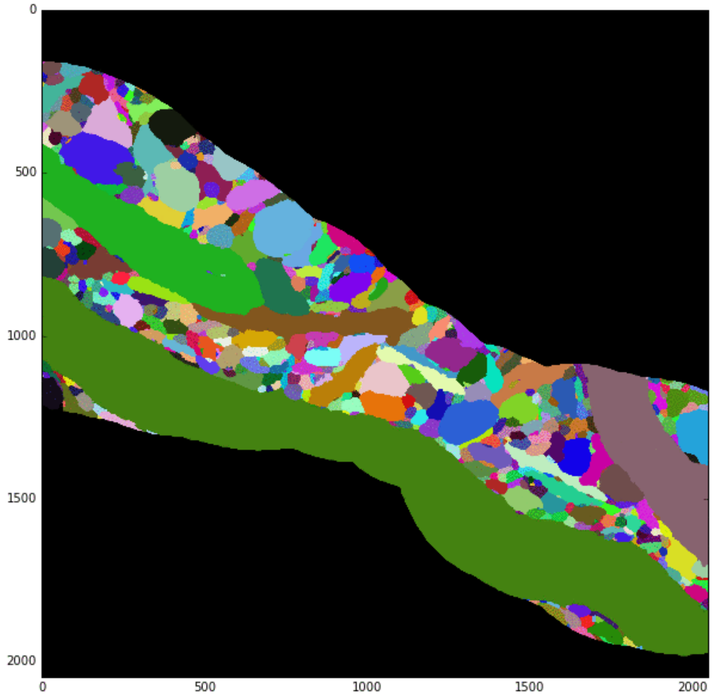




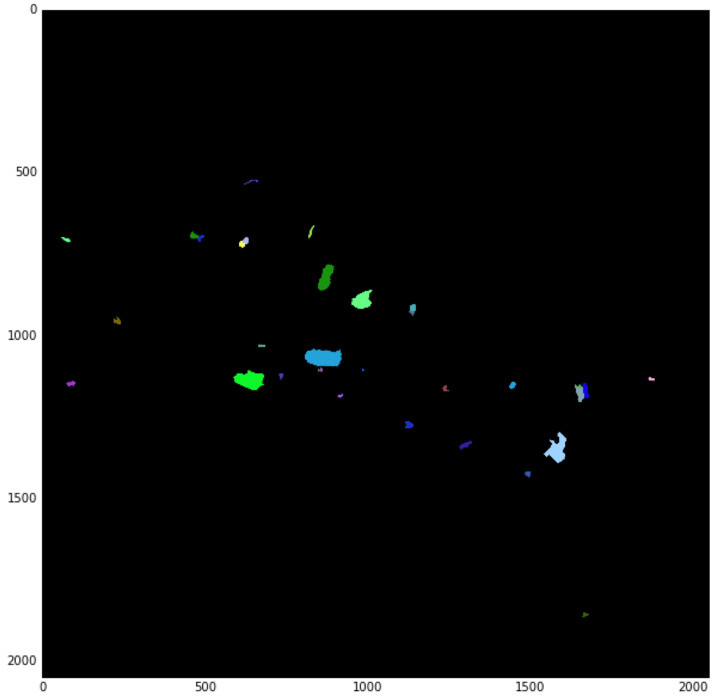
p_t = 0.95

EXPERIMENTS (3)

2k x 2k x 50 voxels, Automatic Corrections (p_t=.95)



Snapshot every 30 corrections



Difference to input segmentation

LIMITATIONS

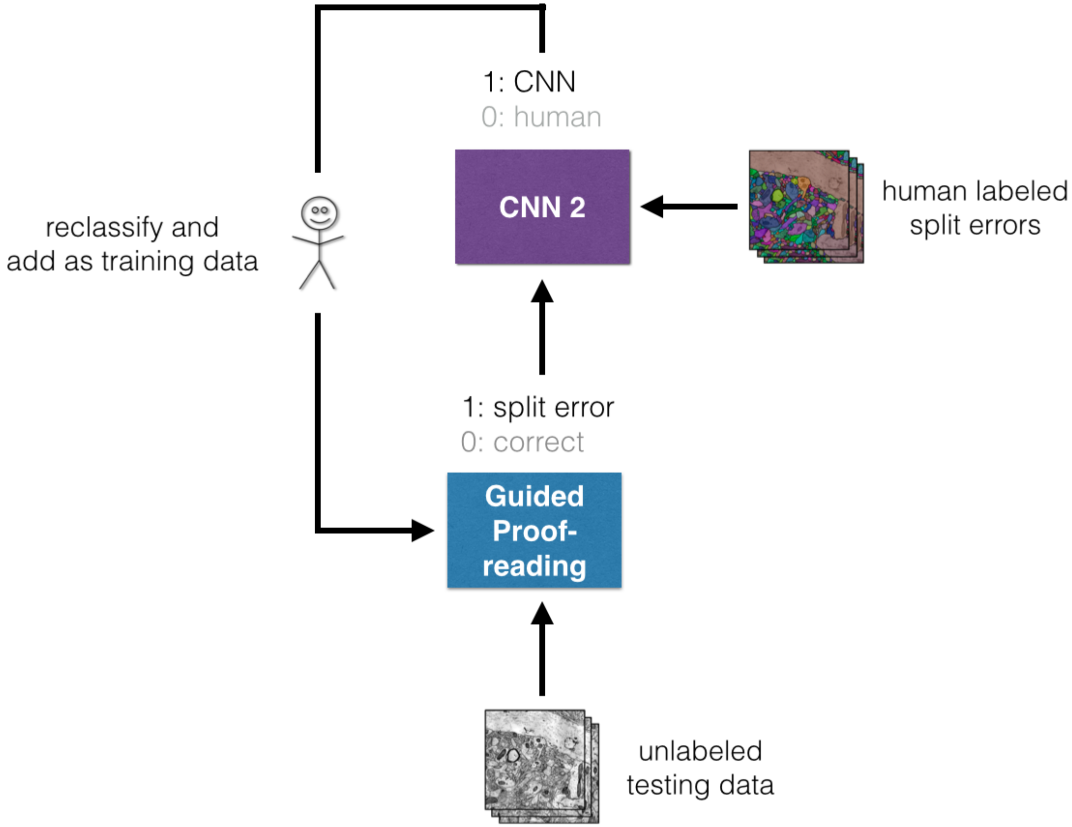
Automatic corrections and user-driven corrections are separate

Classifier training is targeted towards one automatic segmentation method

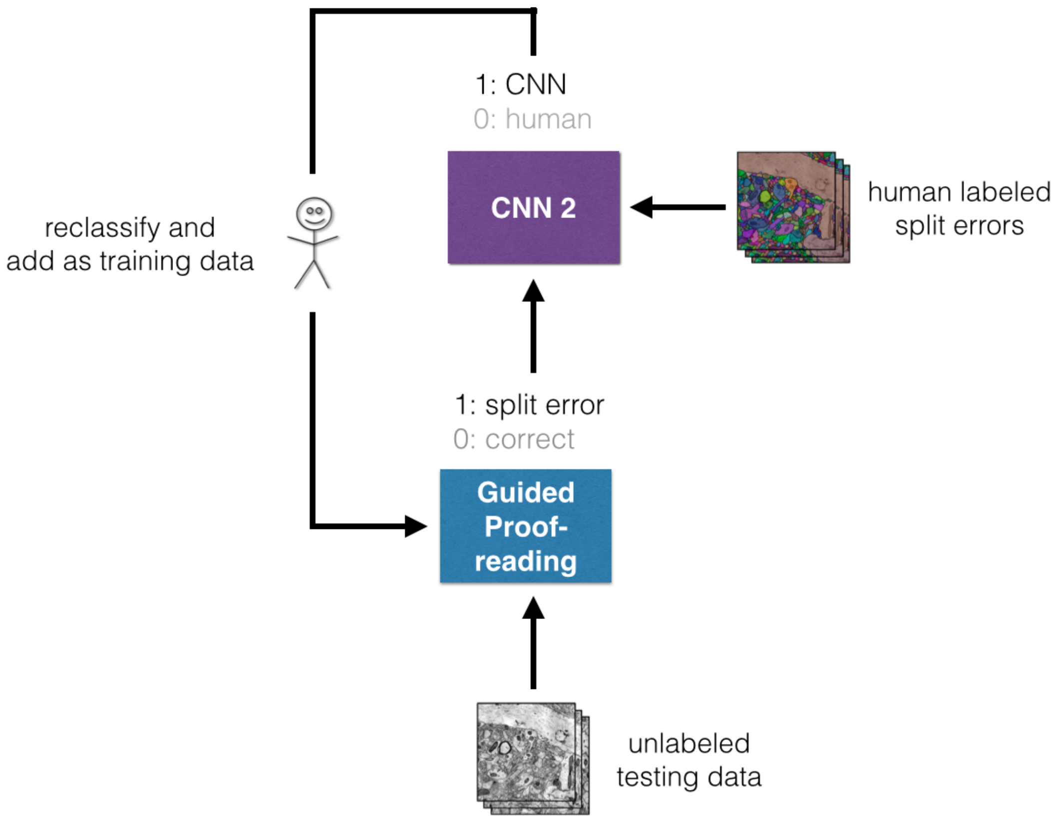
Fusion step is required after 2D proofreading

FUTURE WORK

Active Learning, inspired by
Goodfellow et al.: **Generative Adversarial Nets**, NIPS 2014



FUTURE WORK (2)



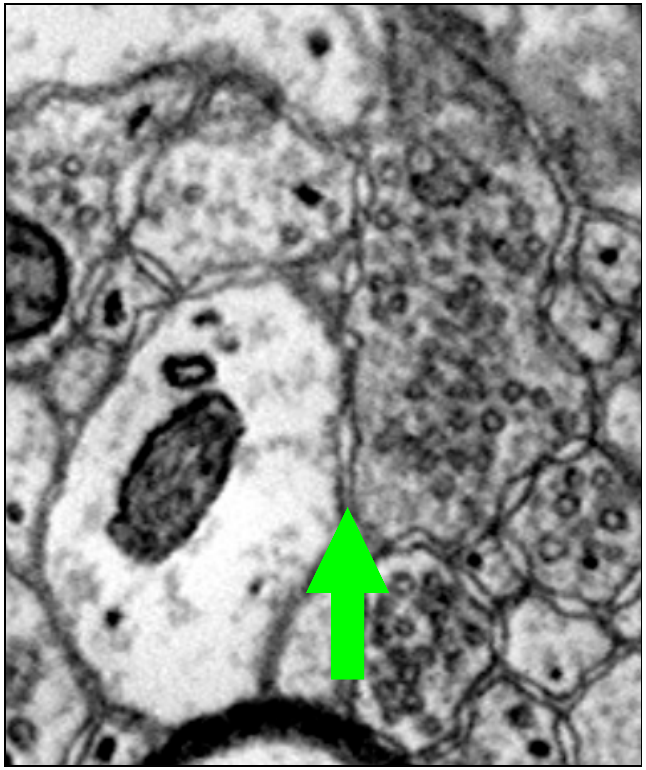
Find pattern of wrong classifications in Guided Proofreading

Idealy, CNN2 is not able to distinguish input from human or our classifier

Fine-tune our classifier

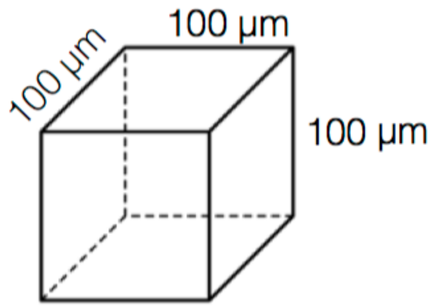
FUTURE WORK (3)

Proofreading of Synapses

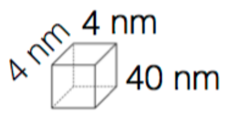


FUTURE WORK (4)

IARPA Project: Analyze 100 microns of data



Voxel size



25000 x 25000 x 2500 voxels
= 1.5625 teravoxels

100 Proofreaders

~43 days with Guided Proofreading

CONCLUSION

Proofreading will always be necessary

Humans will be the driving force

Minimizing manual work is the goal

THANK YOU!