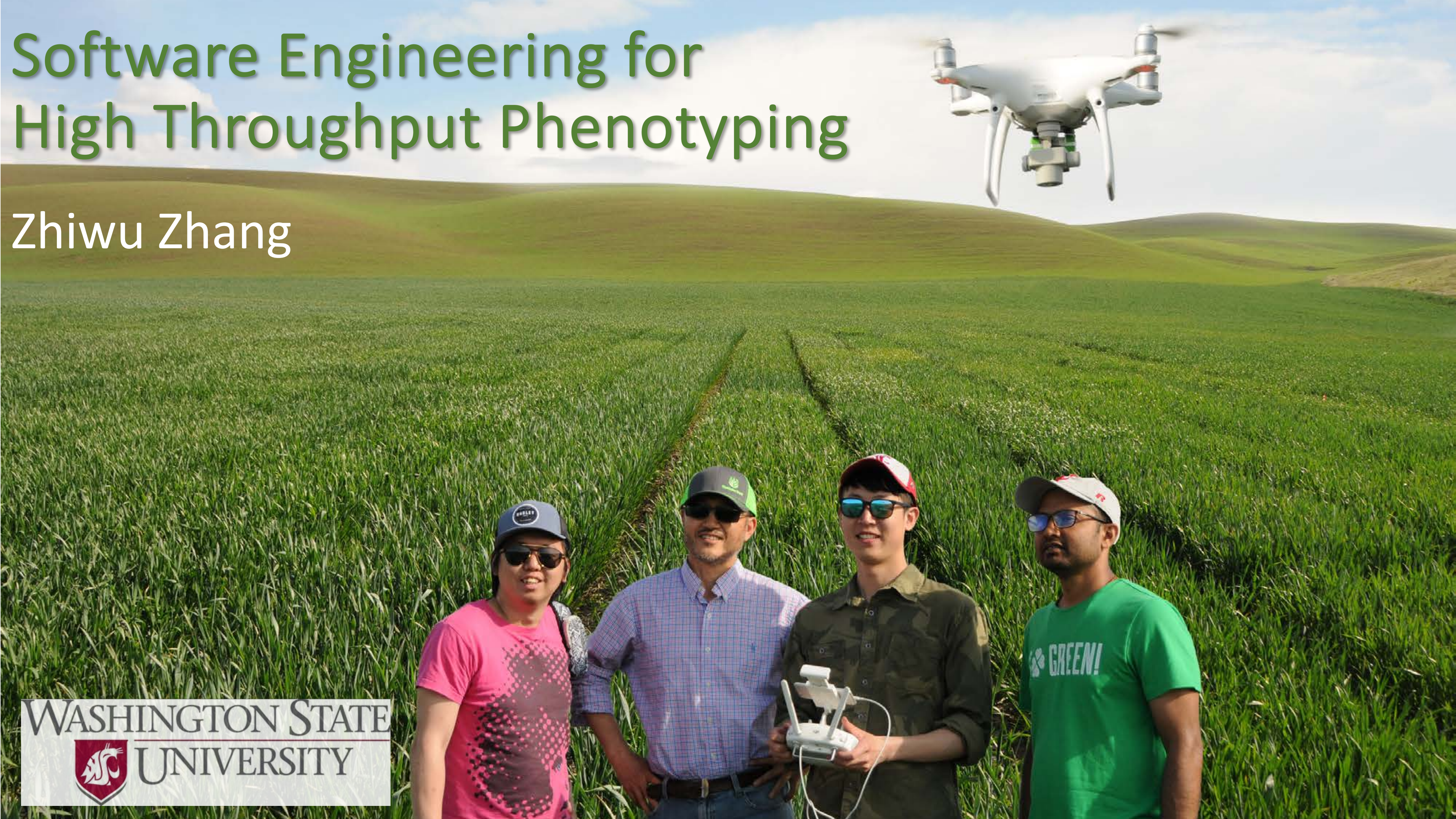
Software Engineering for High Throughput Phenotyping

Zhiwu Zhang



WASHINGTON STATE UNIVERSITY

Zhiwu Zhang Laboratory
for Statistical Genomics

Home People Publication Research Teaching Software Outreach Jobs

Five ingredients to succeed: CS-VMV

Culture: Trying to understand.
Strategy: Solve biological problems with analytical and computational challenges.
Vision: Genomic and phenomic stream data is stationary water for organisms.
Mission: You get data, we help with our analytical methods, tools, and expertise.
Value: Every idea makes sense.

zzlab.net/share

Zhiwu Zhang Laboratory

for Statistical Genomics

Genomics

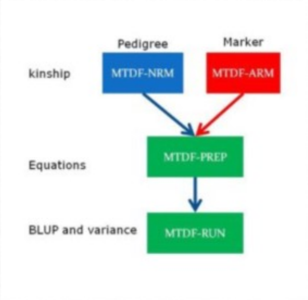
GAPIT

Blink

iPat

FarmCPU

mMAP



MTDFREML

Phenomics

GRID

GridFree

Rooster

LADDER

Audio4EDU

AI4EVER

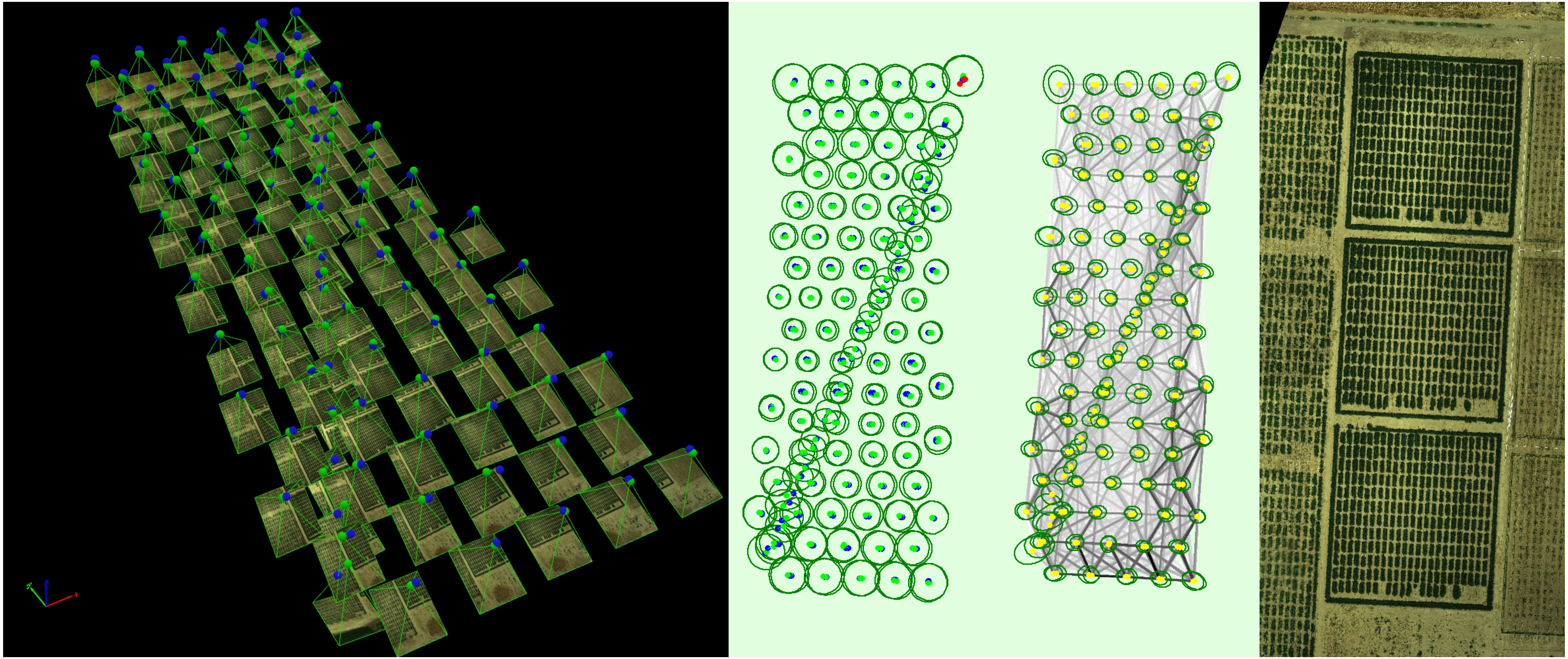
Zhiwu Zhang Laboratory
for Statistical Genomics

Home People Publication Research Teaching Software Outreach Jobs

AI4 Everyone

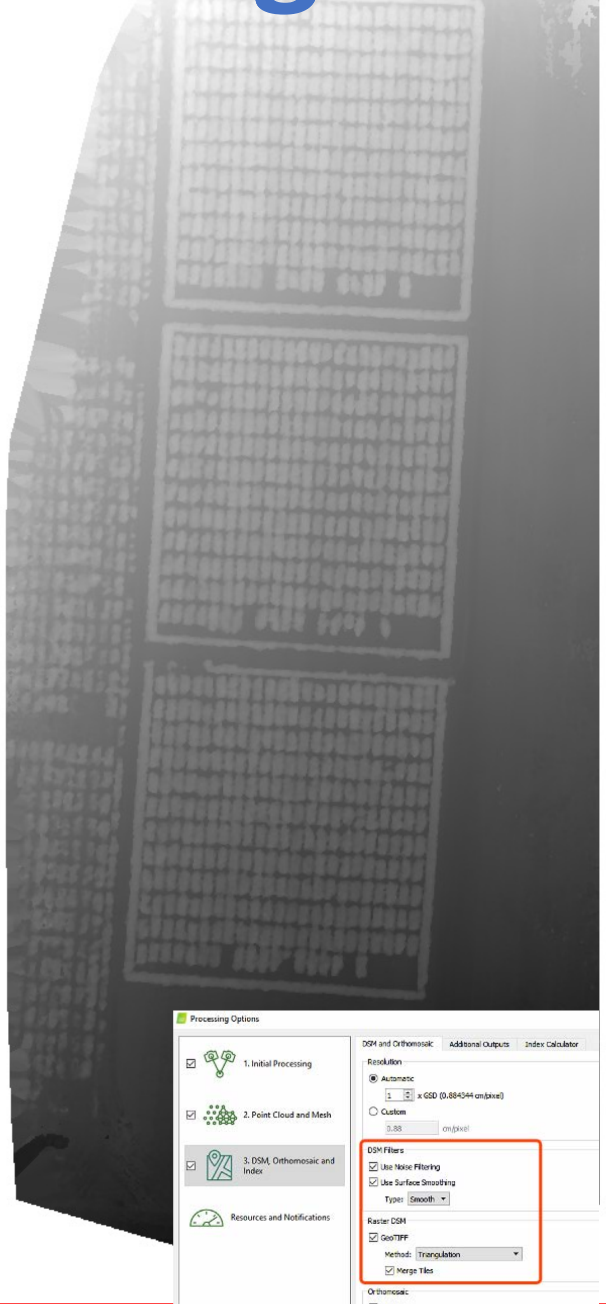
Phenotyping yield is labor and time expensive

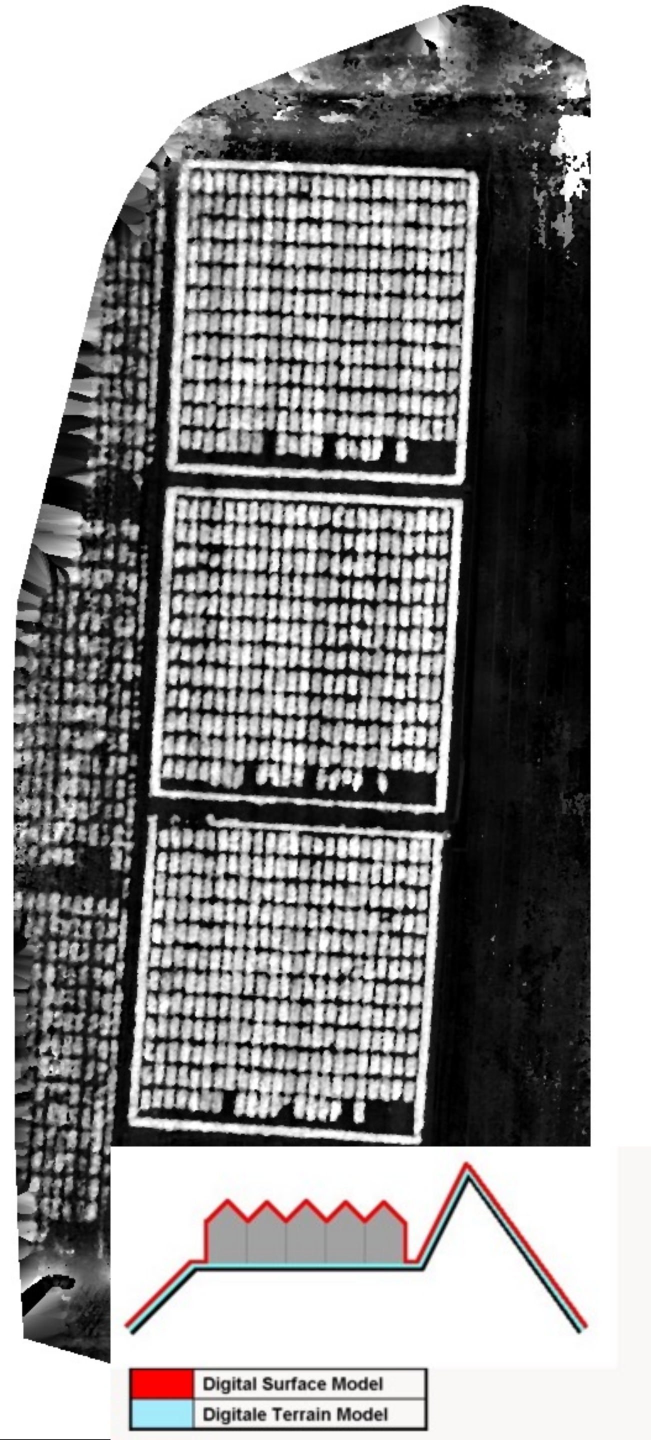
Orthomosaic image using PIX4D



Original images ➡ Layout ➡ Matching ➡ Orthomosaic

Height

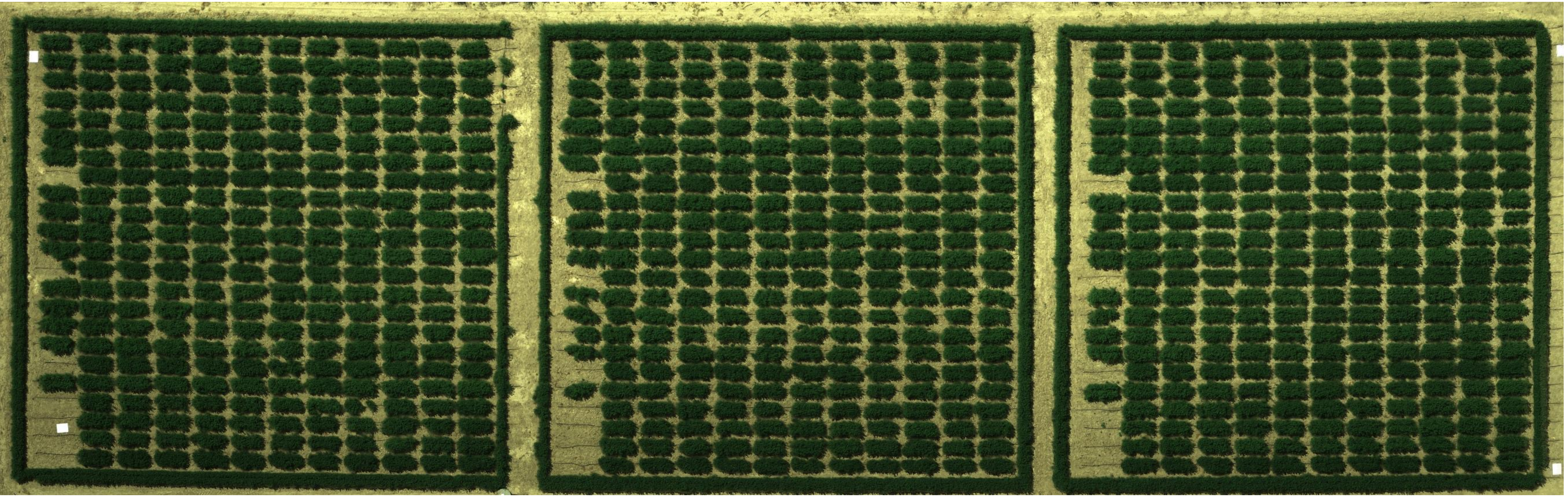




Drone images (RGB)

(May 5, 2019)

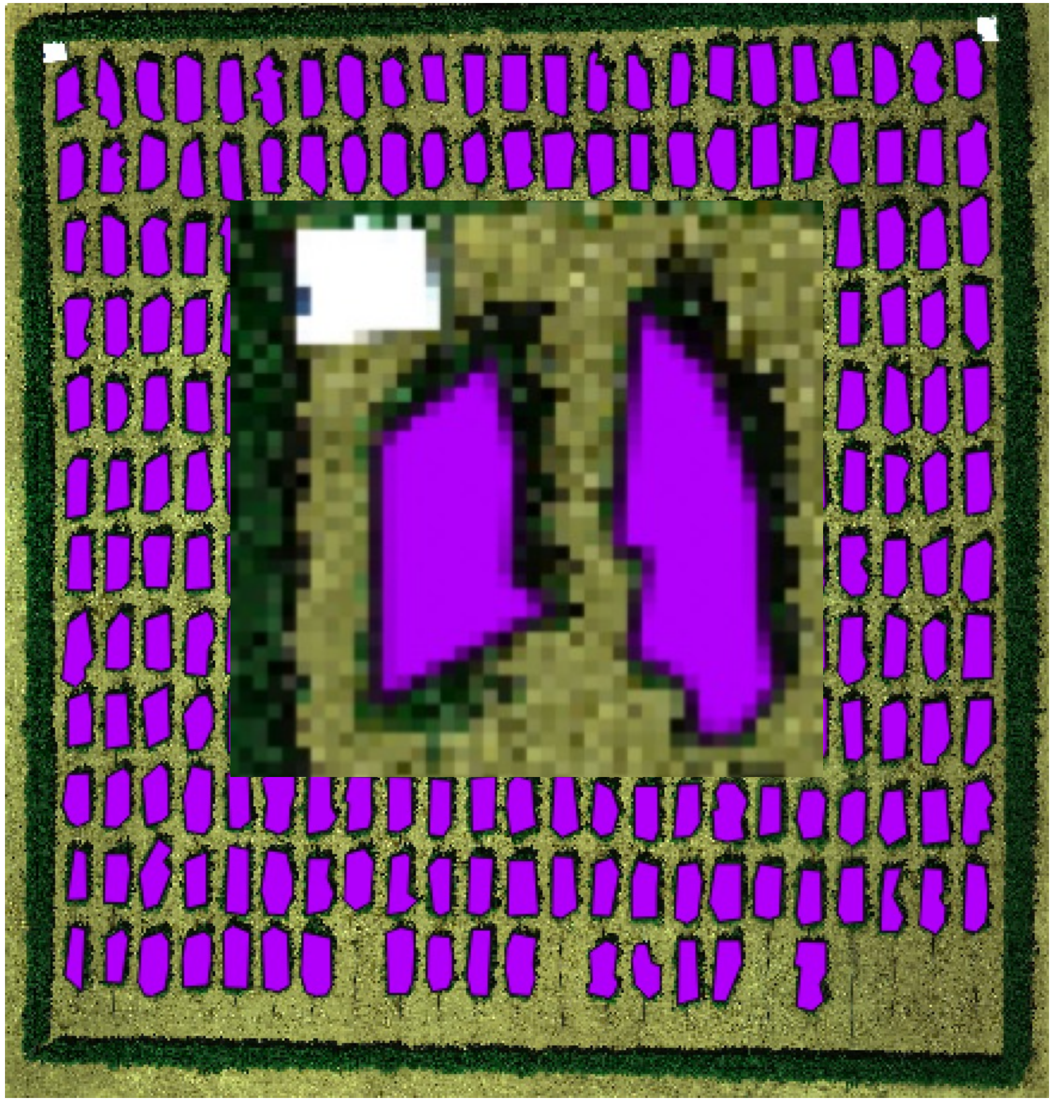
Replicate 3 Replicate 2 Replicate 1



100 feet, ten minutes, ~400 images, joined by PIX4D

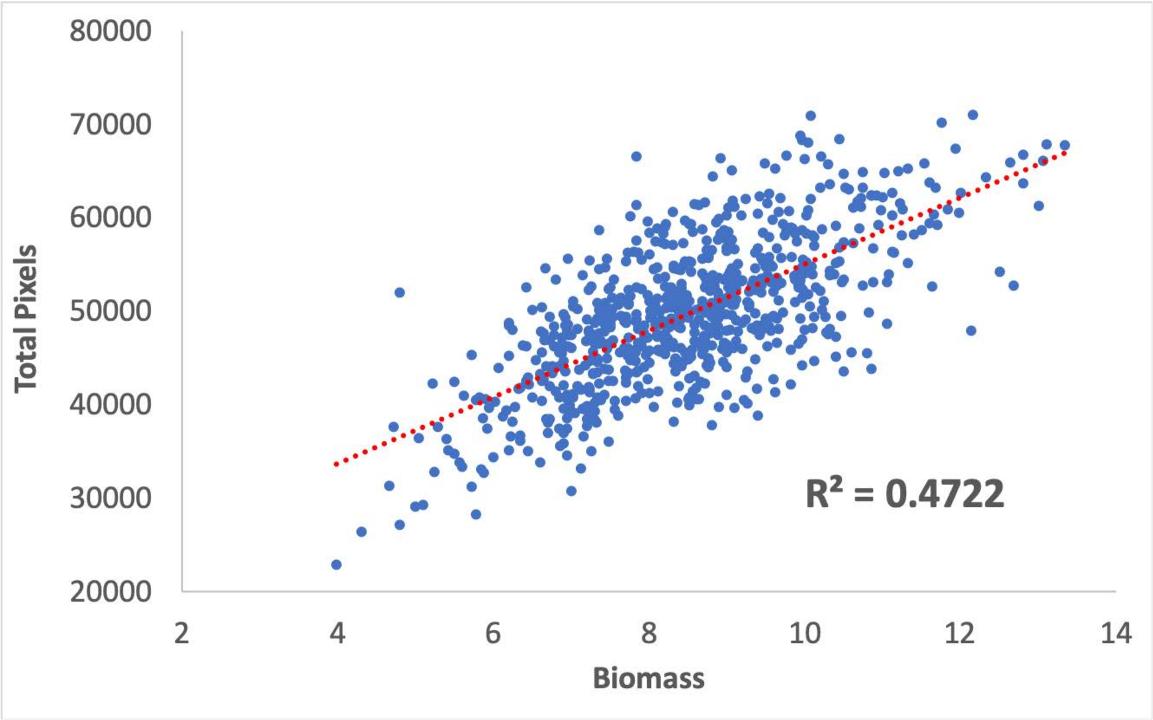
Manual Curation of shape file (QGIS)

Samuel Revolinski
sr.revolinski@uky.edu



	count
1	48567
2	49278
3	46793
4	45305
5	53862
6	45842
7	46060
8	52505
9	57204
10	53092
11	50121
12	54065
13	36114
14	38371
15	41297
16	46786
17	43989

Canopy area explained 50% of biomass variation

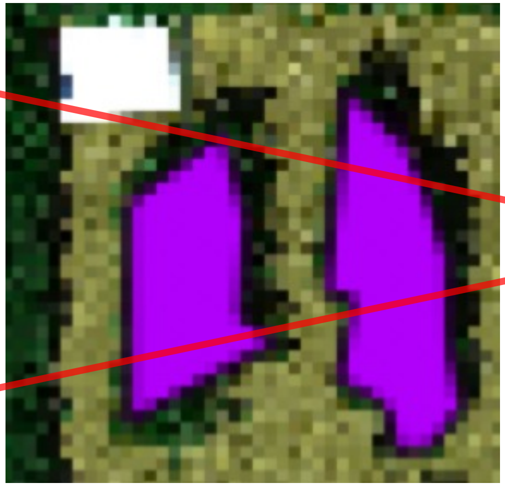


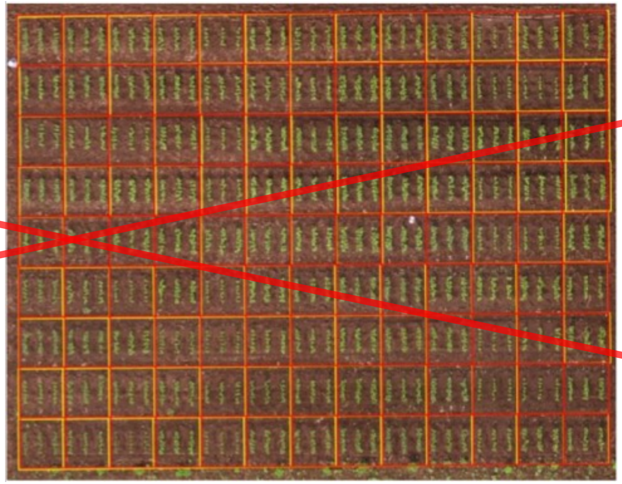
Unpublished data

Four Roadblocks for Using UAV Images

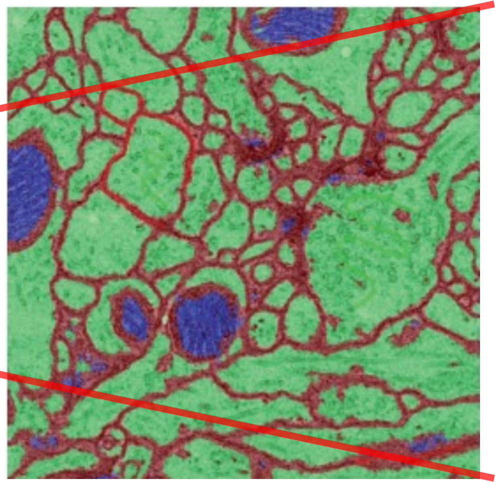
- Depend on ground devices for geographical information
- Manually draw polygons
- Manually draw lines
- Intensive training to extract pixels of interest

https://www.pix4d.com/blog/large-drone-map-yangtze

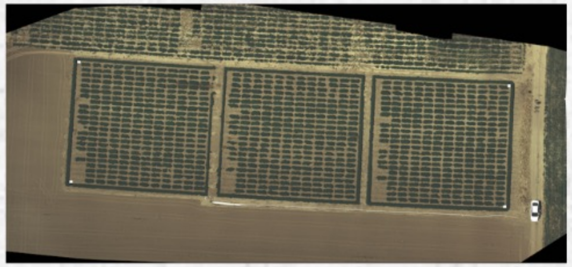


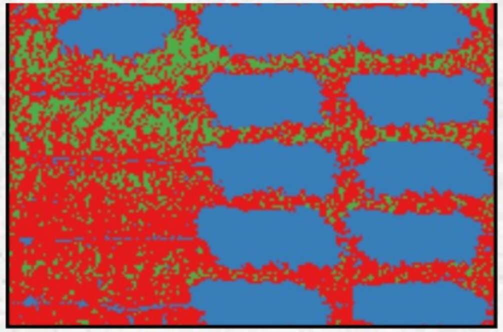


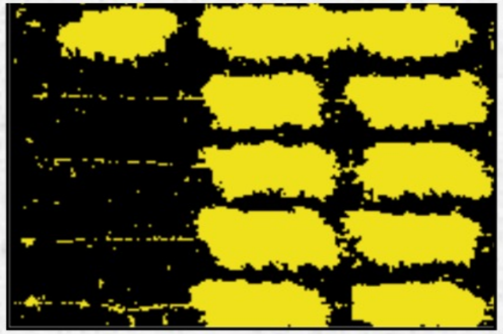
https://spj.sciencemag.org/plantphenomics/2019/2591849/



https://academic.oup.com/view-large/figure/118774504/btx180f1.tif







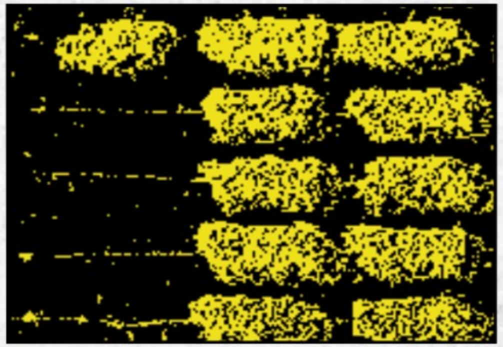
GRID

Easy Way to Extract Info. from **Aerial Images**

DISCOVER MORE

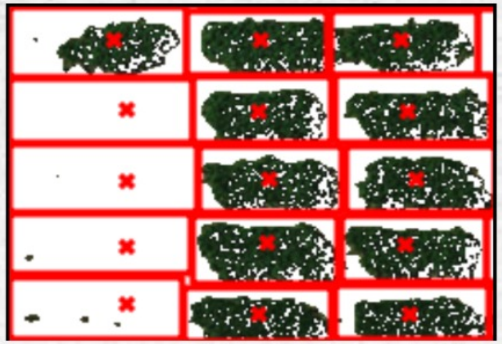
James Chen

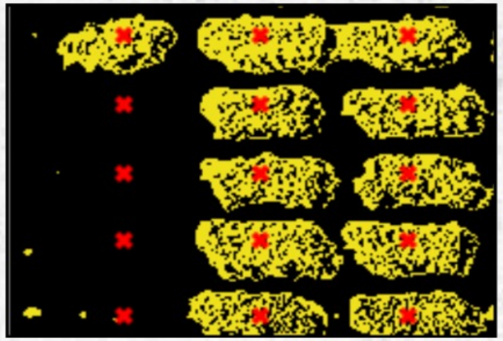
niche@vt.edu

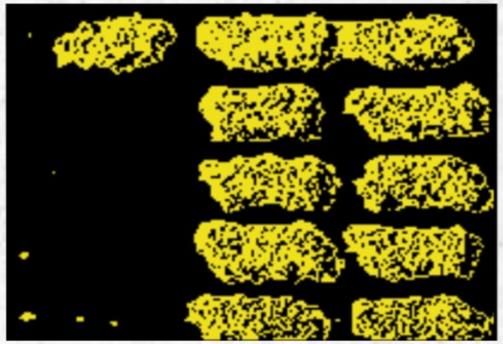


https://doi.org/10.3390/rs12111697

var	row	col	ch_0	ch_1	ch_2	area_veg	NDVI
ID_01	0	0	25.803	87.813	41.814	10144	0.551
ID_02	0	1	25.397	91.118	41.580	7018	0.569
ID_03	1	1	22.636	89.053	39.887	7090	0.598
ID_04	2	1	26.187	89.989	40.921	6465	0.555
ID_05	3	1	24.617	87.876	41.833	9786	0.567
ID_06	4	1	23.870	84.696	40.129	4979	0.568
ID_07	0	2	27.664	87.648	42.068	12526	0.525
ID_08	1	2	21.540	91.220	38.632	14689	0.625
ID_09	2	2	24.423	83.188	40.538	11962	0.552

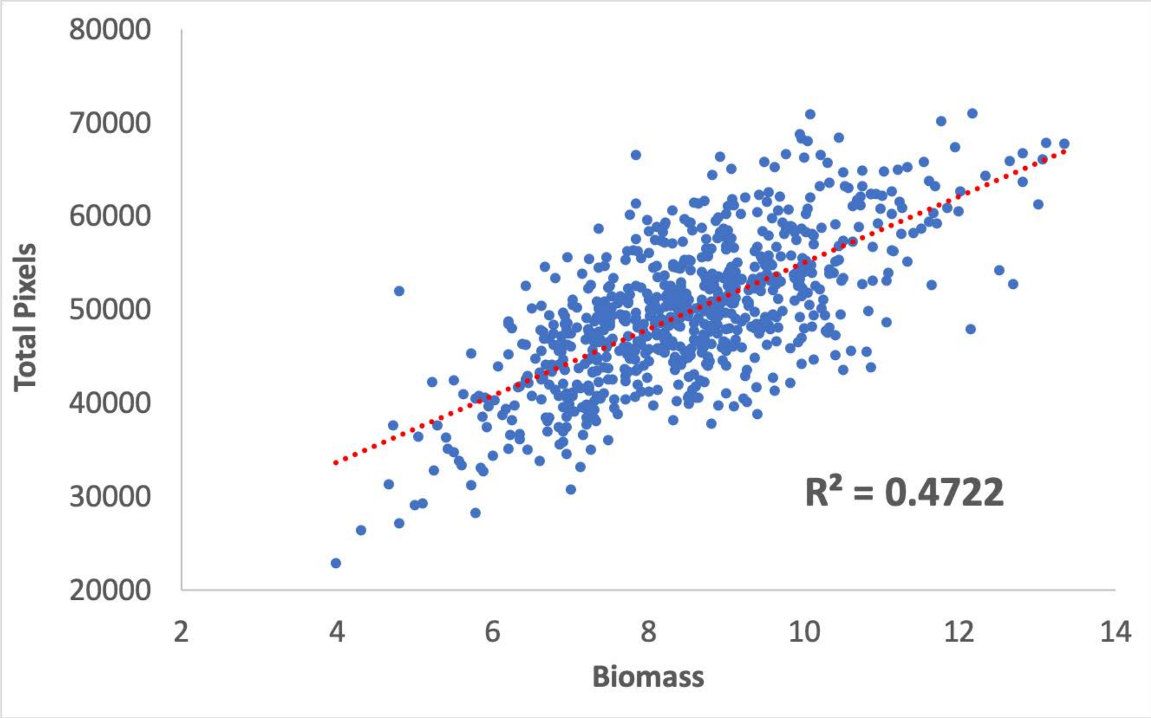




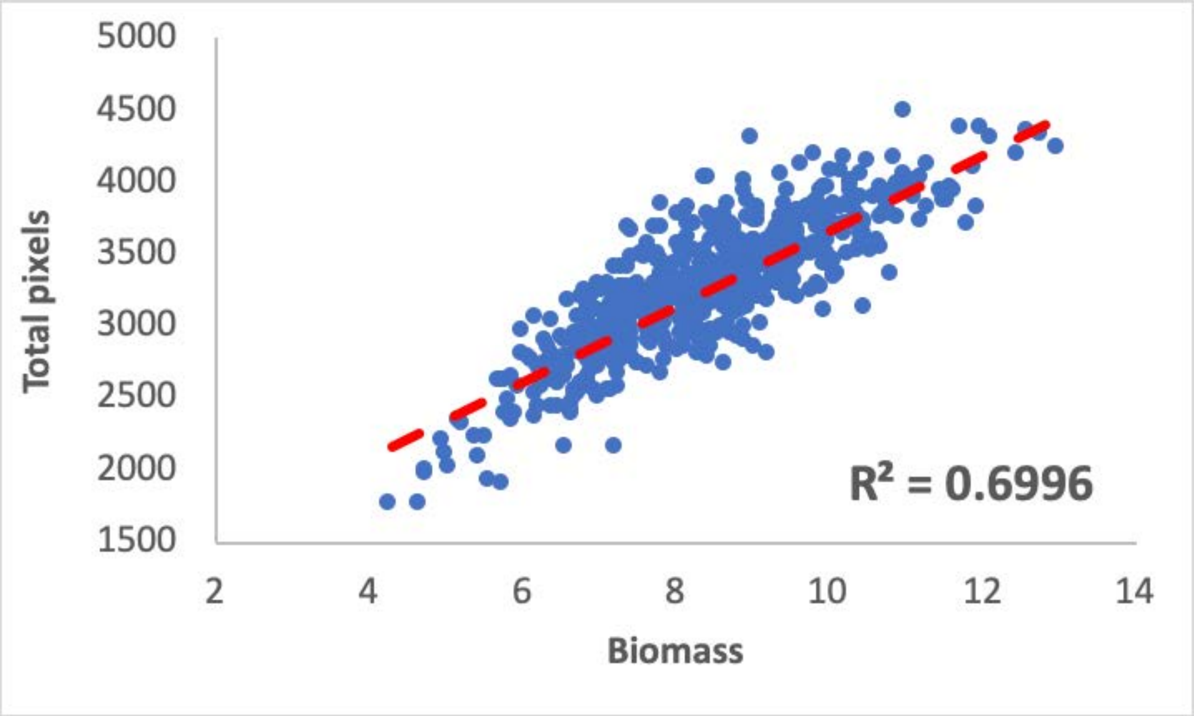


Canopy area explained 70% of biomass variation

Manual

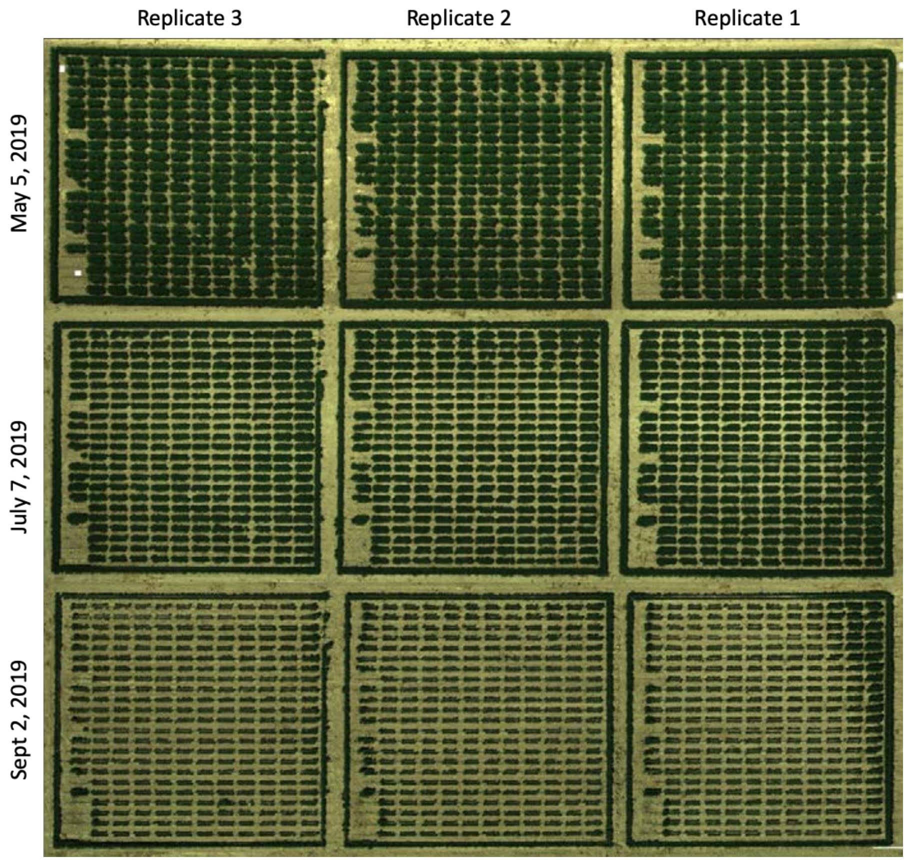


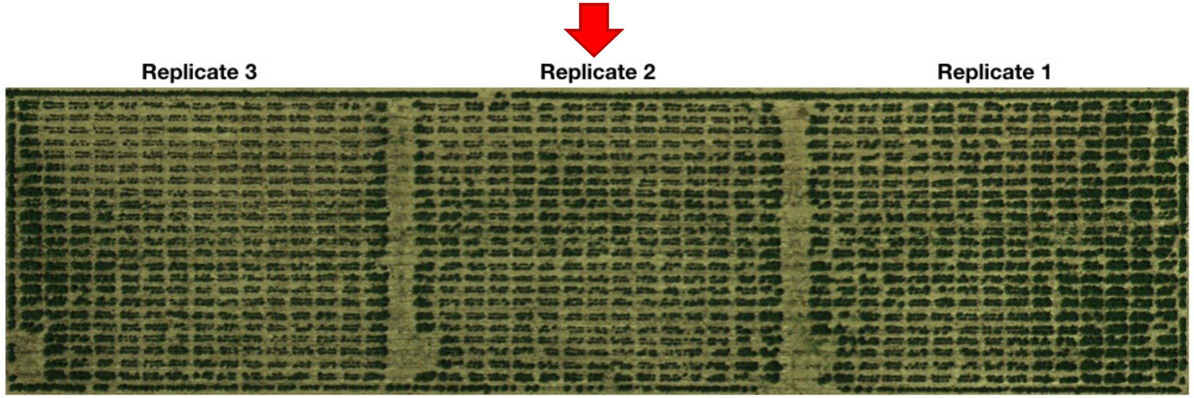
Automatic

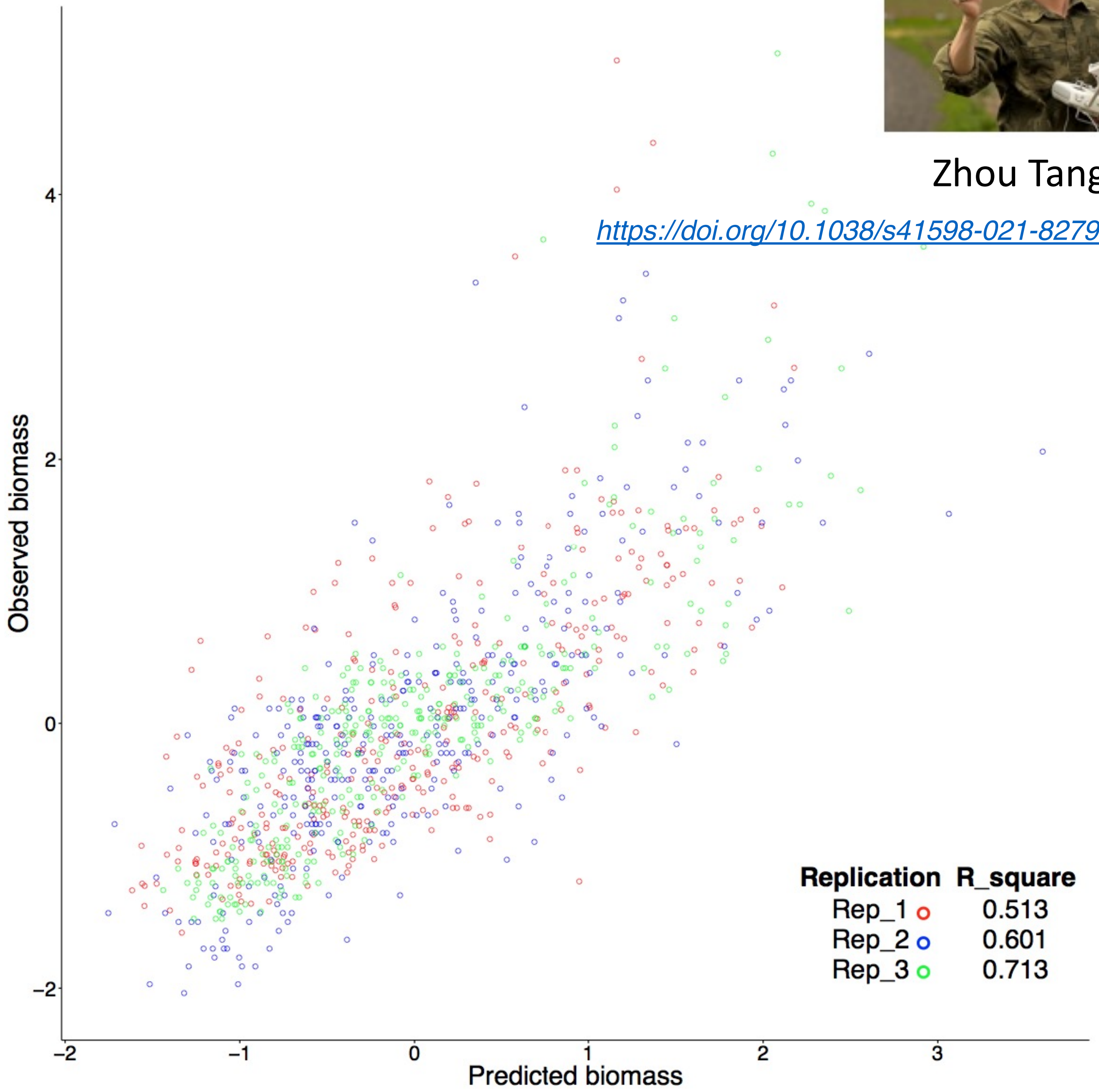


Unpublished data

Independent validation

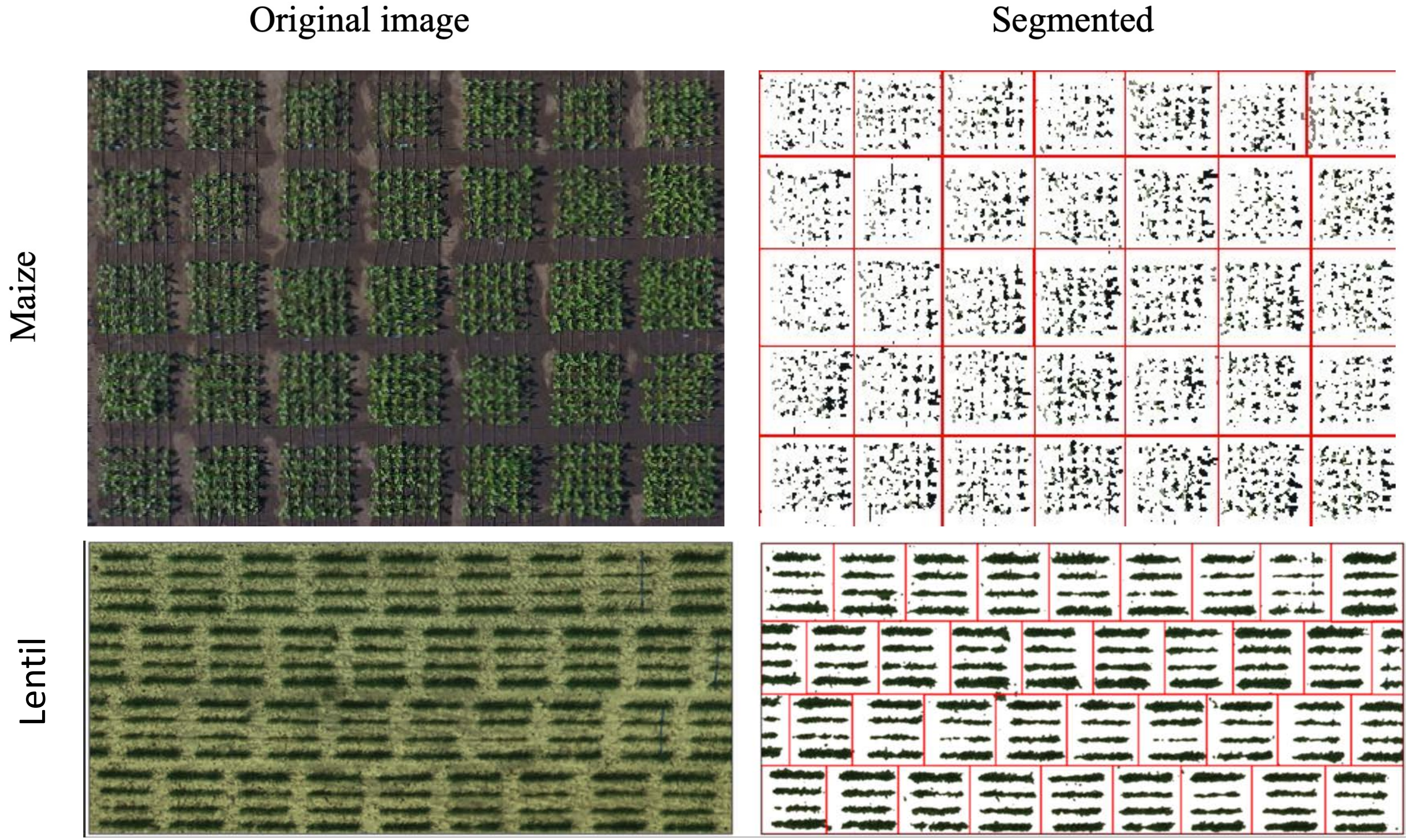




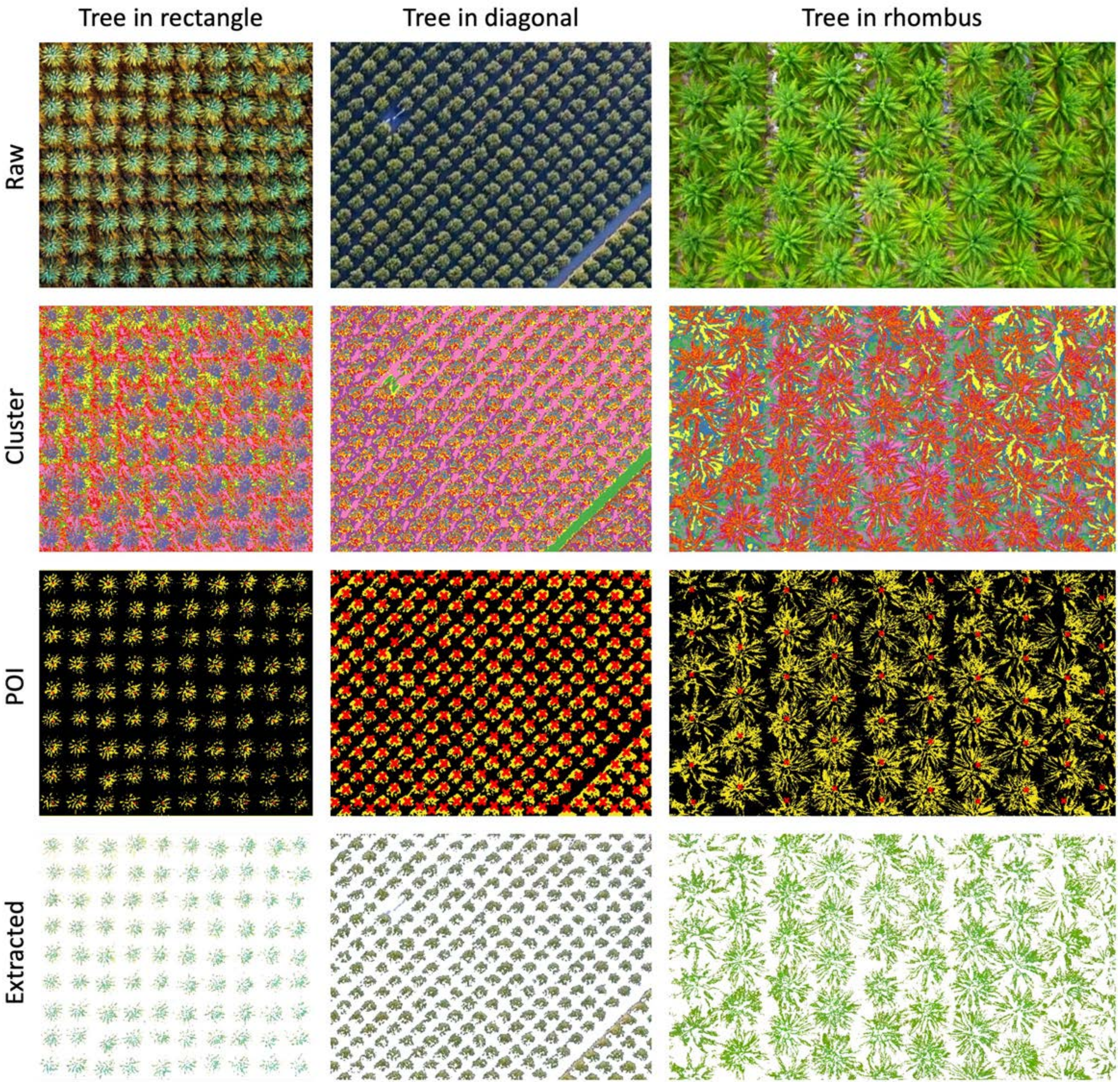


Zhou Tang

https://doi.org/10.1038/s41598-021-82797-x



Original image
Segmented
Maize
Lentil

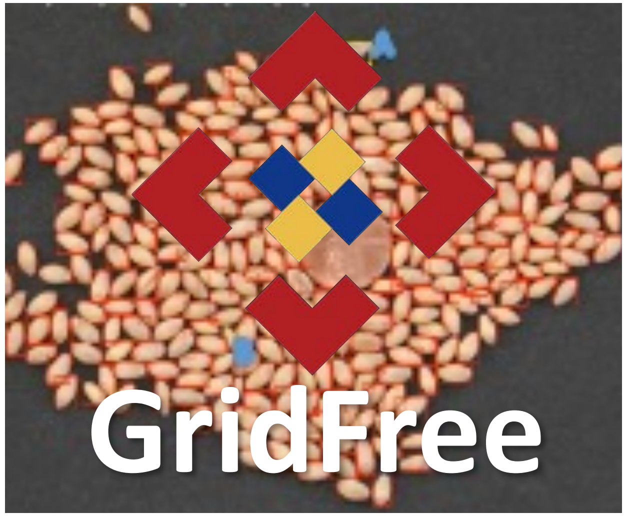


Tree in rectangle
Tree in diagonal
Tree in rhombus
Raw
Cluster
POI
Extracted

Zhiwu Zhang Laboratory
for Statistical Genomics

Home People Publication Research Teaching Software Outreach Jobs

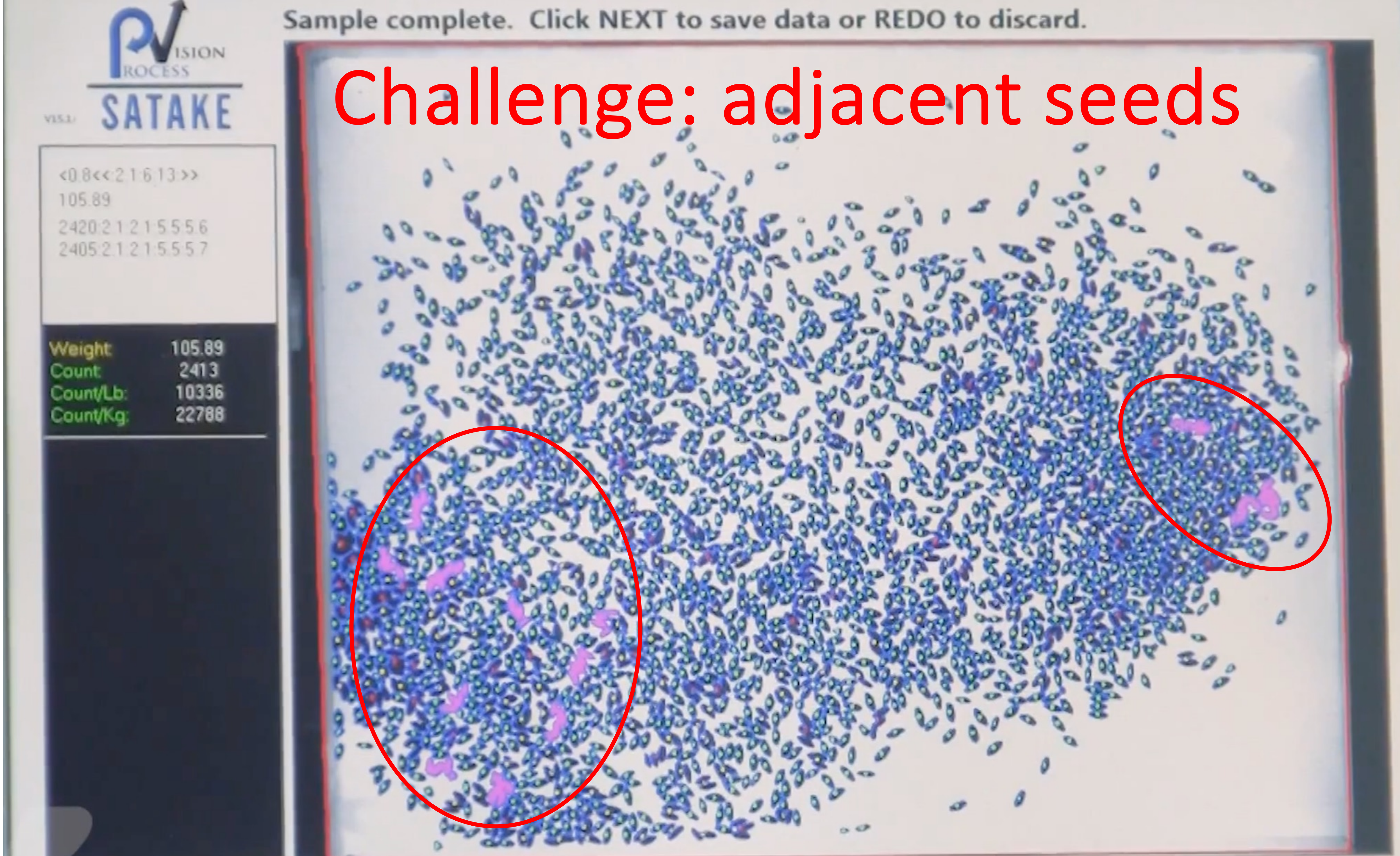
AI4 Everyone



Motivation from counting seeds

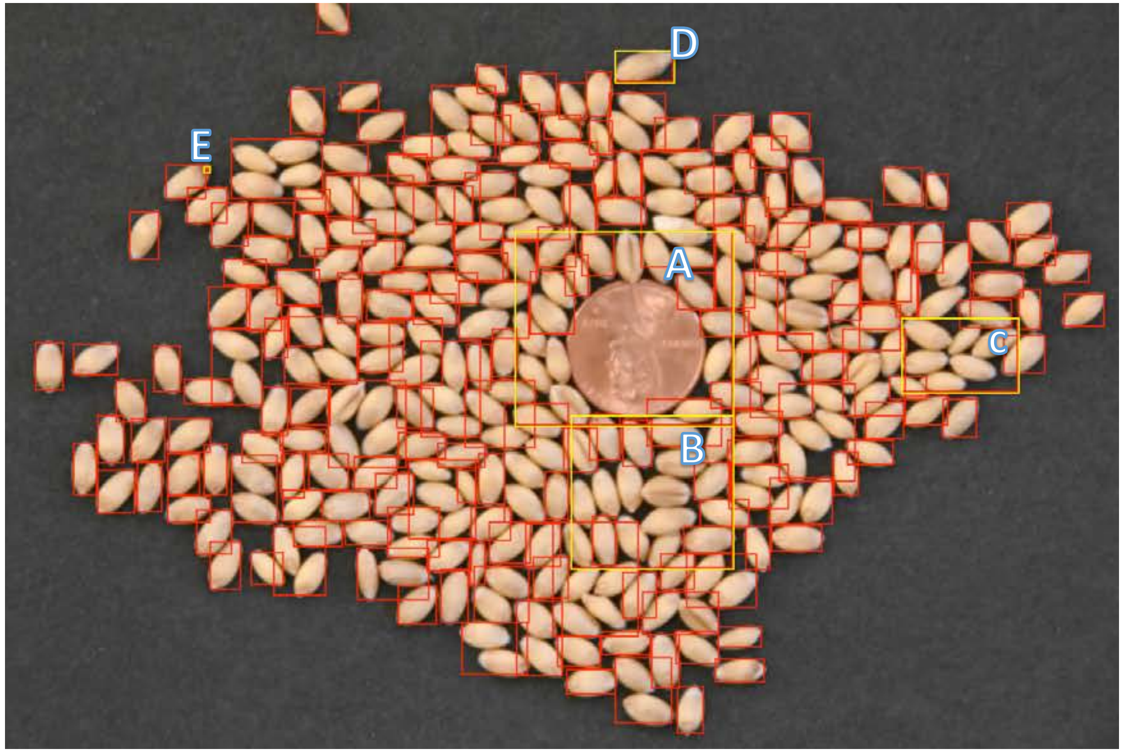
I: Dark age   II: Mechanic age   III: Electronic age   IV: Computer age

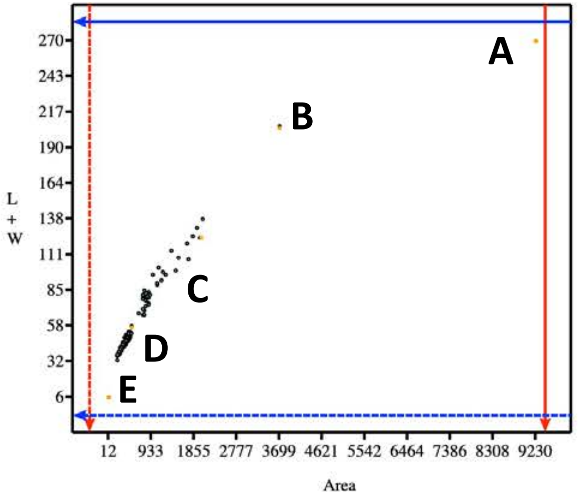
Challenge: adjacent seeds



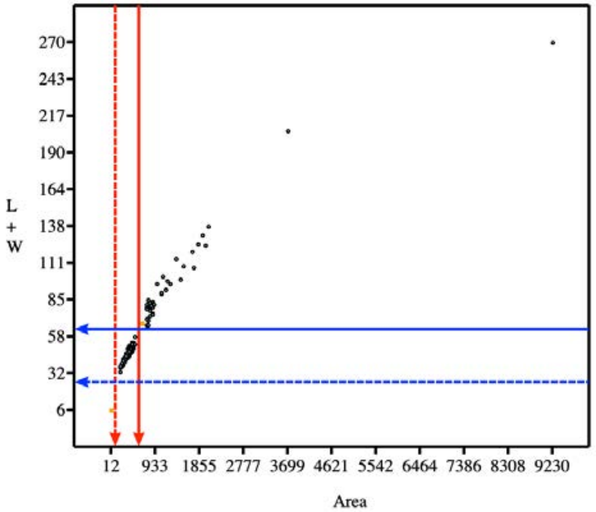
User interaction in GridFree

Yang Hu

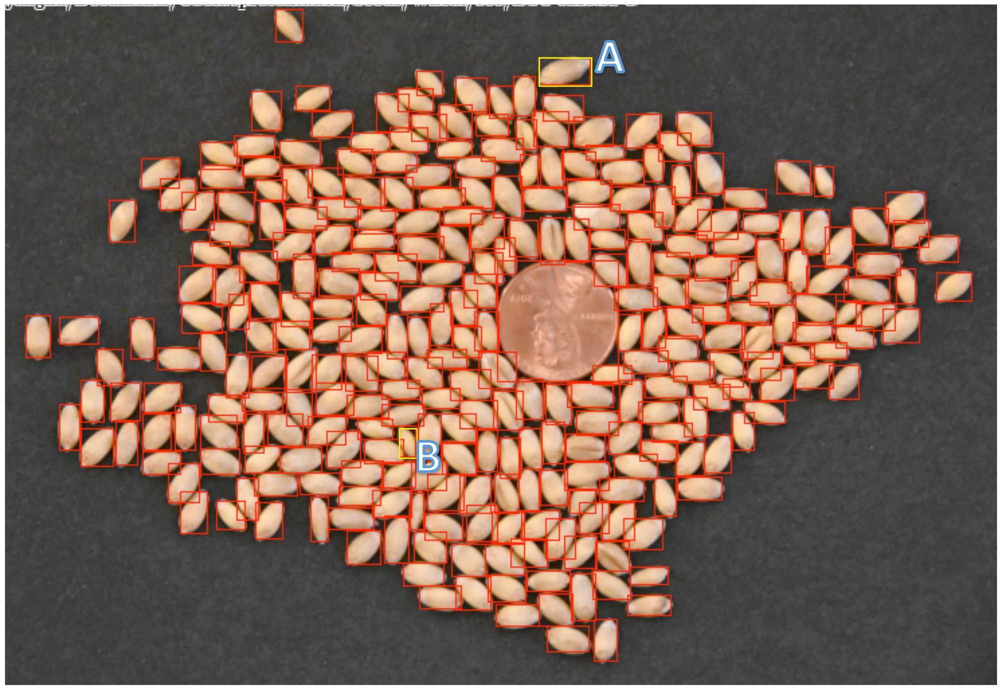


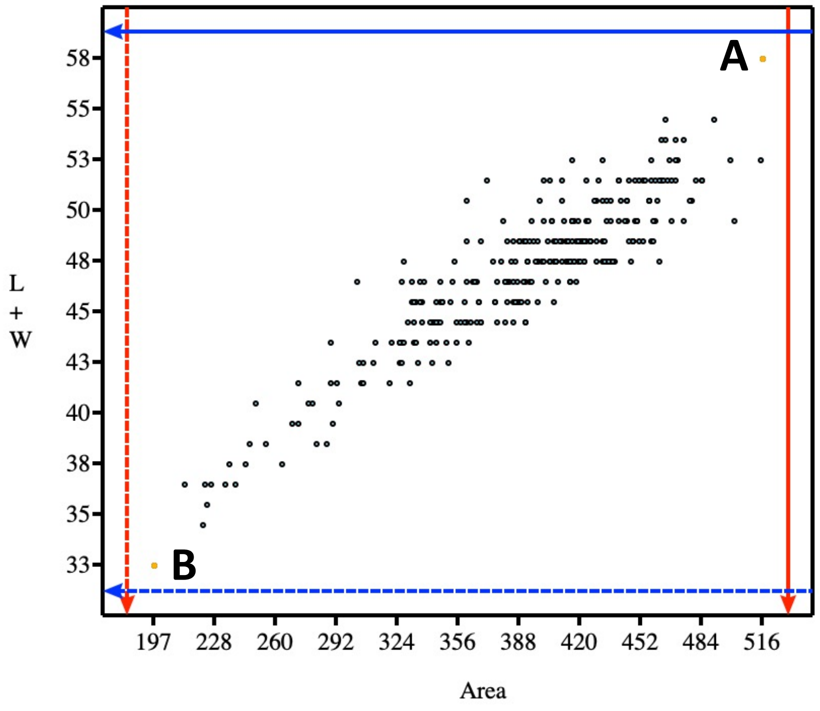


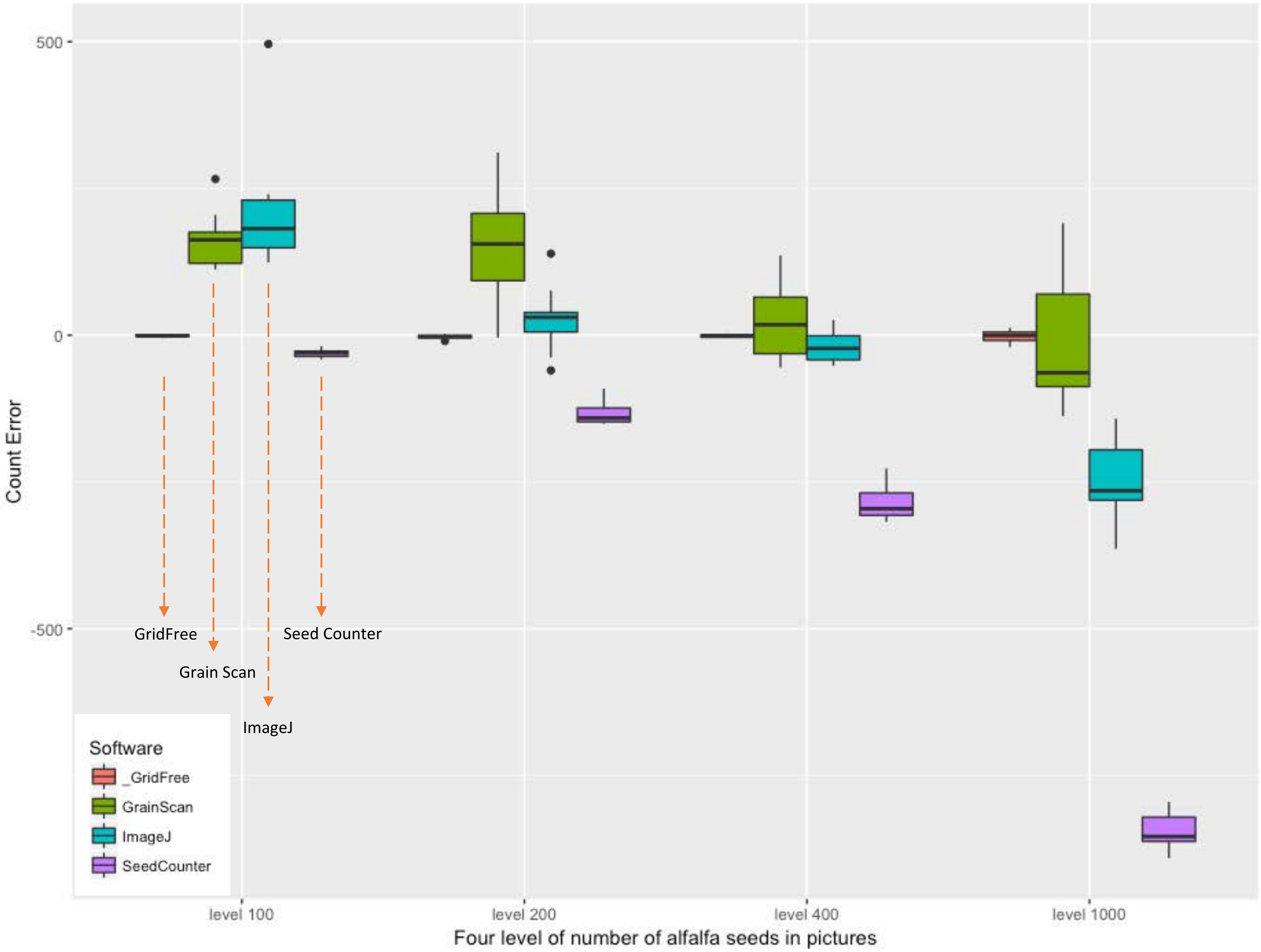
User resets thresholds

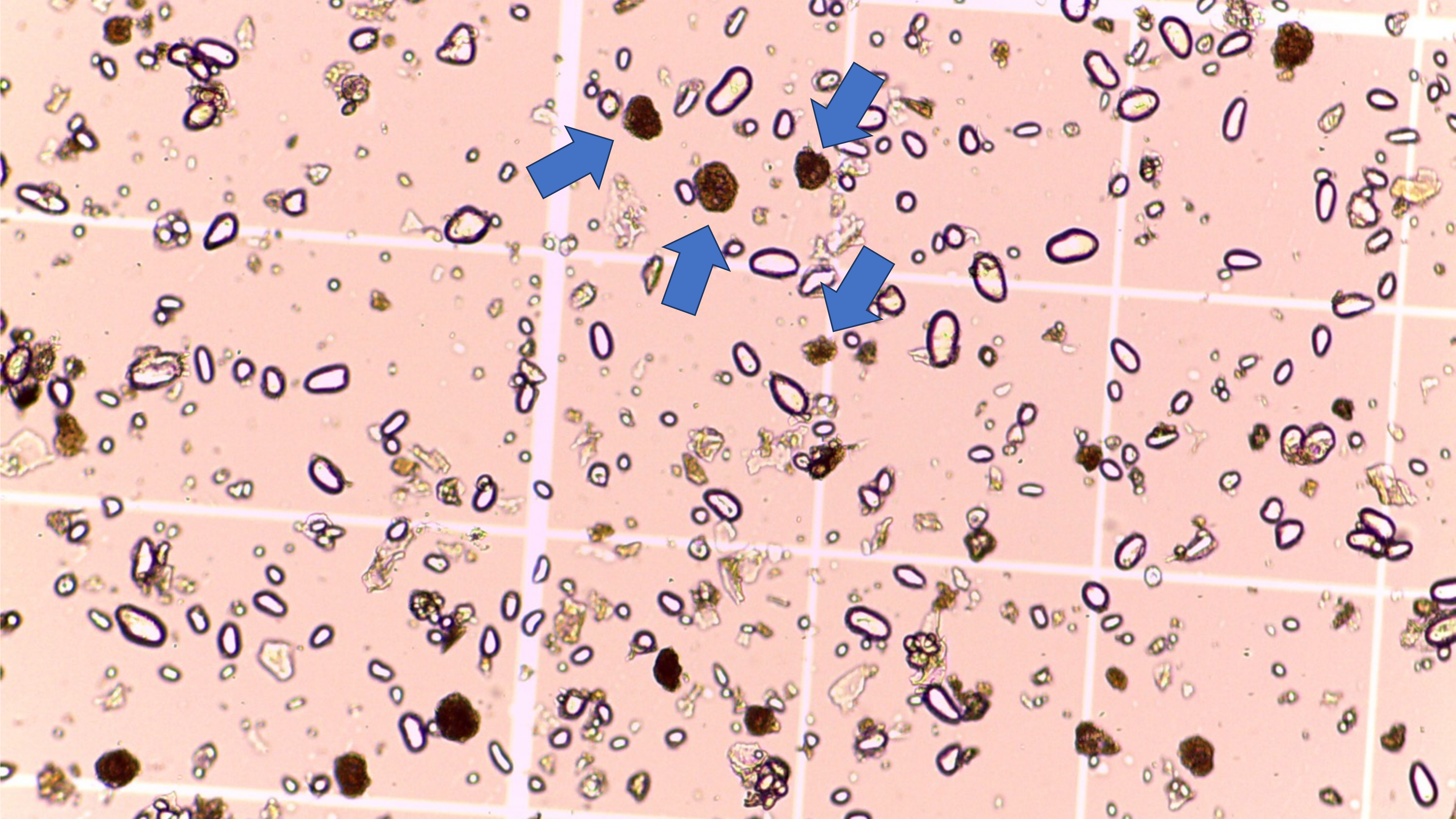


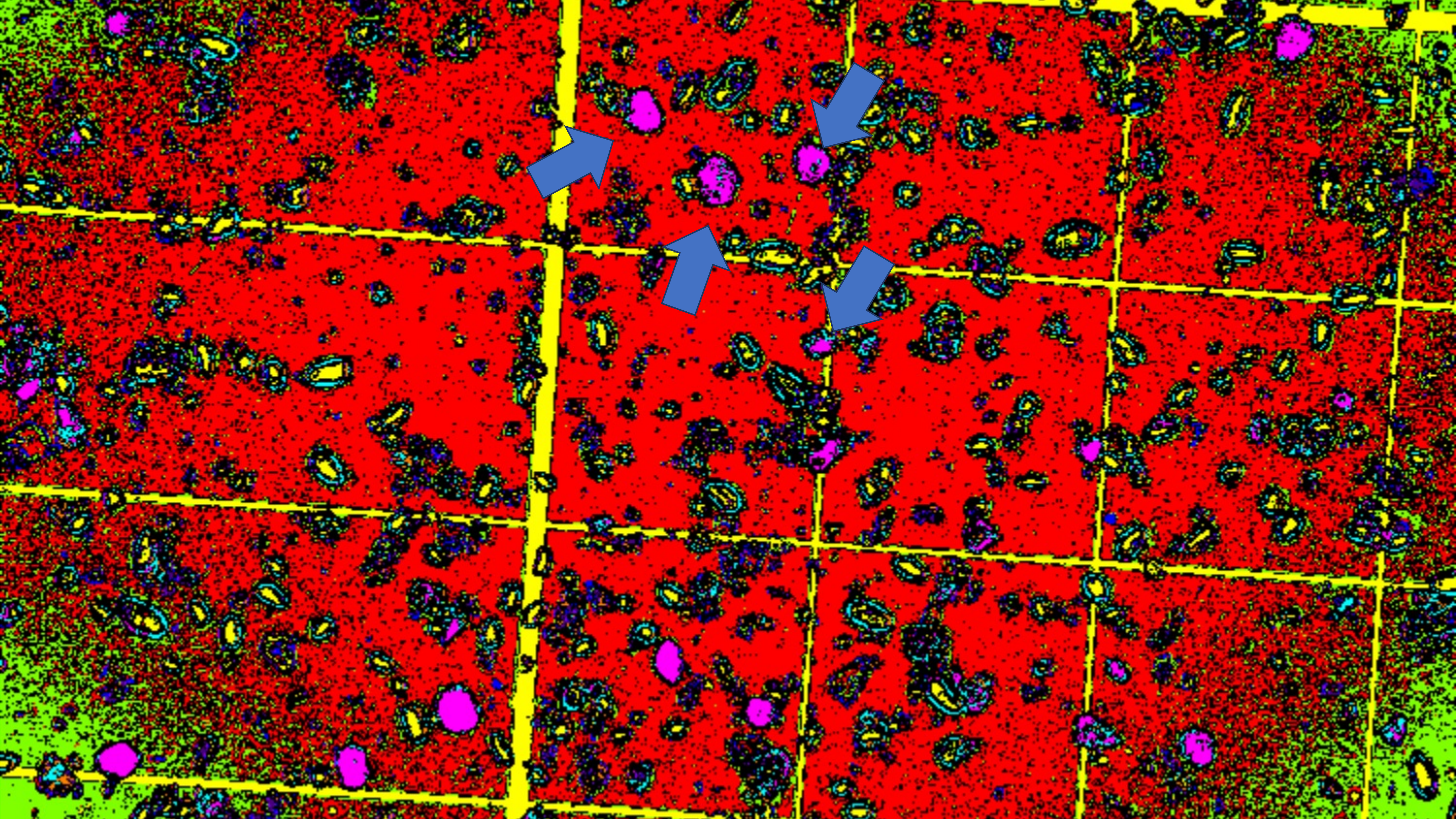
Results of user interaction

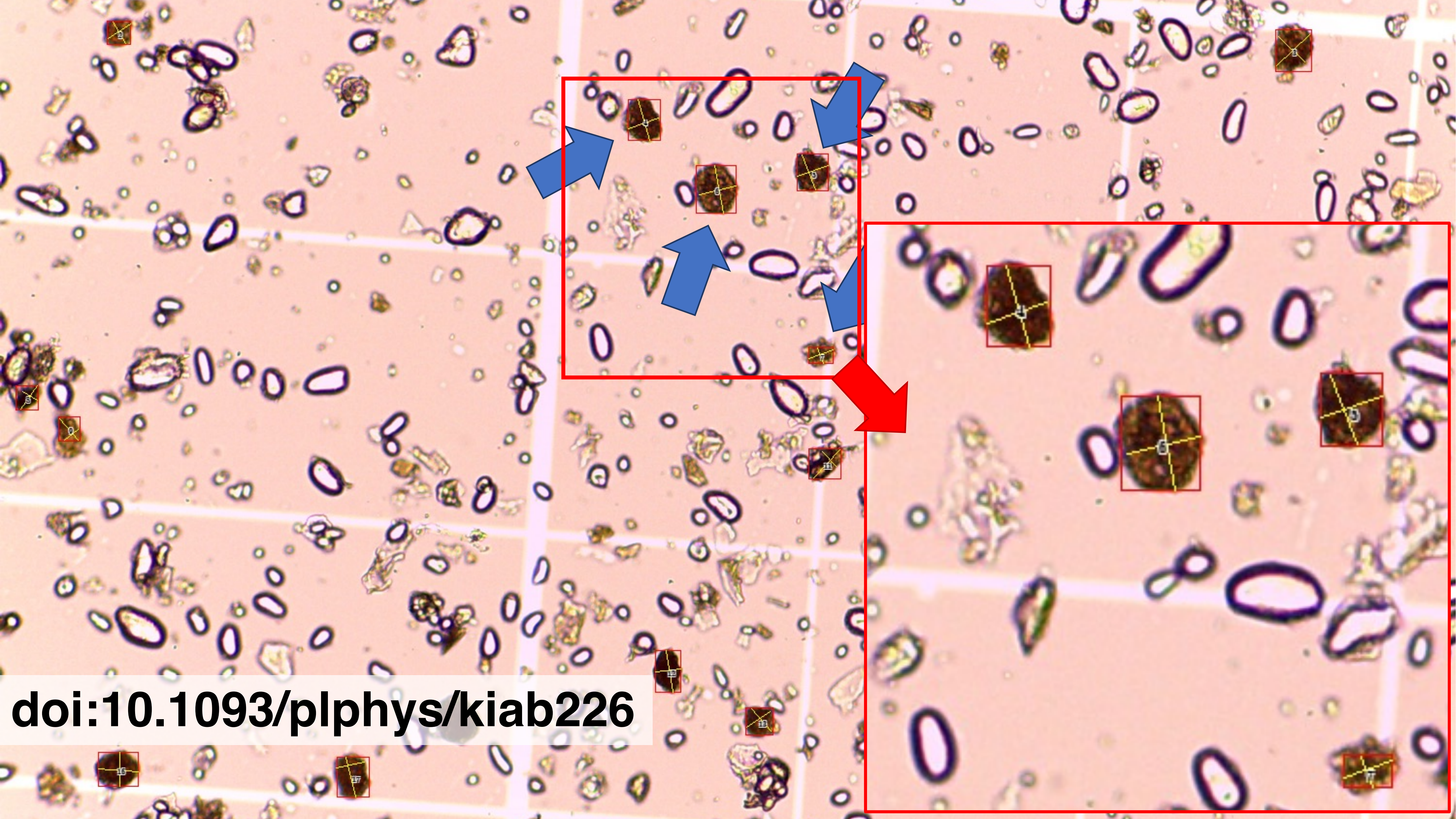





500
0
-500
Count Error
GridFree
Grain Scan
ImageJ
Seed Counter
Software
_GridFree
GrainScan
ImageJ
SeedCounter
level 100
level 200
level 400
level 1000
Four level of number of alfalfa seeds in pictures







doi:10.1093/plphys/kiab226

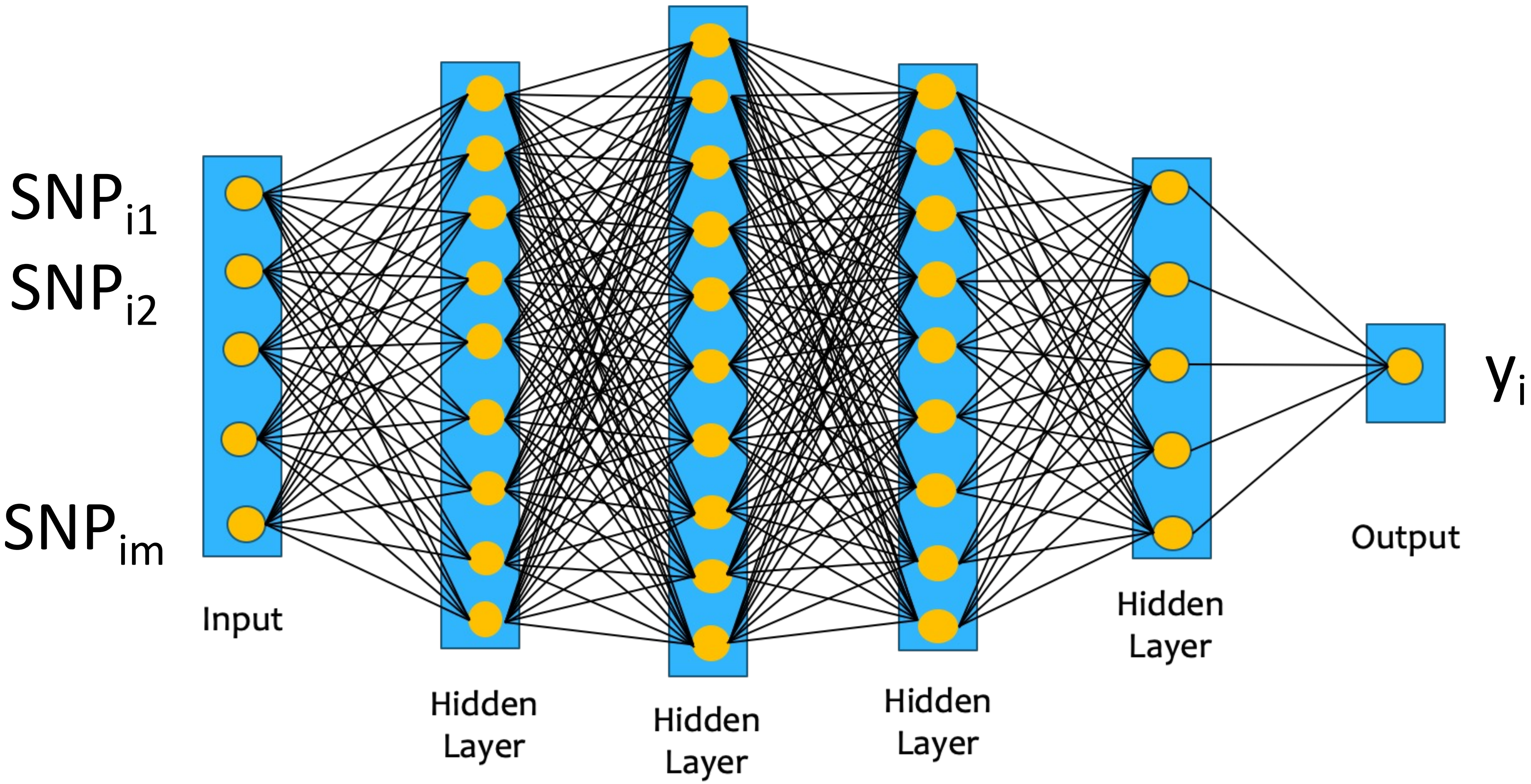
Zhiwu Zhang Laboratory
for Statistical Genomics

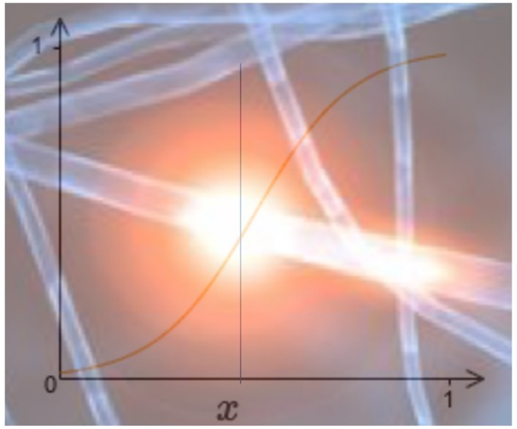
AI4 Everyone

What stories does this image tell?

Early Detection is critical to control stripe rust

Neural Networks





$$\sum w_i x_i \qquad \frac{1}{1 + e^{-(x+b)}}$$

Input Output

Google Scholar

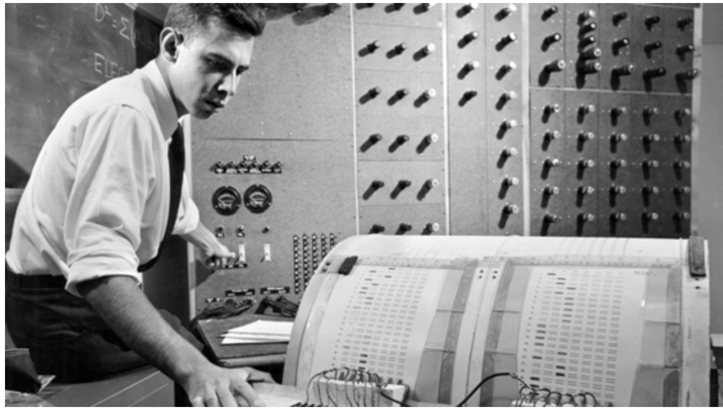
Geoffrey Hinton

FOLLOW

Emeritus Prof. Comp Sci, U.Toronto & Engineering Fellow, Google
Verified email at cs.toronto.edu - Homepage

machine learning psychology artificial intelligence cognitive science computer science

TITLE	CITED BY	YEAR
Imagenet classification with deep convolutional neural networks A Krizhevsky, I Sutskever, GE Hinton Advances in neural information processing systems 25	138063 *	2012
Deep learning Y LeCun, Y Bengio, G Hinton Nature 521 (7553), 436-44	66972	2015
Dropout: a simple way to prevent neural networks from overfitting N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov The journal of machine learning research 15 (1), 1929-1958	44640	2014
Visualizing data using t-SNE L van der Maaten, G Hinton Journal of Machine Learning Research 9 (Nov), 2579-2605	35659	2008
Learning representations by back-propagating errors DE Rumelhart, GE Hinton, RJ Williams Nature 323 (6088), 533-536	34206	1986



Frank Rosenblatt (Cornell)

Marvin Minsky (MIT)

Geoffrey Hinton
Godfather of AI

B.A., Psychology, Cambridge
Ph.D., AI, Edinburgh

CARNEGIE MELLON UNIVERSITY
PITTSBURGH PENNSYLVANIA 1900

UNIVERSITY OF TORONTO

George Cybenko's Theorem

Math. Control Signals Systems (1989) 2: 303–314

Impact factor 1.518
Cited by 17,751

Mathematics of Control, Signals, and Systems

Simple neural networks can *represent* a wide variety of interesting functions

Approximation by Superpositions of a Sigmoidal Function*

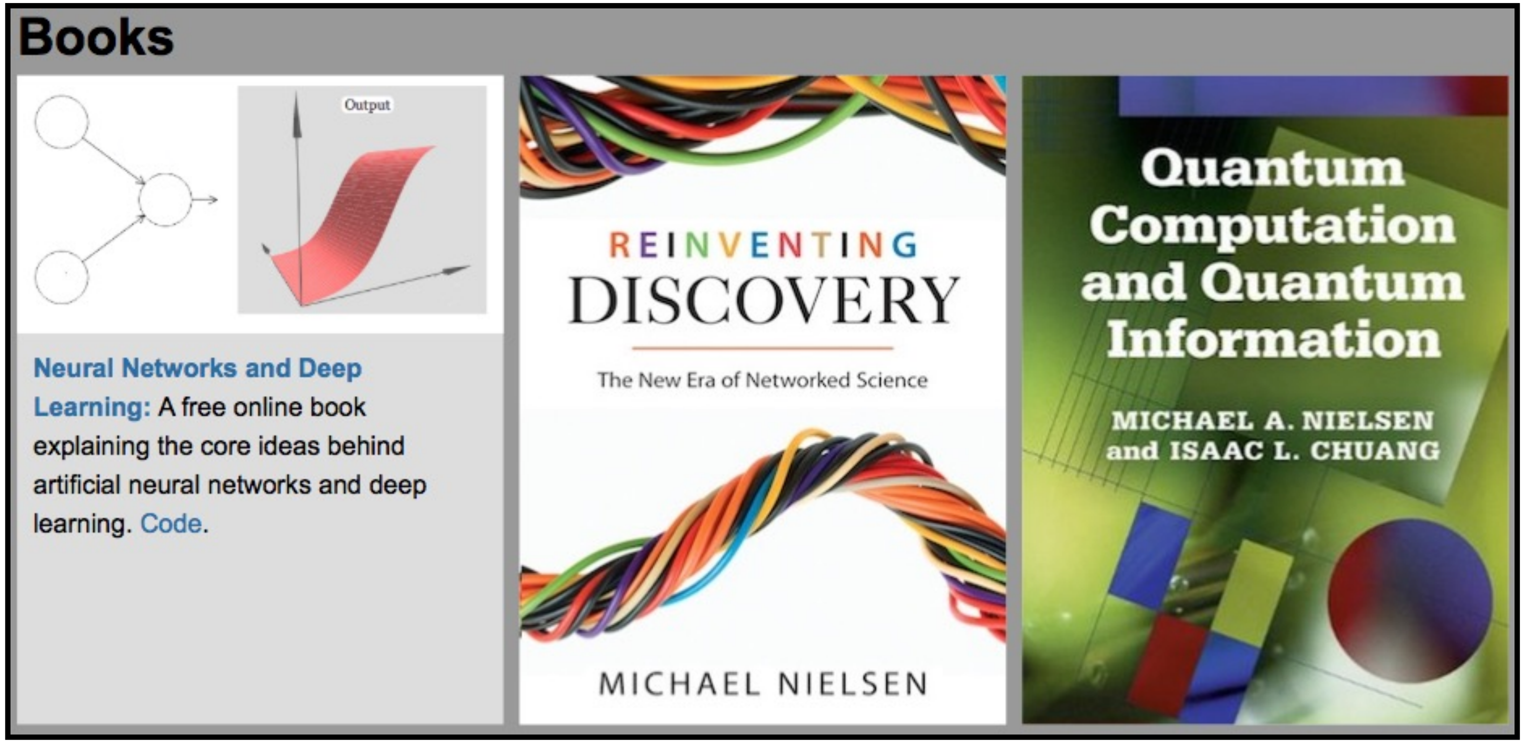
G. Cybenko†

Abstract. In this paper we demonstrate that finite linear combinations of compositions of a fixed, univariate function and a set of affine functionals can uniformly approximate any continuous function of n real variables with support in the unit hypercube; only mild conditions are imposed on the univariate function. Our results settle an open question about representability in the class of single hidden layer neural networks. In particular, we show that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single internal, hidden layer and any continuous sigmoidal nonlinearity. The paper discusses approximation properties of other possible types of nonlinearities that might be implemented by artificial neural networks.

Computer demonstration

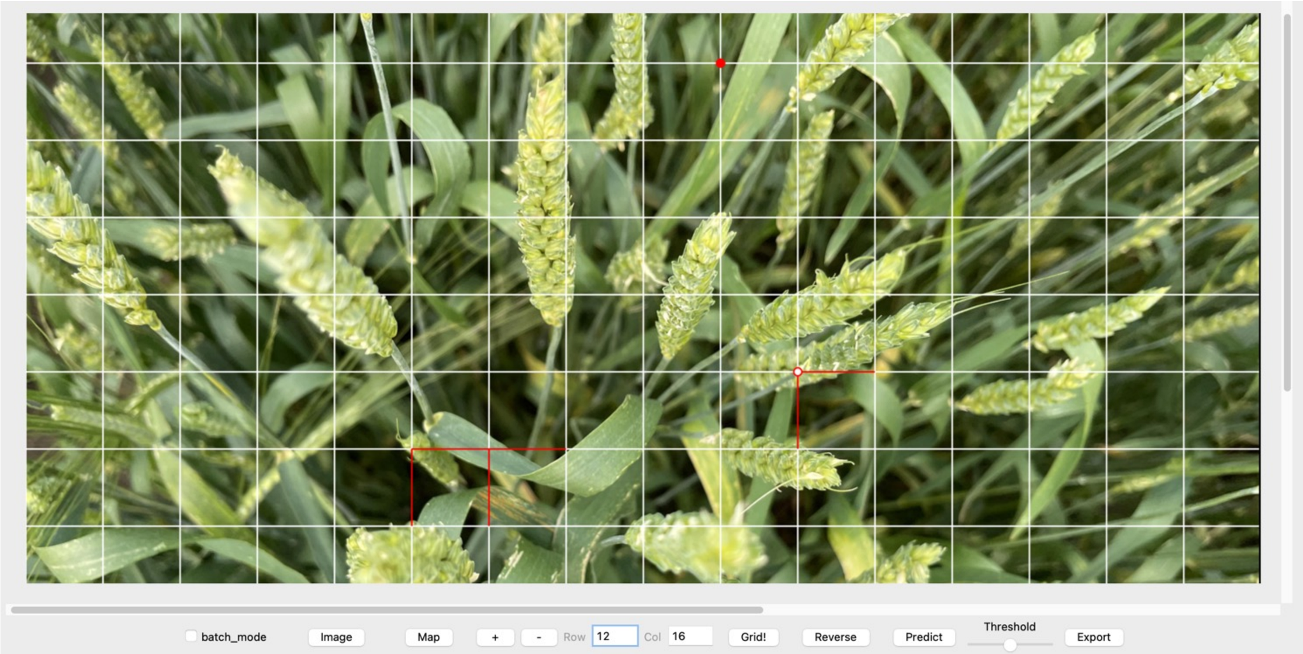
CHAPTER 4

A visual proof that neural nets can compute any function



By Michael Nielsen / Dec 2019

Wheat stripe rust early detection



Yang Hu

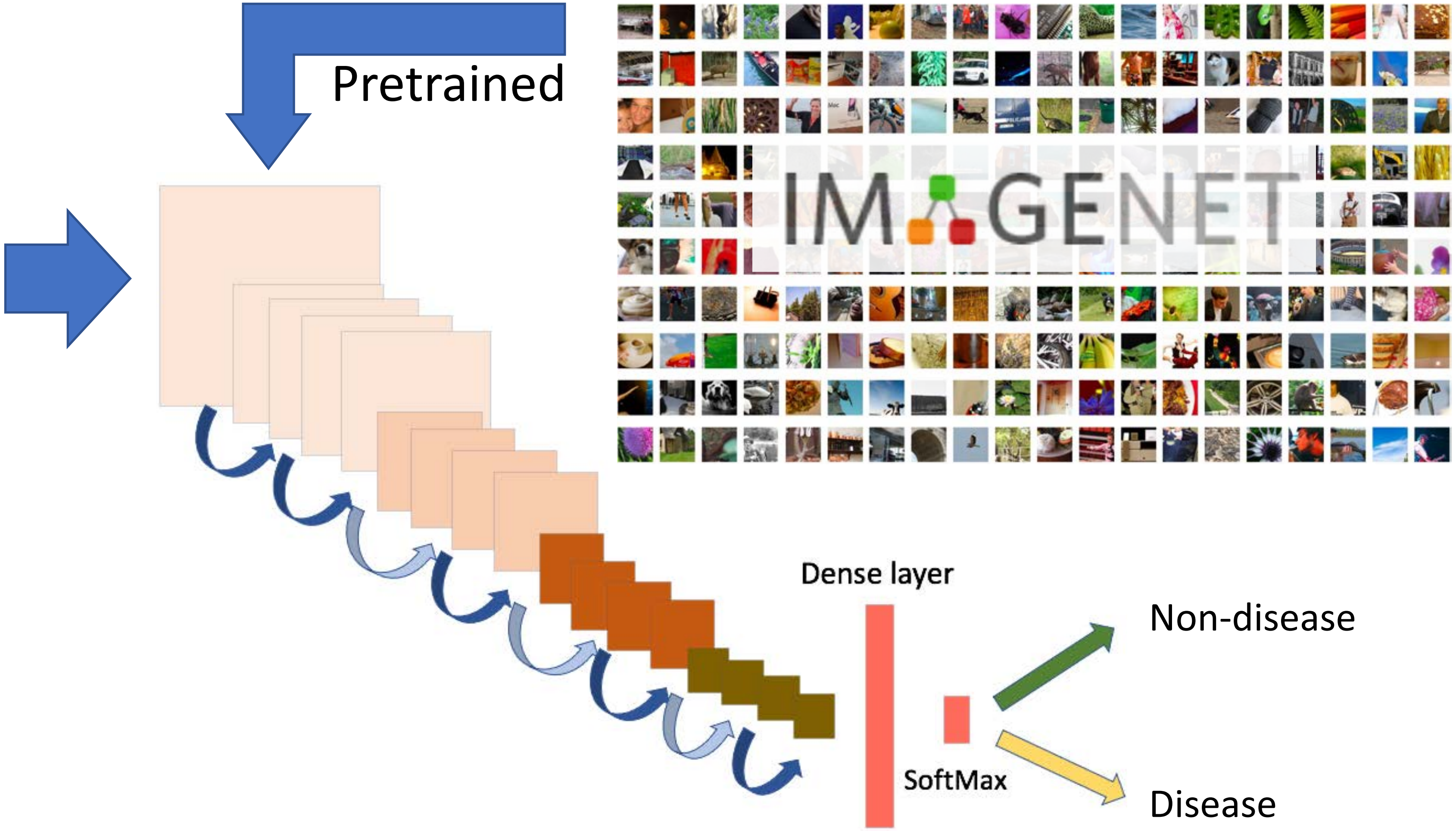
Zhou Tang

https://zzlab.net/Rooster

Affordable High Throughput Field Detection of Wheat Stripe Rust Using Deep Learning Based on Semi Automatic Image Labeling

Zhou Tang[1], Meinan Wang[2], Michael Schirrmann[3], Karl-Heinz Dammer[3], Robert Brueggeman[1], Xianran Li[1], Sankaran, Sindhuja[4], Mike Pumphrey[1], Yang Hu[1*], Xianming Chen[2,5*], and Zhiwu Zhang[1*]

Residual Neural Network Architecture



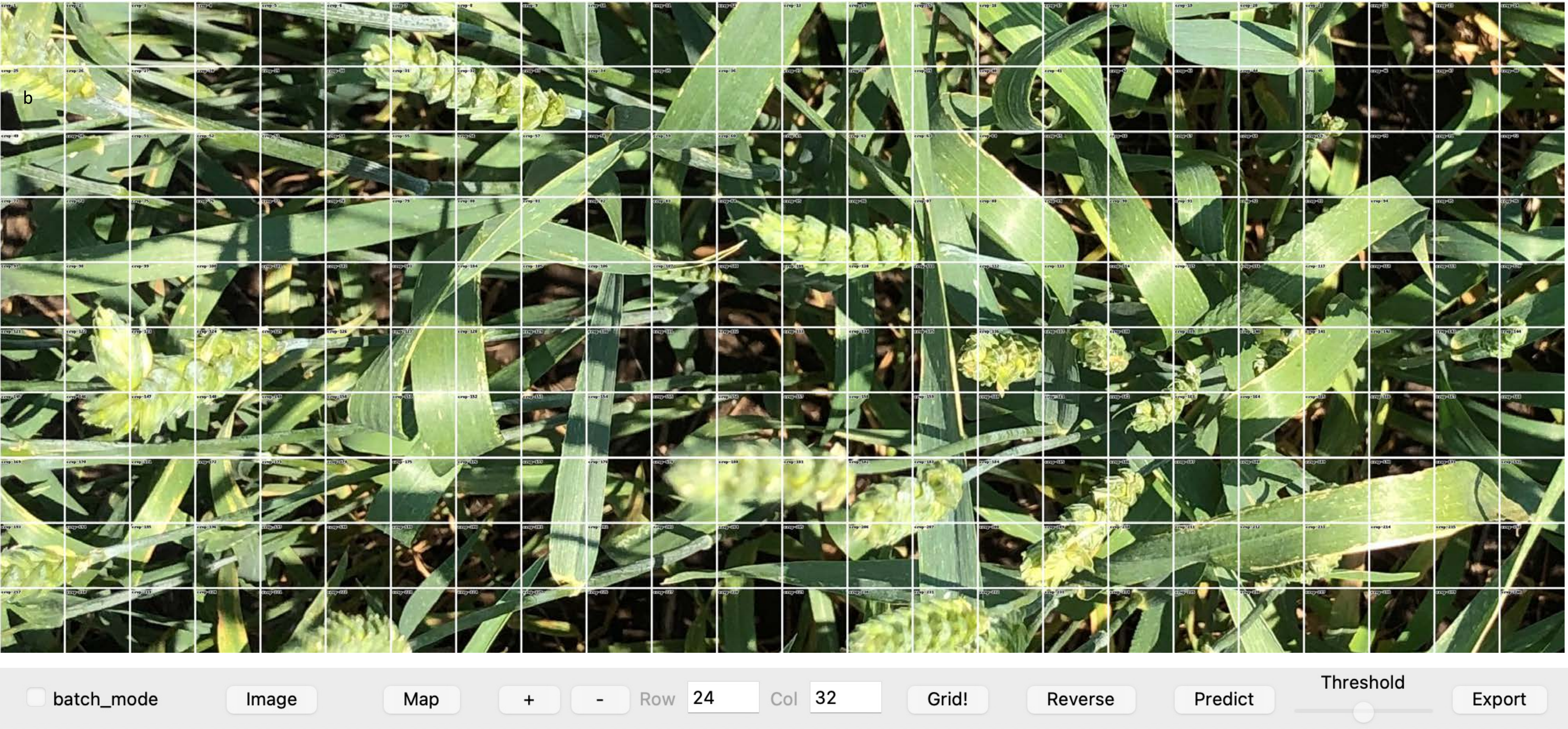
Zhou Tang

Customized ResNet18: **RustNet**

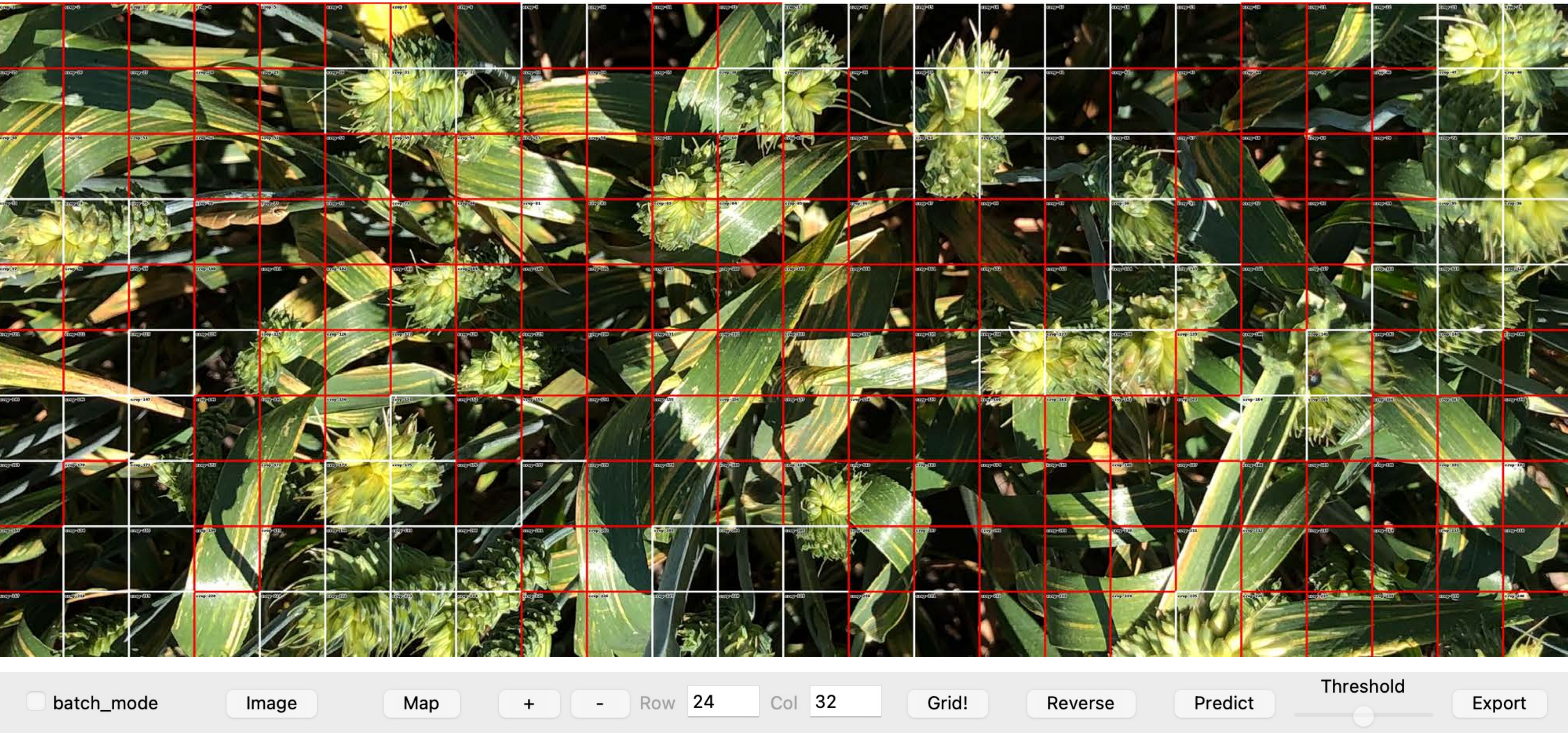
https://doi.org/10.1016/j.compag.2023.107709

Yang Hu

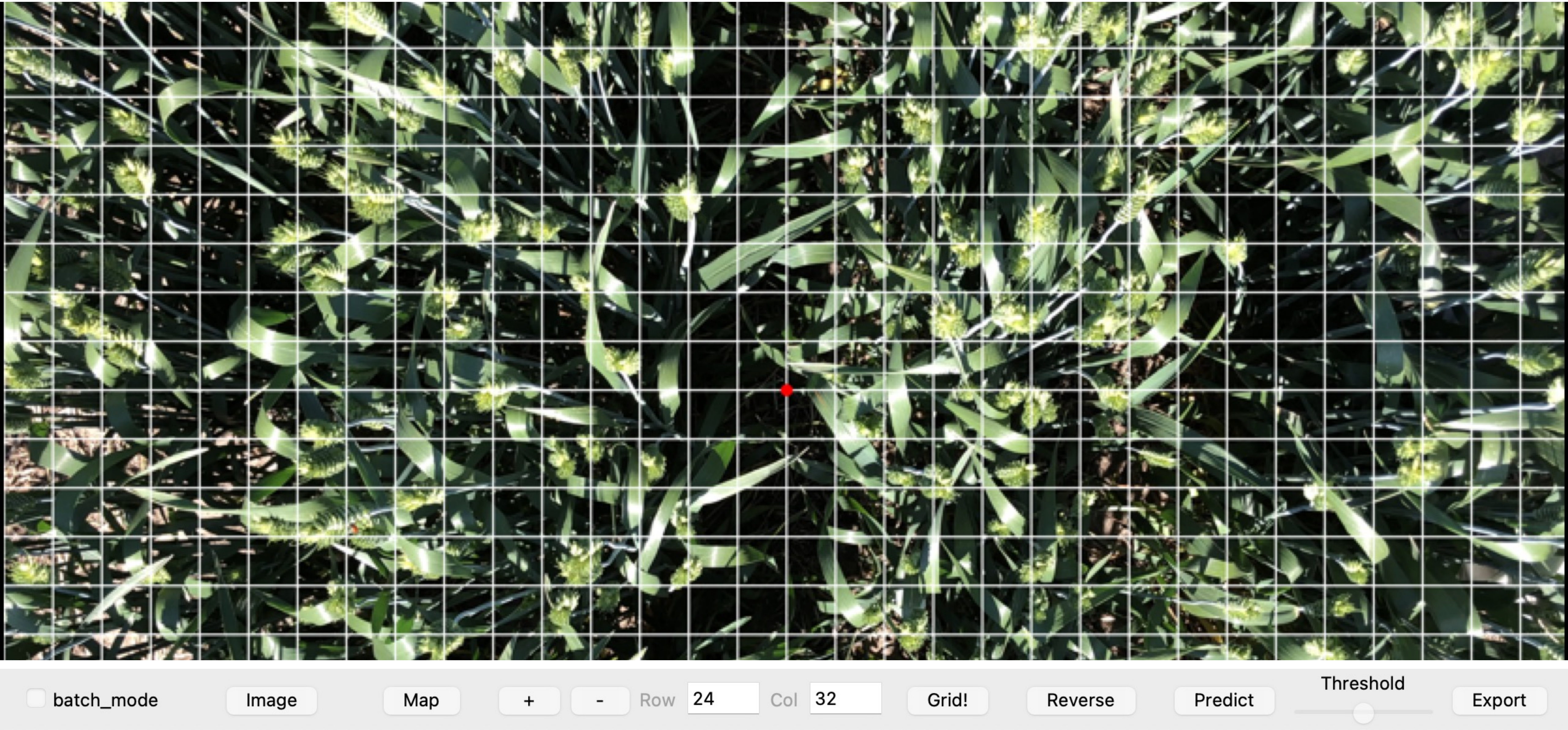
ROOSTER



Revers and customization



Prediction with RustNet

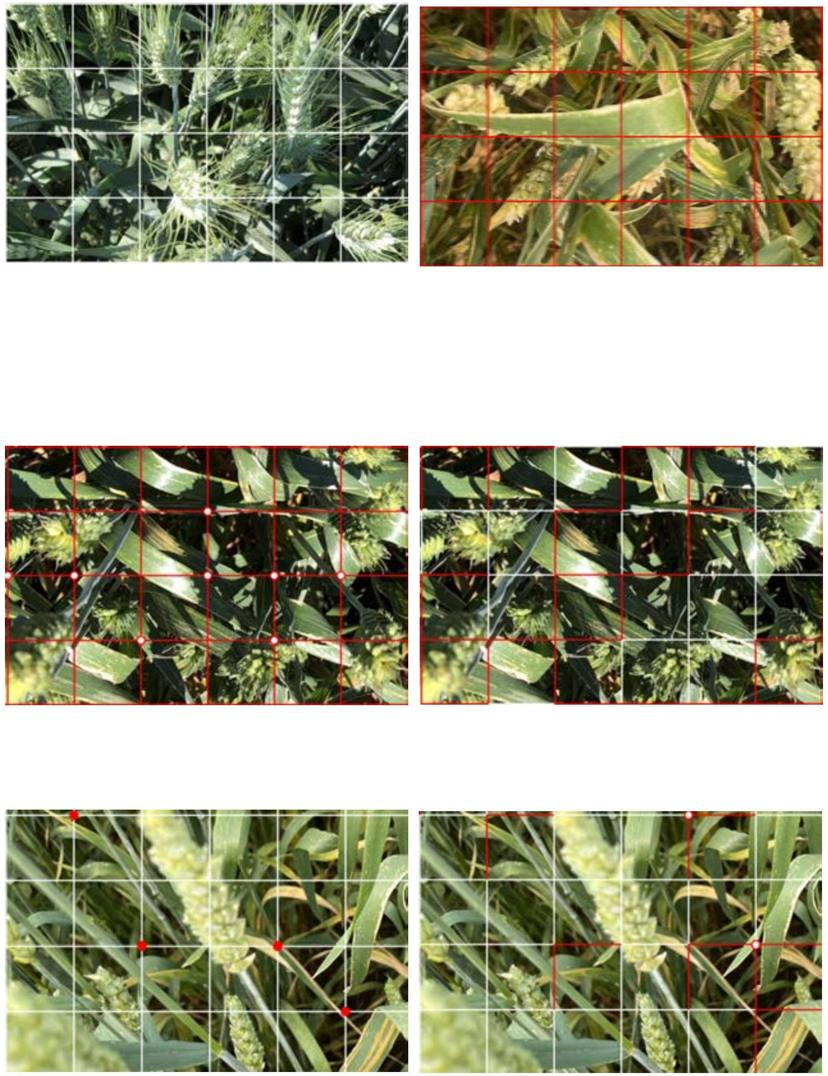


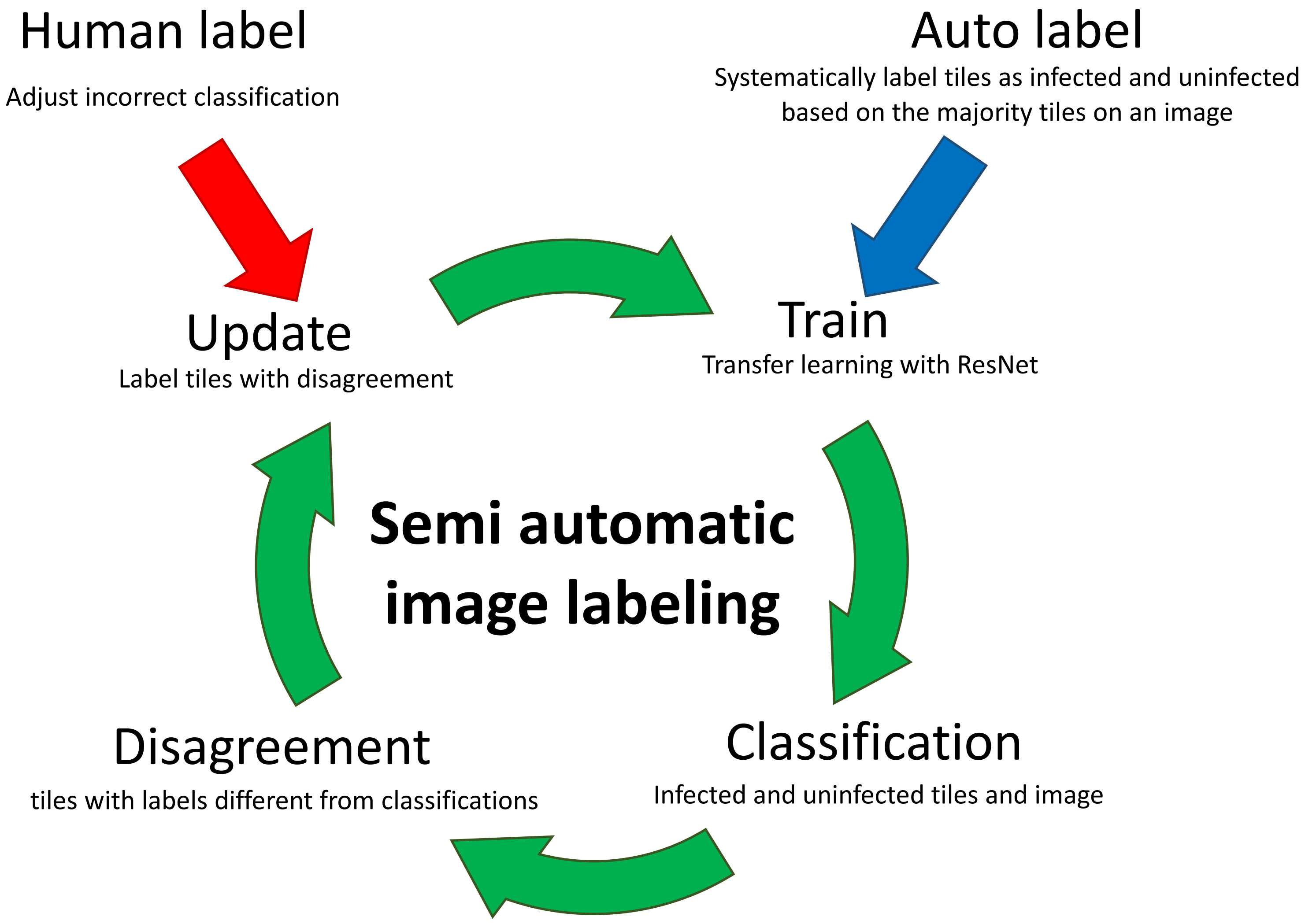
Online learning

Step 1

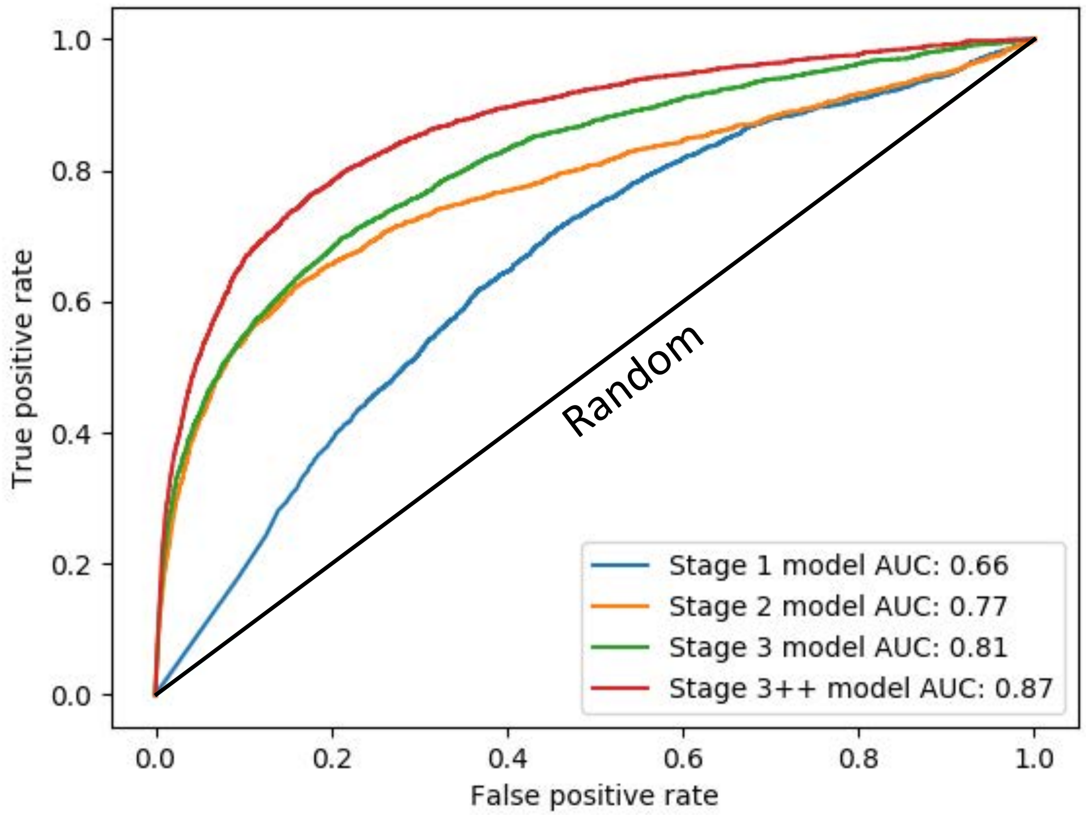
Step 2

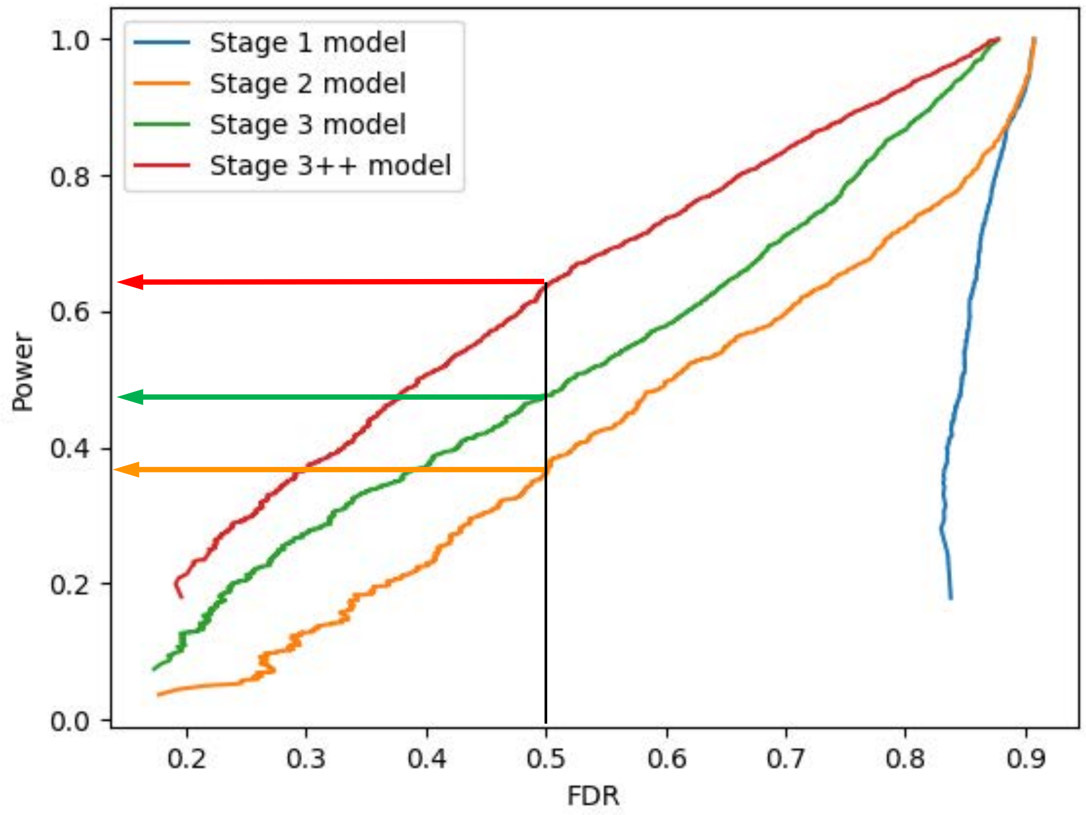
Step 3





Prediction accuracy





https://zzlab.net/Rooster

What does it mean for farmers?

Stage	FDR	False+	Power	Images	Time (m)
2	50%	10	35%	1,429	48
3	50%	10	50%	1,000	33
3++	**50%**	**10**	**65%**	**769**	**26**

Assumptions

- To find ten infected leaves on average
- Confidence to find at least one: 99.99%
- Taking images every two seconds
- Leaves per image: 200
- Frequency: 0.01%

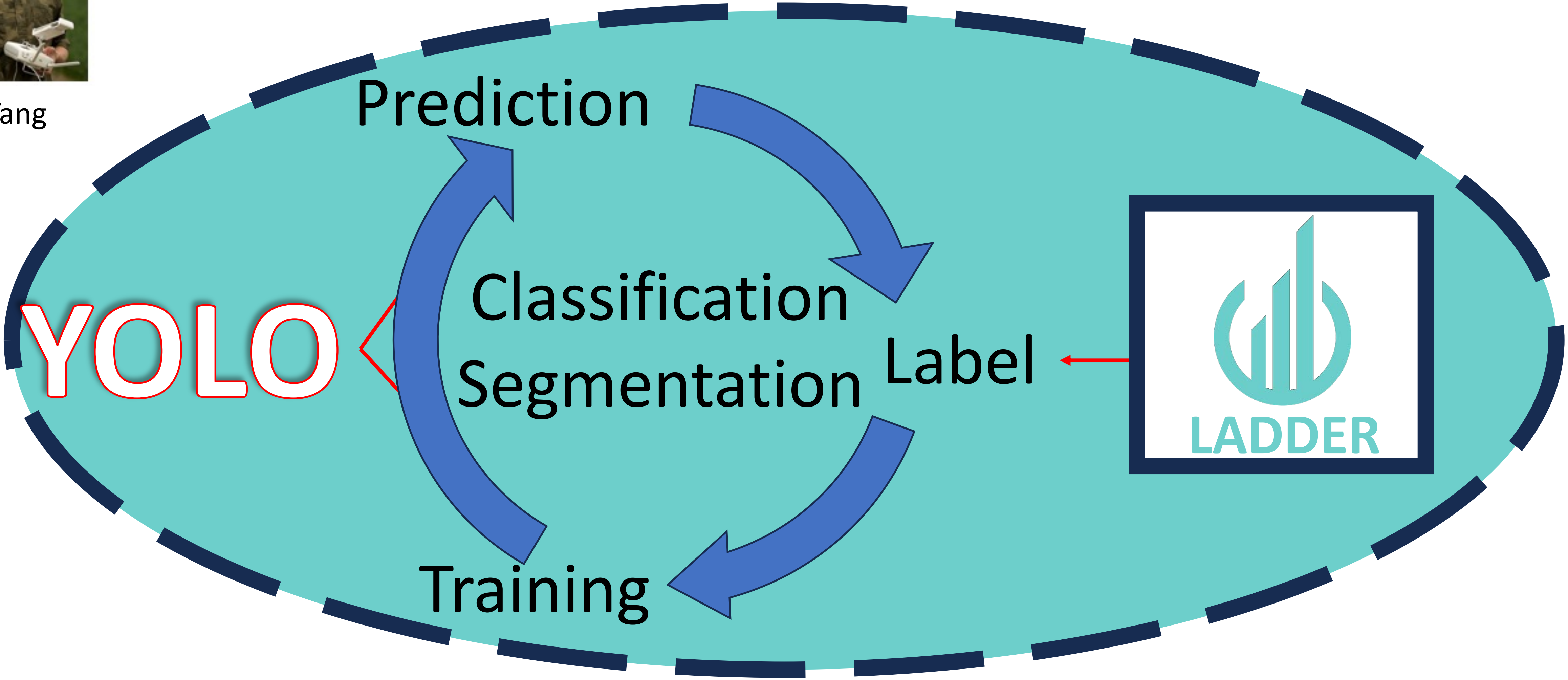
Zhiwu Zhang Laboratory
for Statistical Genomics

Home People Publication Research Teaching Software Outreach Jobs

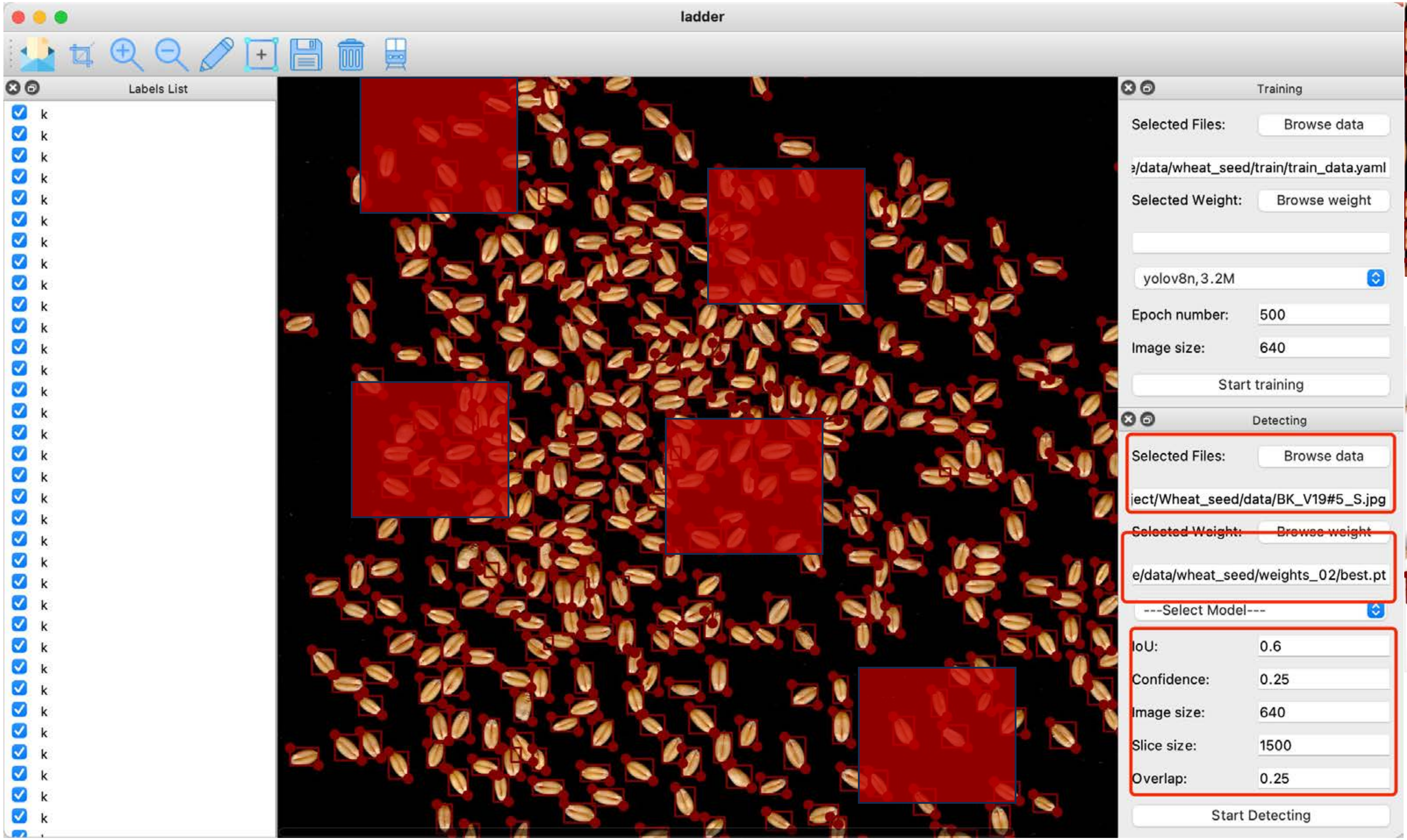
AI4
Everyone

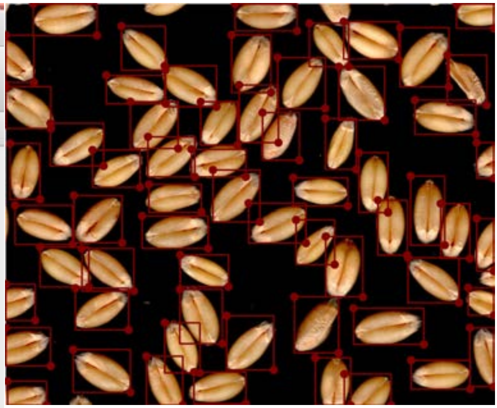

Harv 1
GSD230015
Rabbit holes!

Zhou Tang

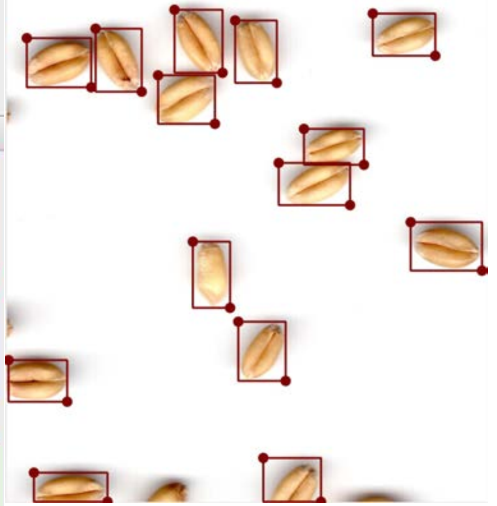

Prediction
Classification
Segmentation
Label
Training
YOLO
LADDER

Build a computer vision system within an hour





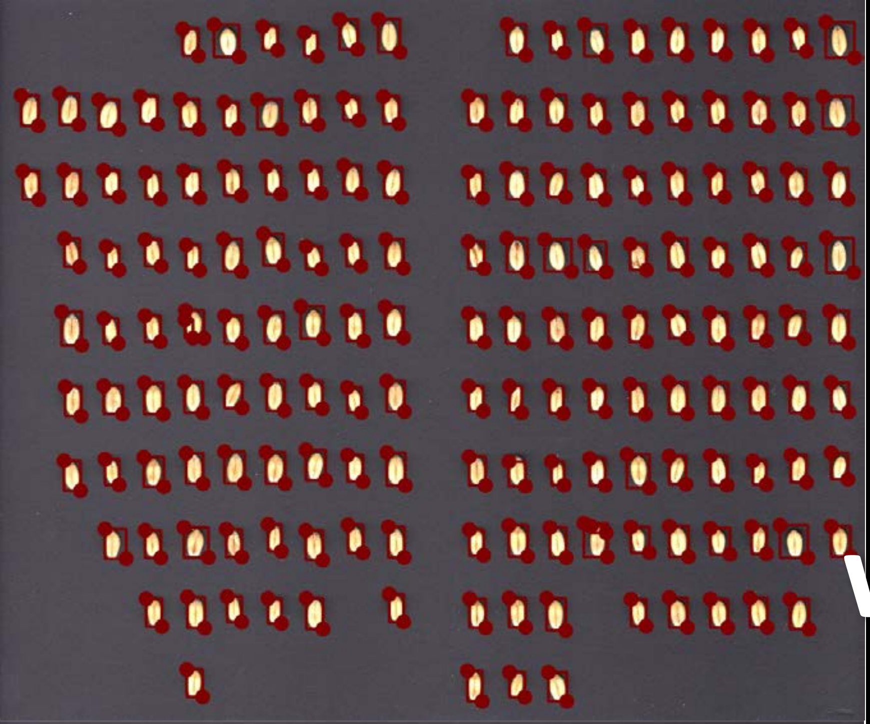
5 tile

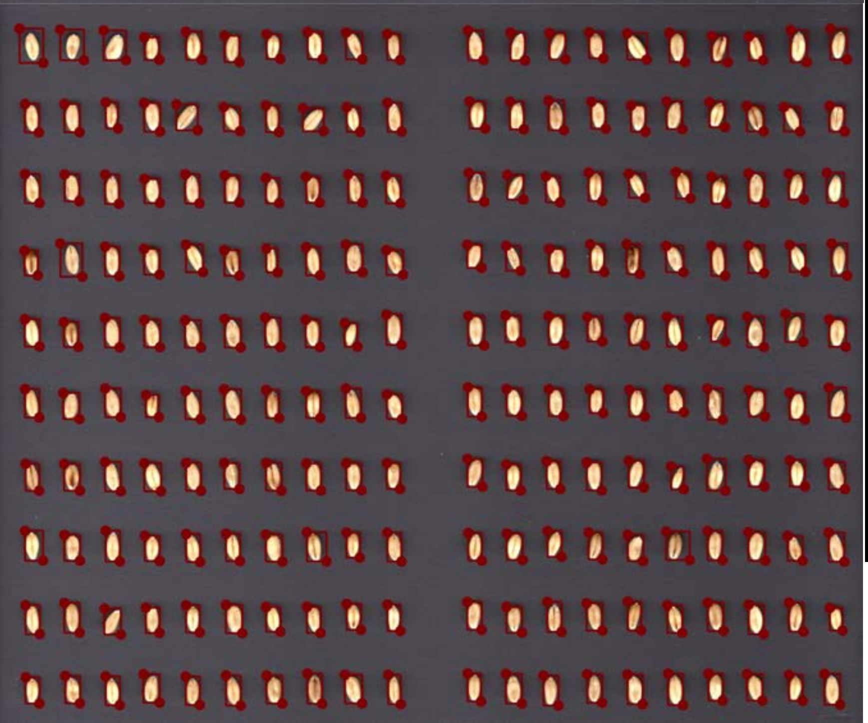


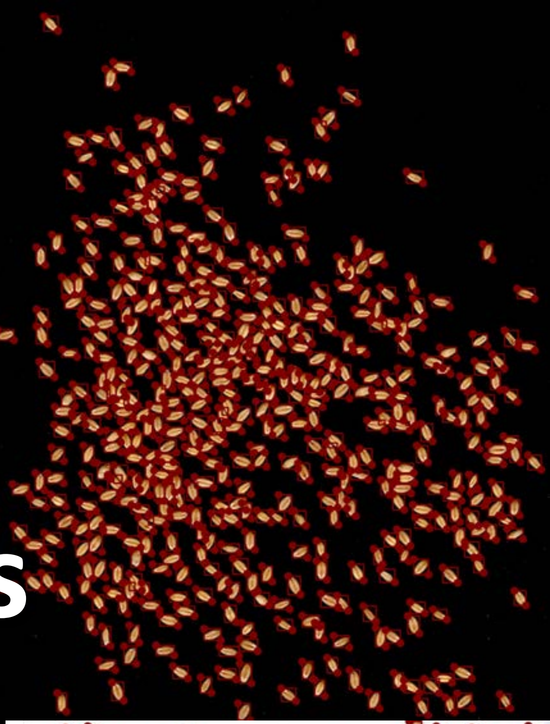
1 tile

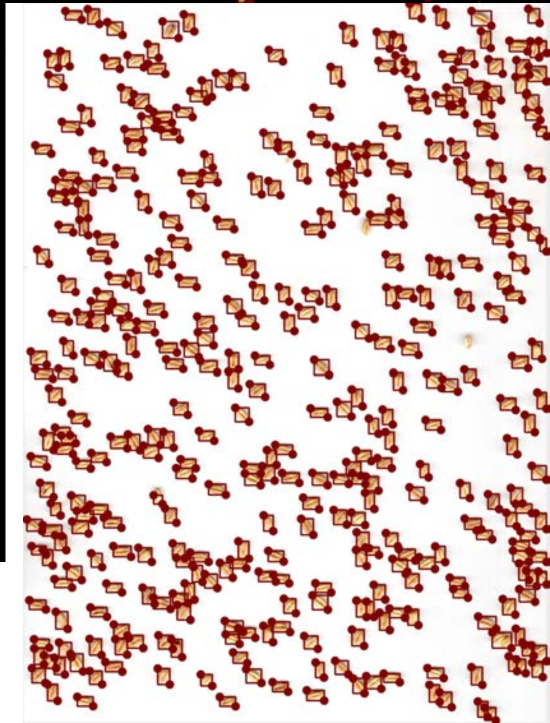
2 - 5 mins to train a tile

Works on wheat with different settings



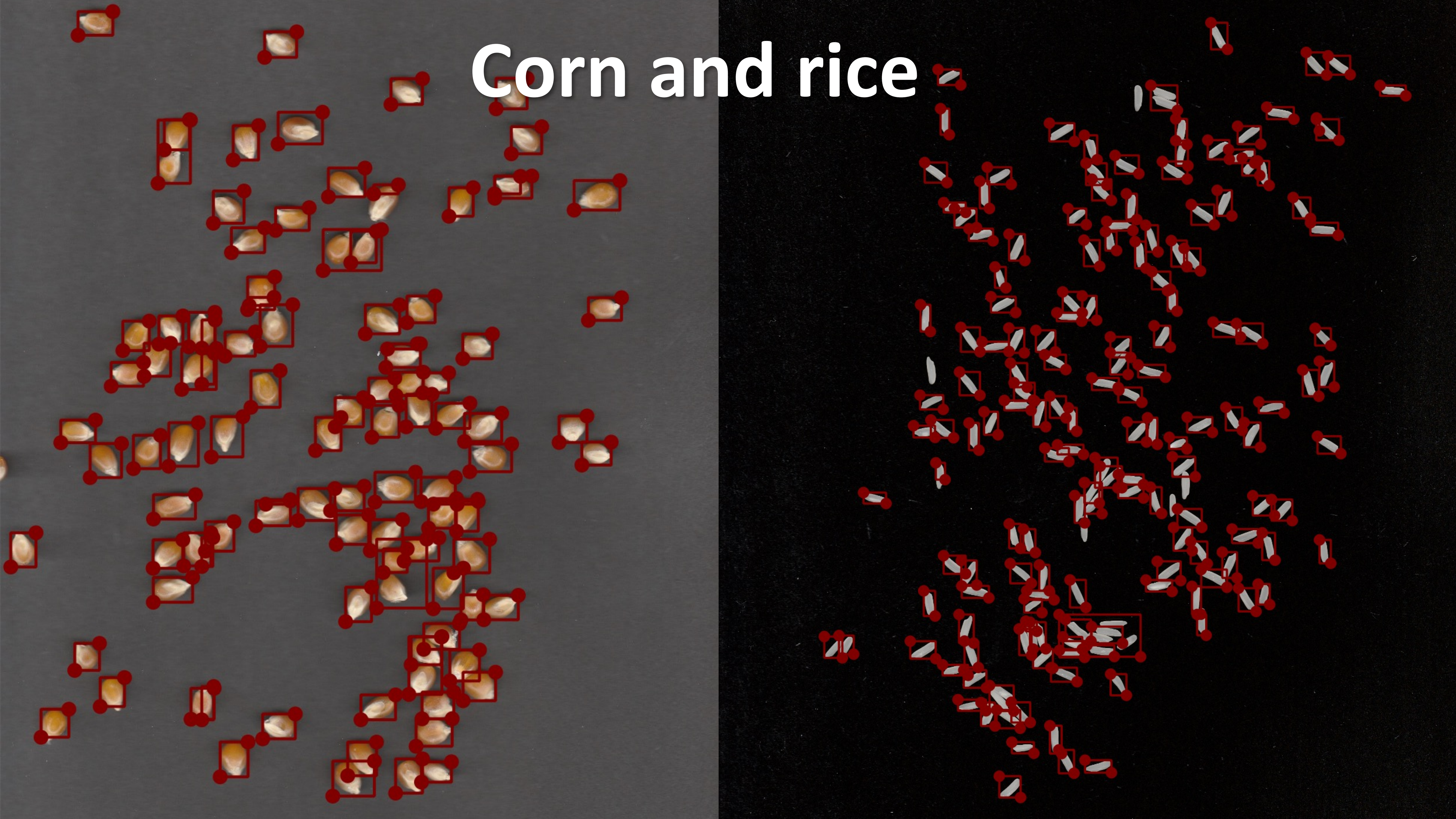




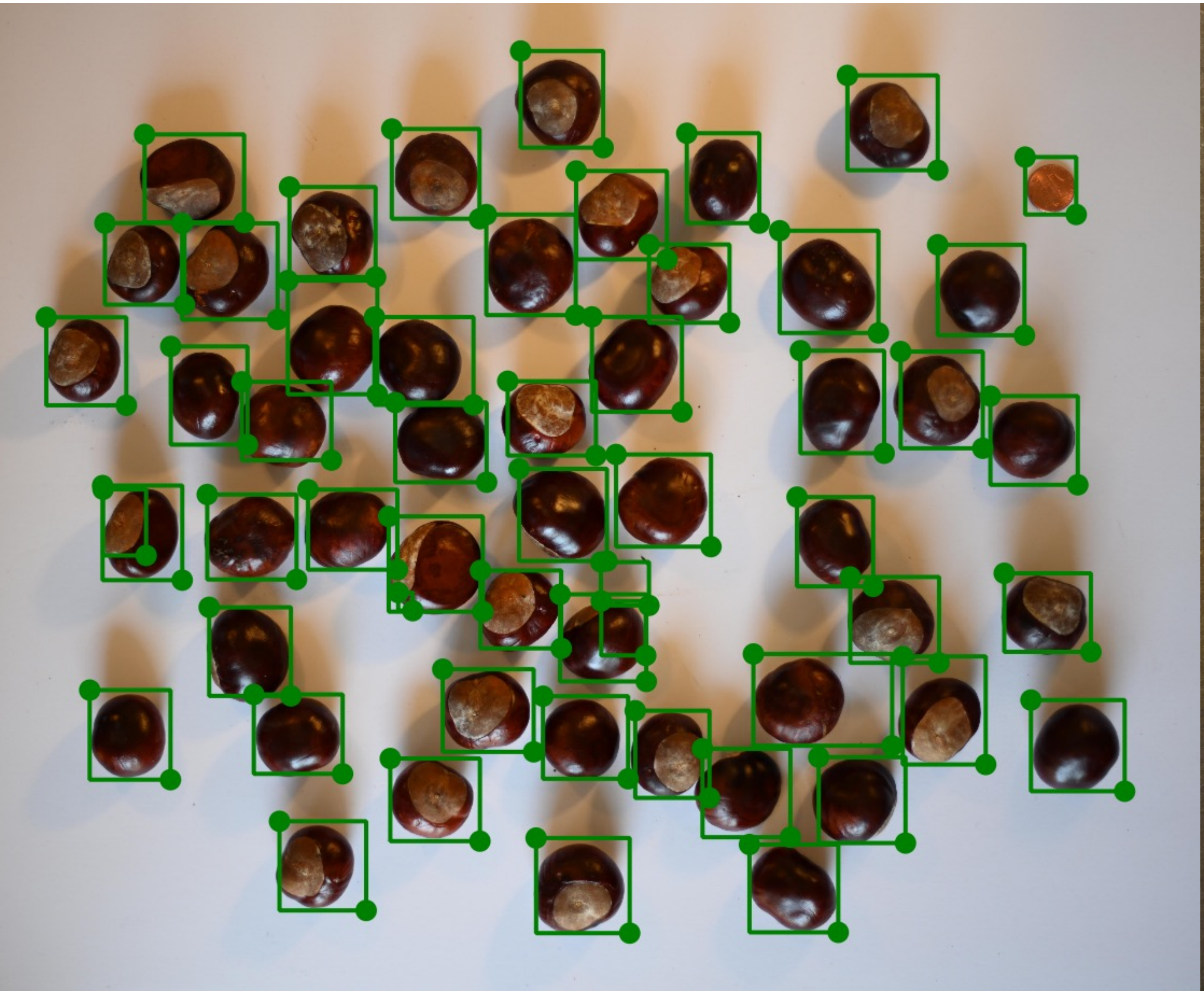


Each image take ~ 2 mins

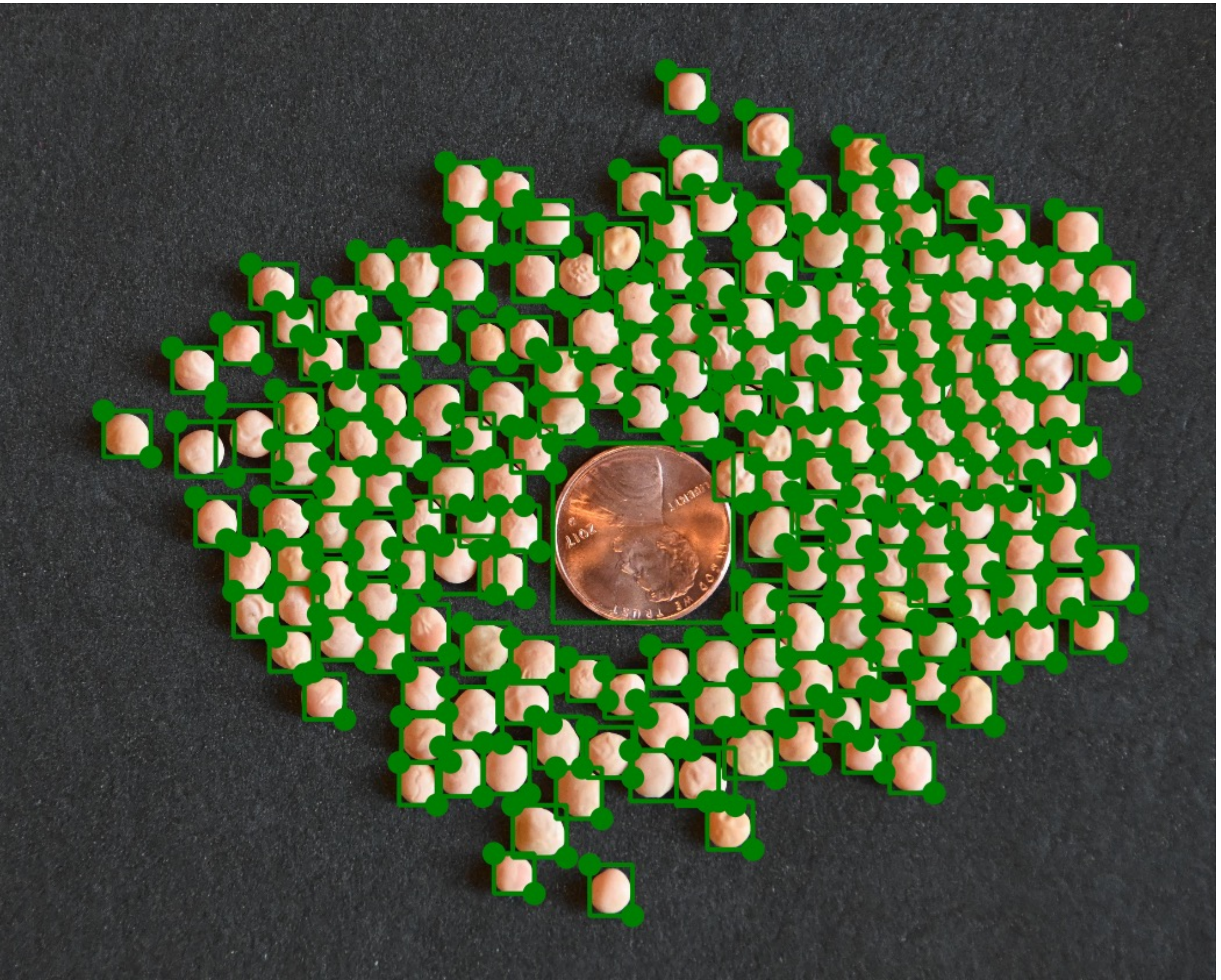
Corn and rice



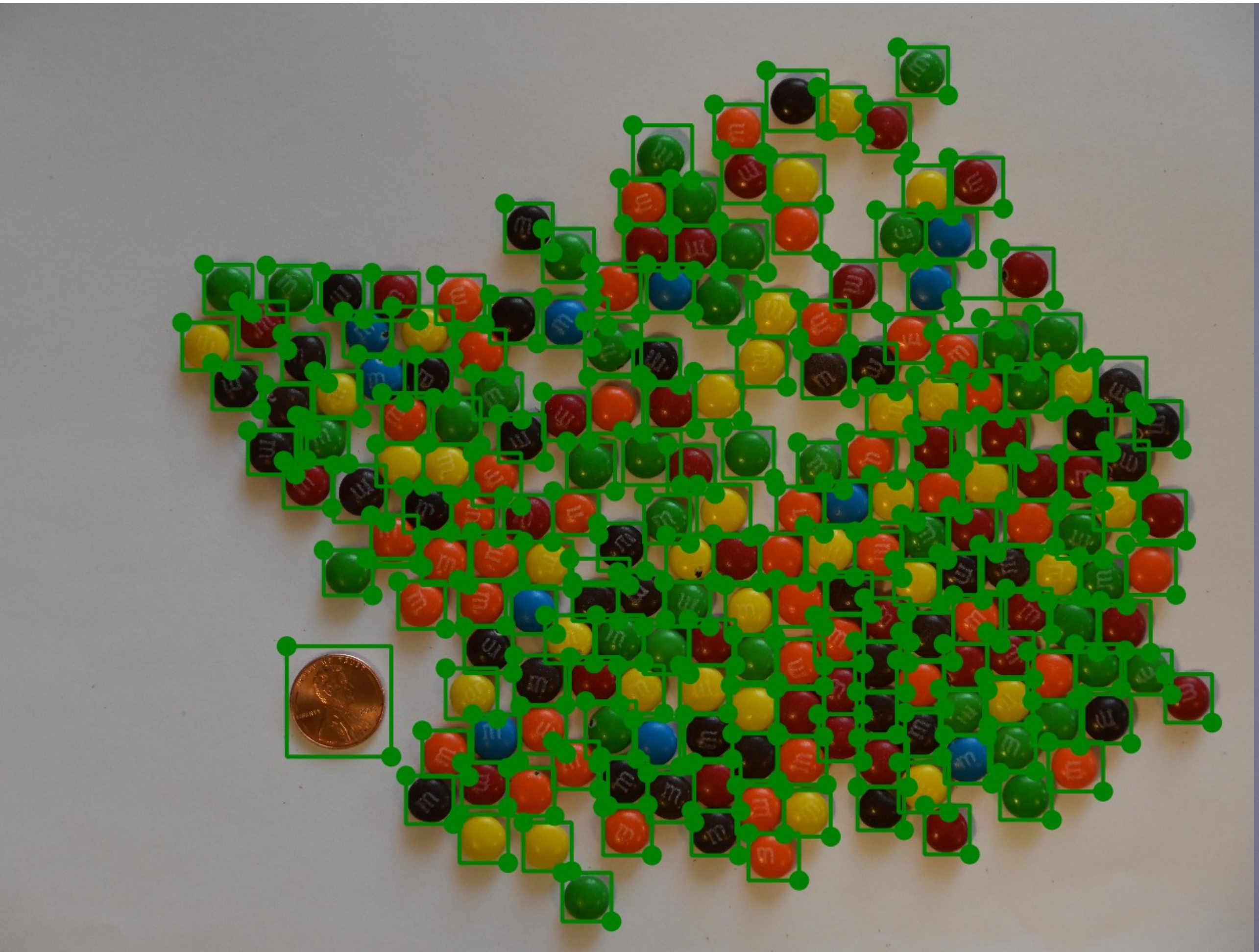
Chesnutt and chickpea

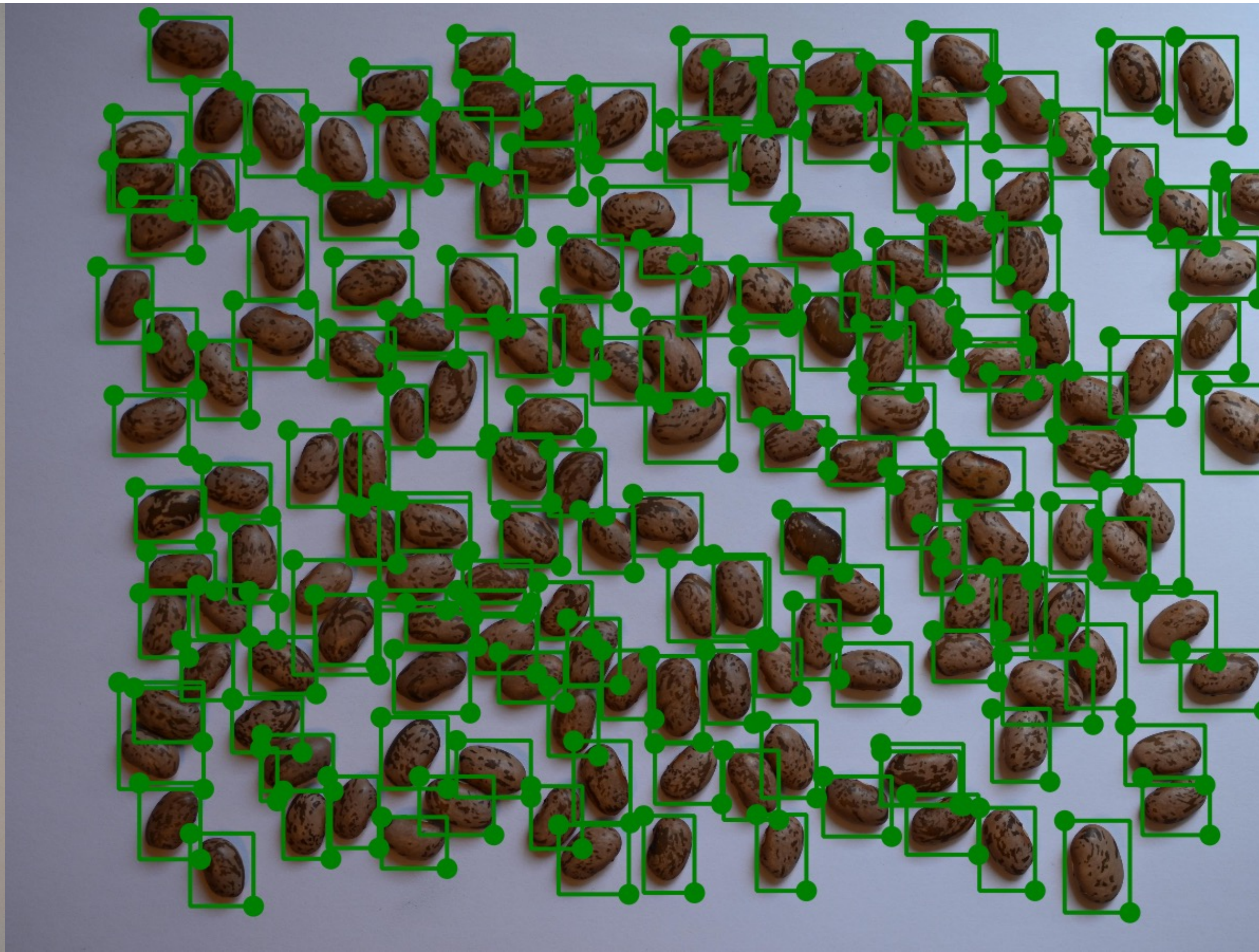


Lentil and Apricot

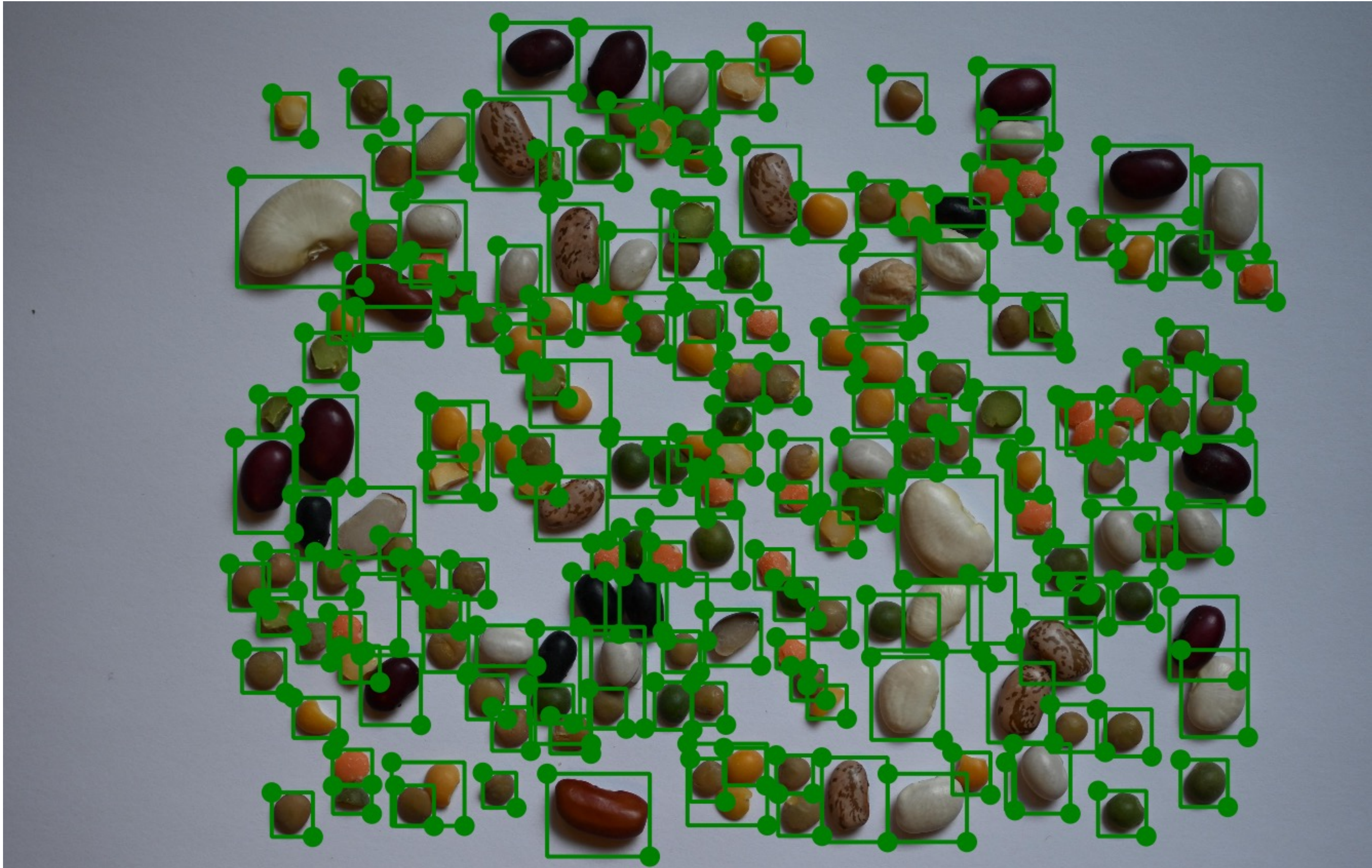


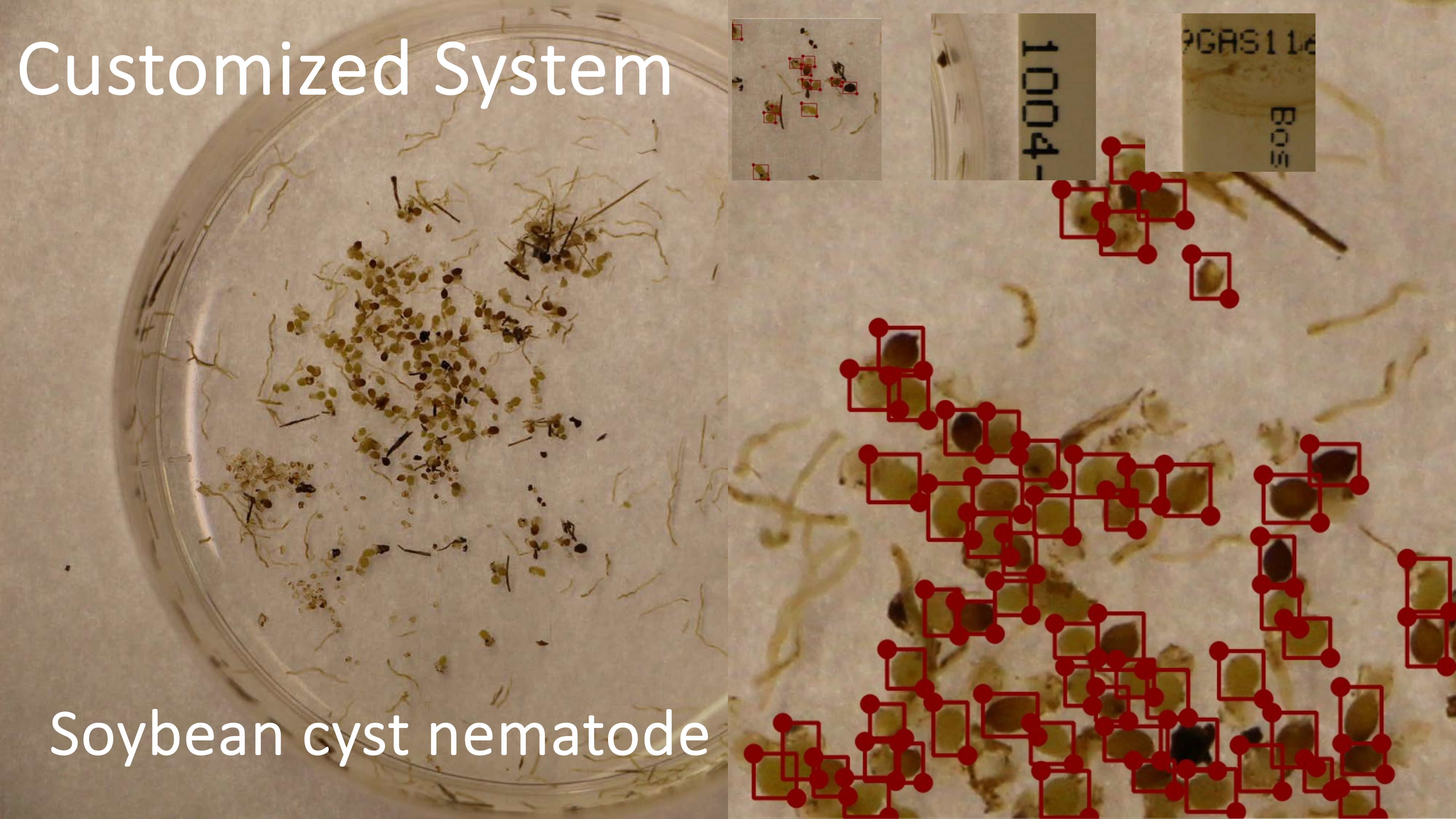
Cands and Beans

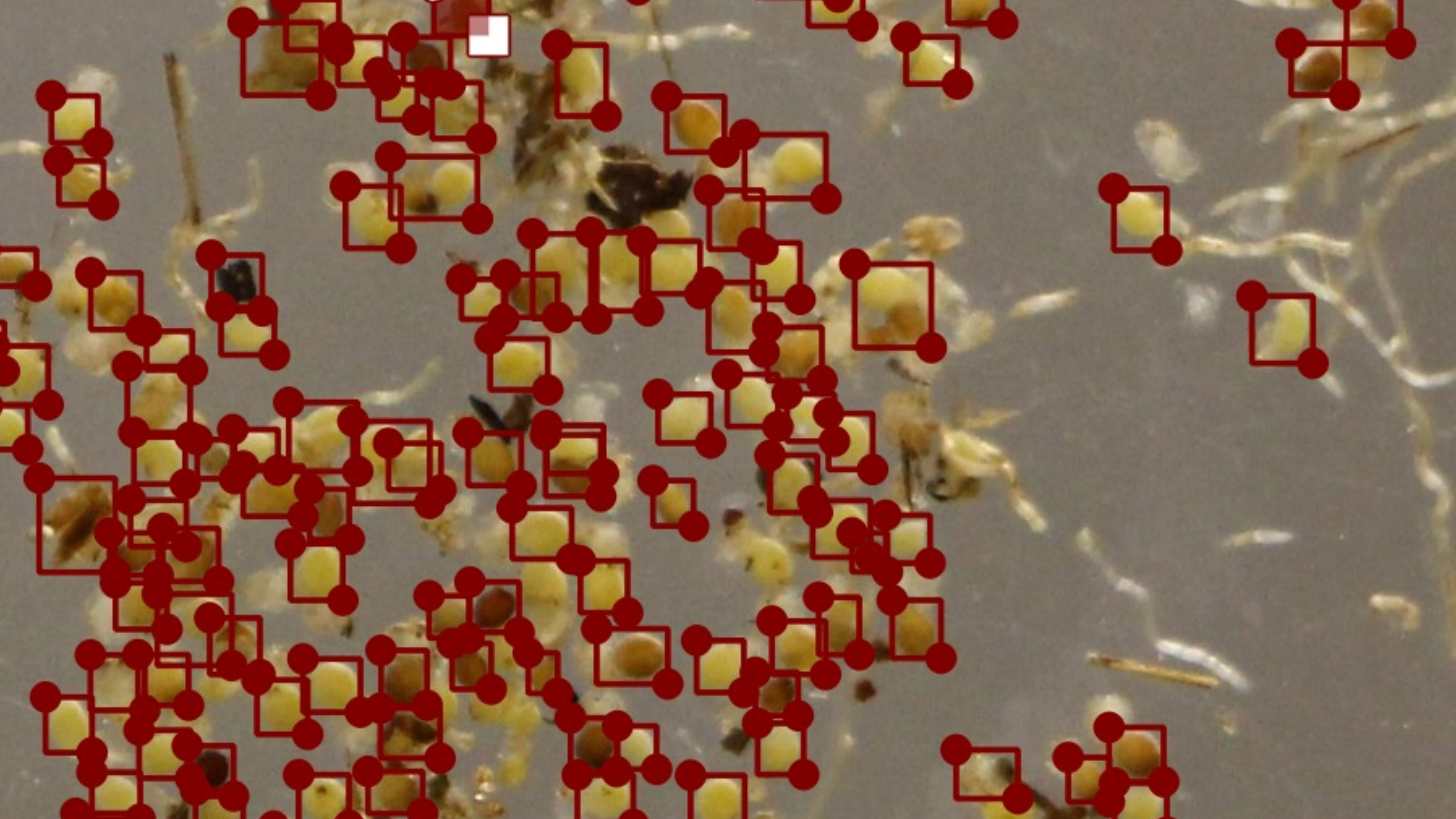




Mixed beans

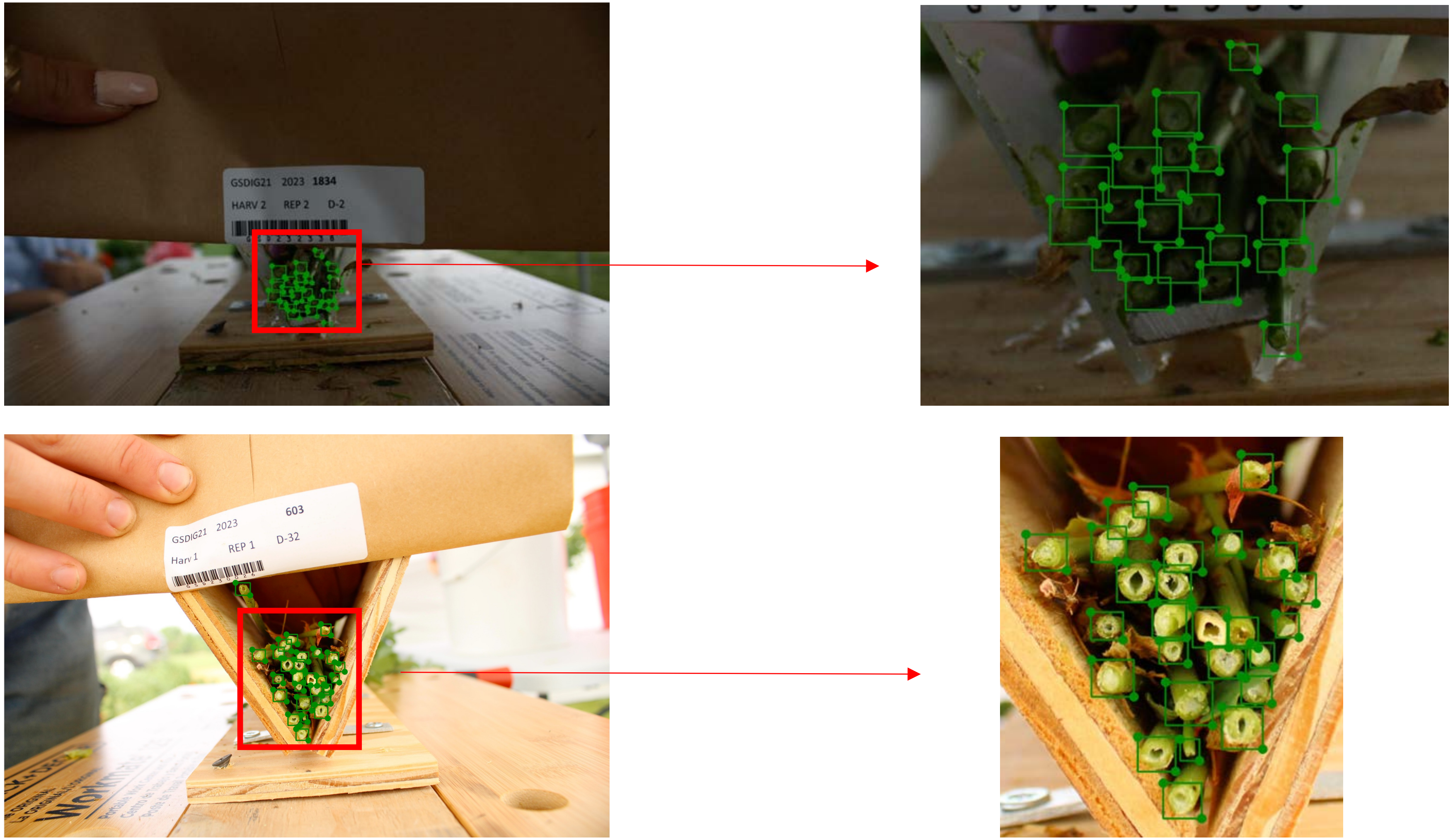



Customized System
Soybean cyst nematode



	Training set	Testing set
Alfalfa flower	3 (2 dark, 1 light)	2 (1 dark, 1 light)

	Training set	Testing set
Alfalfa Stem	6 (3 dark, 3 light)	4 (2 dark, 2 light)



Zhiwu Zhang Laboratory for Statistical Genomics

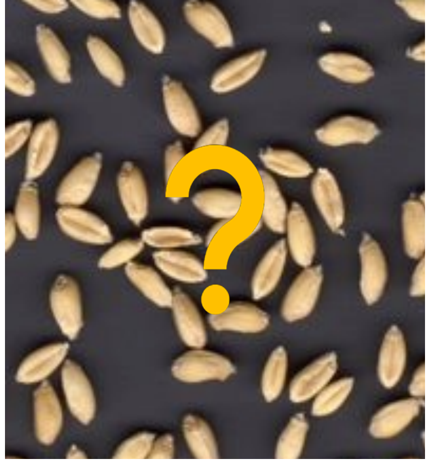
Home People Publication Research Teaching Software Outreach Jobs

AI4 Everyone

Plug & Play


python
TensorFlow
AI4
Convolutional Neural Network
Everyone

- Bran separates more cleanly from the endosperm for milling
- High flour yield
- Low water absorption
- Low gluten strength



AI4 Everyone

80%

Zhiwu Zhang Laboratory

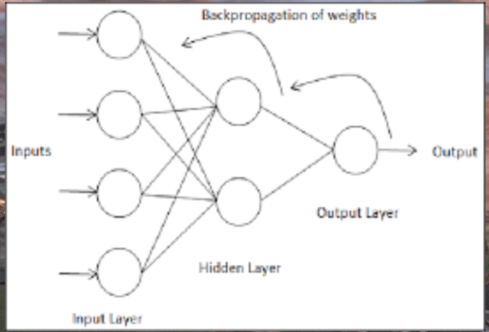
for Statistical Genomics

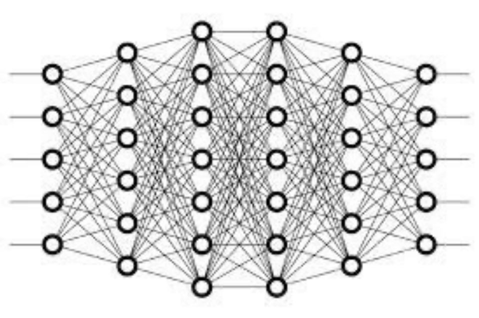
Build a path for everyone to succeed in AI

AI4 Everyone

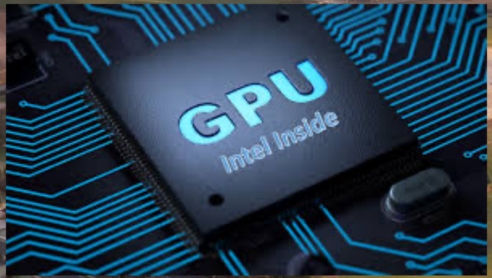
Plug&Play

python



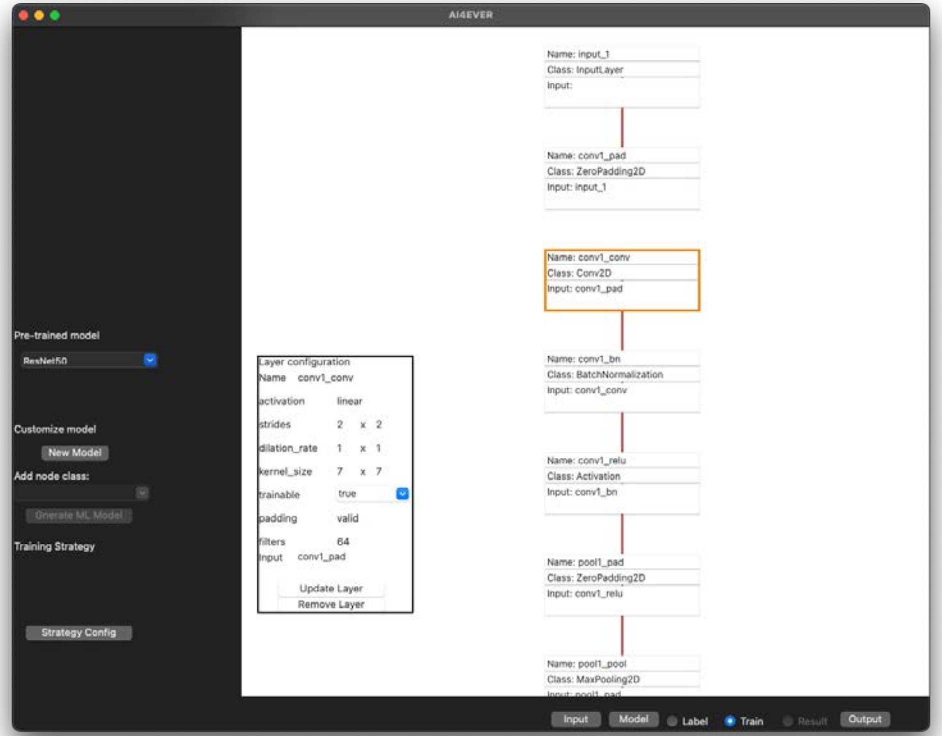


TensorFlow



Convolutional Neural Network

Models for Transfer Learning



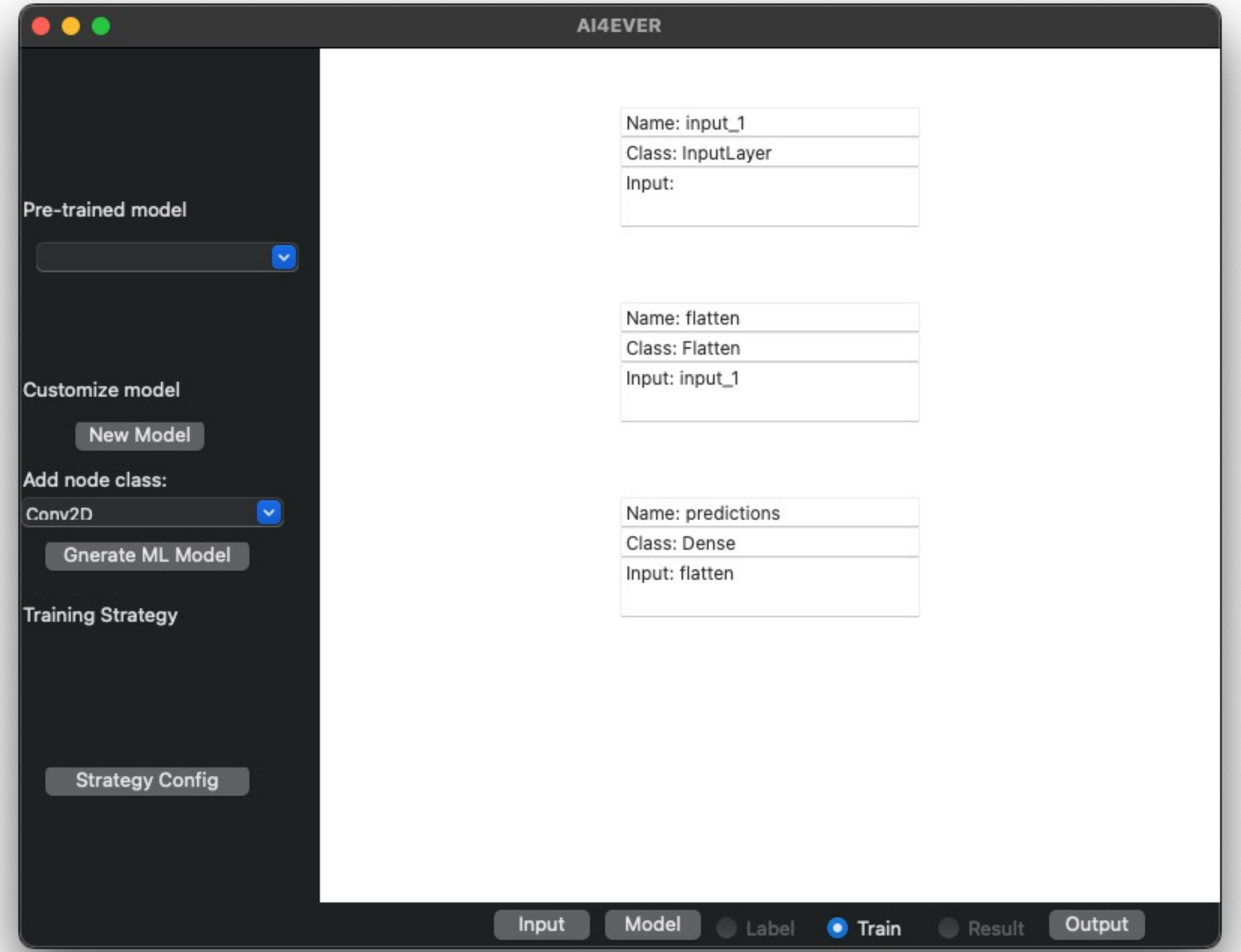
Click “Model” click “OK” do not choose any model
Load pre-trained model “ResNet50”
Neural Network Model editing:
Mouse click on layer component to display layer configuration
*Pre-trained model layers cannot be removed

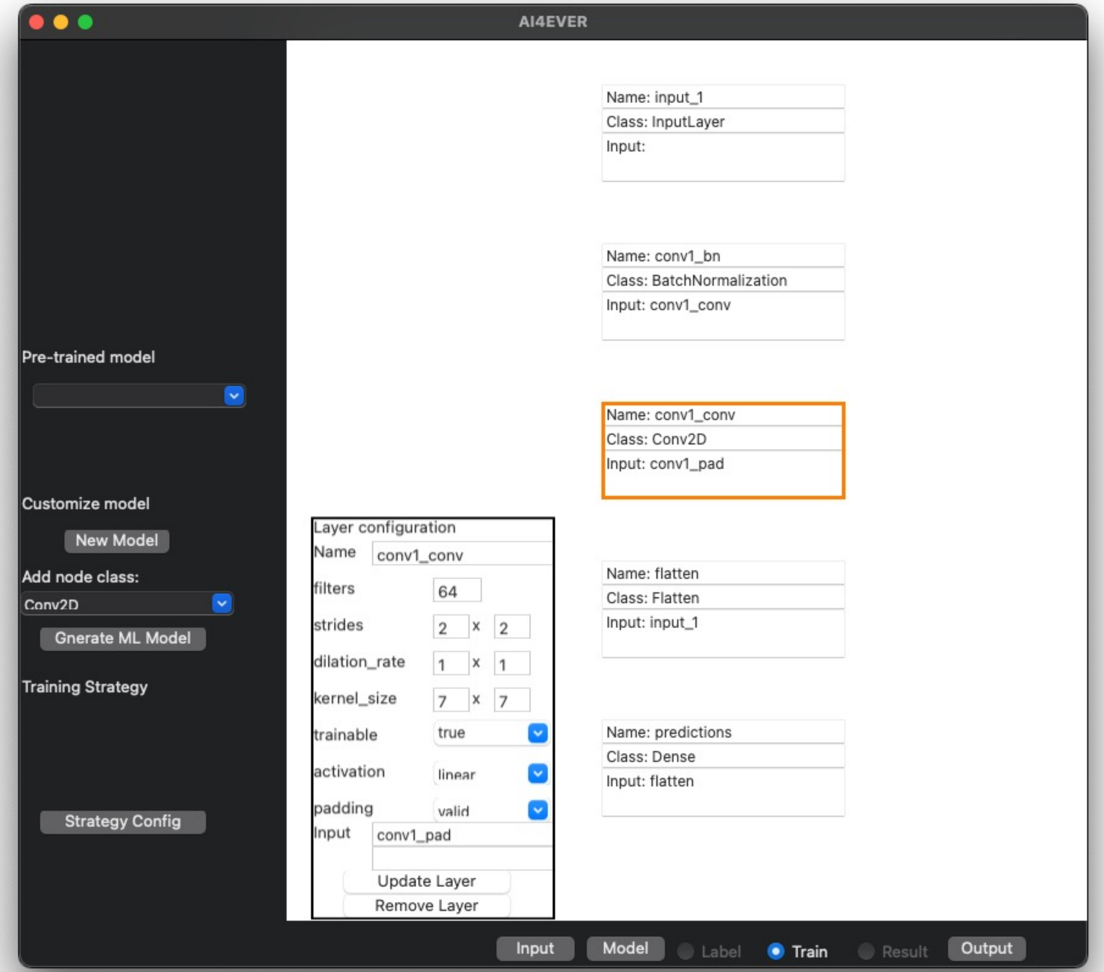
AI4EVER pre-trained (ImageNet) neural-network models

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
Xception	88	79.0%	94.5%	22.9M	81	109.4	8.1
VGG16	528	71.3%	90.1%	138.4M	16	69.5	4.2
VGG19	549	71.3%	90.0%	143.7M	19	84.8	4.4
ResNet50	98	74.9%	92.1%	25.6M	107	58.2	4.6
EfficientNetB0	29	77.1%	93.3%	5.3M	132	46.0	4.9
EfficientNetB1	31	79.1%	94.4%	7.9M	186	60.2	5.6
EfficientNetB2	36	80.1%	94.9%	9.2M	186	80.8	6.5
EfficientNetB3	48	81.6%	95.7%	12.3M	210	140.0	8.8
EfficientNetB4	75	82.9%	96.4%	19.5M	258	308.3	15.1
EfficientNetB5	118	83.6%	96.7%	30.6M	312	579.2	25.3
EfficientNetB6	166	84.0%	96.8%	43.3M	360	958.1	40.4
EfficientNetB7	256	84.3%	97.0%	66.7M	438	1578.9	61.6

* Pre-trained models are from Keras and TensorFlow

Neural Network Model Architecture



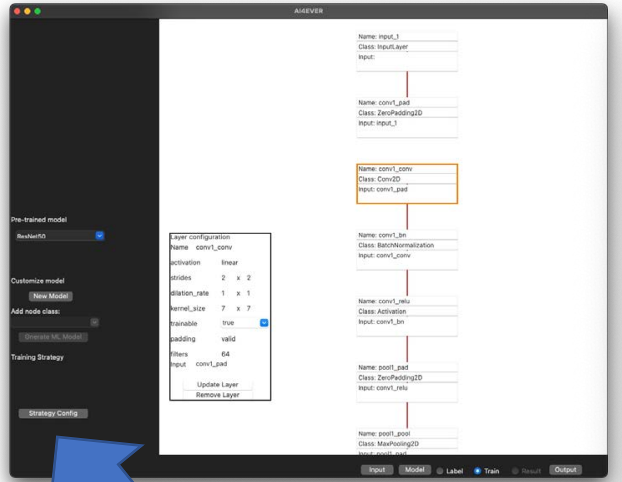


Click “New Model”

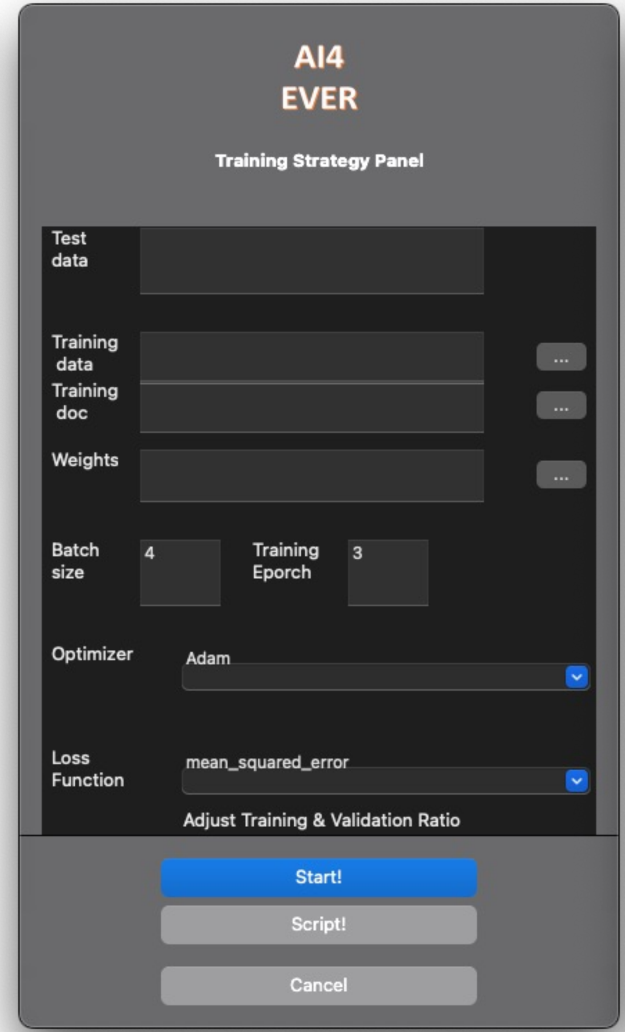
- Default three layers: input, flatten, dense(output)

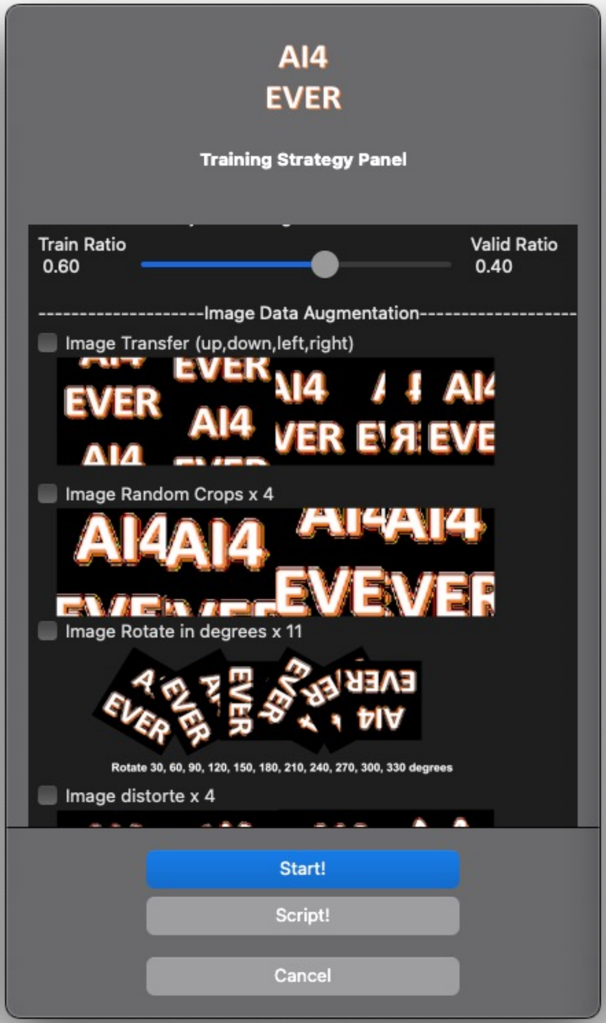
Select neural-network layers from “Add node class”
Click on layer component to trigger layer editor
*Drag layer component to change their positions
* Need to type layer names for input and output

Training



Button to trigger training strategy





- Start: To train at the background
- Script: To generate command line script and all files to run on server or other computer

Collaborators and funding


Arron Carter


Mike Pumphrey


Karen Sanguinet


Kawamu Tanaka


Sindhuja Sankaran


Longxi Yu


Jack Brown


Ananth Kalyanaraman


Kim Campbell

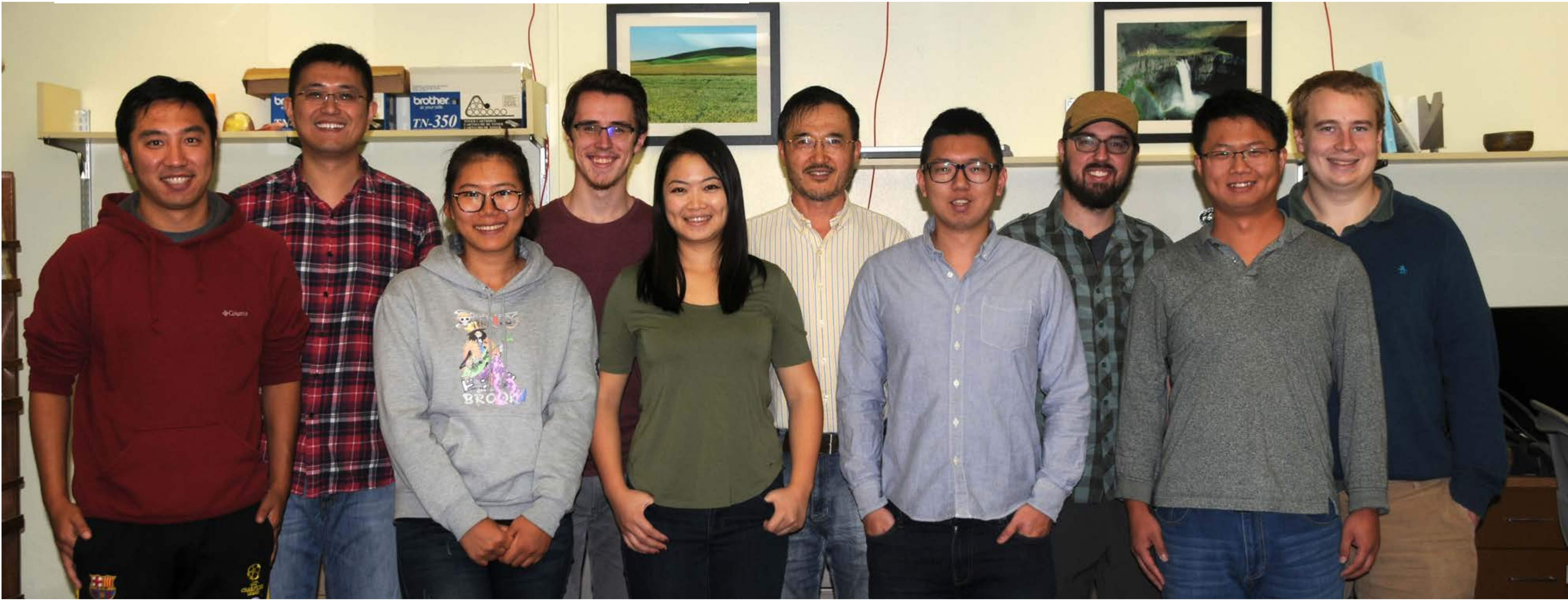

Deven See


Camille Steber


Mike Peel

Zhiwu Zhang Laboratory
for Statistical Genomics

WASHINGTON STATE UNIVERSITY



WASHINGTON STATE UNIVERSITY
Department of Crop and Soil Sciences

张志武教授在中国定向招收博士研究生数名

学习地点： 美国华盛顿州立大学，Pullman, WA, USA
招生对象： 国家留学基金委(CSC)奖学金(四年)获得者
资助内容： 提供学费、医疗保险，年度生活费补贴五千美金、GRE 豁免与协助 CSC 奖学金申请
资助来源： 美国农业部、华盛顿州谷物协会和华盛顿州立大学研究生院
研究方向： 图像处理或核酸数据分析
偏重技能： AI, GIS, Fixed and Random Effect Mixed Model, Bayesian Analysis 与计算编程
申报条件： 英语 TOEFL 80 分或 IELTS 7 分以上(英语成绩和申请推荐信可后补)
递交申请： http://css.wsu.edu/graduate-studies：Financial aid 填写"Pending CSC Application"，CSC 申请与咨询致信(Zhiwu.Zhang@WSU.EDU)或微信(zhiwu-zhang)
导师信息： Zhiwu Zhang Laboratory (http://zzlab.net)

PO Box 646420, Pullman, WA 99164-6420
509-335-3475 • Fax: 509-335-8674 • http://css.wsu.edu