

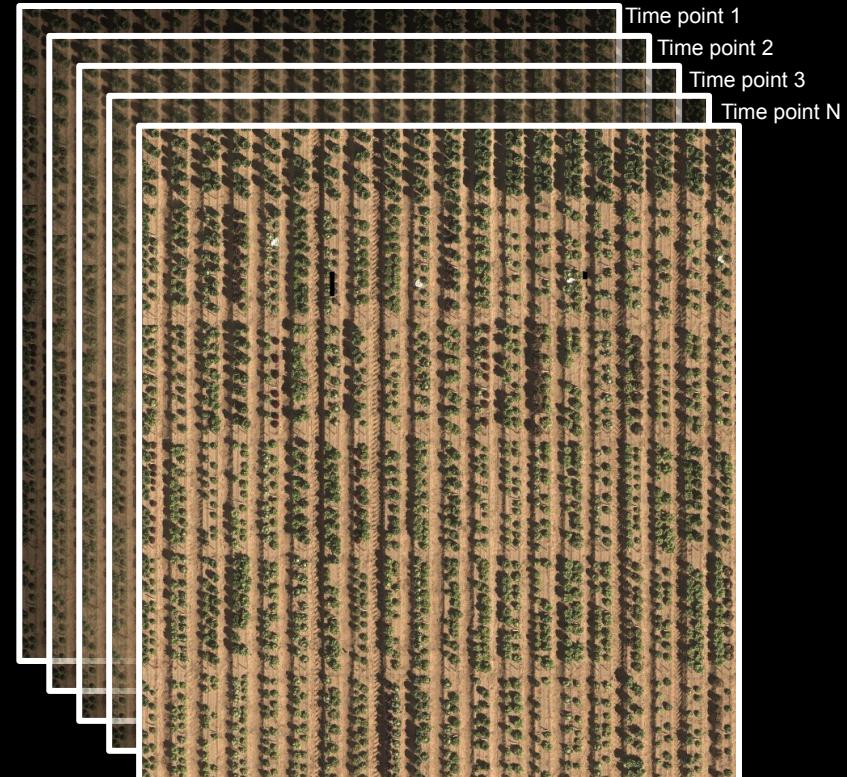


As easy as 1-2-3: Machine learning in real-world projects

ACIC 2021

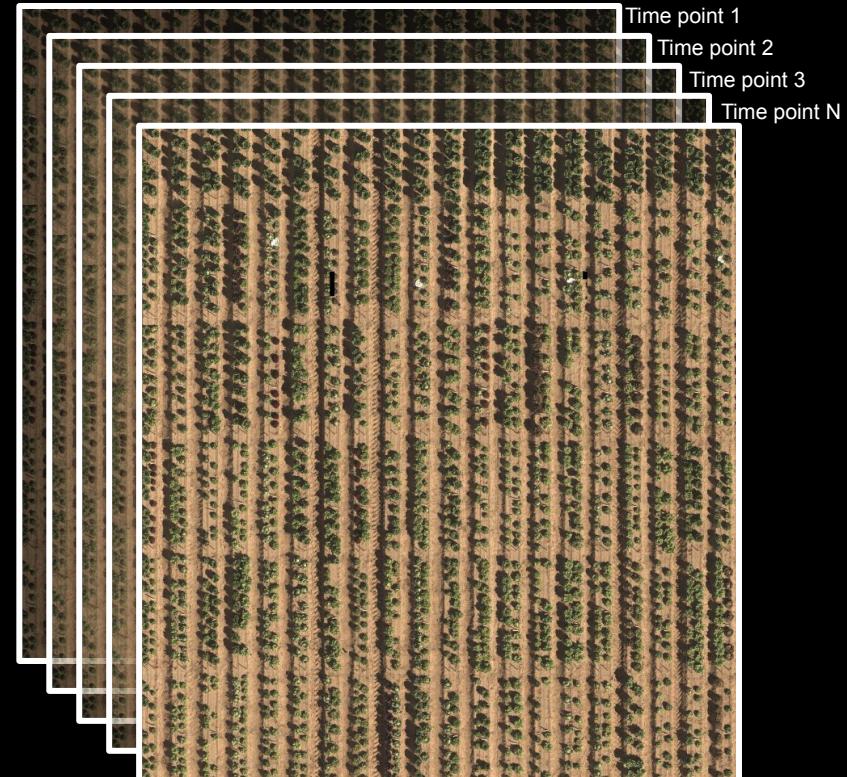
Problem Framing: Tracking thousands of plants

- **Plant Science perspective**
 - Interested in assessing phenotypic variation
 - Can we use our data to predict yield ?
 - Can we use our data to improve crop performance?
 - What other data and/or ML approaches will allow us to understand plant growth?

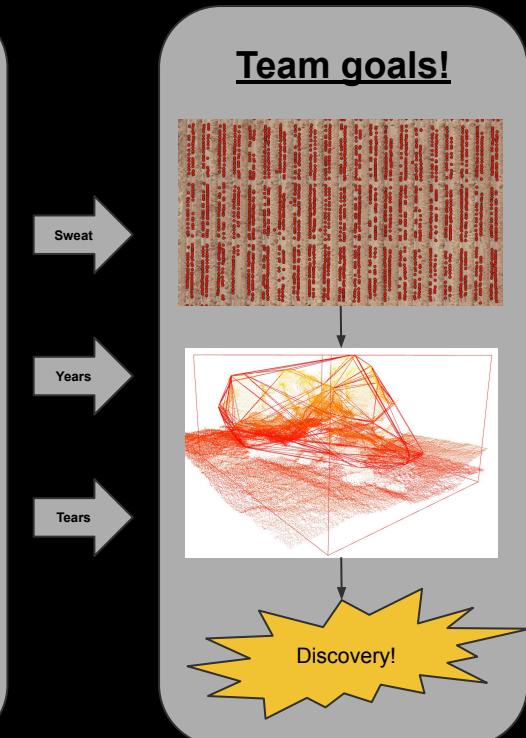
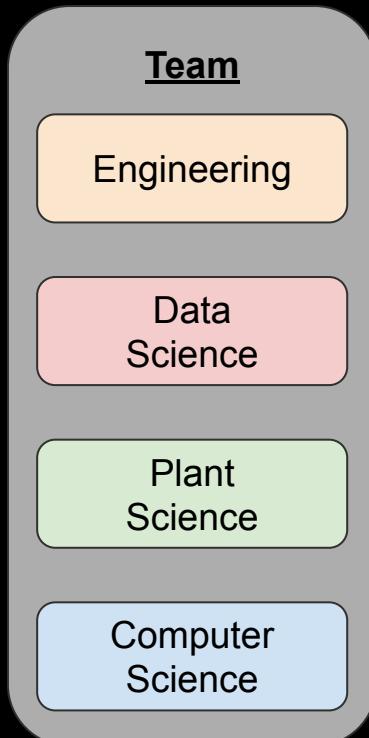
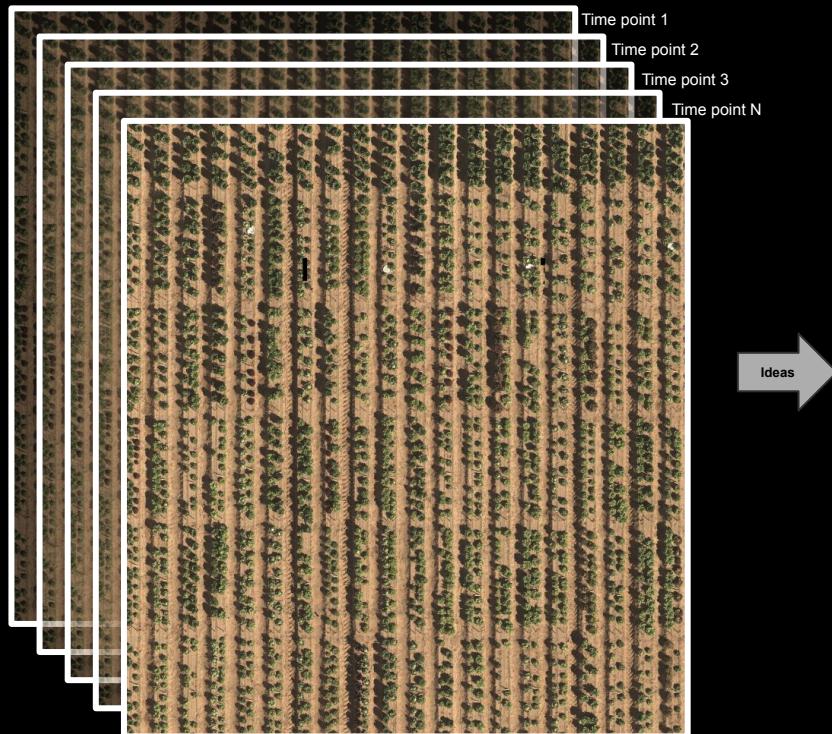


Problem Framing: Tracking thousands of plants

- Data Science perspective
 - Multiple data time points throughout growing season
 - Lots of training data = good!
 - Image processing techniques have been unsuccessful
 - Shift to ML methods



Problem Framing: Tracking thousands of plants

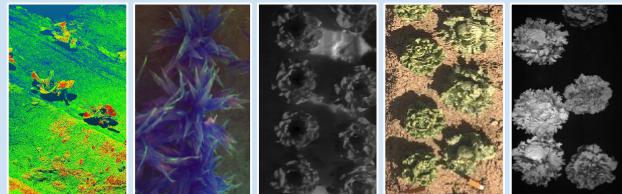


Data Collection: Converting raw data into gold

Collection method



Data management



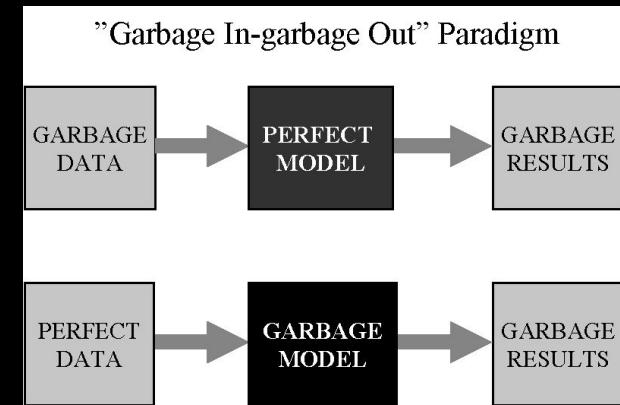
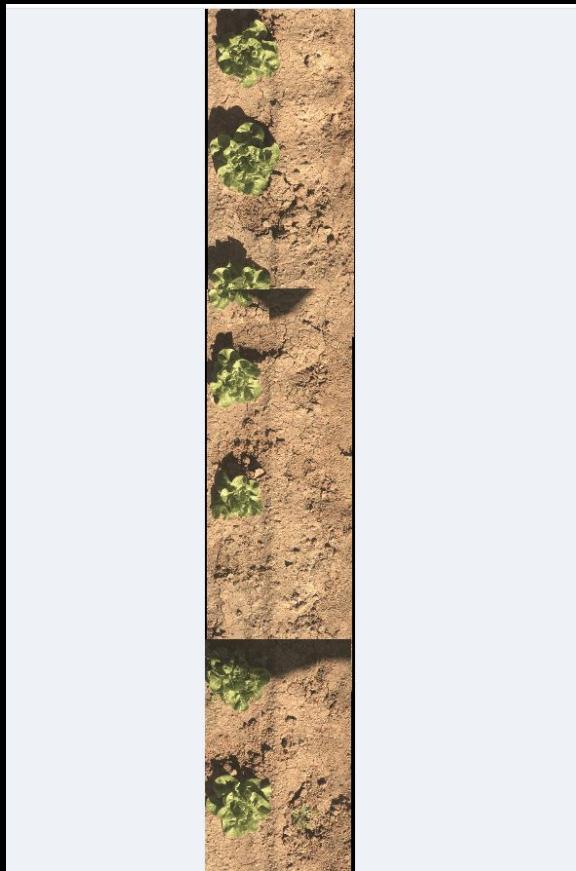
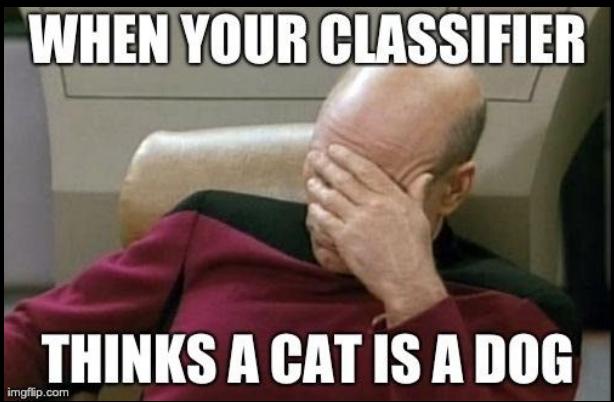
Data volume:

- Max: 10 TB/day
- Typical: 1.5 TB/day

Data curation



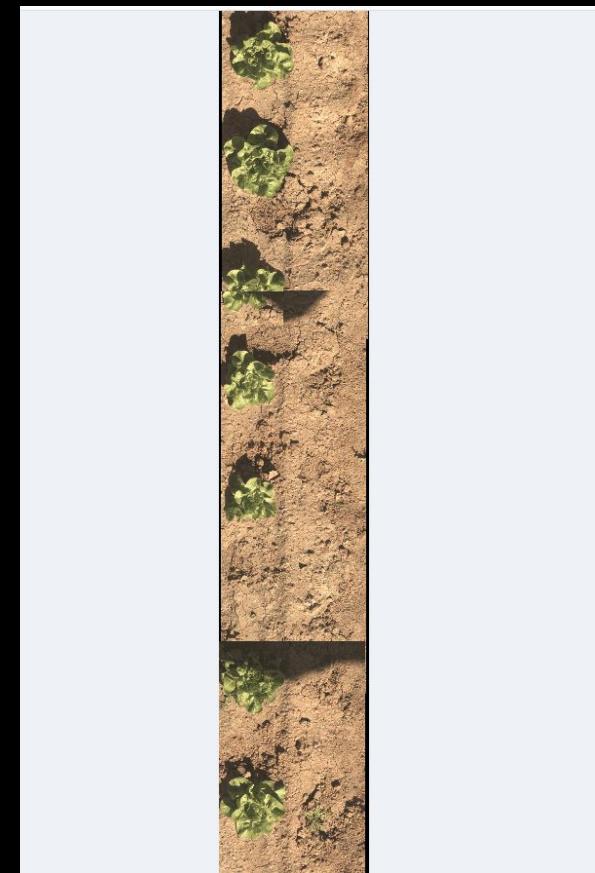
Data Preparation: Garbage in, garbage out



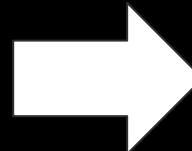
Data Preparation: Garbage in, garbage out

Let's label some images: [Labelbox](#)

The screenshot shows the Labelbox web interface for a project titled "acic_2021_plant_detection". At the top, there are navigation links: Projects, Models, Datasets, and Catalog. Below the title, a subtitle reads "Detection of lettuce in red-green-blue (RGB) images." A horizontal navigation bar includes tabs for Overview (which is underlined in blue), Labels, Performance, Issues, Export, and Settings. The main content area displays the message "Your project setup is complete." followed by "Next steps" and the instruction "You can now start labeling the data or add members to this project and then start labeling." At the bottom left is a blue button labeled "Start labeling".



Data Preparation: Garbage in, garbage out



Data Preparation: Garbage in, garbage out

In groups of 4-5:

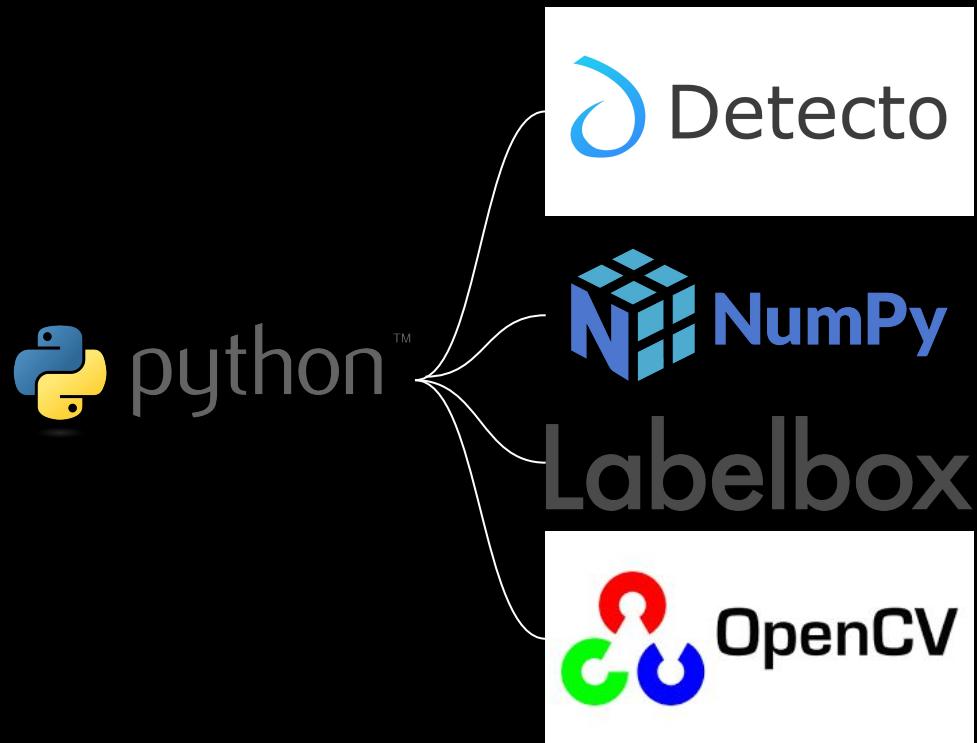
- Export labels
- Convert JSON file to XML
- Visualize labels for quality control

Labels:
[{"label": "Objects", "objects": [{"featuredId": "ckgd0jvjs042d0y9r8kx4fpf5", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "top left"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "top center"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "top right"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "top bottom"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "top middle"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "bottom right"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "bottom left"}, {"label": "Objects", "objects": [{"featuredId": "ckgdn0jw0q0000000000000000000000", "schemaId": "ckgd0d9zjl0530yashahesh5s", "color": "#1c1eff", "title": "plant", "value": "plant", "bbox": "bottom middle"}]}]



Model training: The relatively easy part

Launch Notebook: [Google Collab](#)



Model training: The relatively easy part

Labeling

- [Labelbox](#)

Machine learning

- [Detecto](#)

Array/image manipulation

- [NumPy](#)
- [OpenCV](#)

