Classifying the Histomorphology of Prostatic Adenocarcinoma with Deep Neural Networks

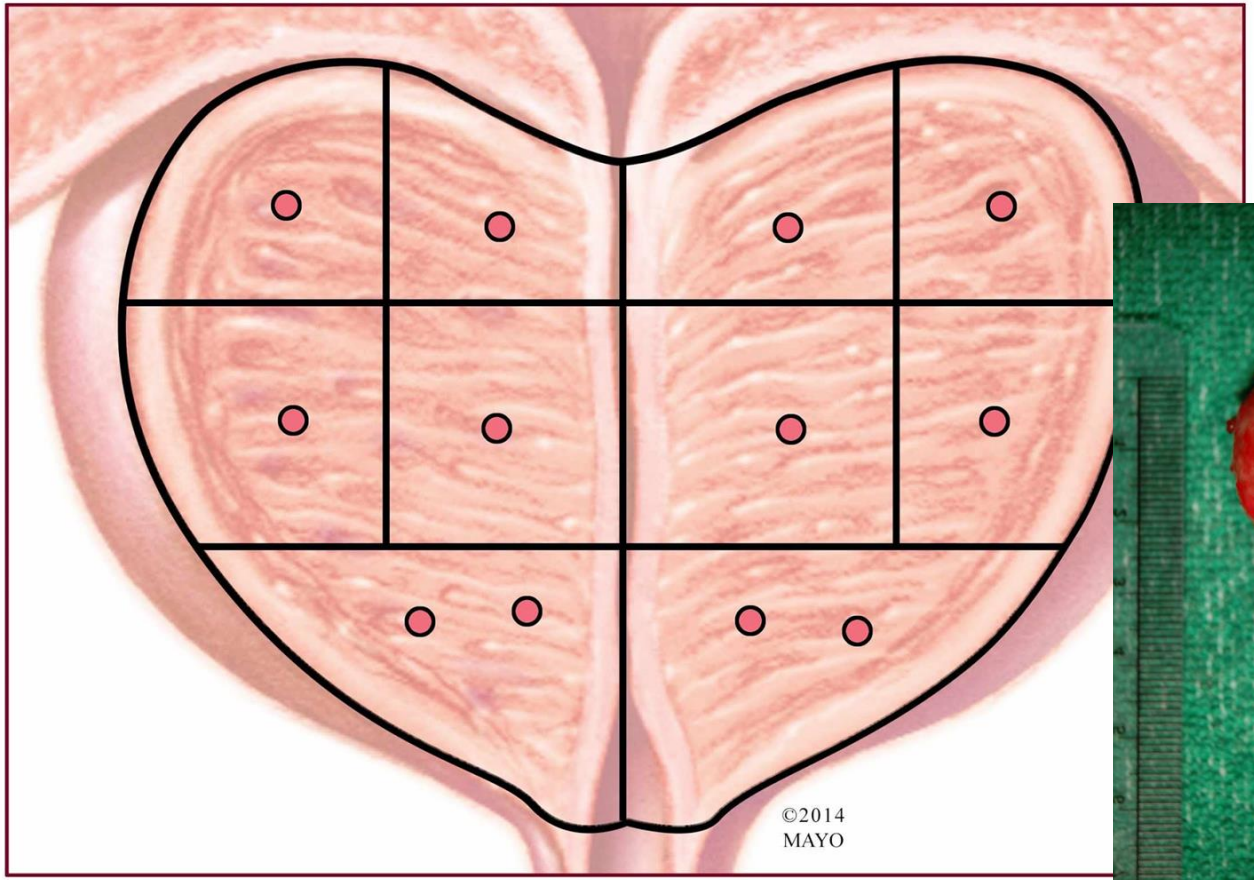
Teaching a Machine to Diagnose Cancer

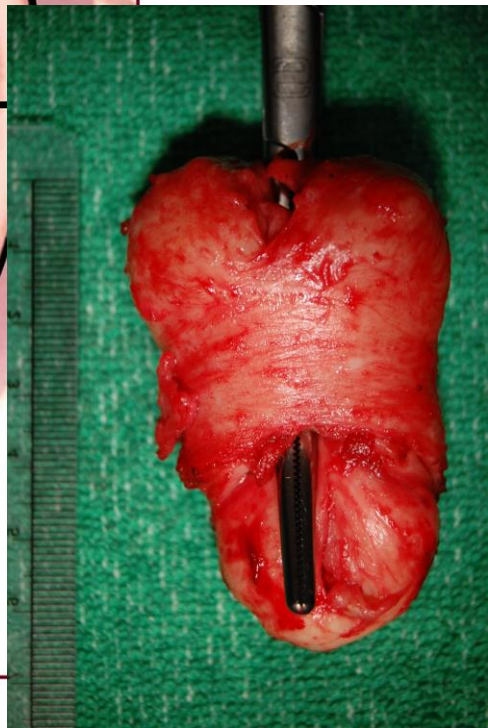
I have no conflicts of interest

The views expressed are those of the author and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the United States Government.

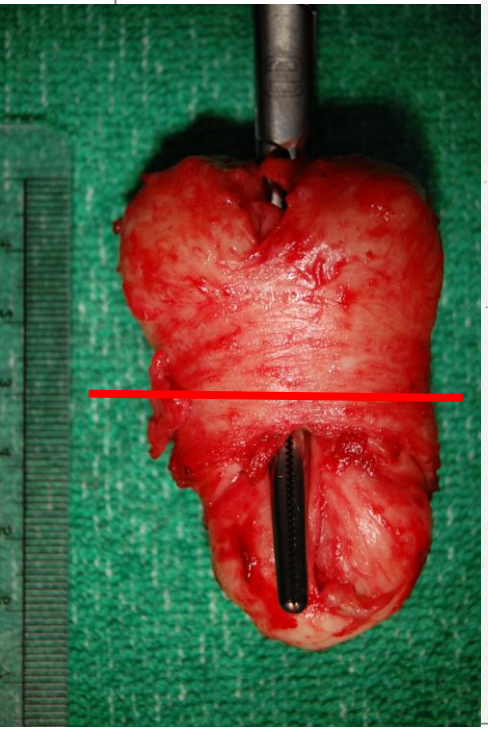
The Problem

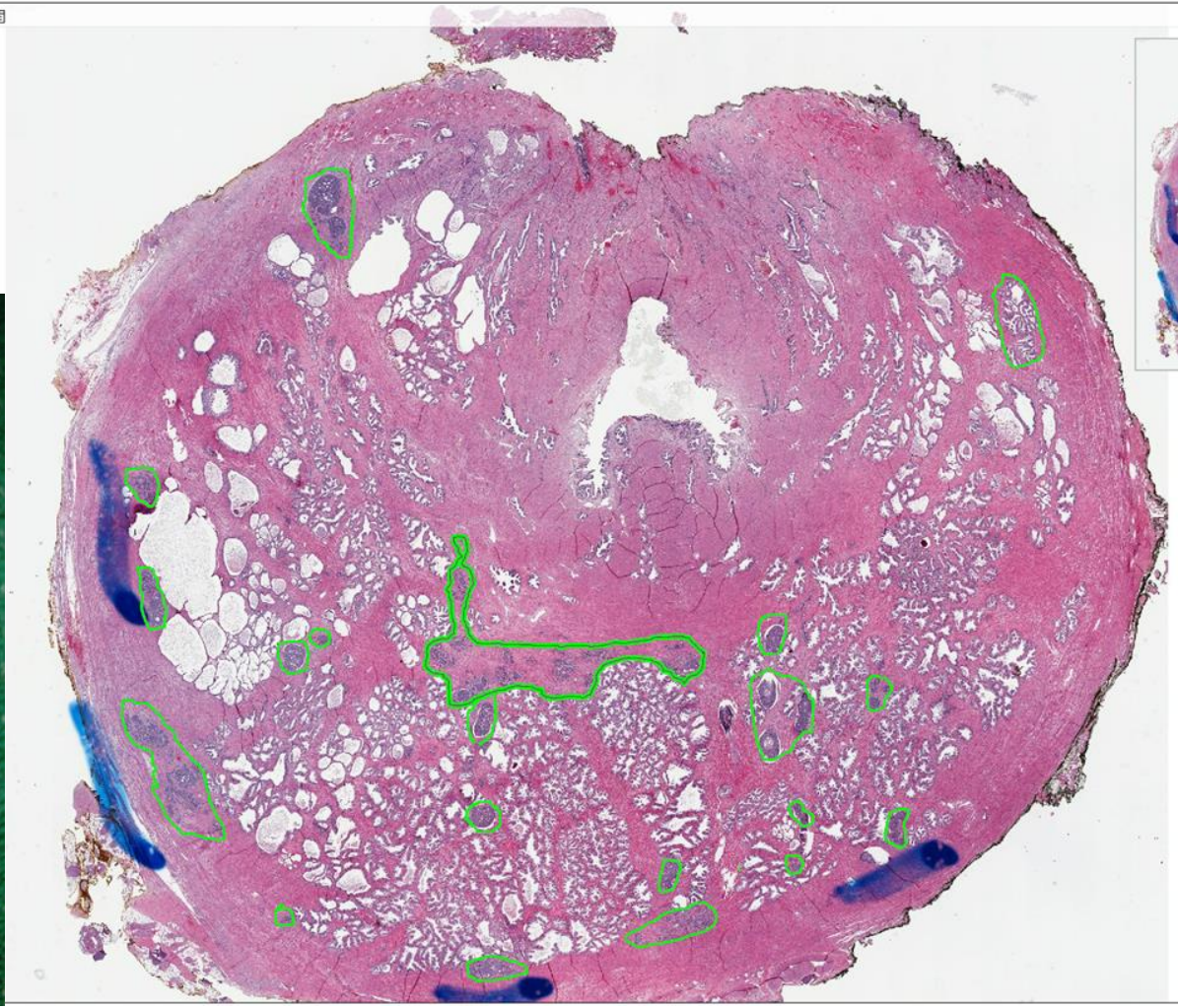
Prostate biopsy pattern

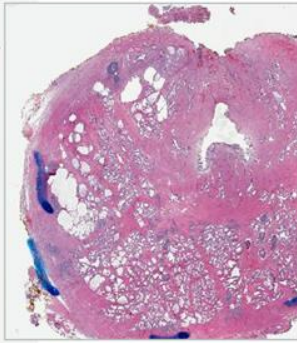


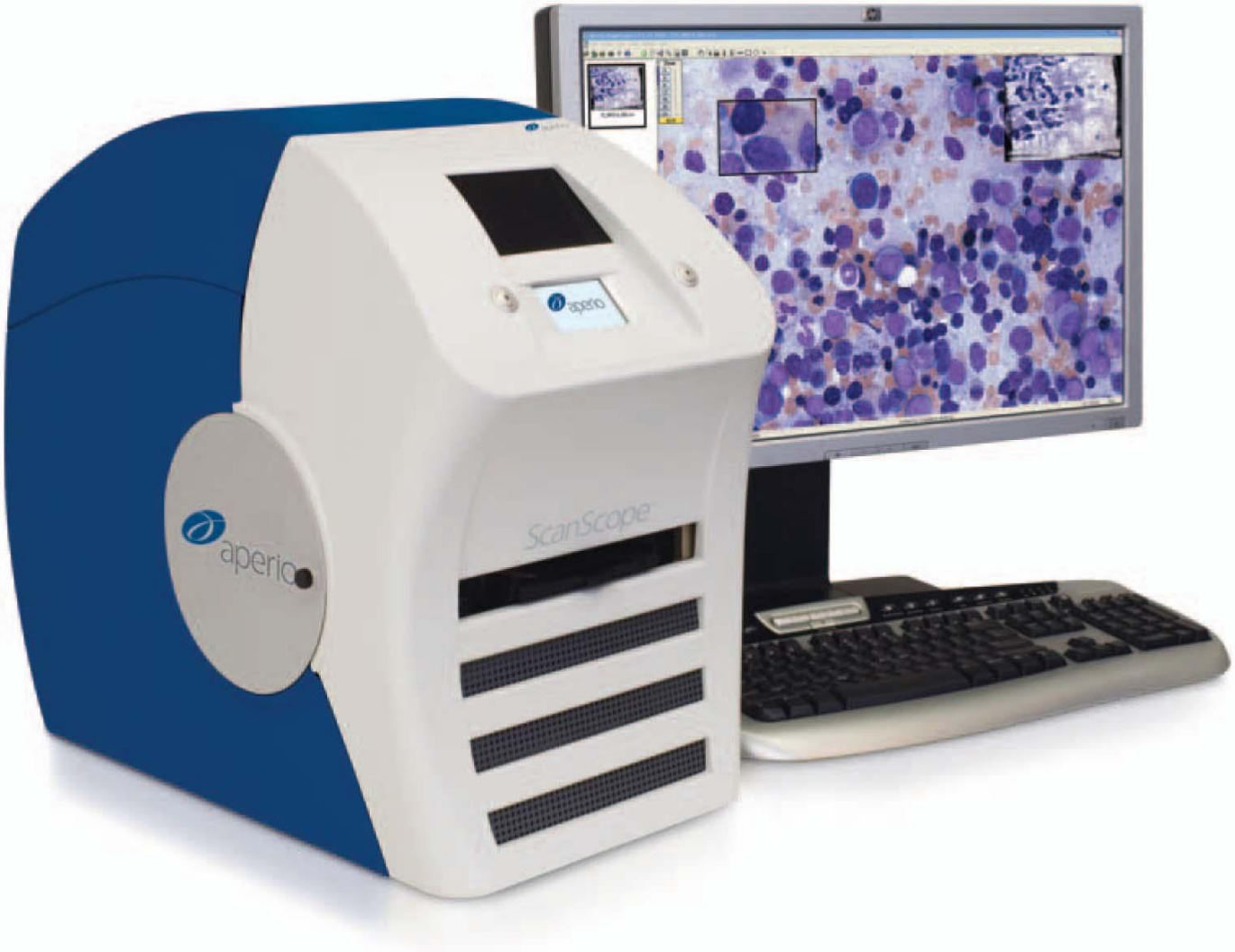


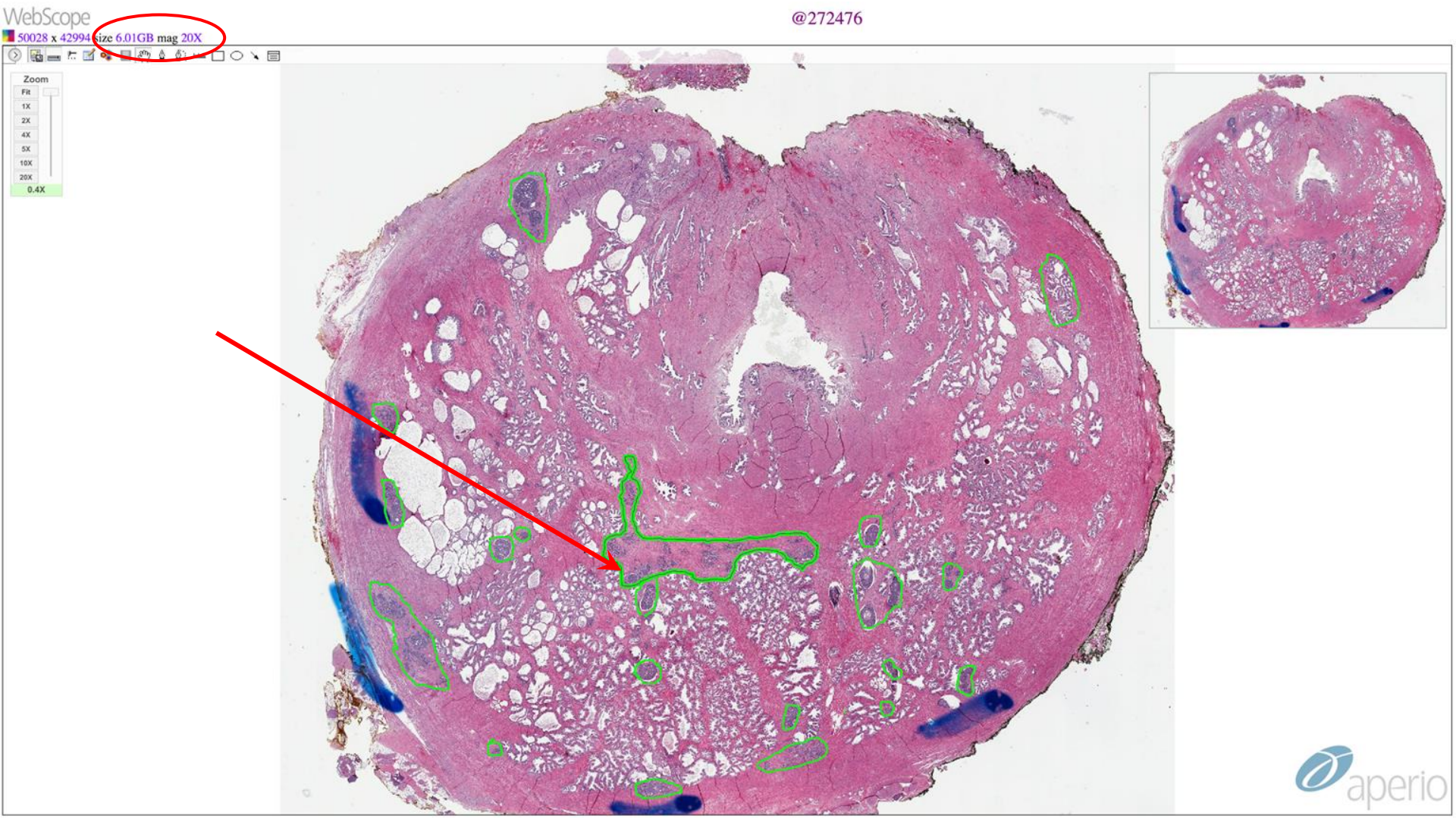
Zoom
Fit
1X
2X
4X
5X
10X
20X
0.4X

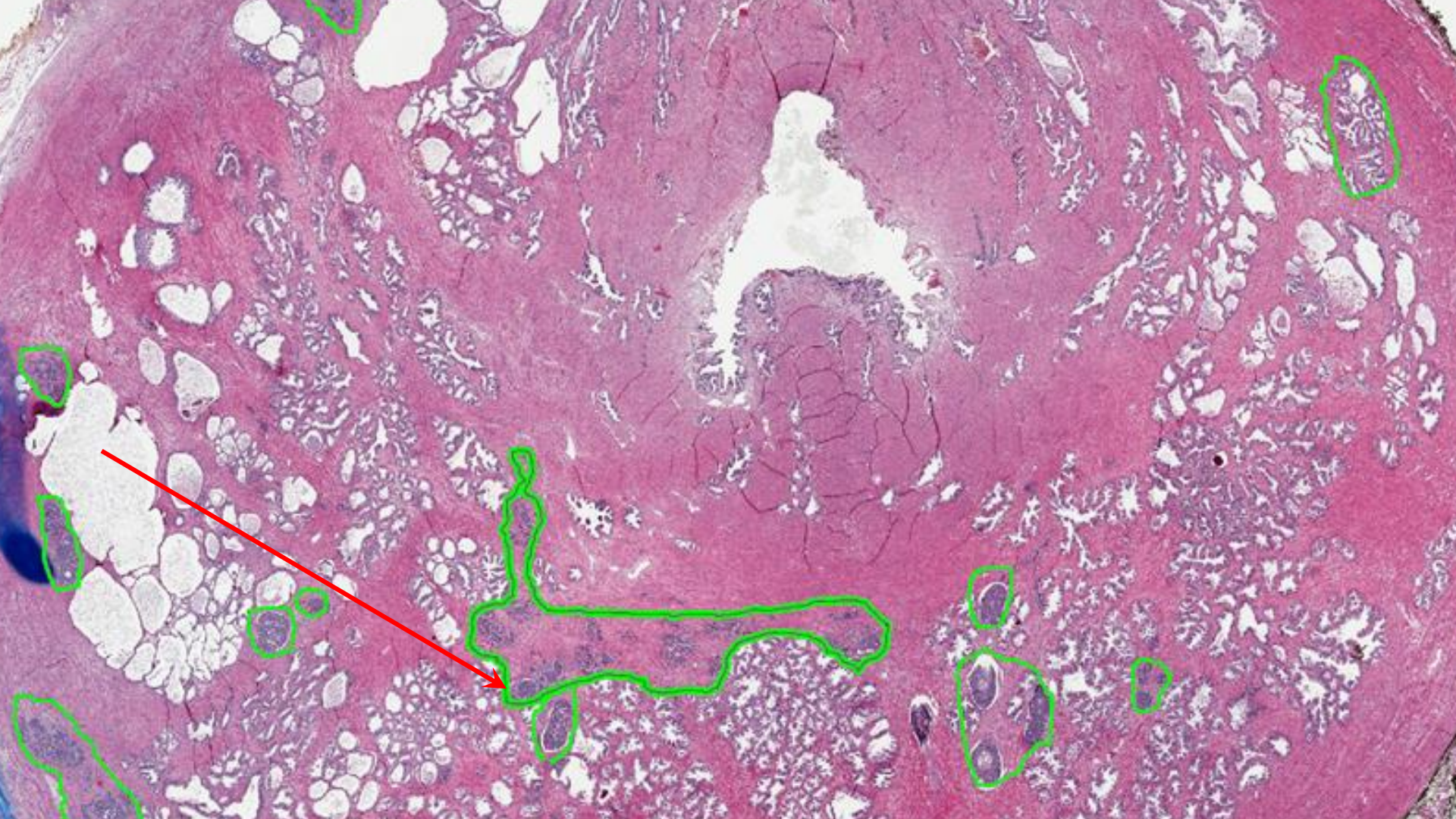




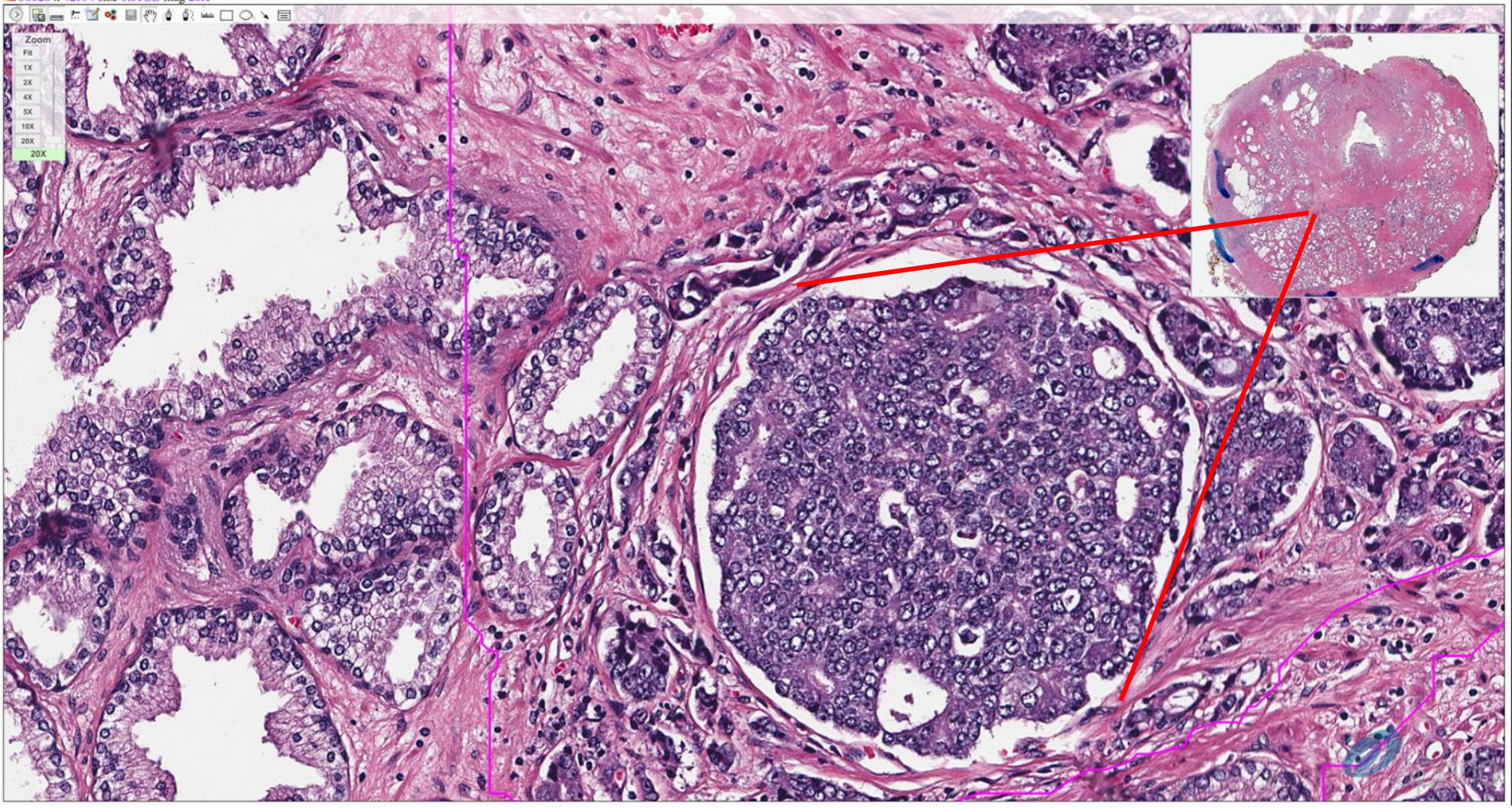


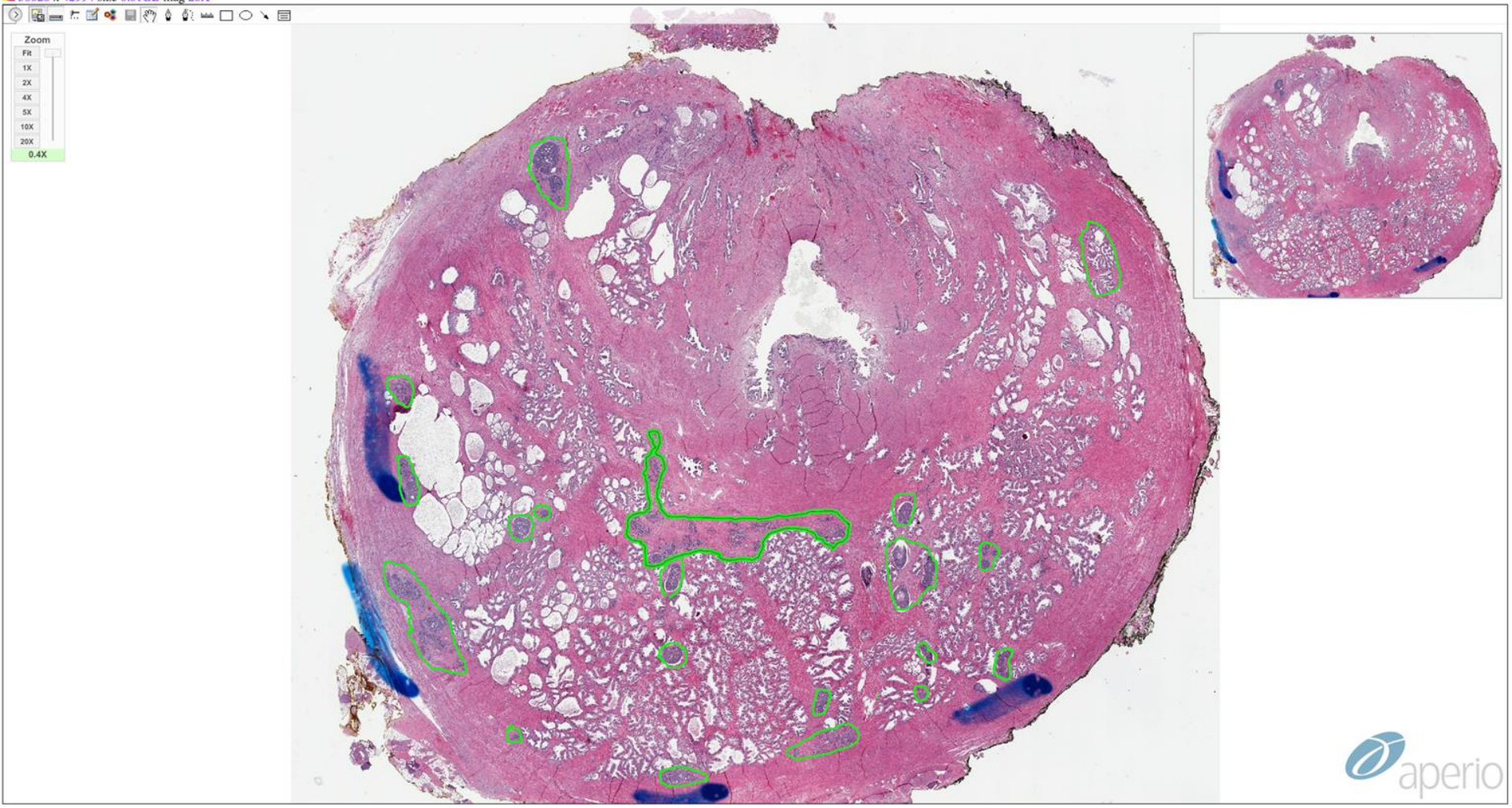

aperio
ScanScope
aperio

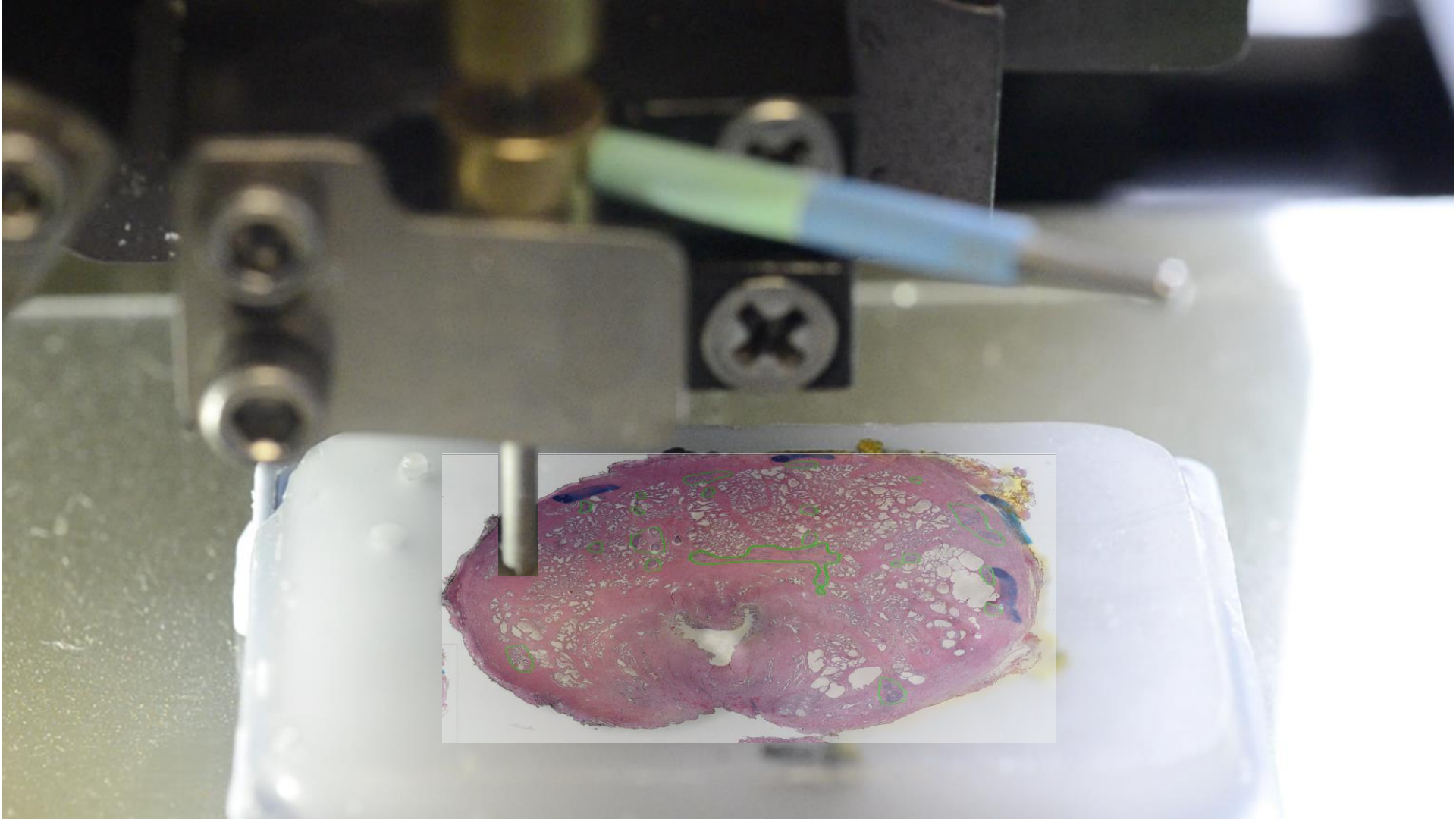

WebScope
50028 x 42994 size 6.01GB mag 20X
@272476
Zoom
Fit
1X
2X
4X
5X
10X
20X
0.4X
aperio

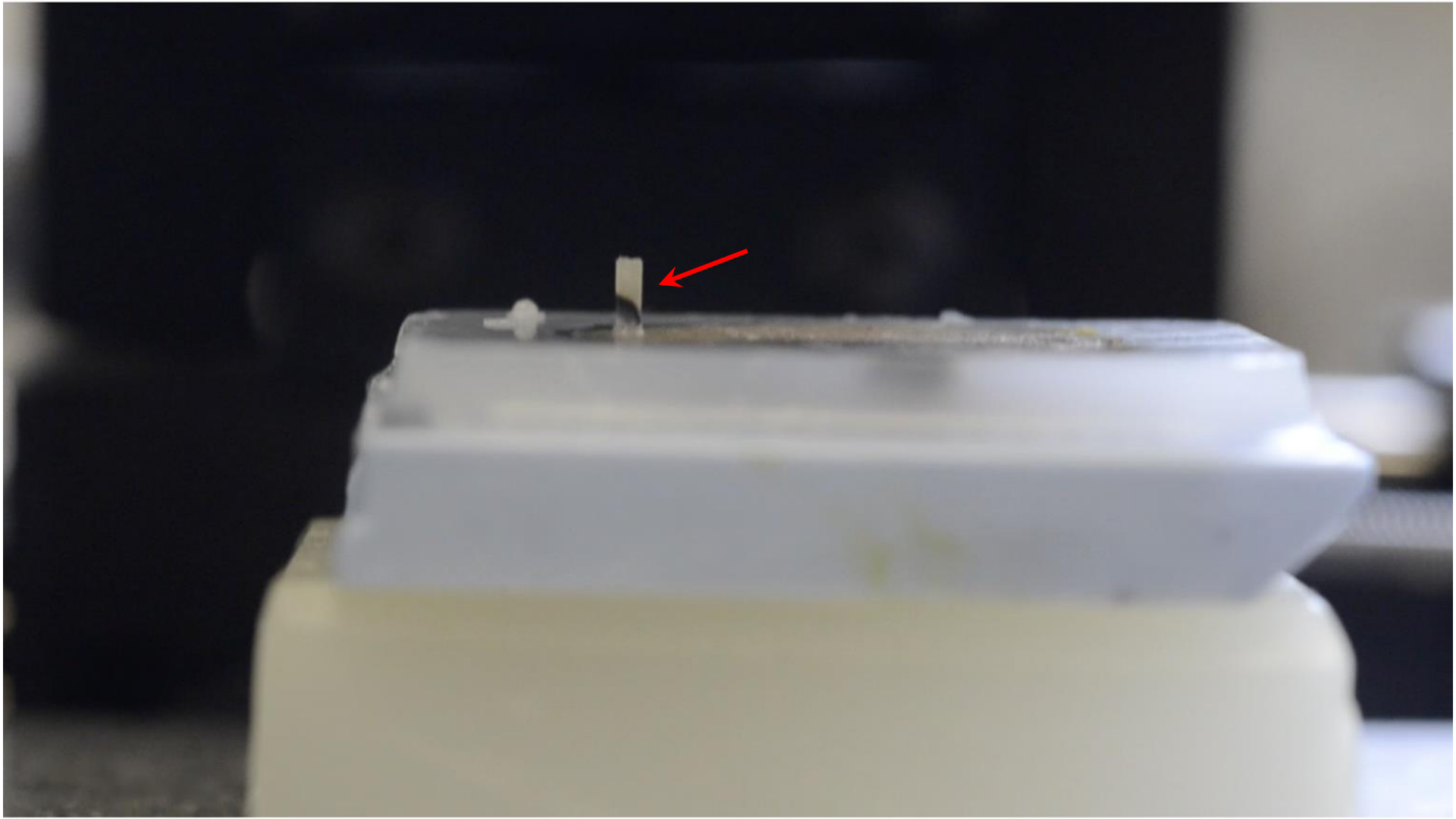


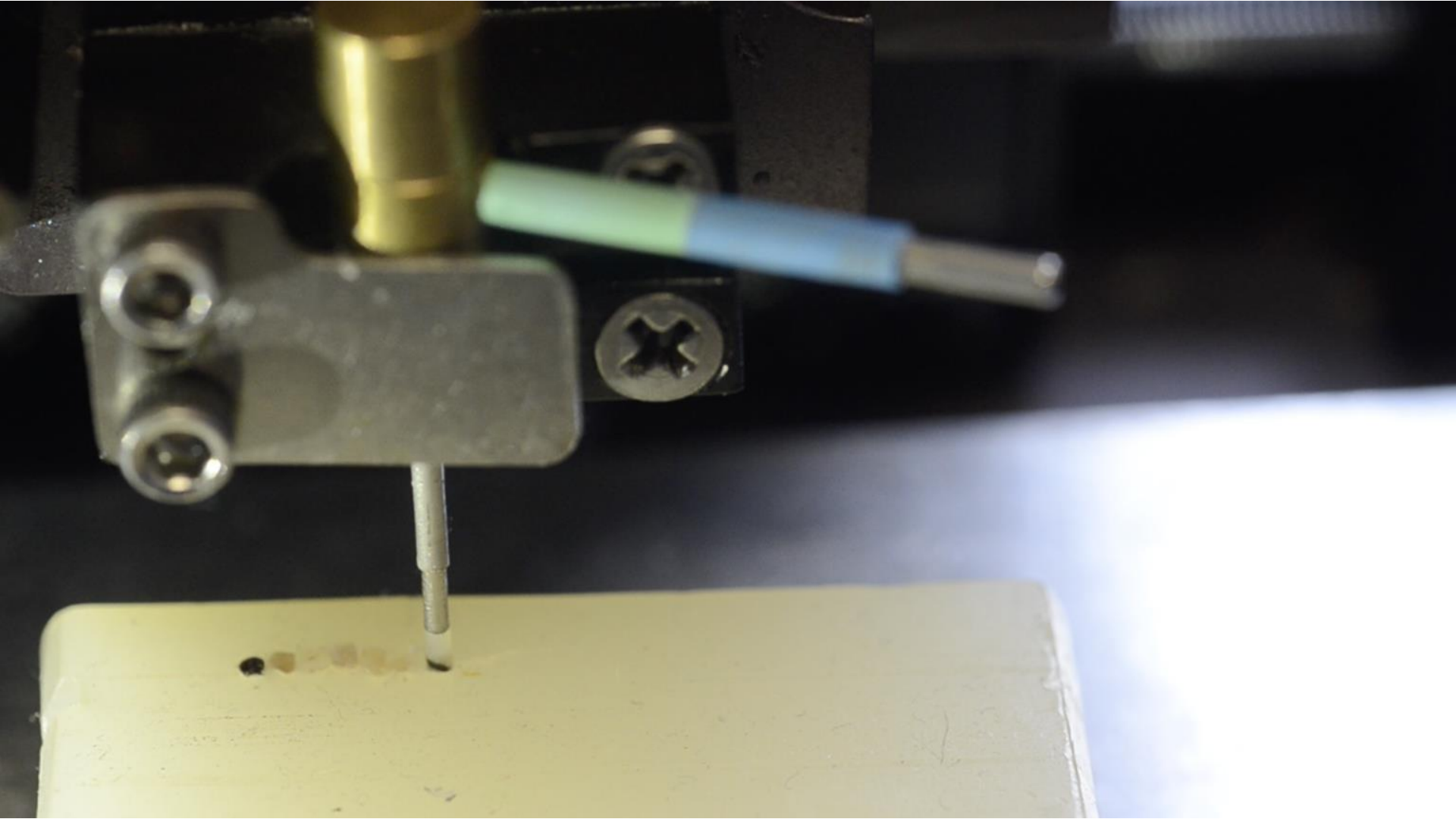
50028 x 42994 size 6.01GB mag 20X

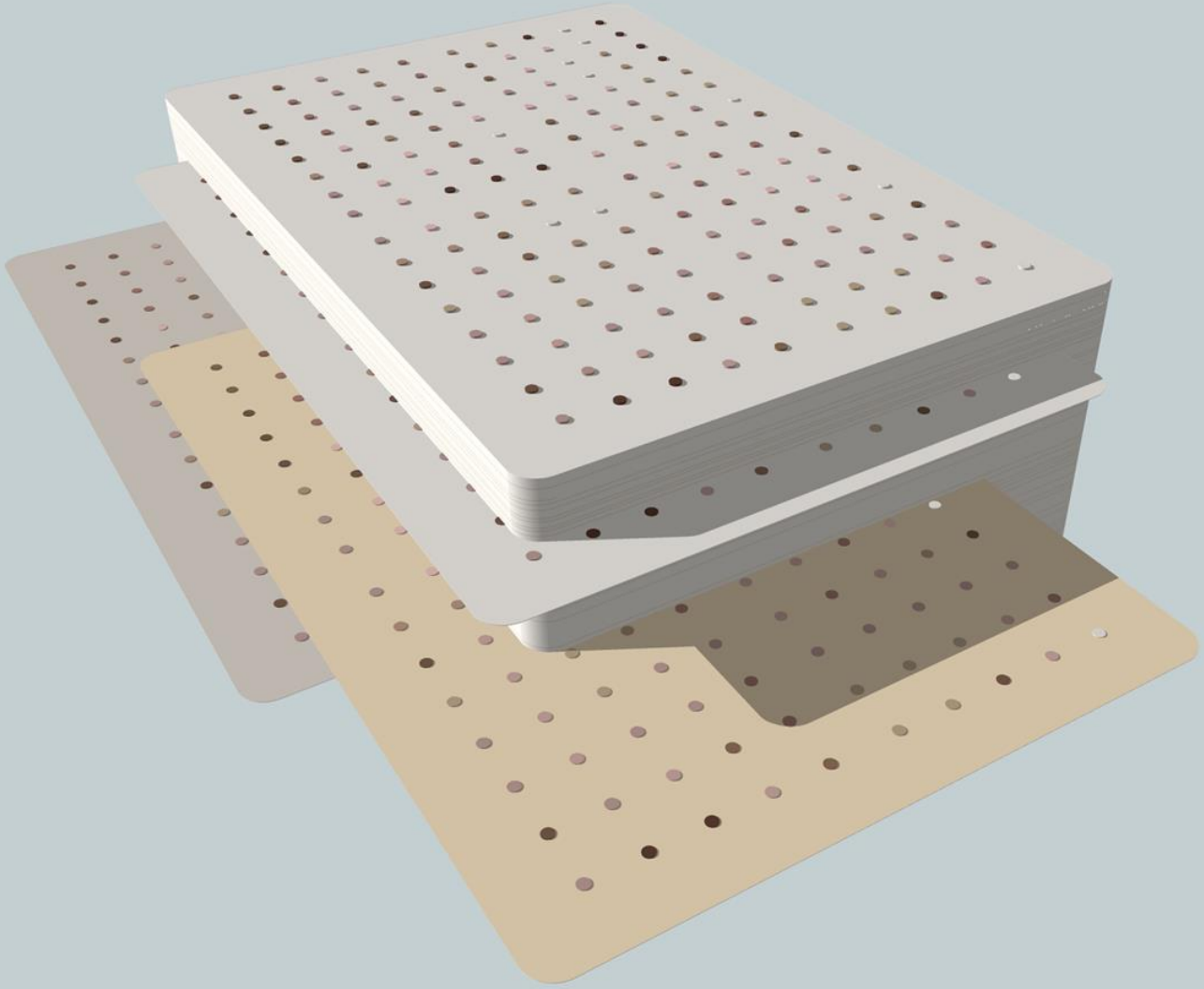


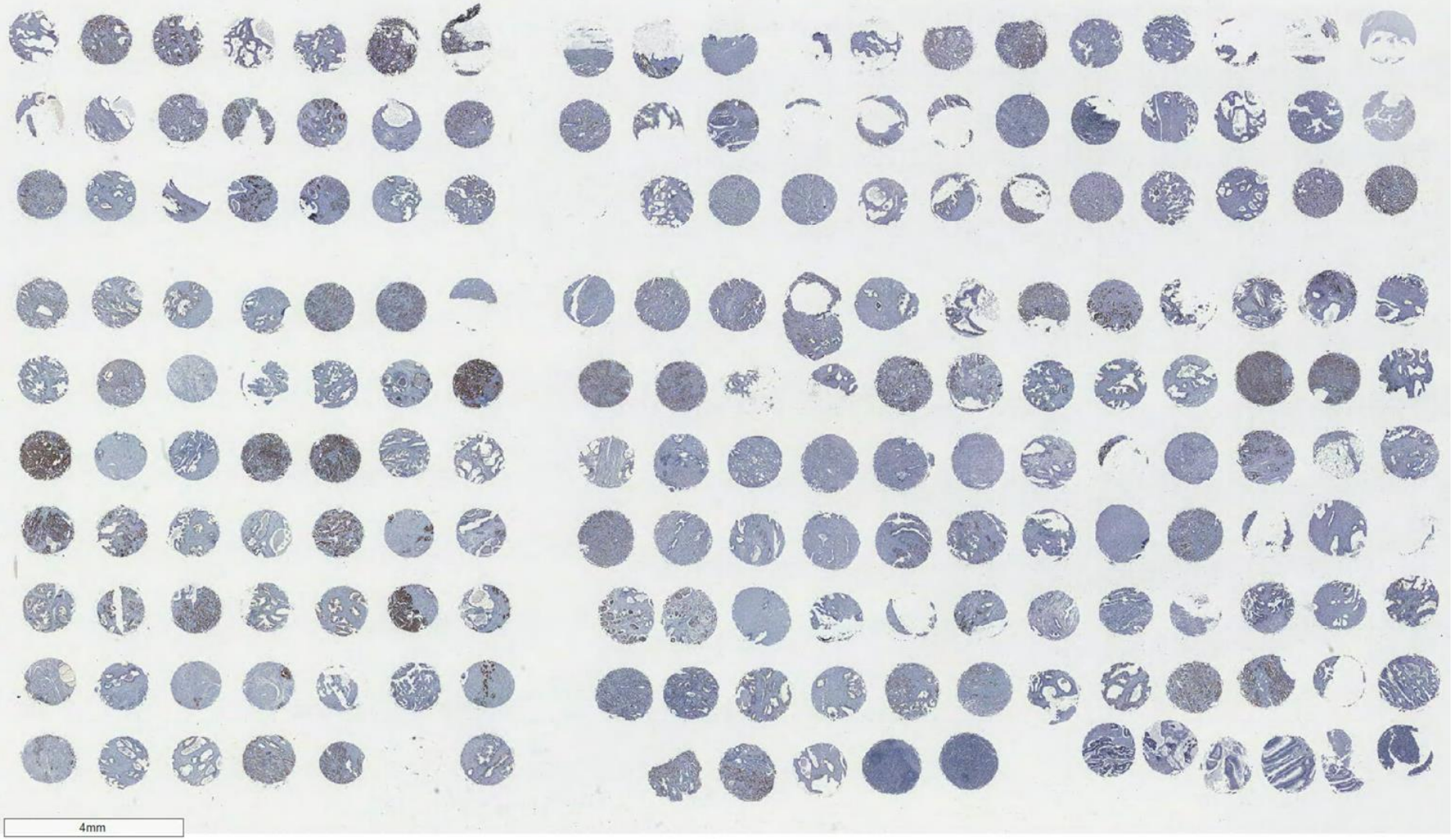

Zoom
Fit
1X
2X
4X
5X
10X
20X
0.4X
aperio



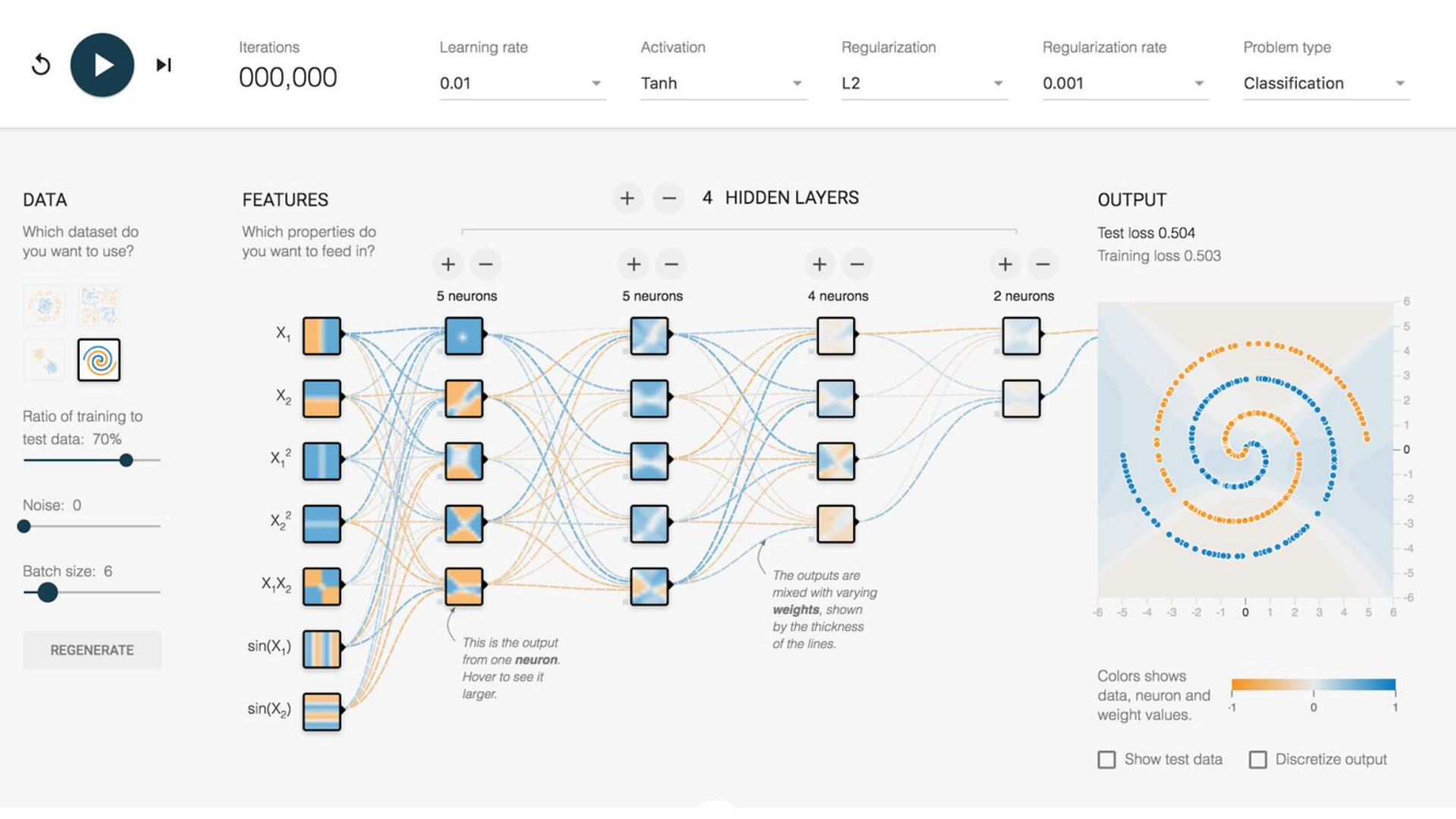


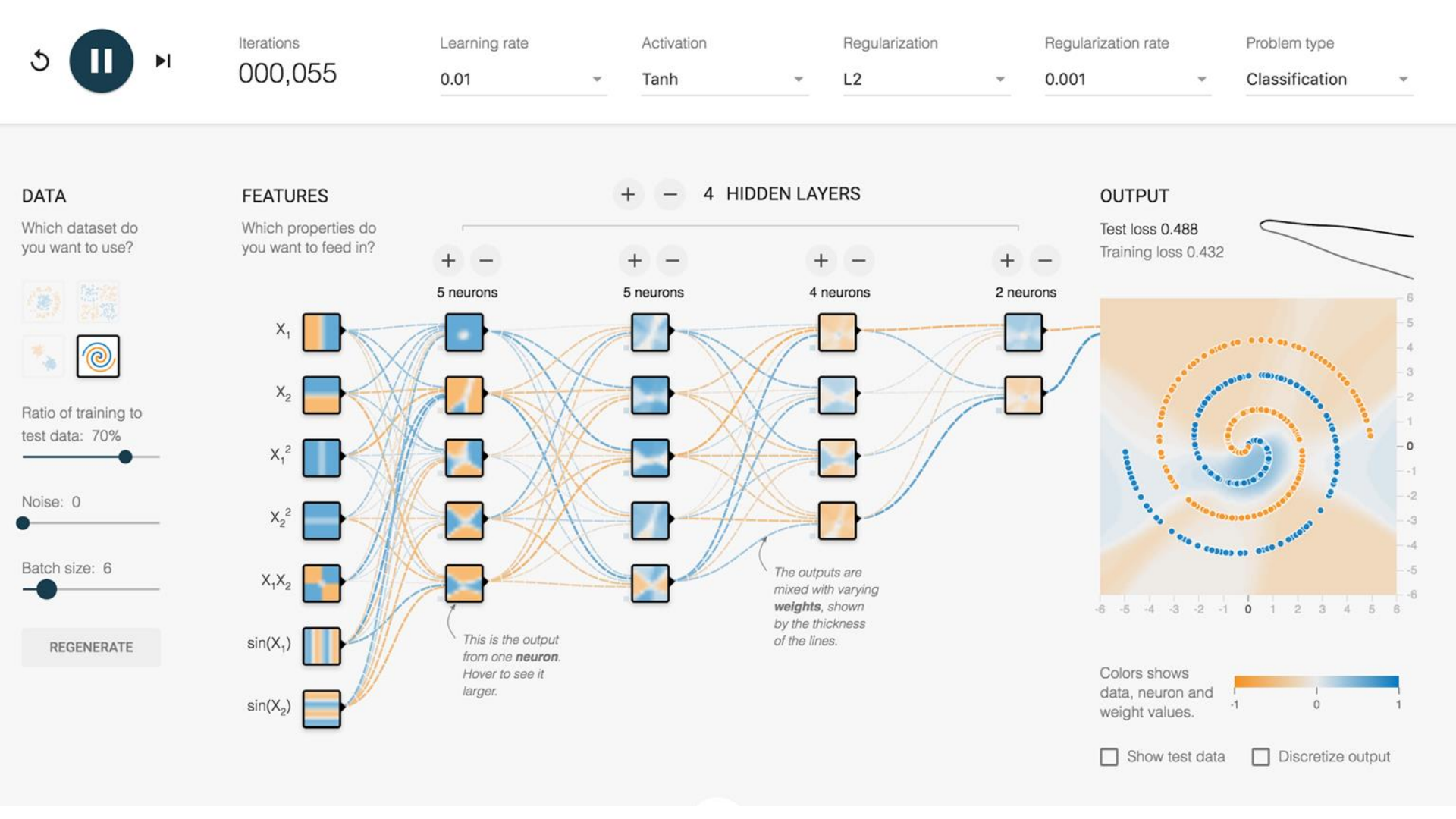





4mm

Teaching a Machine to Diagnose Cancer


Iterations
000,000
Learning rate
0.01
Activation
Tanh
Regularization
L2
Regularization rate
0.001
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 70%
Noise: 0
Batch size: 6
REGENERATE
FEATURES
Which properties do you want to feed in?
4 HIDDEN LAYERS
5 neurons
5 neurons
4 neurons
2 neurons
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
This is the output from one neuron. Hover to see it larger.
The outputs are mixed with varying weights, shown by the thickness of the lines.
OUTPUT
Test loss 0.504
Training loss 0.503
Colors shows data, neuron and weight values.
-1
0
1
Show test data
Discretize output


Iterations
000,055
Learning rate
0.01
Activation
Tanh
Regularization
L2
Regularization rate
0.001
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 70%
Noise: 0
Batch size: 6
REGENERATE
FEATURES
Which properties do you want to feed in?
4 HIDDEN LAYERS
5 neurons
5 neurons
4 neurons
2 neurons
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
This is the output from one neuron. Hover to see it larger.
The outputs are mixed with varying weights, shown by the thickness of the lines.
OUTPUT
Test loss 0.488
Training loss 0.432
Colors shows data, neuron and weight values.
-1
0
1
Show test data
Discretize output

Iterations
000,164

Learning rate: 0.01

Activation: Tanh

Regularization: L2

Regularization rate: 0.001

Problem type: Classification

DATA

Which dataset do you want to use?

Ratio of training to test data: 70%

Noise: 0

Batch size: 6

REGENERATE

FEATURES

Which properties do you want to feed in?

X_1

X_2

X_1^2

X_2^2

X_1X_2

$\sin(X_1)$

$\sin(X_2)$

4 HIDDEN LAYERS

5 neurons | 5 neurons | 4 neurons | 2 neurons

*This is the output from one **neuron**. Hover to see it larger.*

*The outputs are mixed with varying **weights**, shown by the thickness of the lines.*

OUTPUT

Test loss 0.235

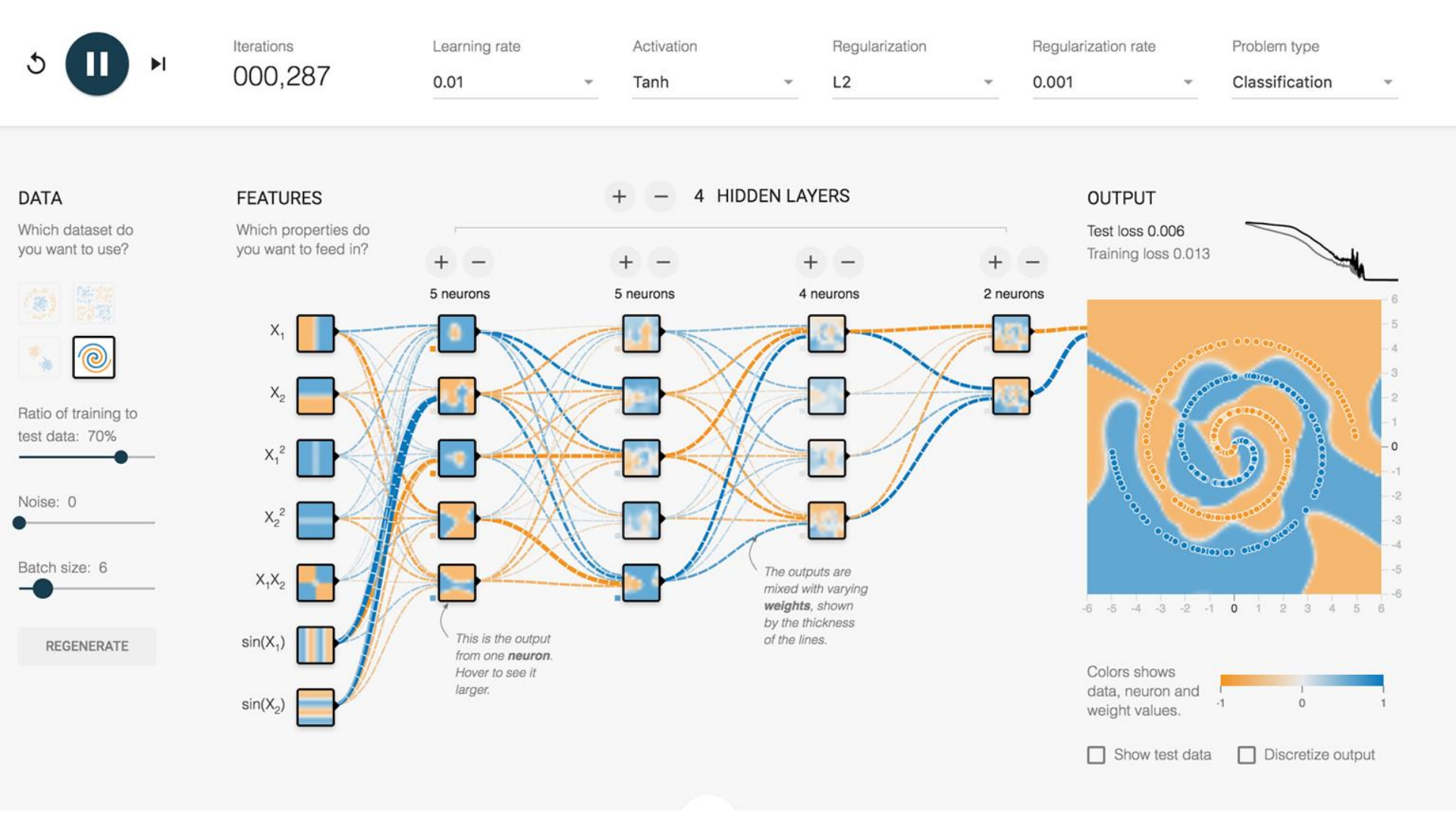
Training loss 0.174

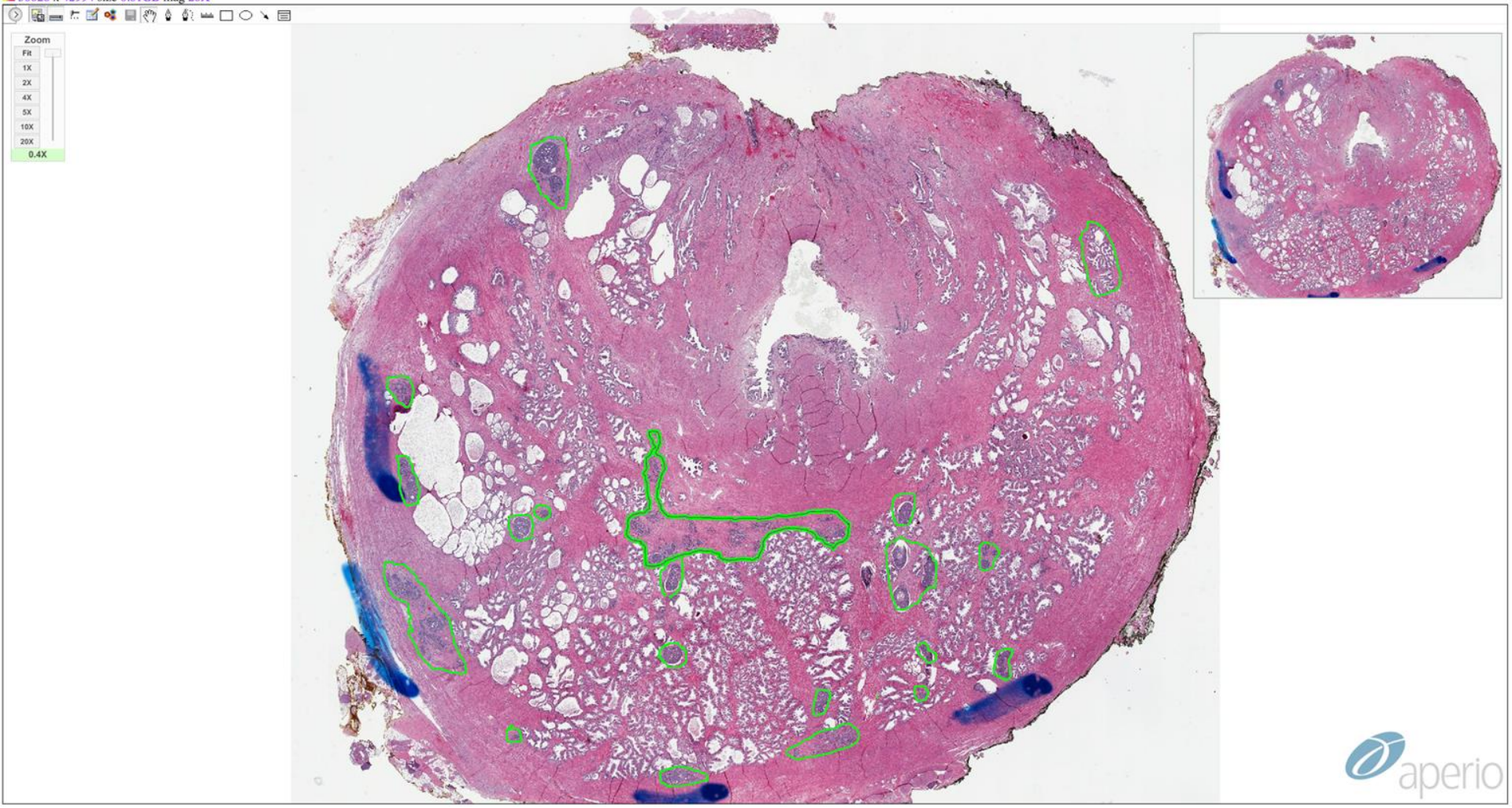
Colors shows data, neuron and weight values.

-1 0 1

Show test data

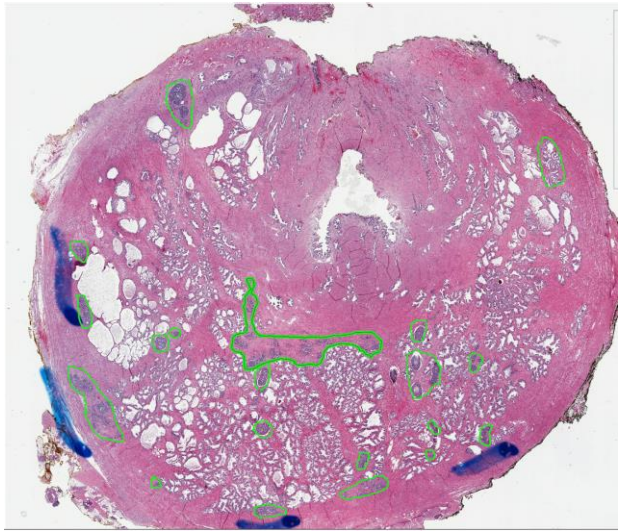
Discretize output


Iterations
000,287
Learning rate
0.01
Activation
Tanh
Regularization
L2
Regularization rate
0.001
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 70%
Noise: 0
Batch size: 6
REGENERATE
FEATURES
Which properties do you want to feed in?
4 HIDDEN LAYERS
5 neurons
5 neurons
4 neurons
2 neurons
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
This is the output from one neuron. Hover to see it larger.
The outputs are mixed with varying weights, shown by the thickness of the lines.
OUTPUT
Test loss 0.006
Training loss 0.013
Colors shows data, neuron and weight values.
-1
0
1
Show test data
Discretize output

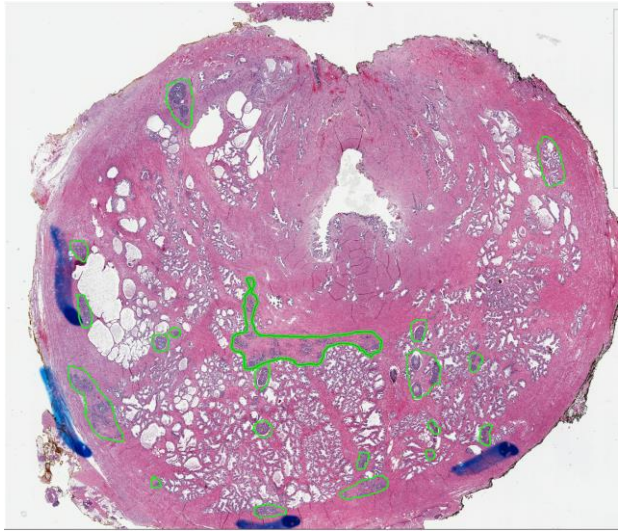

Zoom
Fit
1X
2X
4X
5X
10X
20X
0.4X
aperio


GTX980
NVIDIA

NAVAL MEDICAL CENTER
SAN DIEGO



Methods and Materials



Data

- 202 annotated color Aperio SVS images
- Scanned with an Aperio CS-1 at 20x
- 106,024 image chips at 10x magnification
- Average full size image ~ 845 MB
- Annotated by Navy pathologists
- Manually specifying additional categories: ink, white space, fat, and stroma

System

- NVIDIA GeForce GTX980 GPU (single card) via Intel Haswell-E PCIe 3.0
- Maxwell architecture, 2048 CUDA cores, 4GB memory, NV driver 352.63
- 6-core Intel Xeon E5-2603 v3 at 1.60 GHz with 16GB DDR4
- Ubuntu 14.04, DIGITS 3.0-rc3, CUDA 7.5, cuDNN v4, NVCaffe 0.14

Results

March 2016

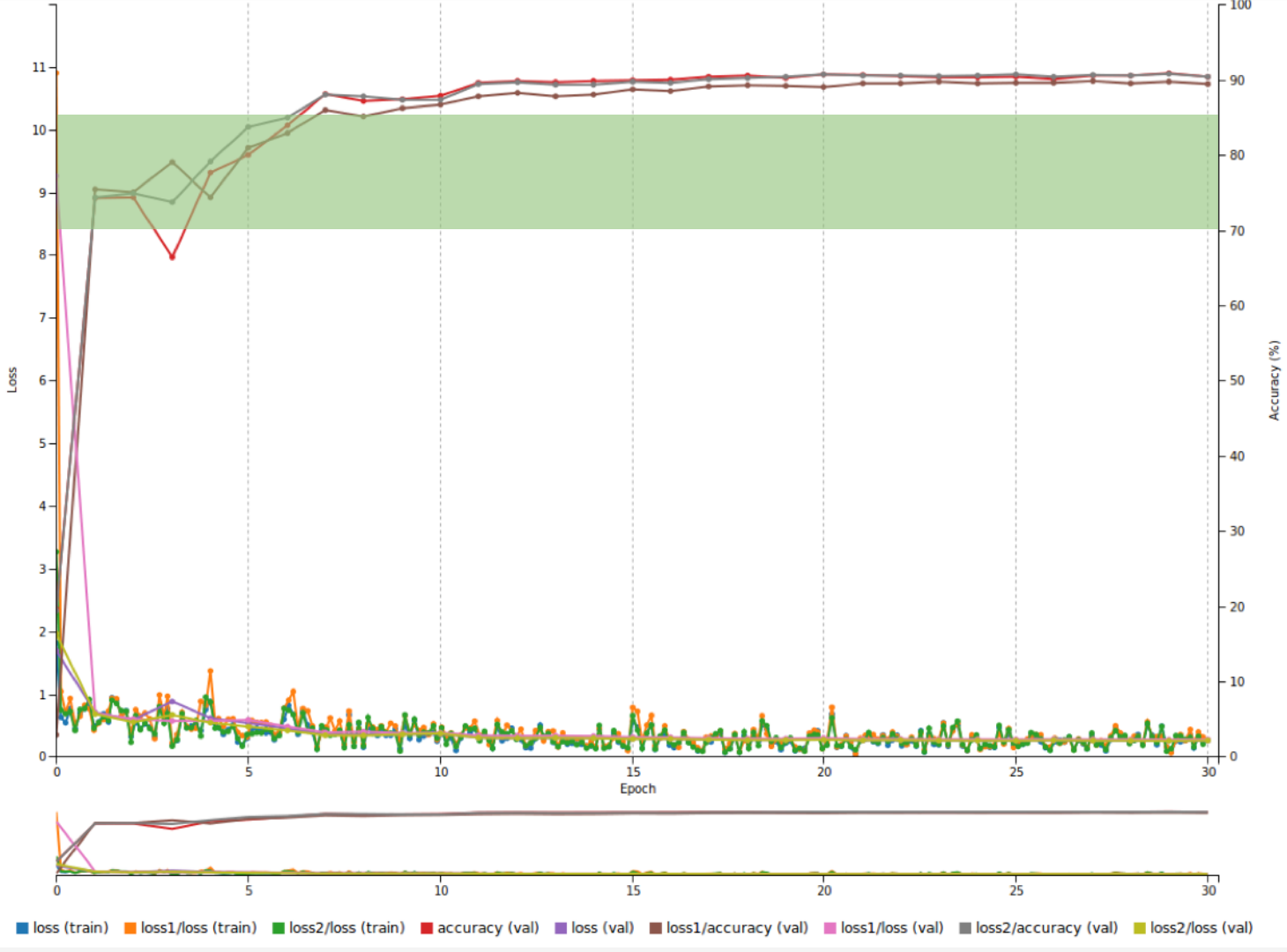
256x256

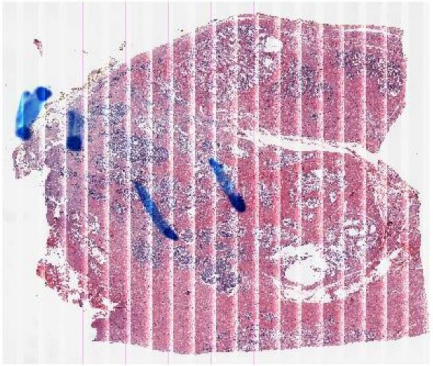
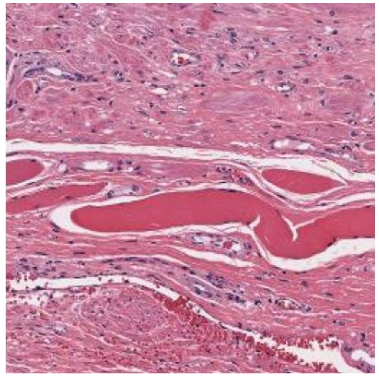
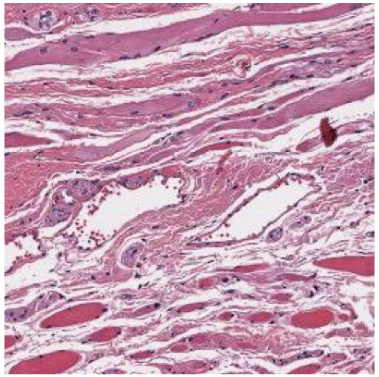
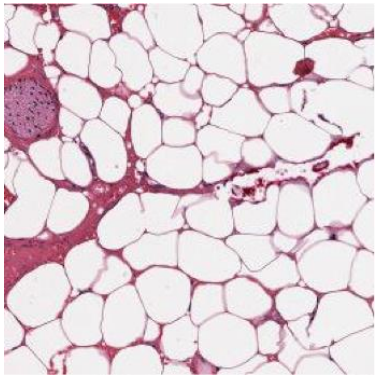
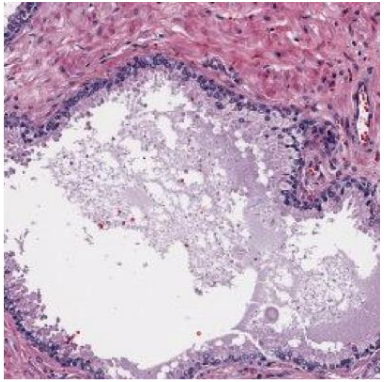
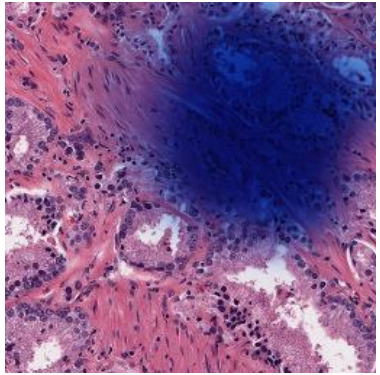
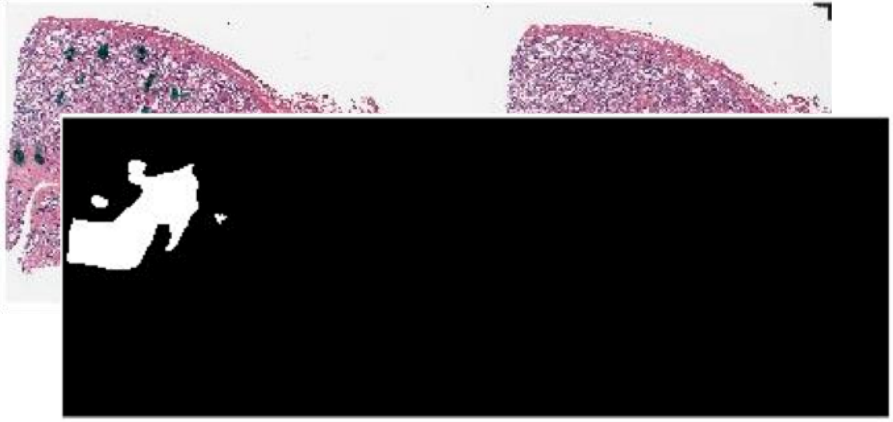
Full color

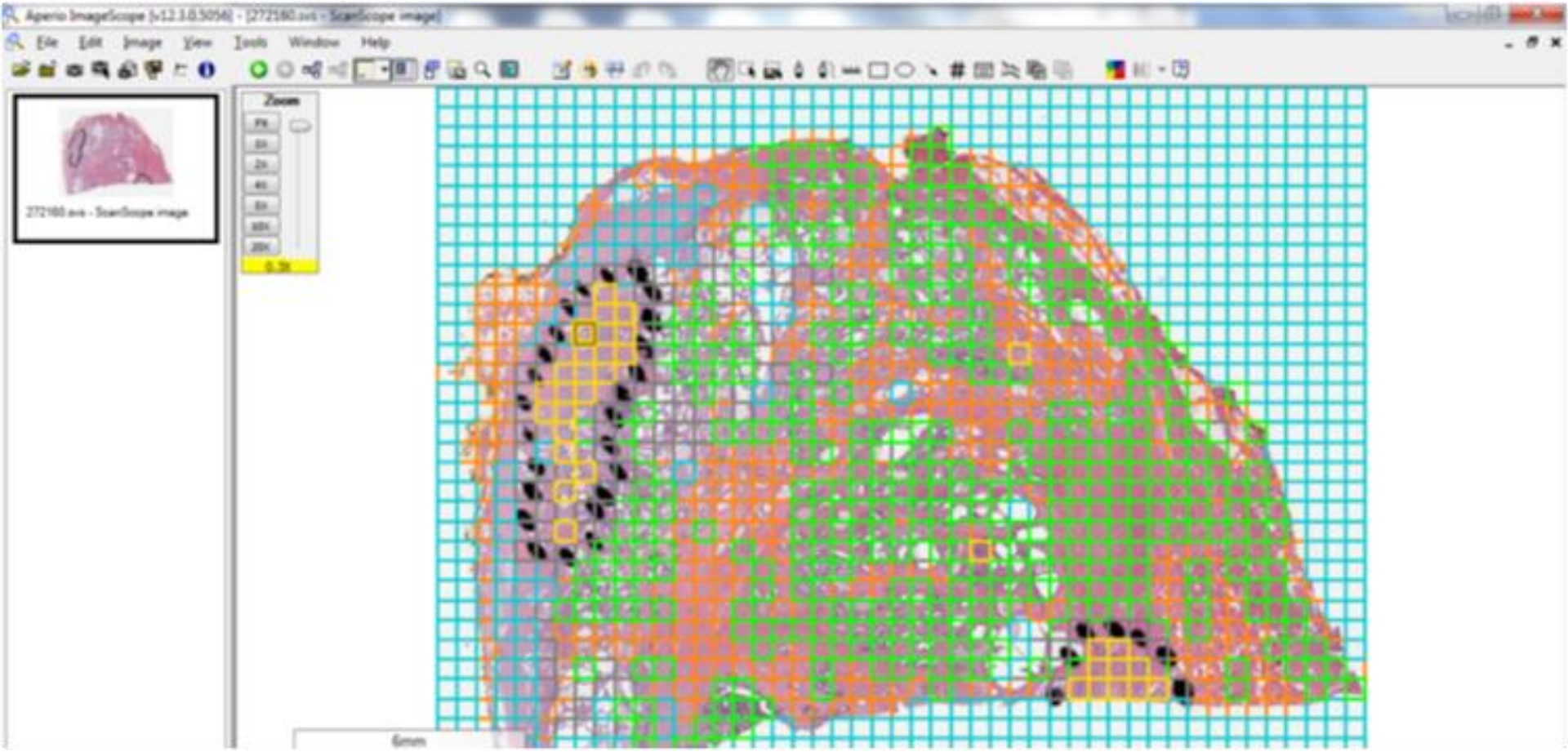
GoogLeNet

60 epochs

Additional annotation of previously unrelated image contents (fat, surgical ink, marking ink).



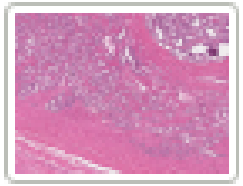








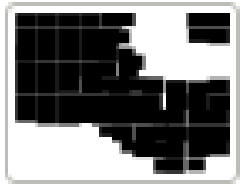

Aperio ImageScope
File
Edit
Image
View
Tools
Window
Help
Zoom
272160.svs - ScanScope image
6mm

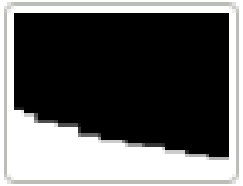

3_orig.jpg

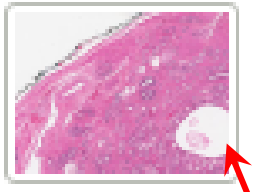

3_pred.jpg

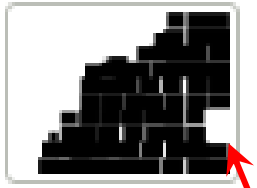

3_true.jpg

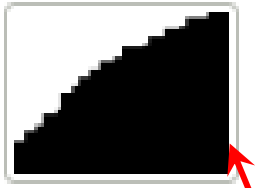

5_orig.jpg

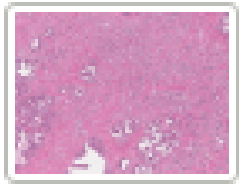

5_pred.jpg

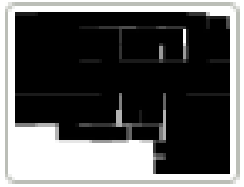

5_true.jpg

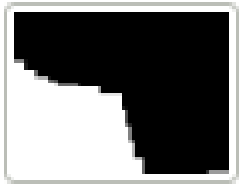

7_orig.jpg


7_pred.jpg


7_true.jpg


9_orig.jpg


9_pred.jpg


9_true.jpg

Future Work

Y **Hacker News** new | comments | show | ask | jobs | submit login

1. Change Your Name (paulgraham.com)
   104 points by franze 1 hour ago | 81 comments
2. Nvidia desktop devbox four GPUs (nvidia.com)
   39 points by hendler 1 hour ago | 18 comments
3. Porting LibreOffice to GTK3 [pdf] (skynet.ie)
   45 points by tomkwok 3 hours ago | 11 comments
4. Ask HN: Who here would be interested in a free daily updated GeoIP database?
   106 points by dutchbrit 3 hours ago | 41 comments
5. How to make a book with Pollen (pollenpub.com)
   20 points by lobo_tuerto 2 hours ago | 5 comments
6. VPaint: An experimental vector graphics editor (vpaint.org)
   255 points by jestinjoy1 12 hours ago | 45 comments
7. Linux-insides: Introduction to deferred interrupts (github.com)
   46 points by 0xAX 4 hours ago | 2 comments
8. C++ Internals: STL Vector, Part 3 (gahcep.com)
   23 points by ingve 4 hours ago | 3 comments
9. Building Netflix Playback with Self-Assembling Components (2014) (netflix.com)
   20 points by coffeedan 3 hours ago | 3 comments
10. Let's Build a Simple Interpreter, Part 2 (ruslanspivak.com)
    25 points by danso 4 hours ago | 4 comments


Home
Datasets
New Dataset
Images
In progress
dataset_imagenet@256x256
Submitted: 05:29:57 PM (53 seconds ago)
Status: Running
Completed
dataset_minst_10k@256x256
Submitted: 05:21:27 PM
Status: Done after 31 seconds
voc_cropped@256x256
Submitted: 05:14:31 PM
Status: Done after 2 minutes, 26 seconds
Models
New Model
Images
In progress
model_minist10k
Submitted: 09:30:48 PM (3 seconds ago)
Status: Running
LeNet_model_voc_cropped@256x256
Submitted: 05:18:43 PM (12 minutes, 7 seconds ago)
Status: Running
Completed
GEFORCE GTX
GEFORCE GTX
GEFORCE GTX
GEFORCE GTX
TITAN


FISMA
FEDERAL INFORMATION SECURITY MANAGEMENT ACT


NAVY MEDICINE
INFORMATION SYSTEMS SUPPORT ACTIVITY


DEFENSE HEALTH AGENCY
PRO CURA MILITIS
DEPARTMENT OF DEFENSE


FedRAMP


HIPAA


DISA


NIST

DOD HPC MODERNIZATION PROGRAM

SPAWAR

Systems Center PACIFIC

HPC Modernization Program

DREN

850,000