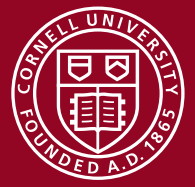
Cornell AgriTech

New York State Agricultural Experiment Station



Cornell CALS

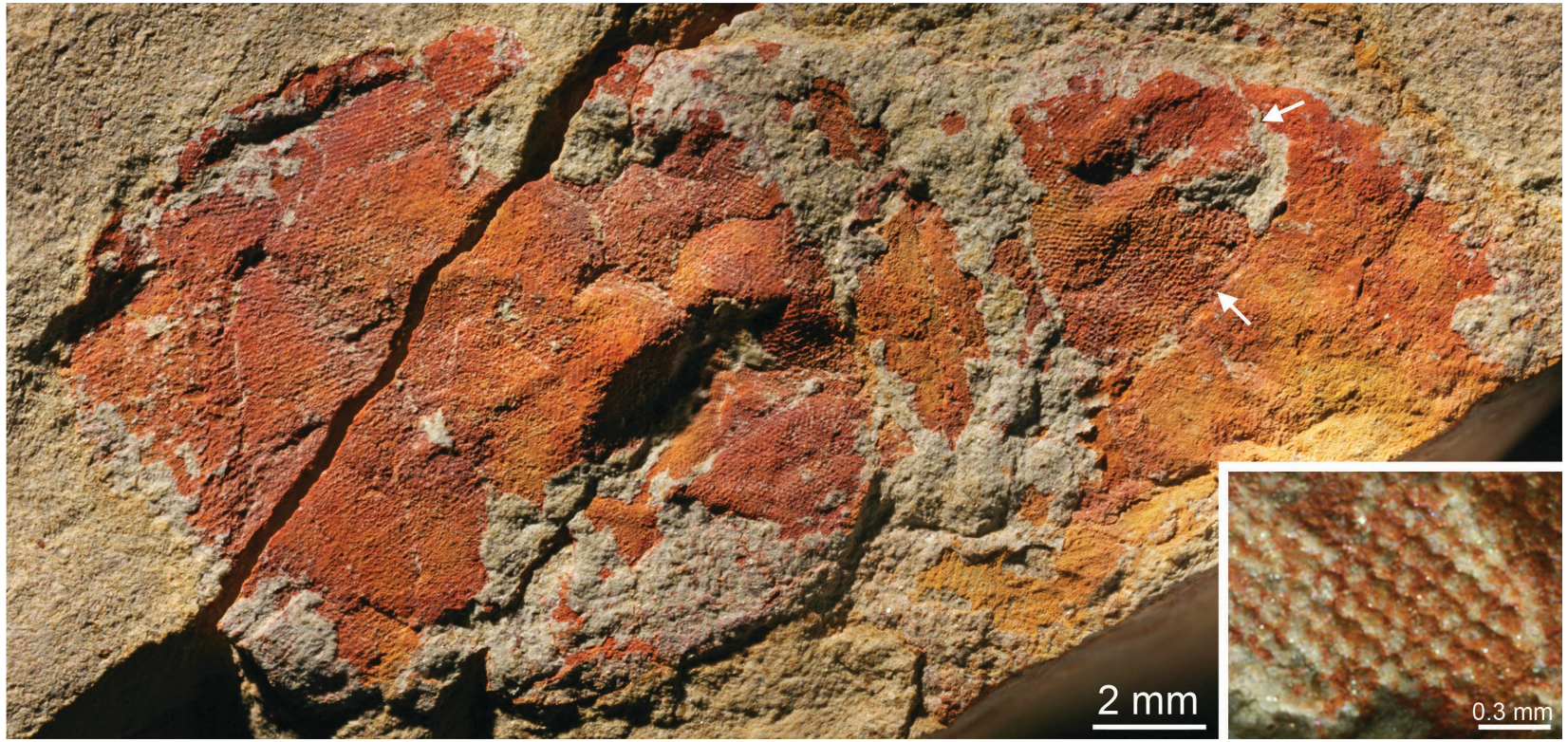
College of Agriculture
and Life Sciences

Computer Vision for Phenotyping: A Hands-on Workshop

***Yiyuan Lin* 林翊源**
Center for Agricultural Intelligence and Robotics (CAIR)
School of Electrical and Computer Engineering
Cornell University

Oct 30, 2026
VG3 and VxM Annual Meeting

What changes when an organism can see?



515 million years ago

Anomalocaris
A compound eye with ≥16,000 lenses

Cambrian seas

Fossil evidence

Seeing may have changed the rules of survival.

Vision as a driver of diversification: a hypothesis. Photo: UNE Photos · CC BY 2.0

What could we learn if we could measure every image?



01 Is the target present?

02 Where is it?

03 How much is there?

Image ➡ **Prediction** ➡ **Phenotype** ➡ **Biological question**

Computer vision: extracting structured information from images.

Phenotypes

Image



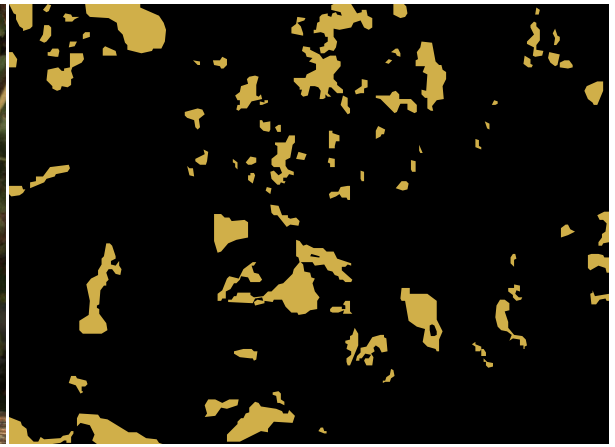
Mask - Canopy (C)



Image



Mask - Downy Mildew (D)



Image



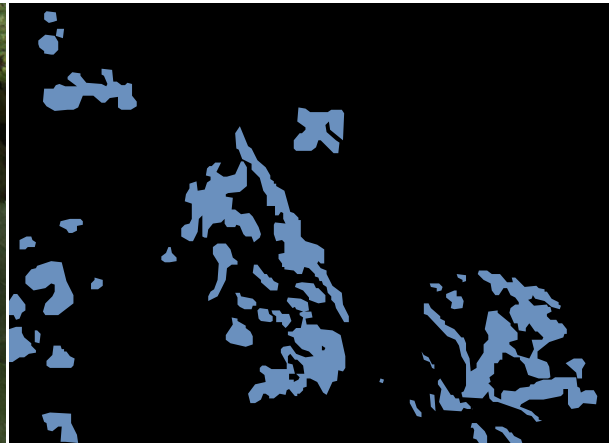
Mask - Berry (B)



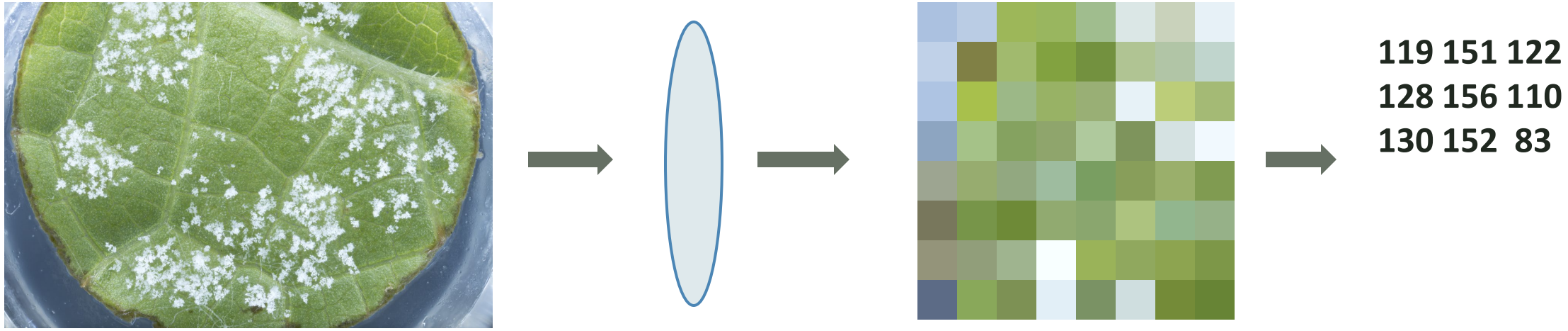
Image



Mask - Powdery Mildew (P)



A digital image begins with light

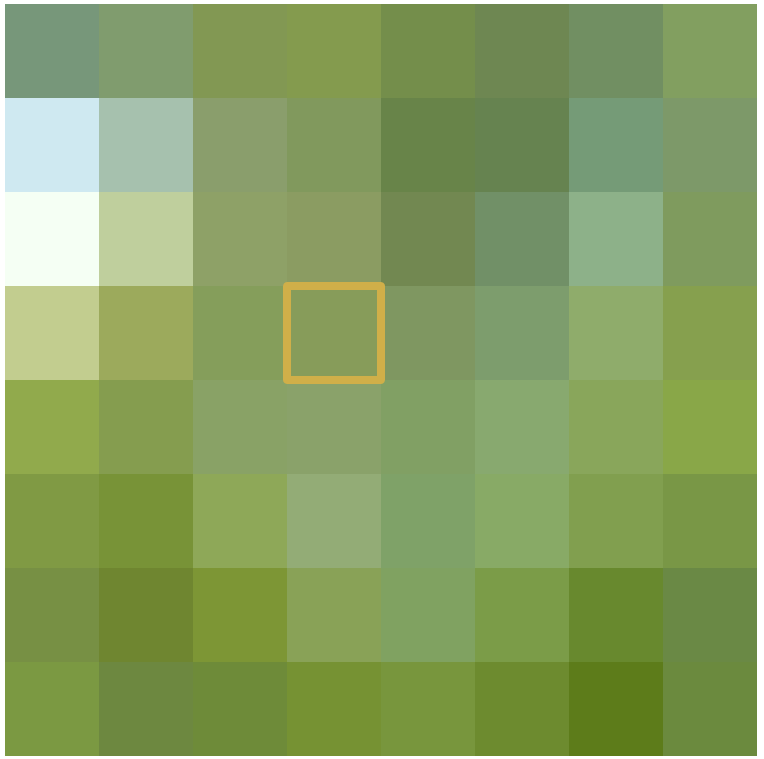


Specimen **Lens** **Sensor** **Numbers**

Light → electrical signal → sampled, digitized pixel values

Zoom in far enough, and an image is a grid





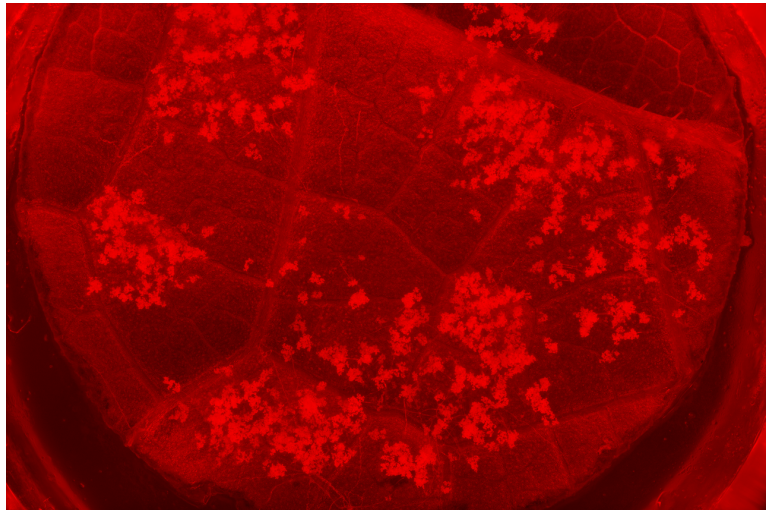
4,127 columns × 2,751 rows

8 × 8 real pixels

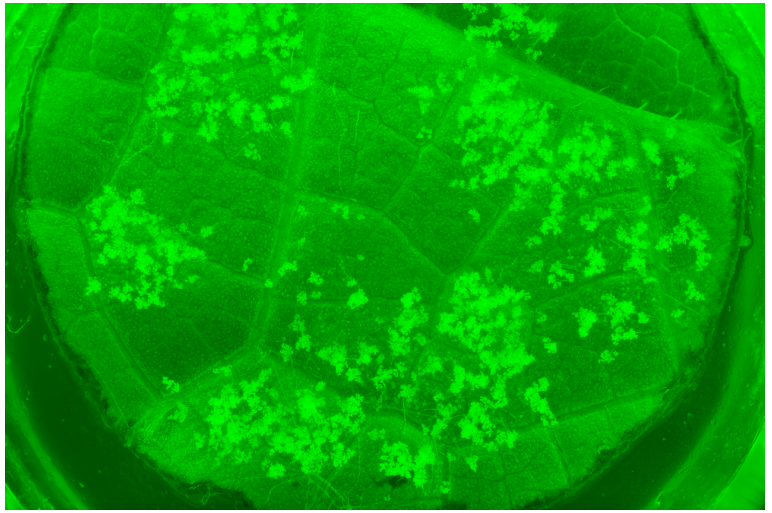
A pixel stores numbers. Its location is described by a row and a column.

RGB channels combine to form a color image

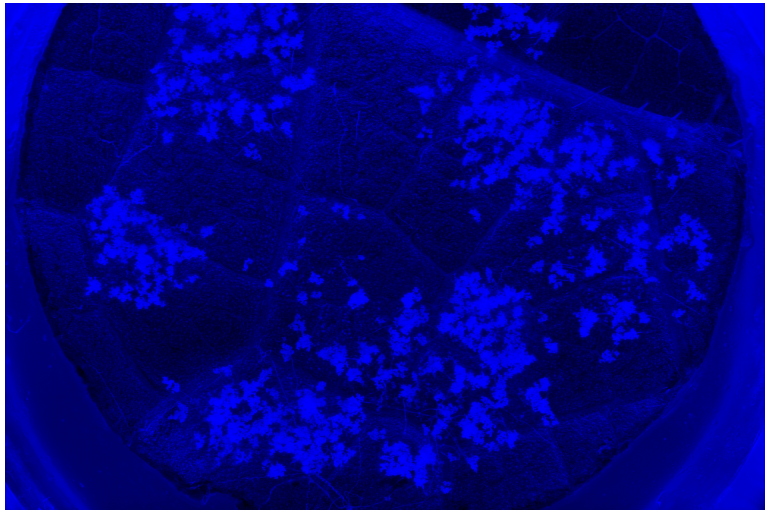
Red channel



Green channel



Blue channel



Each channel describes the same pixel positions.

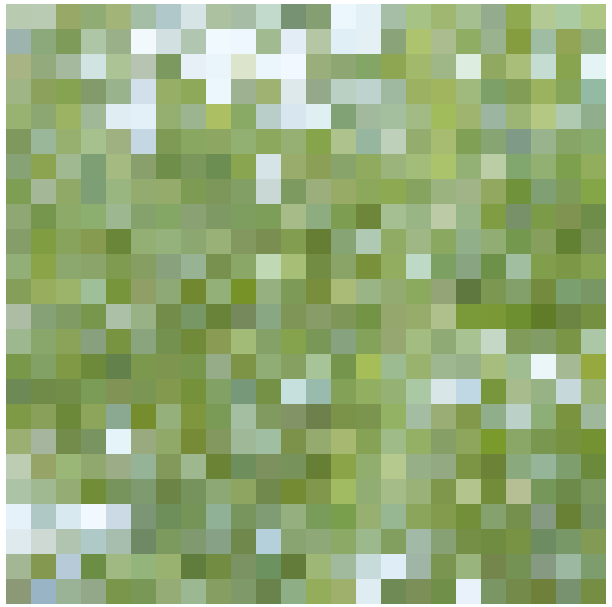
One pixel: **R = 135** **G = 156** **B = 90** = RGB color

Click to combine the channels into one RGB image.

Pixels need context: scale, detail, and channels



12 × 12 samples



24 × 24 samples



Original detail

Physical area needs a scale.

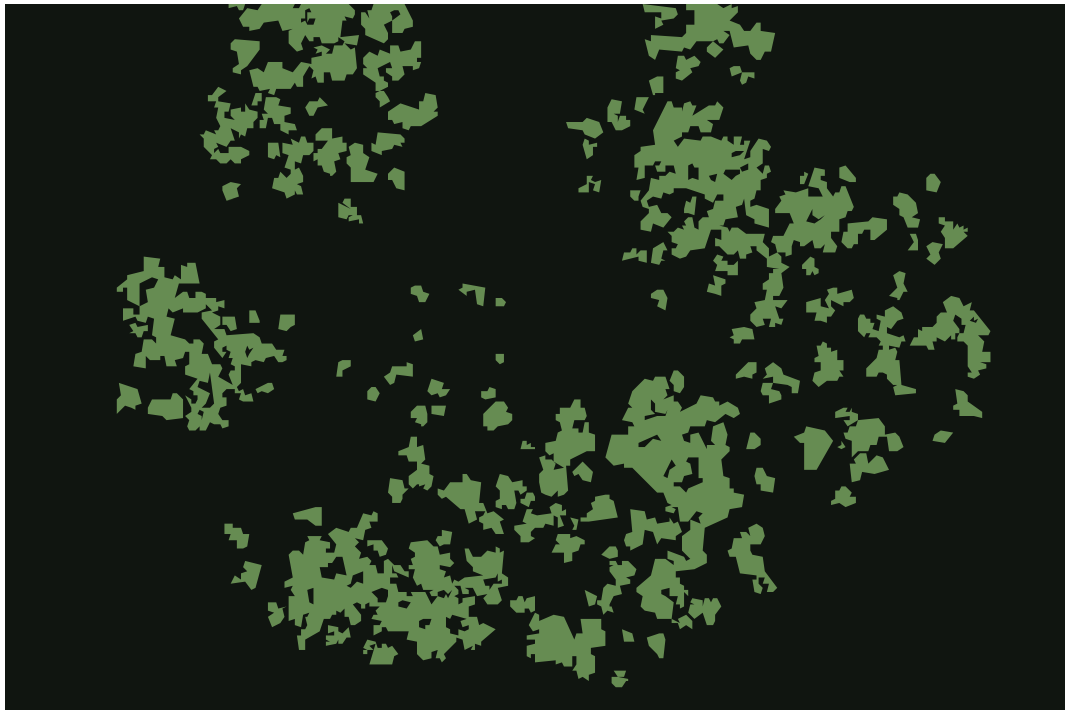
Microscopy may have C channels.

More pixels ≠ more optical detail. Display color ≠ biological identity.

A mask is an image of labels



RGB image: what the camera measured



Label mask: what the annotator marked

0 = background

1 = target region

Green is a display choice.

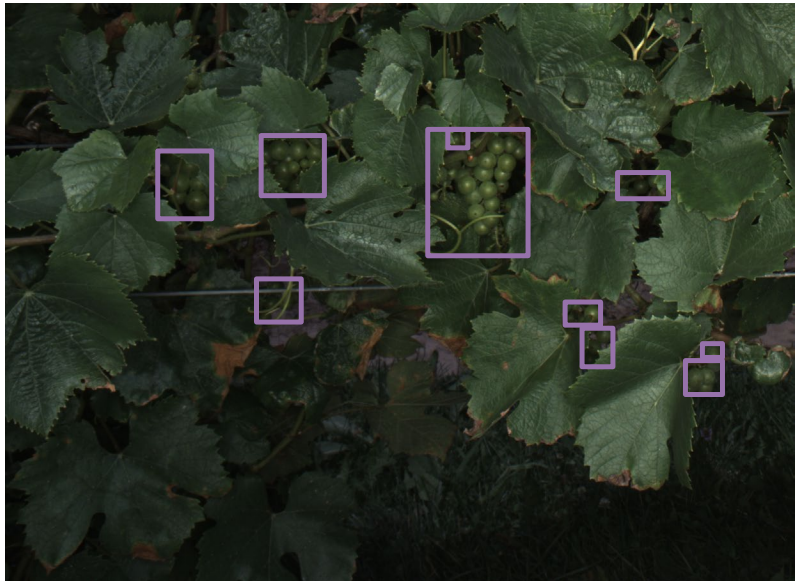
Same image, different questions and outputs

Classification



Berry region present

Detection



Boxes locate regions

Semantic segmentation

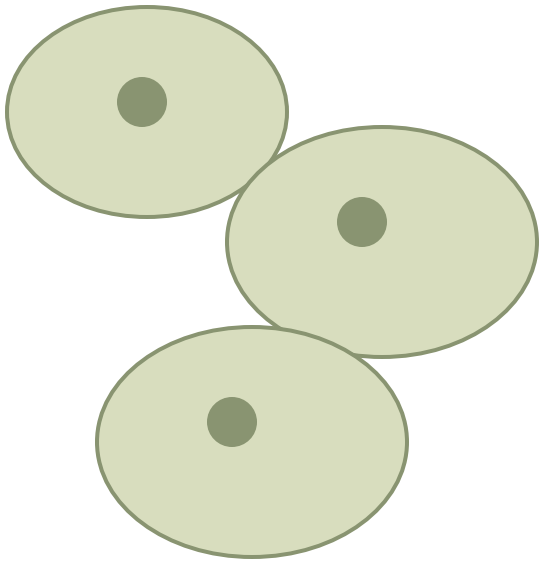


Pixels define boundaries

Choose the output that contains the information your biological question needs.

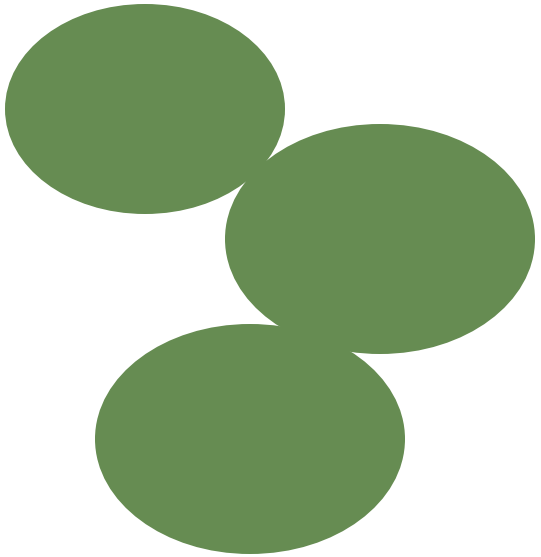
Class is different from object identity

Image (schematic)



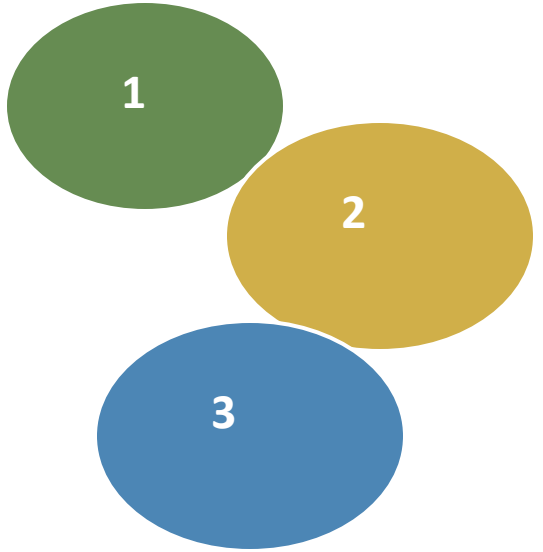
What is visible?

Semantic mask



Which class?

Instance masks

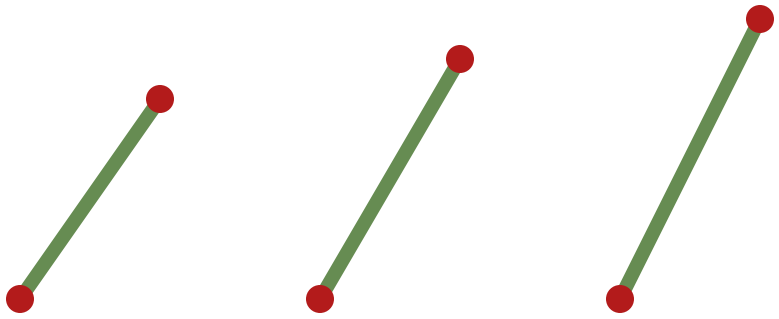


Which individual?

Our exercise: semantic segmentation of disease regions.

Some questions require time or 3D information

Keypoints + tracking



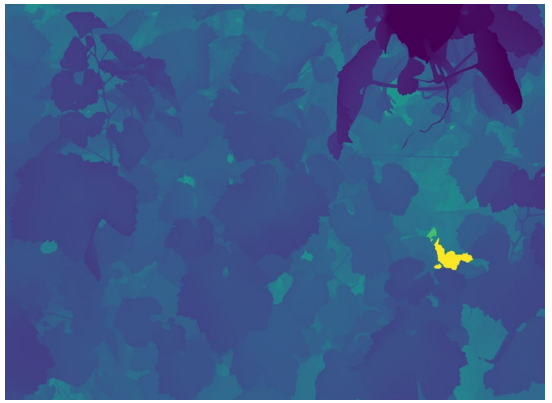
ID 7 ID 7 ID 7

t0 t1 t2

Posture and growth (schematic)

Image + depth





Geometry needs calibrated acquisition.

Tracking adds identity over time; depth adds spatial information.

A boundary contains more information than a box



Mask → bounding box

Where is the region?

Mask → pixel area

How much projected area?

An image label or a box cannot recover the missing boundary.

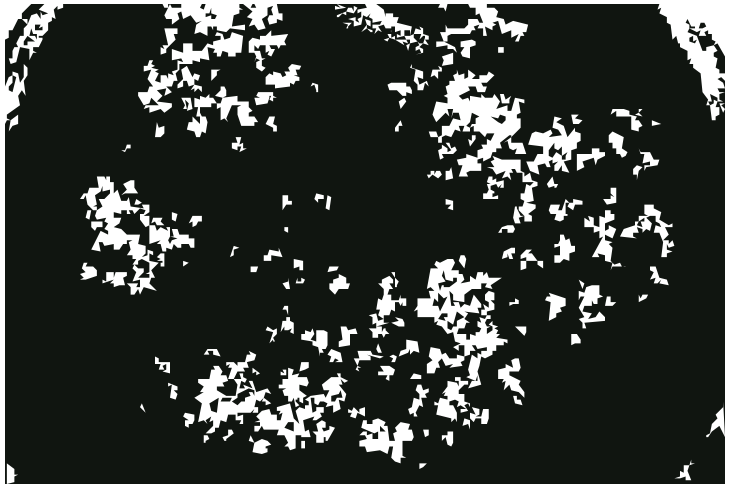
Traditional CV: write a rule, inspect its errors

Original image



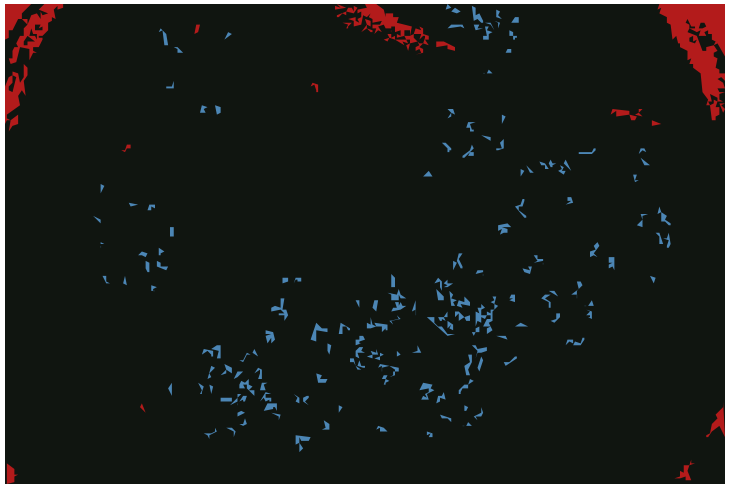
Veins? Glare? Sporulation?

Brightness rule



min(R, G, B) > 180

Errors vs reference



Extra pixels

Missed pixels

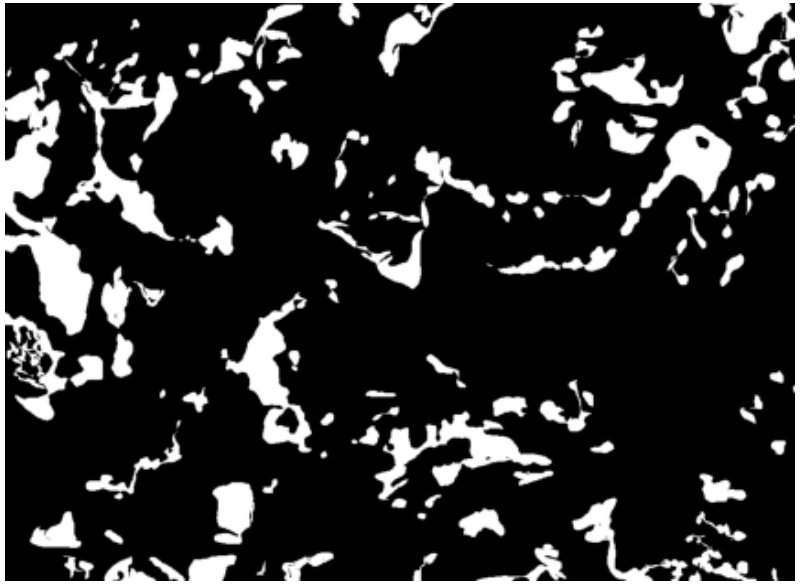
A rule follows the numbers we specify. Biological context may be missing.

Deep learning learns from image–label pairs

Image



Reference label



Model prediction

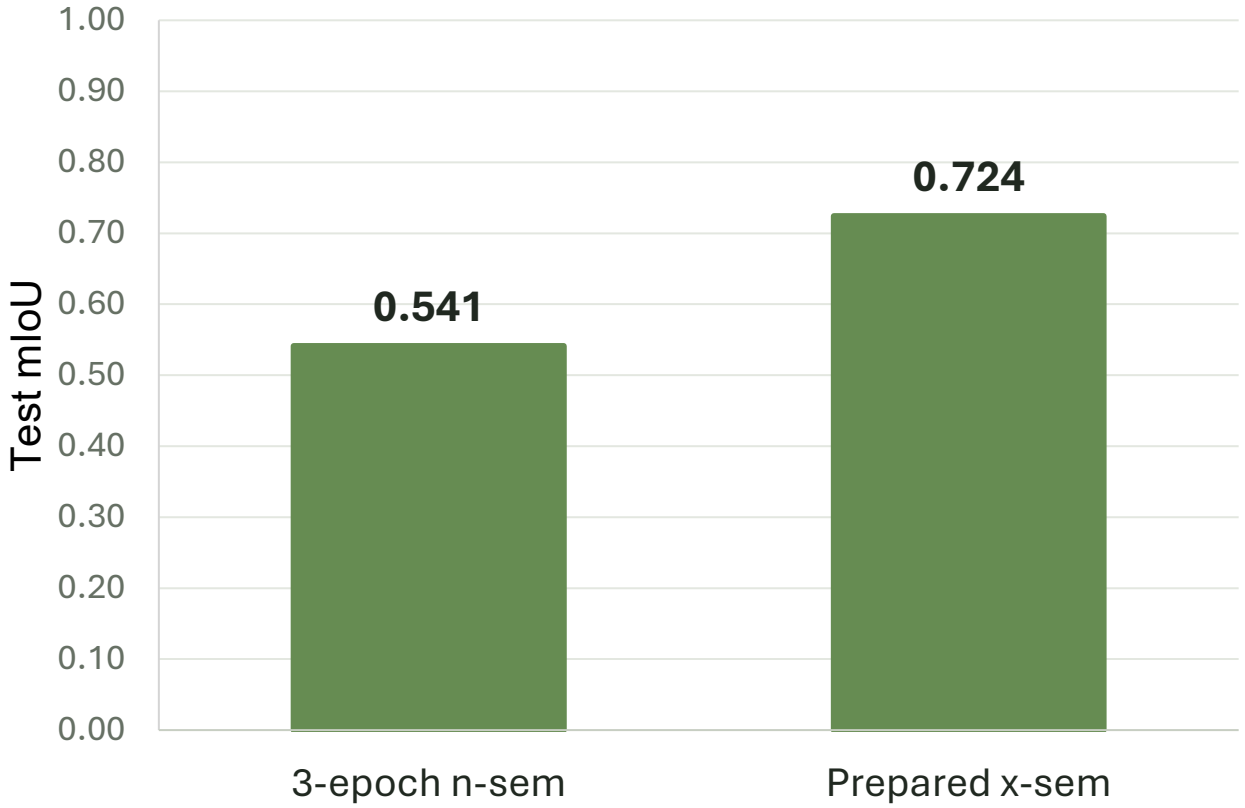


Compare prediction with reference

Loss → update model parameters

Training changes parameters. Inference applies the trained model.

Pretraining gives us a starting point



Broad visual features

↓

Your labeled images

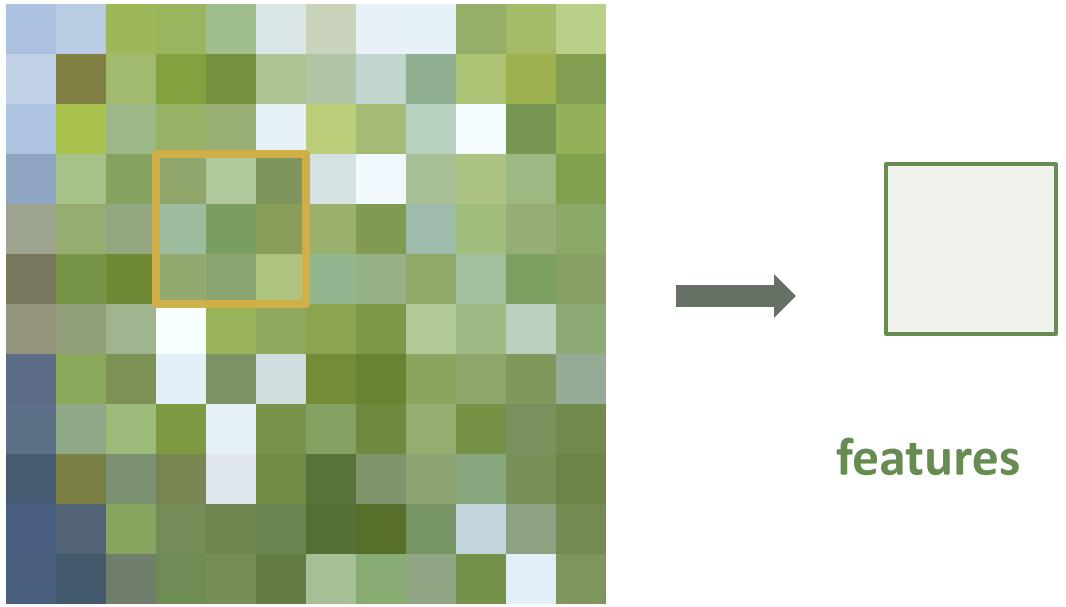
↓

Task-specific model

Recorded APS example. Different model sizes and training histories; not a controlled comparison.

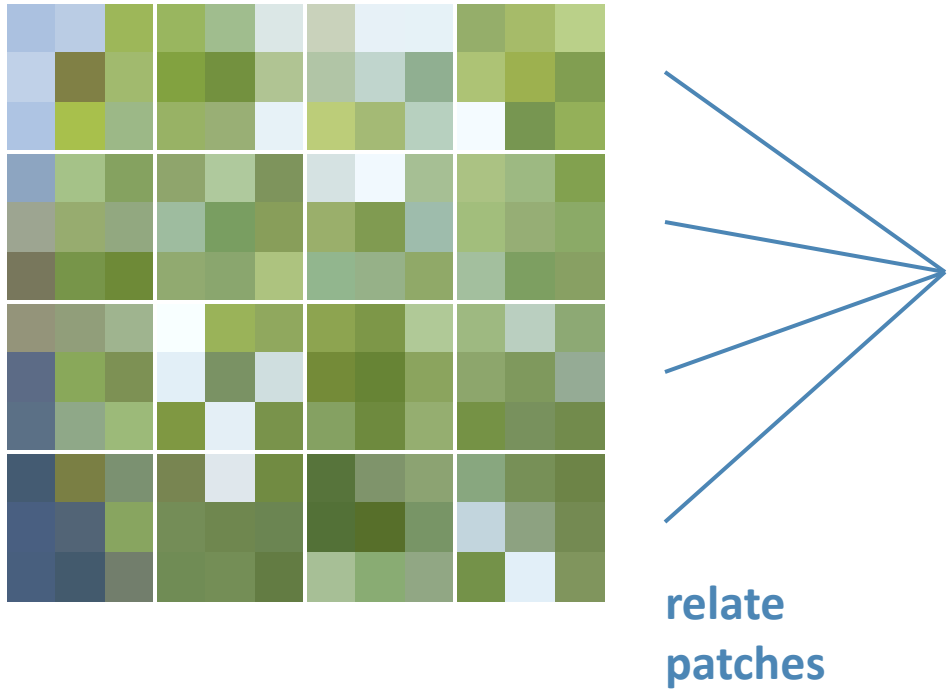
Two ways to learn image features

CNN: local filters



features

Transformer: image patches



relate patches

Local → larger patterns

Context across the image

Architecture learns features; task design determines what the model returns.

Foundation models make segmentation promptable



Official SAM3 demonstration · Meta

Prompt

“a brown Golden Retriever”

Image + prompt → masks

For biology: define the target, inspect failures, validate on your own images.

Your turn: mark the target on a Blackbird image



15-minute exercise

1 Find the target

2 Trace its boundary

3 Review together

visit this website for annotation tool:

https://tinyurl.com/vg3vxm

visit this website for sample image:

https://tinyurl.com/vg3workshop
(download the annotation_demo_img.png)

Good labels begin with a shared biological definition



What counts as the target?



Reference boundary for discussion

Agree on inclusion, exclusion, and uncertainty before annotating at scale.

Evaluate on images the model has never seen

154 Blackbird images

108	23	23
Training	**Validation**	**Test**
Fit model weights	Choose settings	Final evaluation

Keep related biological samples together.

Audit independence by plant, leaf, experiment, and acquisition session.

Hands-on: open the APS YOLO26 notebook



01 Connect Drive + GPU

02 Check image / mask pairs

03 Train → inspect → predict

Open the YOLO26 notebook

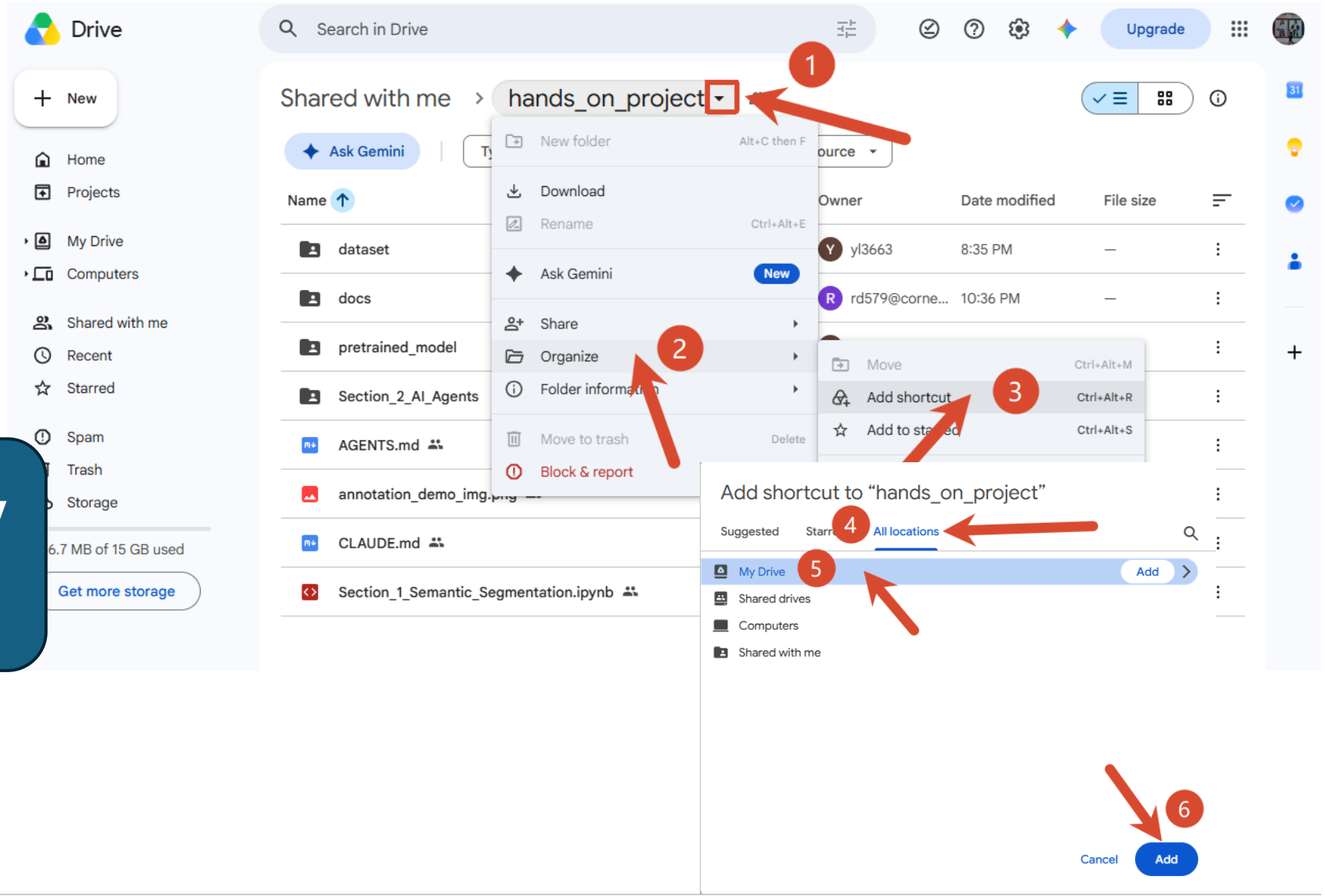
yolo26n-sem.pt · semantic masks

Prepared dataset: blackbird_dataset. Fallback: the notebook's saved outputs.

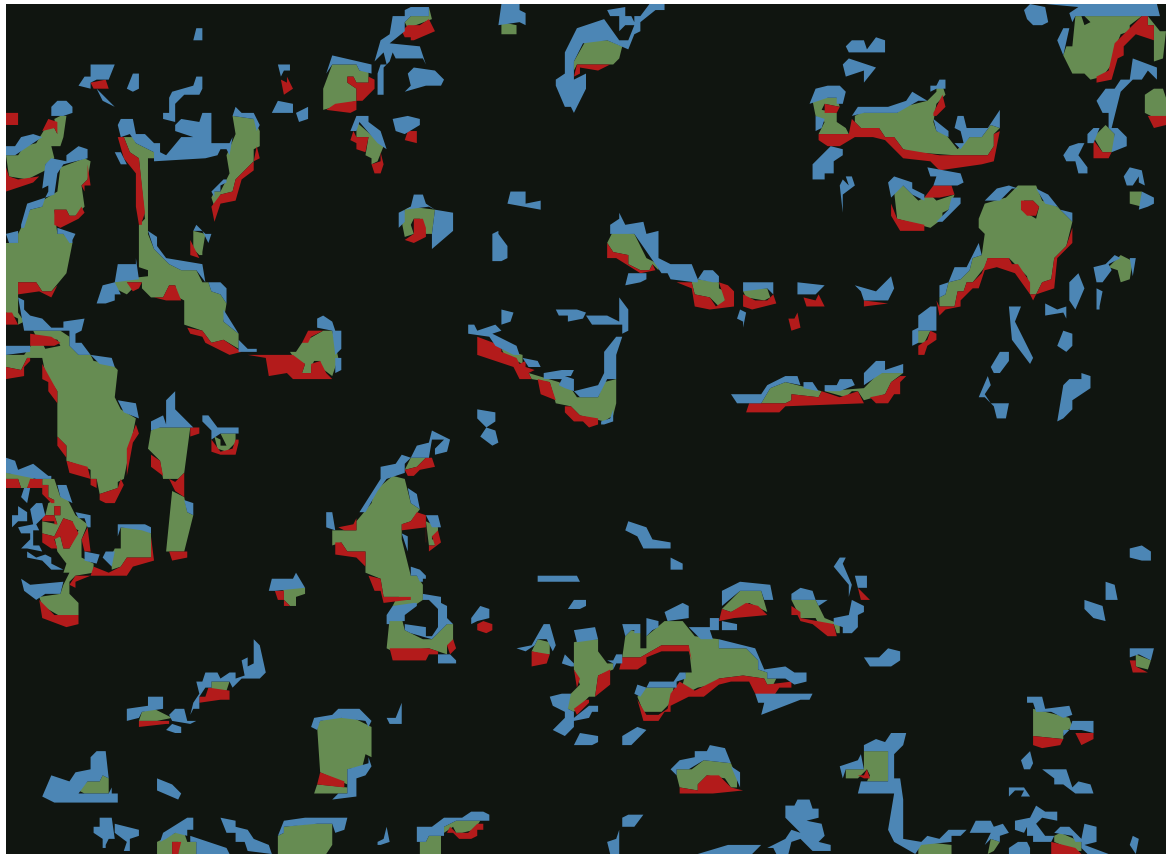
Project 1: Blackbird Downy Mildew Sporangia Segmentation

Visit this link for all the materials in Google Drive:

https://tinyurl.com/vg3workshop



IoU measures agreement at the pixel level



$$\frac{\text{Intersection}}{\text{Union}}$$

$$TP / (TP + FP + FN)$$

Correct foreground

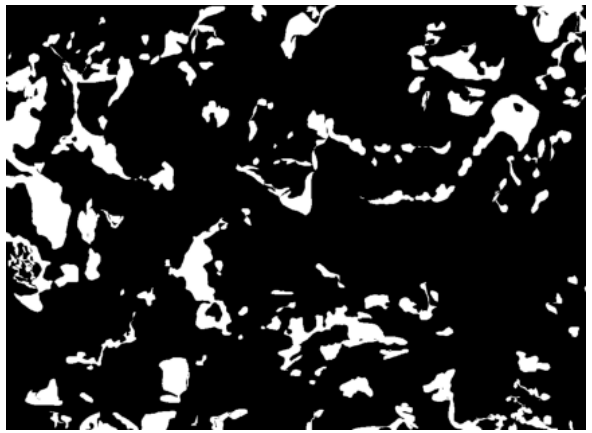
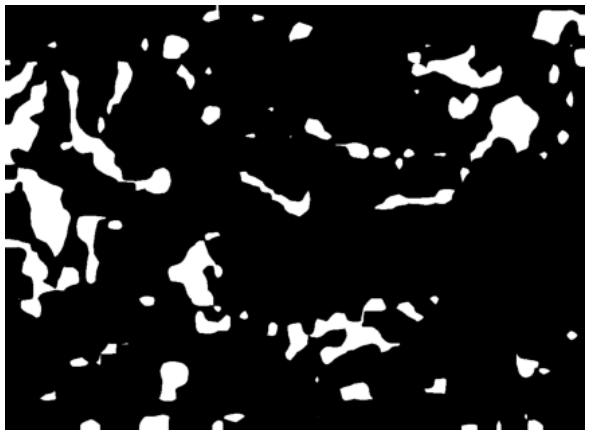
Extra prediction

Missed target

Compare the predicted mask with an independent reference, especially on difficult images.

Inspect a prediction before trusting its measurement

Image | **Reference** | **Prediction** | **Overlay**


What is missing?

What was added?

Does the error change the biological conclusion?

A disease mask still needs a denominator





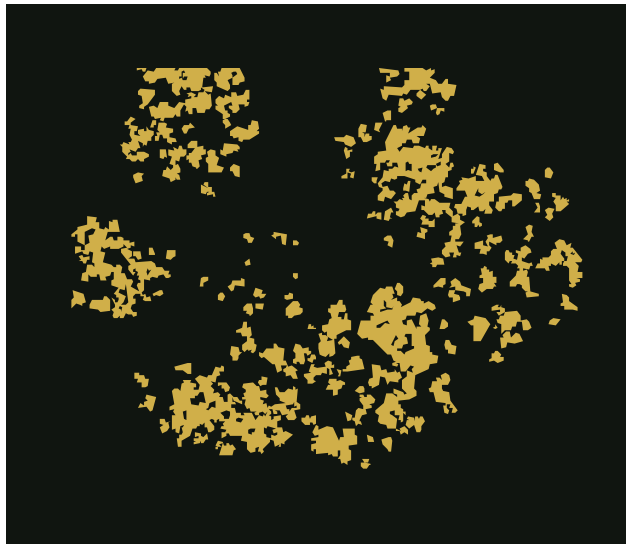
Target area

Which tissue area?

Visible affected fraction = target pixels within tissue ROI / visible tissue pixels

A phenotype must be useful for the biological question





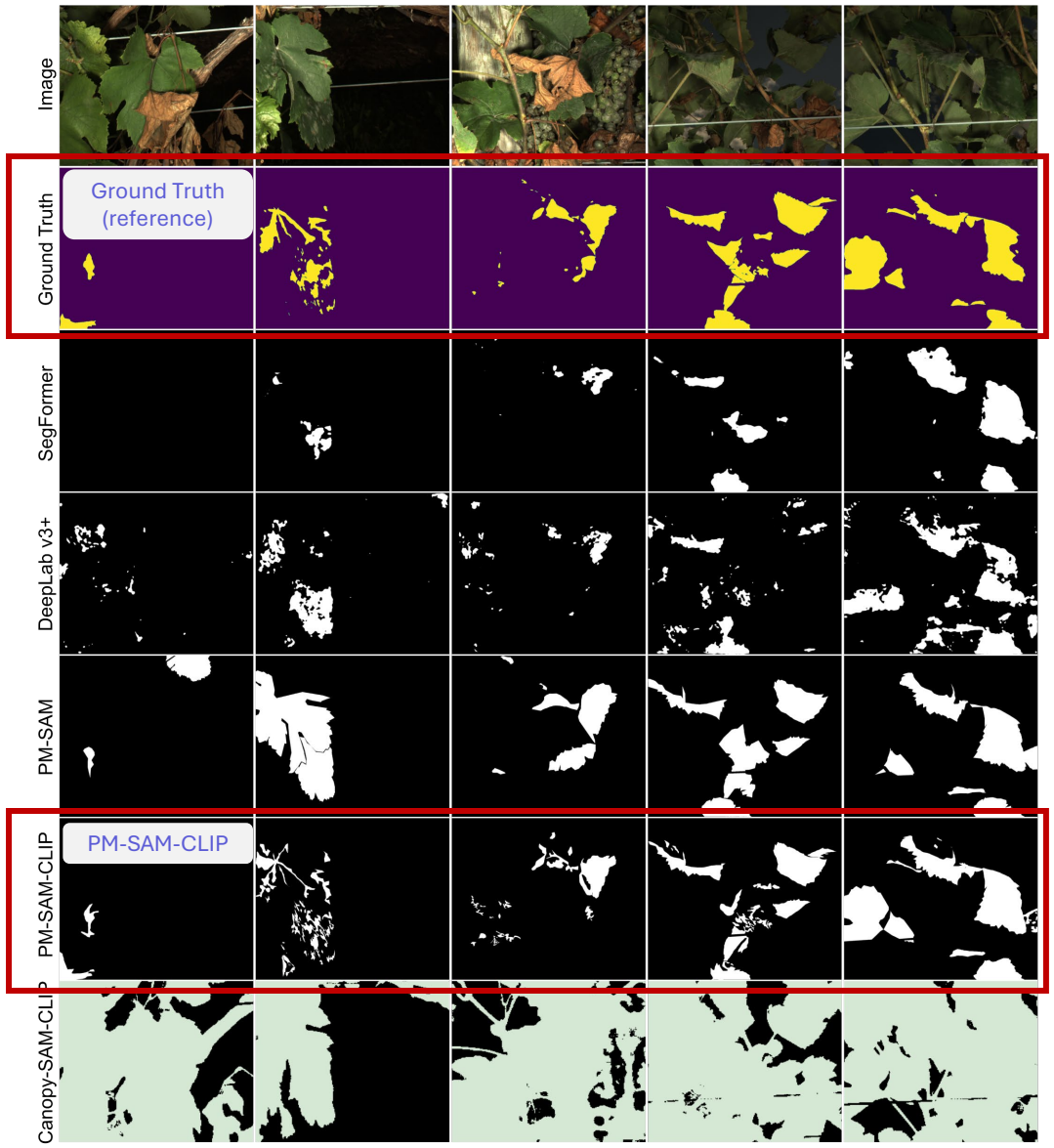
Plant ID
Time point
Trait + units
Uncertainty

Image **Target mask** **Phenotype record**

Validate → compare biological replicates → test the scientific claim

Pipeline illustration. The presenter's research examples follow.

SAM-CLIP Model Performance

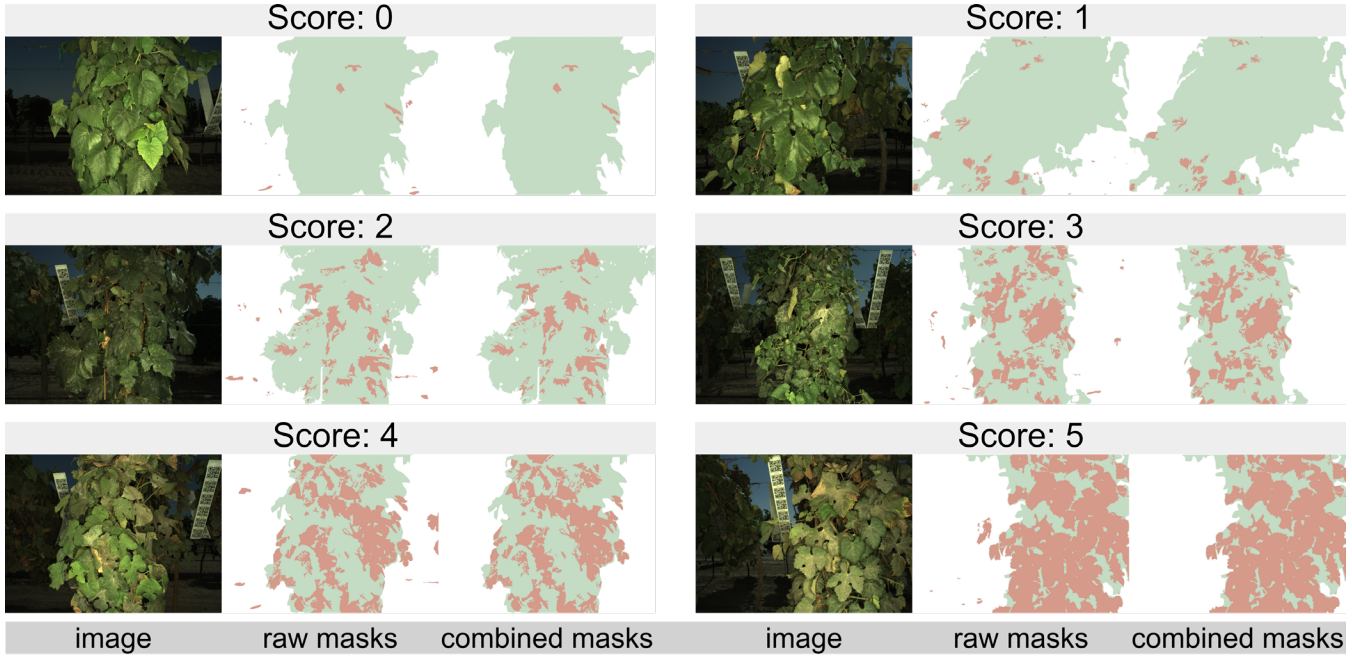


Model performance of SAM-CLIP and baseline

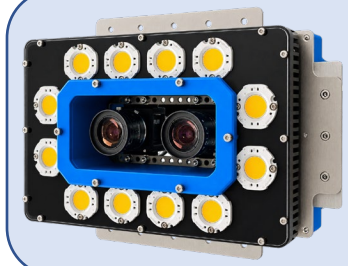
Model	Dataset	$mIoU^I$	$mIoU^D$	IoU^{C5}	IoU^{C10}	IoU^{C15}
PM-SAM-CLIP	PM_NY_2019	59.63	64.12	21.51	26.40	30.25
PM-SAM	PM_NY_2019	46.69	50.11	3.46	6.45	14.03
DeepLab v3+	PM_NY_2019	25.89	29.12	0.00	0.44	0.63
SegFormer	PM_NY_2019	21.41	25.54	0.00	0.00	0.00
Canopy-SAM-CLIP	Canopy_NY_2020	95.59	97.15	65.70	77.95	83.02

Human Scouting: 0-5 score

Robot Scouting: % to 0-5 score



PPBv2 Imaging System and Sample Data



PPBv2 Imaging System

(Strobe light「PhenoStereo」
+ stereo RGB cameras)



image (left) image (right)

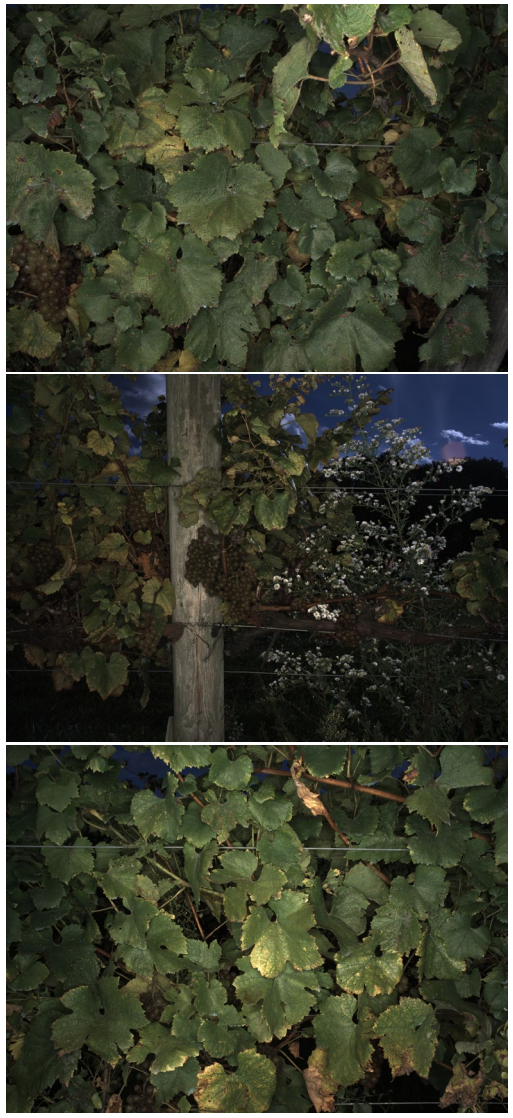
FoundationStereo



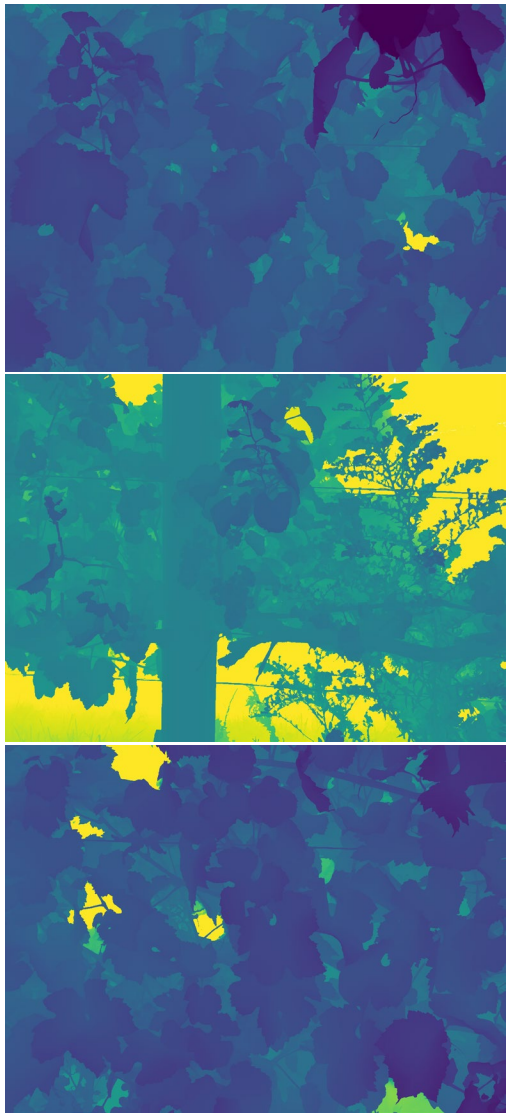
SuperPoint + LightGlue + PnP + GNSS



image



Reconstructed Depth

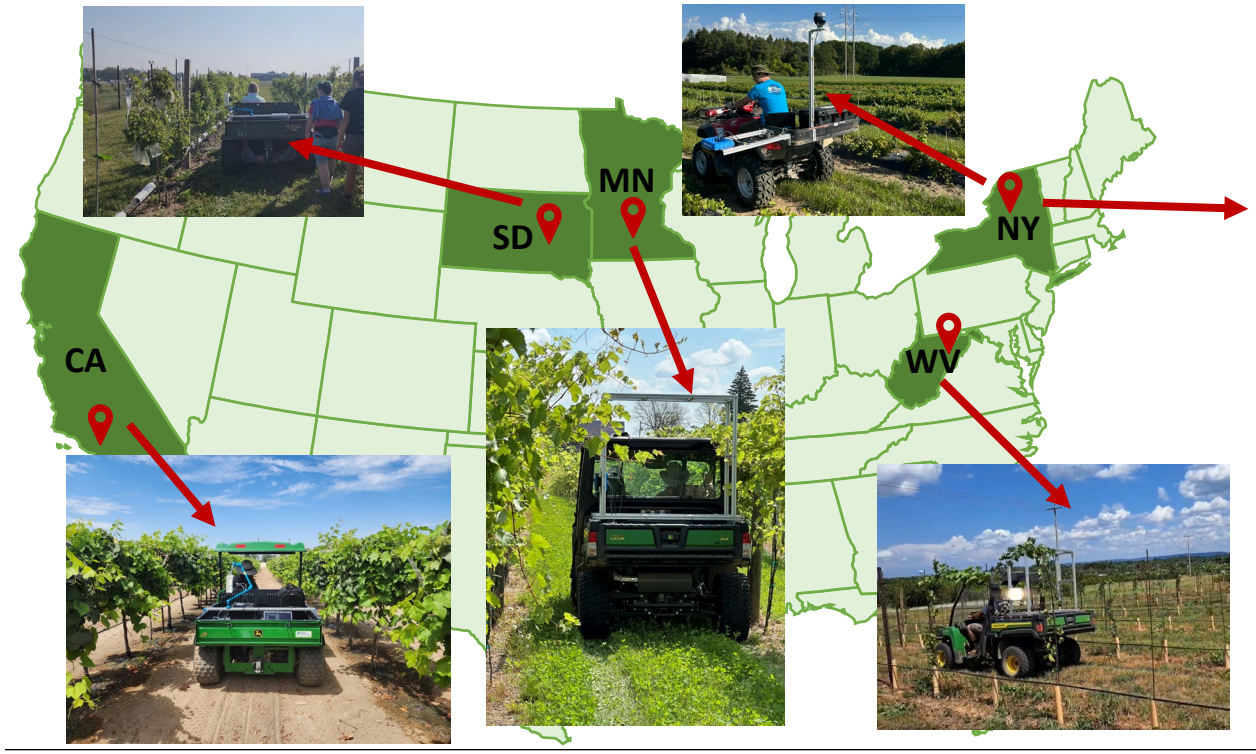


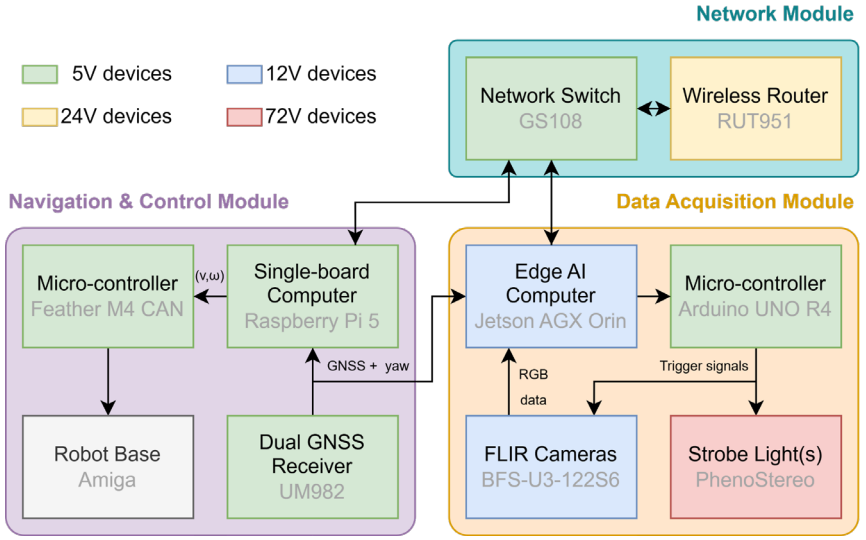
Reconstructed Point Cloud (0-2m)

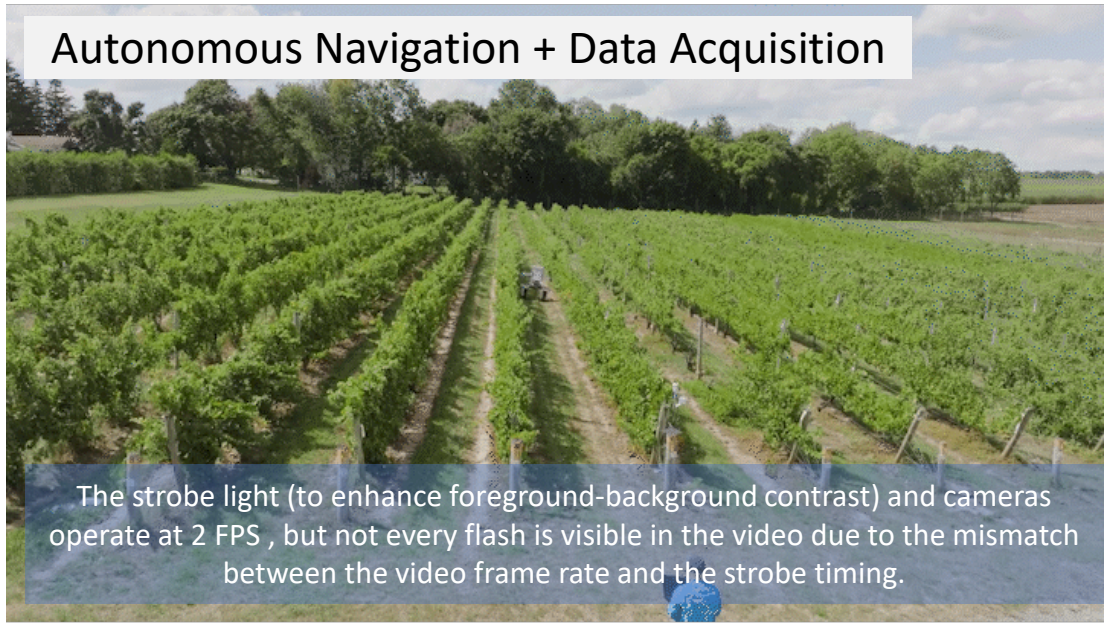


Weekly PPBv2 deployment

- PPBv2 robots based autonomous system
- ATV-based system
  (shared data acquisition module)



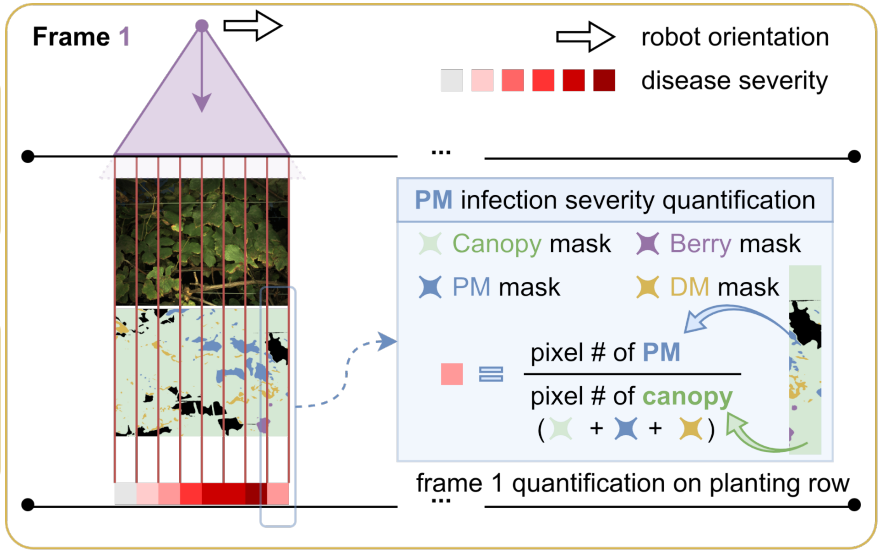


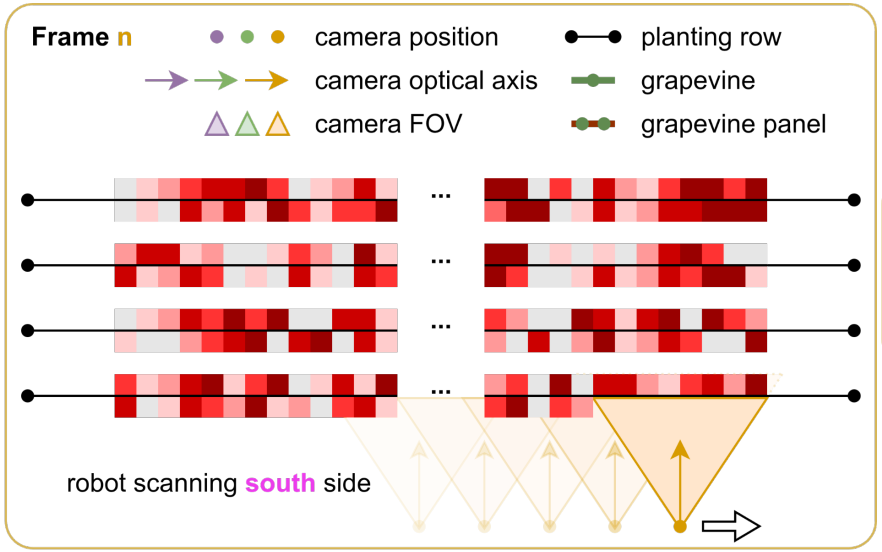


Geo-Referenced Data Mapping

row-based spatial discretization

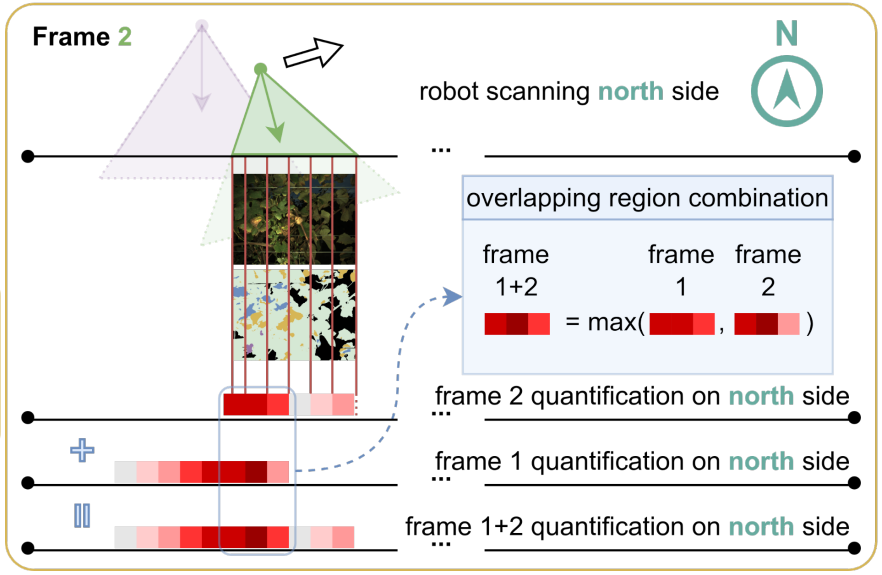
back-projection to image space

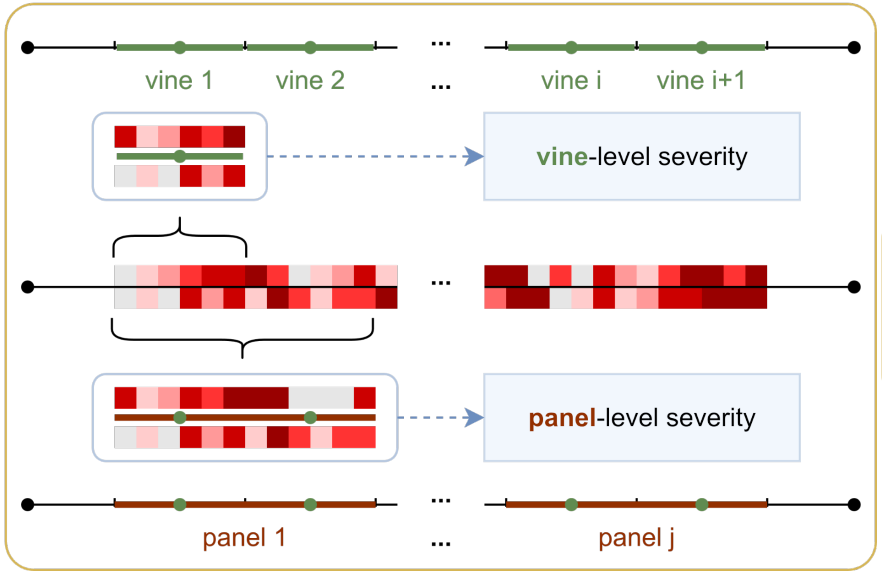




multi-view perception

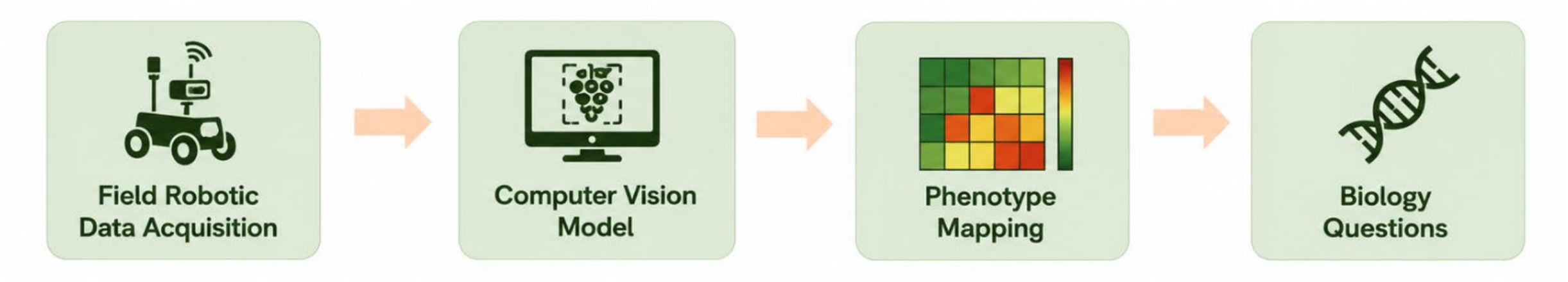
local disease quantification





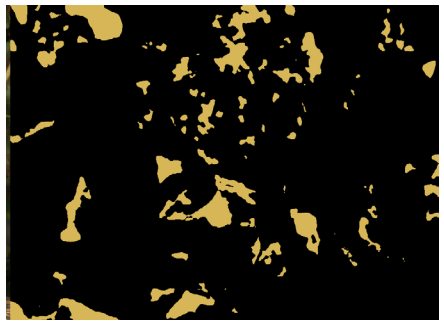
global spatial aggregation

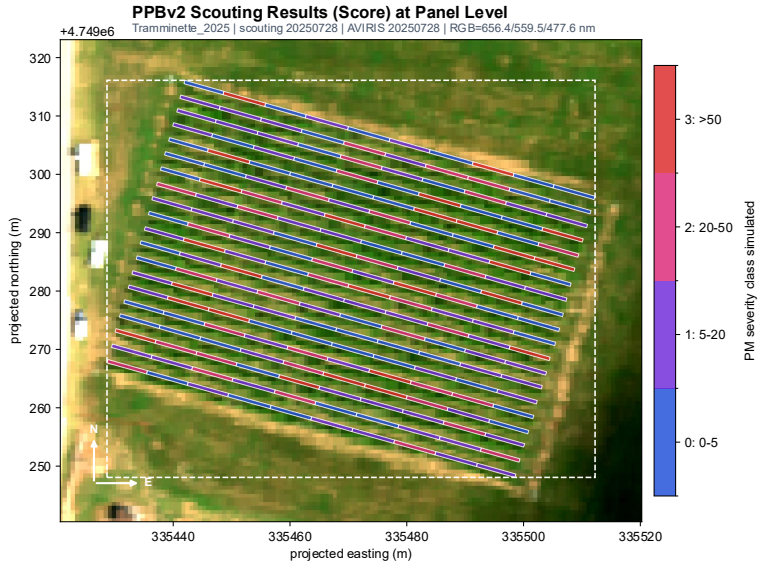
From Phenotyping to Answer Biology Questions

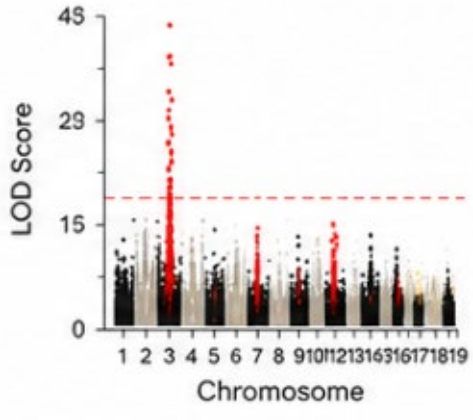




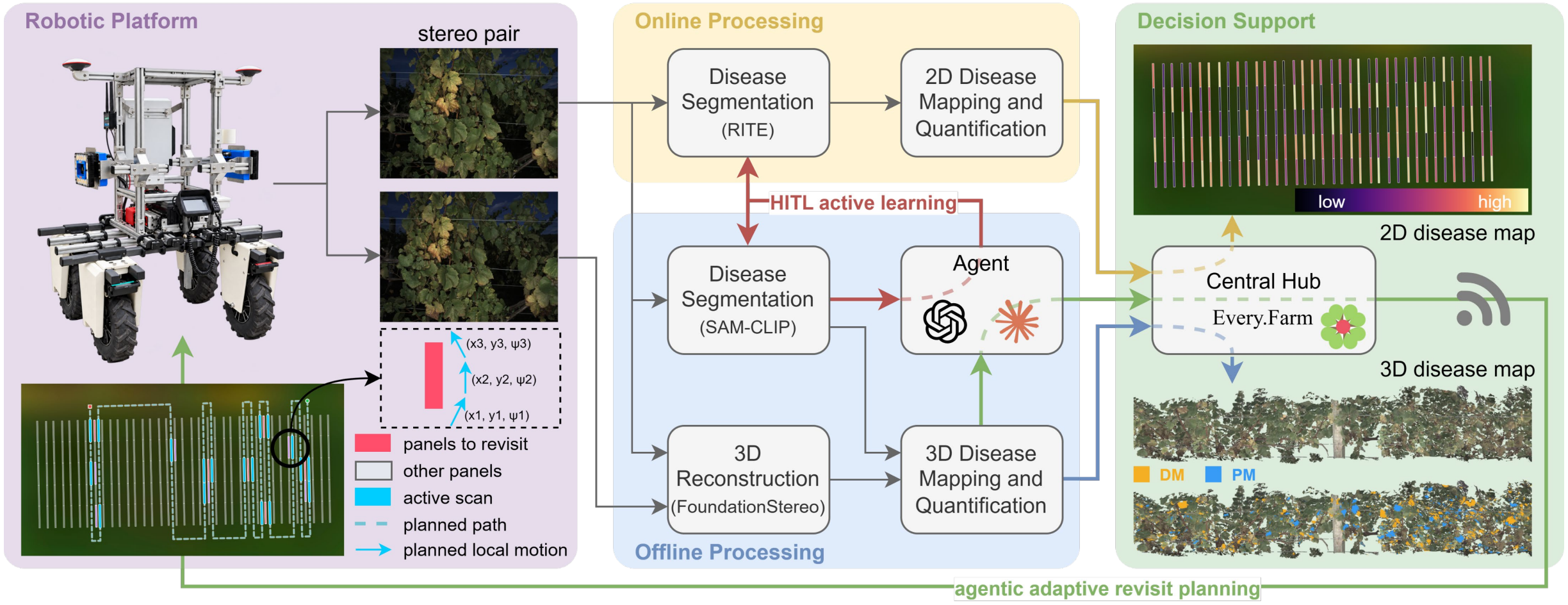








PPBv2 Agentic Workflow



PhytoPatholoBot Next Generation (PPB-NG)

GNSS Antenna
Sparkfun SPK6618H



Dual GNSS Receiver
MJRTK UM982



SWIR Power Control Unit
Specim PCU



Computer for Data Acquisition
Vecow ECX-3200 (M12)



Robot Base
Farm-ng Amiga

Computer for Navigation
Raspberry Pi 5



In-view White Reference (for Hyper-spectral cameras)

Hyper-spectral Cameras
Specim SWIR & Specim FX10e (VNIR)



Stabilization Mount
SOMAG RSM 400



Thermal Camera
FLIR A6701 (MWIR)

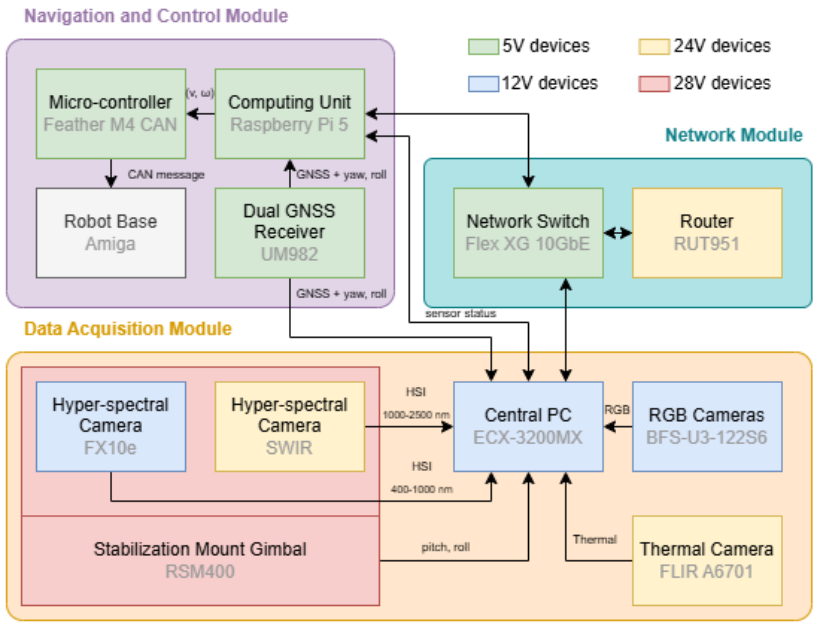


RGB Camera
FLIR BFS-U3-123S6

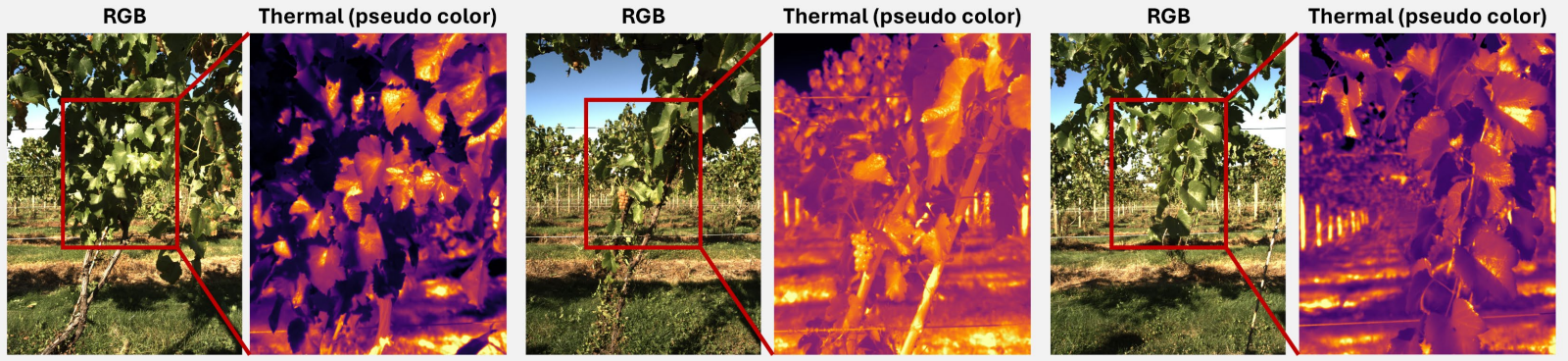




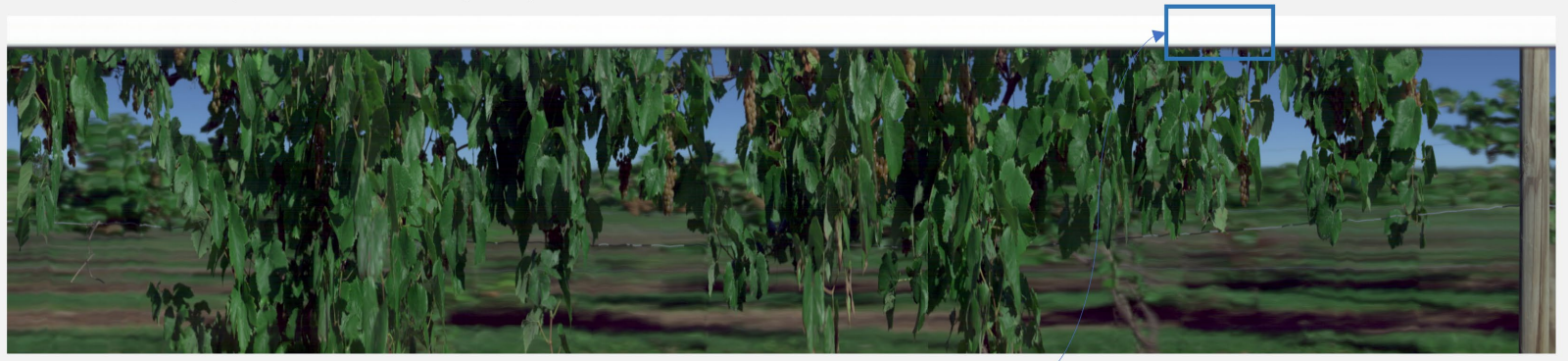
PPB-NG Modular Design & Sample Data





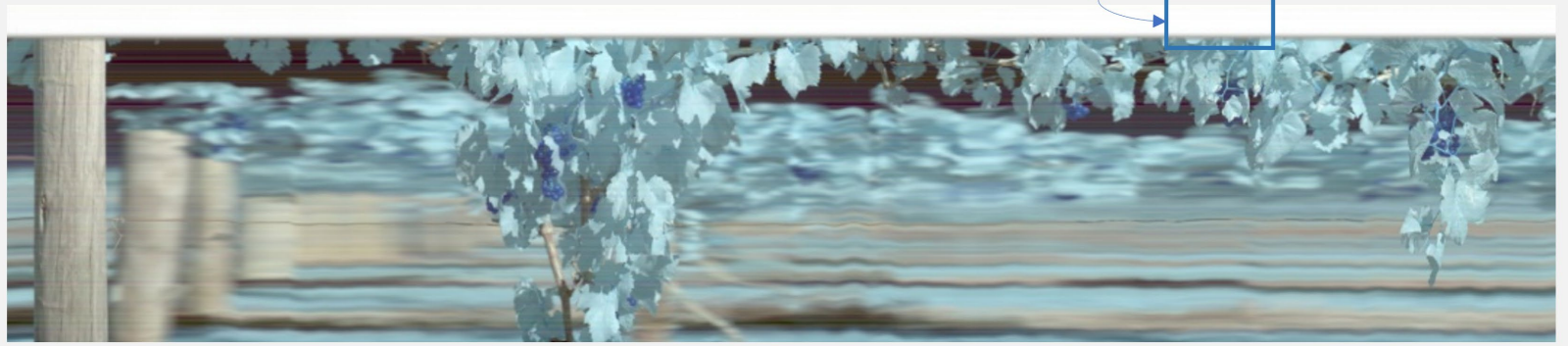


400-1000nm VIS-NIR (visible-band RGB composite)

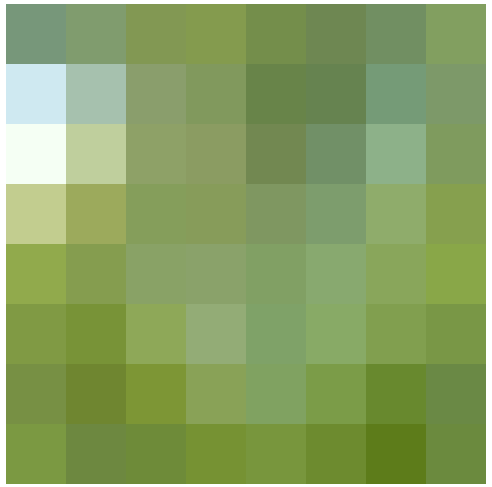


1000-2500nm SWIR (false-color composite)

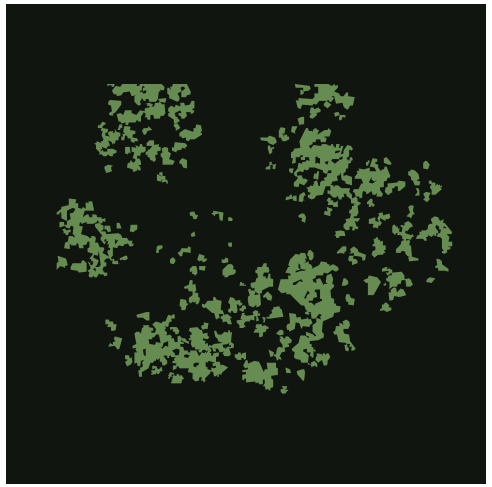
In-view white reference



Four questions for your next imaging experiment



What did the camera measure?



What exactly is the target?



Where does the model fail?

plant_01
day_07
trait ± error

Does the trait answer the question?

From light and pixels to reproducible biological measurements.

Thanks for your attention!



Yiyuan Lin 林翊源

Center for Agricultural Intelligence and Robotics
Email: yl3663@cornell.edu
Personal Website: https://yiyuanlinxx.github.io/
CAIR Website: https://www.cair.cals.cornell.edu/

My Robots

About Me

CAIR