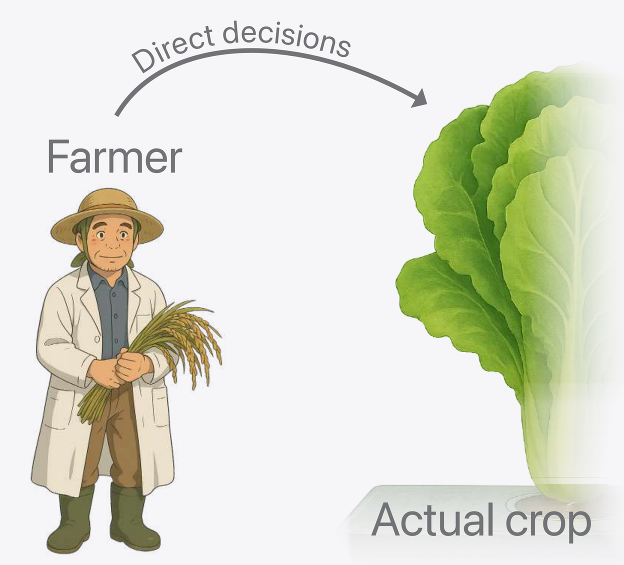
UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Investigating Deep Learning and Computer Vision for Predicting and Simulating Plant Growth Structures: Laying the Groundwork for Digital Twins in Agriculture

John Ivan T. Diaz, Craig Joseph B. Goc-ong, Kaye Louise A. Manilong, Alvin Joseph S. Macapagal, Philip Virgil B. Astillo*
Department of Computer Engineering, University of San Carlos



Current farming practice

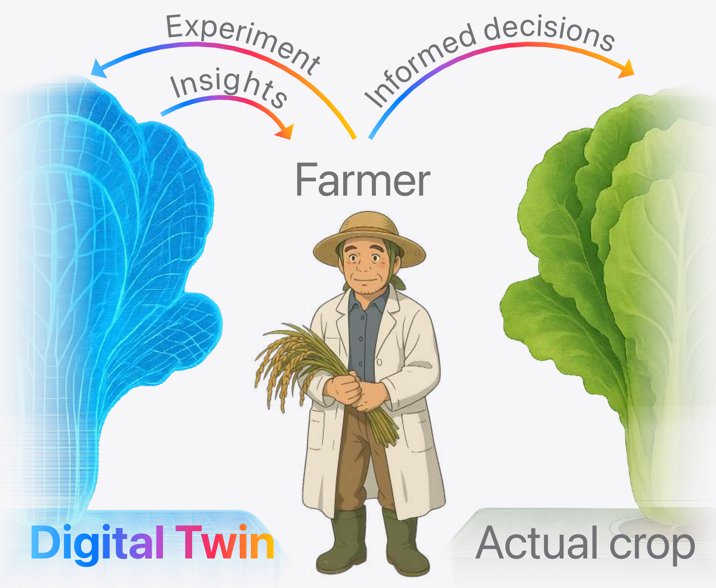
The problem

Vegetation is one of our primary sources of food. Crops grow in response to external conditions around them. A change in pH level, for instance, can influence their growth. We rely on farmers to grow and harvest them. Modern farming practices like vertical farming allow crops to grow in controlled environments, enabling farmers to manipulate external conditions such as pH levels. However, their decisions might not always be favorable to the crops. Applying too little or too much configuration can result in healthy or unhealthy crops. Crops can become damaged or experience stagnated growth, leading to reduced yields and resource wastage. It's risky, inefficient, and unsustainable.

Our Vision

Solution

We envision a future in farming practices where farmers use digital twins to guide them in their decision-making. A digital twin is defined as a digital replica of a physical counterpart. It mimics the behavior and appearance of an object, such as crops. Farmers first apply experimental decisions to the digital twin. Then, the digital twin mimics the crop's response based on those decisions. As a result, farmers receive insights and use them to make informed decisions on their real crops. This leads to no resource wastage, healthier crops, and higher yields. It redefines modern farming practices to be more sustainable and efficient, ultimately benefiting mankind and the planet.



Promotes SUSTAINABLE DEVELOPMENT GOALS

- 2 ZERO HUNGER
- 9 INDUSTRY, INNOVATION, AND INFRASTRUCTURE
- 12 RESPONSIBLE CONSUMPTION AND PRODUCTION

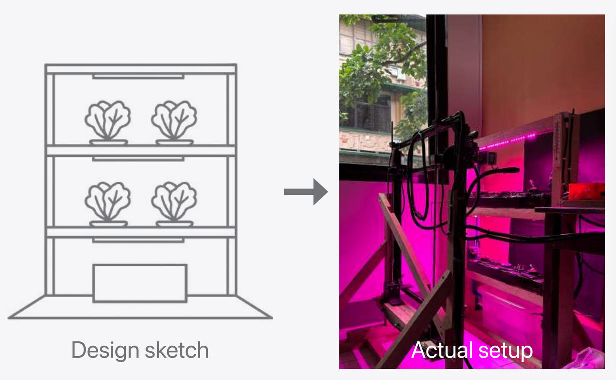
The Technological Impacts

- Farmers can make informed agricultural decisions
- Drives healthier crops and higher yields
- Promotes Sustainable and Efficient farming practices
- Entrepreneurial Opportunities for Developers

Our Foundational Contributions

Developing a Data Collection Platform

A small replica of a hydroponic vertical farm with cameras and sensors to collect plant growth data.



Features

- Plant images are captured in Stereo Vision
- Plants grow on top of rotating lazy susans to capture all four sides of their appearance
- External conditions (e.g., pH level) are recorded using calibrated sensors
- Images and sensor data are captured periodically using Arduino and C, and managed using Python and MySQL

Investigating Multimodal Convolutional Long Short-Term Memory Architecture for Predicting Future Plant Appearance

Multimodal learning

It learns how plants look under a given set of external conditions (e.g., ambient temperature), thus gaining the ability to predict future plant appearances based on those conditions.

Plant appearances (Images) + External conditions (Numerical data) → Pre-processing → Multimodal Convolutional Long Short-Term Memory Architecture → Future plant appearances (Images)

Model architecture

- ConvLSTMs
- Time Distributed
- Reshape
- Normalization
- Lambda
- Concatenation
- 3D Convolution

Proposing Unified Loss Function for Model Training

Three loss functions are explored: Mean Squared Error, Temporal Consistency, and one that combines both. The unified loss function showed slightly better training and testing results than the other two in predicting images of plant appearance.

Methodology flowchart

Training phase: Train Dataset → Preprocessing → Model Training → Trained model
Testing phase: Test Dataset → Preprocessing → Model Evaluation

The dataset contains 64 plant growth sequences captured at 12-hour intervals. The plant used is lettuce, but the model is scalable to other plant types.

Preprocessing

Segmentation → Resize → Normalization → Sequence Length Standardization → Augmentation

Test results

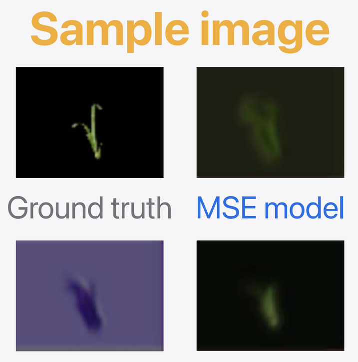
★ Better result

Metric	MSE model	TC model	Unified model (MSE+TC)
MSE	0.0001 ★	0.0326	0.0001 ★
PSNR	45.1848 dB	14.8740 dB	46.9794 dB ★
SSIM	0.9633	0.0162	0.9692 ★
TC	0.0001	0.0000 ★	0.0001
TV	148.1984	239.4765	141.8375 ★

Metrics are Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Temporal Consistency (TC), and Total Variation (TV).

Evaluation process: Actual image from the dataset ← Compare w/ metrics ← Predicted image by the model

Sample image

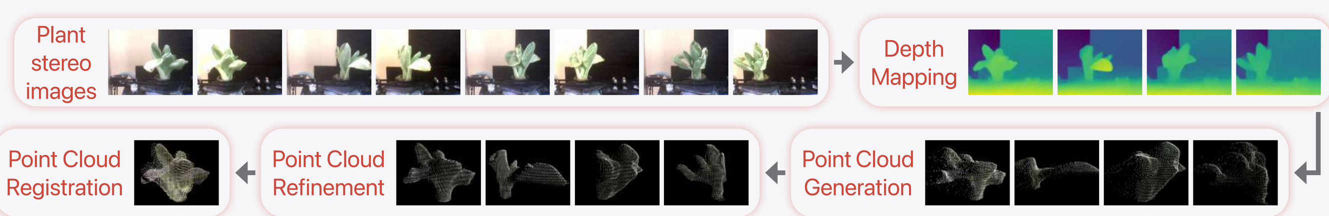


Predicting the 10th future time step based on the past 35 images. The same configuration of external conditions from the dataset is used to generate these predictions. Window size is 30. Stride is 1.

Investigating Computer Vision Techniques to Create a 3D Model of the Plant from Constrained Stereo Images

The challenge is to transform 2D stereo images, captured around an object at 90-degree horizontal rotations, into a 3D model.

Techniques used



The techniques mentioned use OpenCV and Open3D libraries in Python. Point cloud generation uses 3D reprojection. Point cloud refinement uses the .transform function with statistical and radius outlier removal. Point cloud registration uses the .translate, .rotate, .scale, and .transform functions. Mesh generation uses the Ball Pivoting Algorithm.

Investigating Segmentation Architectures Trained Solely for the Plant Class

Correct segmentation of plant objects positively affects succeeding phases such as 3D modeling. YOLOv8 and Detectron2 are the initial architectures chosen for comparison.

Train results

★ Better result

Metric	Test A		Test B		Test C	
	YOLOv8	Detectron2	YOLOv8	Detectron2	YOLOv8	Detectron2
Box AP/50	0.850	0.870 ★	0.830	0.840 ★	0.830 ★	0.740
Mask AP/50	0.840	0.880 ★	0.830 ★	0.800	0.830 ★	0.730
Box Loss	0.400	0.180 ★	0.360	0.190 ★	0.390	0.390
Mask Loss	0.800	0.110 ★	0.660	0.110 ★	0.810	0.150 ★
Train Time	1.400 hrs ★	1.590 hrs	1.570 hrs ★	1.920 hrs	0.870 hrs ★	2.020 hrs
System RAM	5.500 GB	4.600 GB ★	5.900 GB	3.300 GB ★	4.900 GB	4.000 GB ★
GPU RAM	9.200 GB	3.500 GB ★	9.200 GB	3.300 GB ★	5.200 GB	2.500 GB ★
Disk Usage	33.100 GB ★	35.900 GB	33.300 GB	33.200 GB ★	33.200 GB ★	35.900 GB

Decreasing hyperparameters →

The dataset comprise over 1,000 diverse plant images. Epoch sizes are 100, 75, and 50 (Tests A to C). Batch sizes are 16, 8, and 3. Worker counts are 8, 6, and 4. The initial and final learning rates are constant across all tests, at 0.01 and 0.001, respectively.

Methodology flowchart

Training phase: Train Dataset → Annotation → Model Training → Trained model
Testing phase: Test Dataset → Annotation → Model Evaluation

Strengths / Weaknesses

YOLOv8

Strengths:
- Fast training speed
- Stable accuracy with fewer epochs
- Low disk storage usage
- Effective with smaller hyperparameters

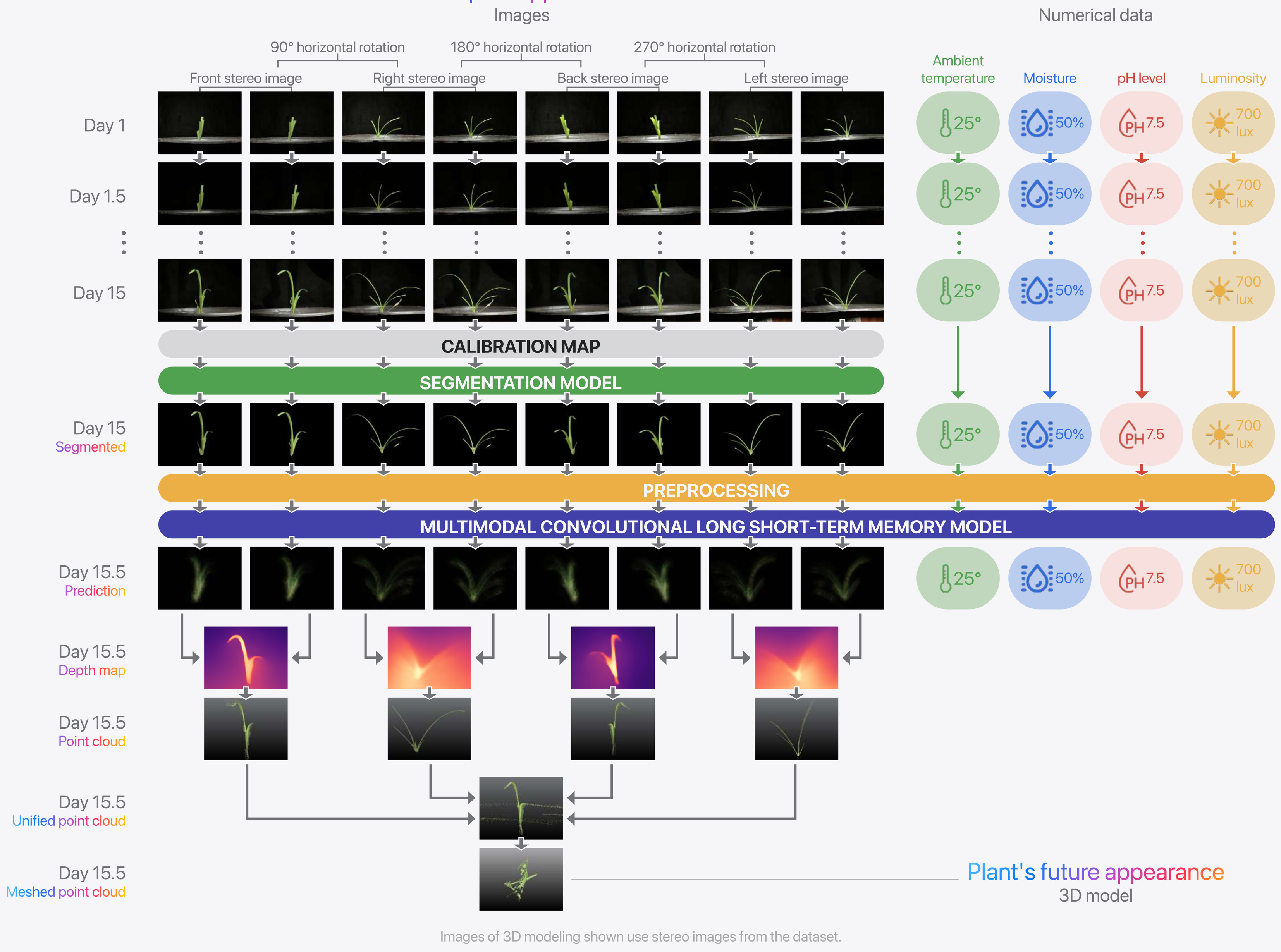
Weaknesses:
- Slightly lower segmentation accuracy
- High GPU RAM consumption with larger sets
- Higher RAM consumption

Detectron2

Strengths:
- Higher segmentation accuracy
- Lower GPU RAM usage
- Low RAM consumption
- Significant accuracy gains with higher hyperparameters

Weaknesses:
- Slower training speed
- High disk usage
- Accuracy greatly affected by fewer epochs
- Poor performance with minimal hyperparameters

Sample of whole process



Images of 3D modeling shown use stereo images from the dataset.

Recommendations

For researchers who believe in the vision and wish to contribute to its pursuit:

- Continue collecting data. Grow more plants. Expand to other plant types, and grow them under different external conditions.
- Explore data collection techniques that capture top and bottom plant views and yield denser point clouds.
- Explore deep learning techniques that can address weaknesses found in current models, such as blurriness in predictions.
- Expand the 3D modeling methodology to include quantitative evaluation.
- Implement an automatic point cloud merging technique in 3D modeling to replace the current manual process.
- Continue future R&D stages, such as developing augmented reality experiences to view 3D plant models in physical space.