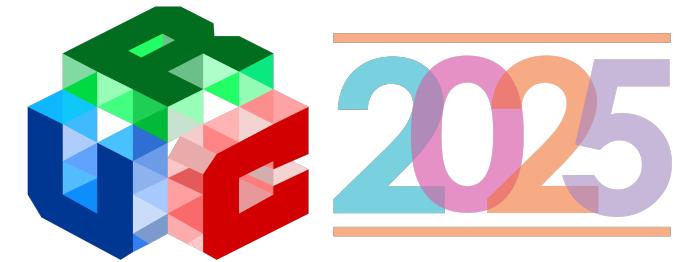
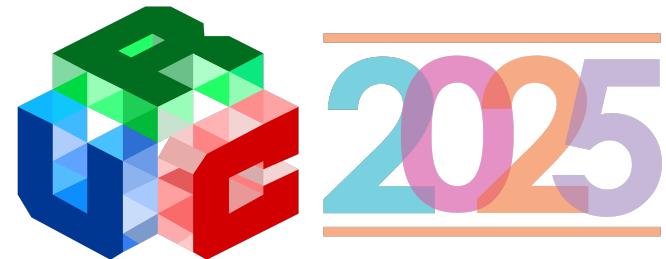
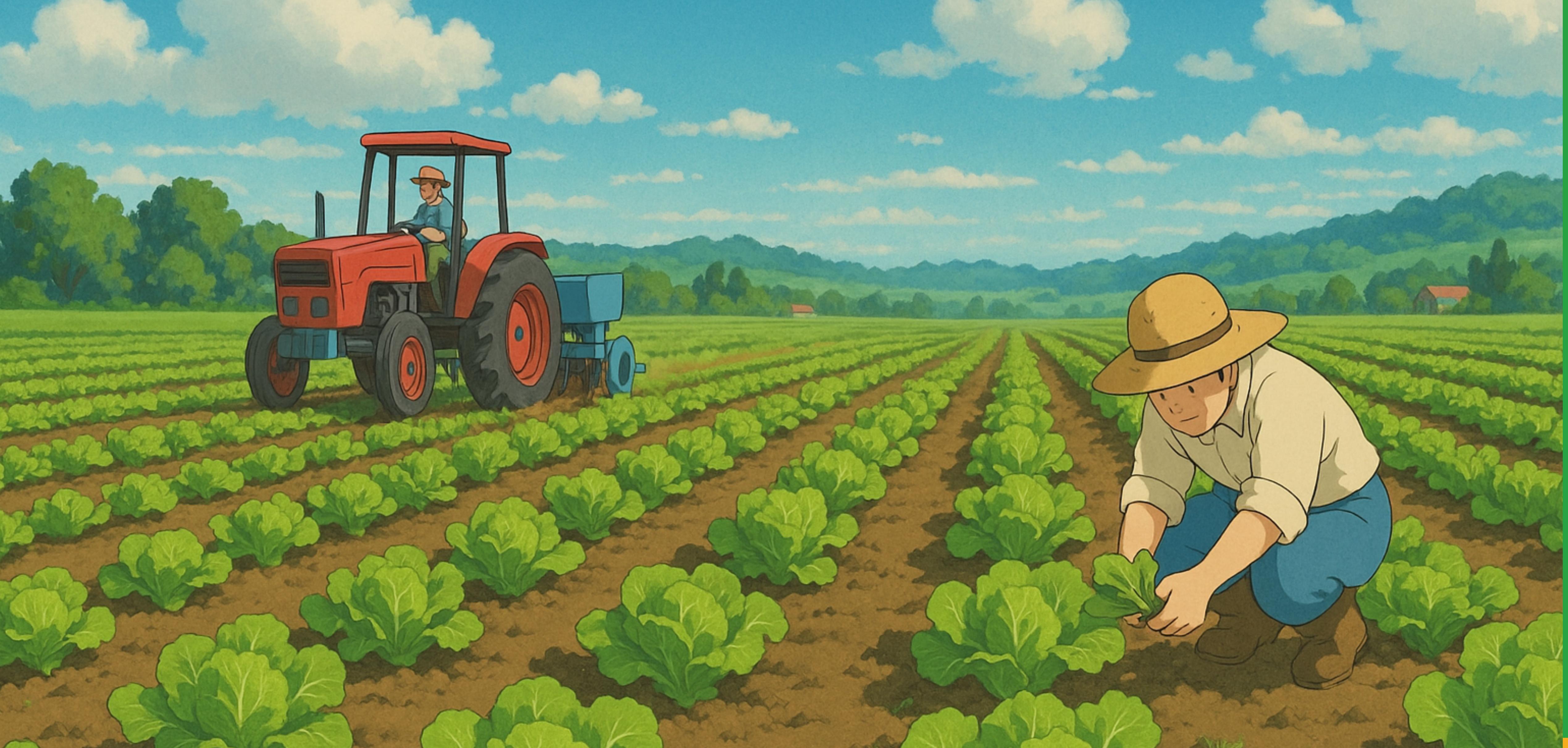


Unified Temporal Consistency and Mean Squared Error Loss Function for Predicting Plant Growth Structures through Multimodal Convolutional Long Short-Term Memory: Laying the Groundwork for Digital Twins in Agriculture

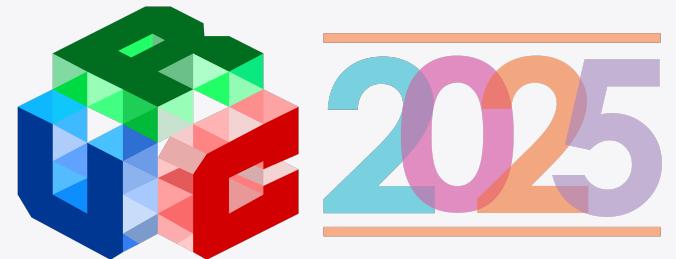
John Ivan T. Diaz, Craig Joseph B. Goc-ong, Kaye Louise A. Manilong
Alvin Joseph S. Macapagal, Philip Virgil B. Astillo*
Department of Computer Engineering, University of San Carlos

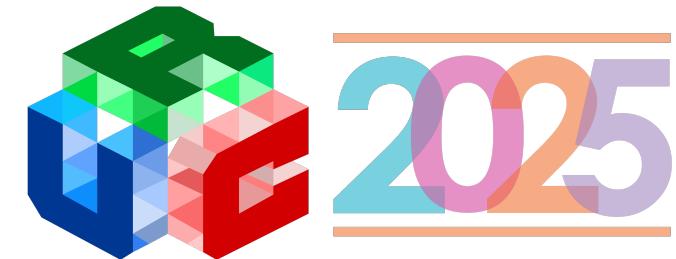


UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO



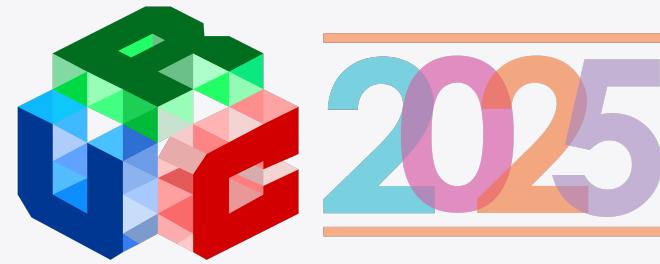
UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO





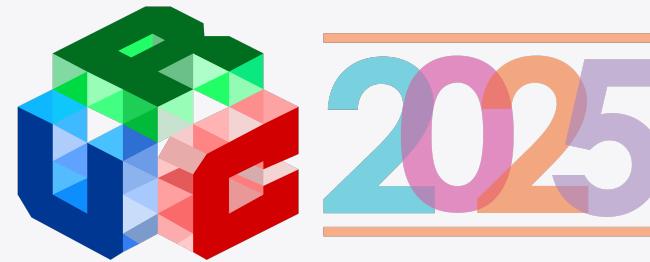
UNIVERSITY *of* SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Digital Twin



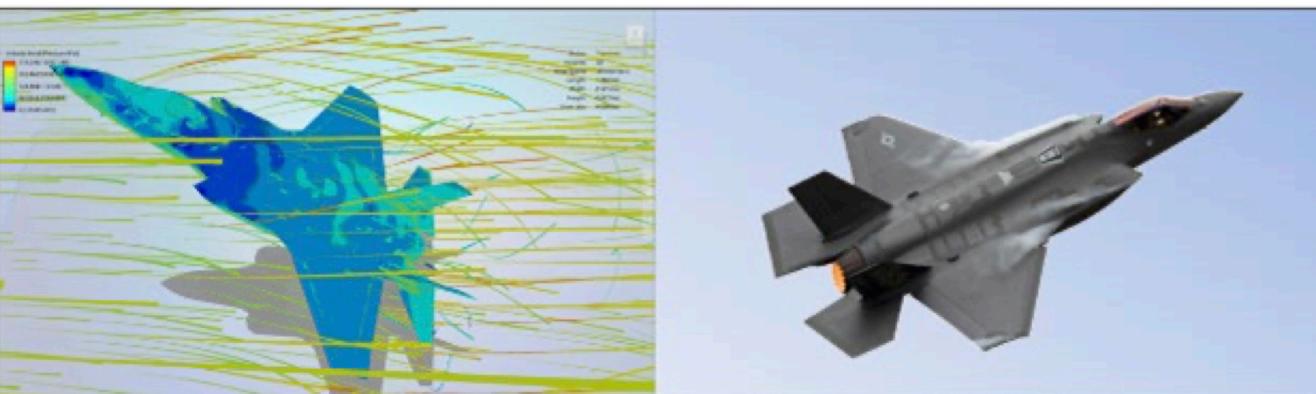
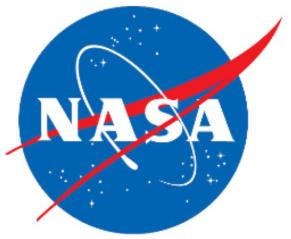
UNIVERSITY *of* SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Digital Twin?



UNIVERSITY *of* SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Persistent “TwinSim”



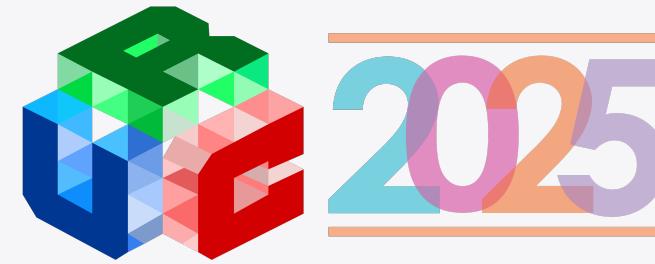
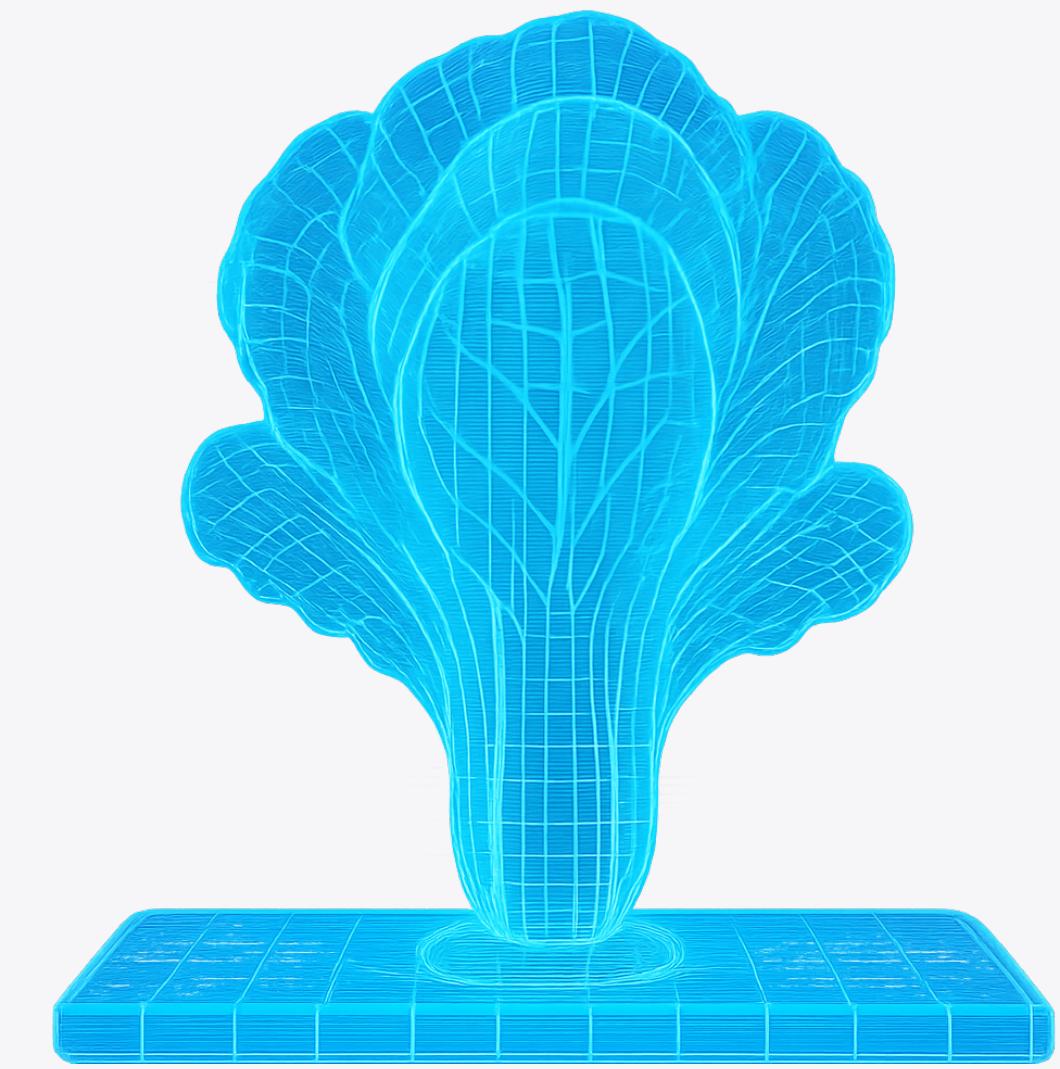
Digital Twin Artifacts/Environments



Credit: Natalia Alexandrov

“Digital Twin” Ecosystem

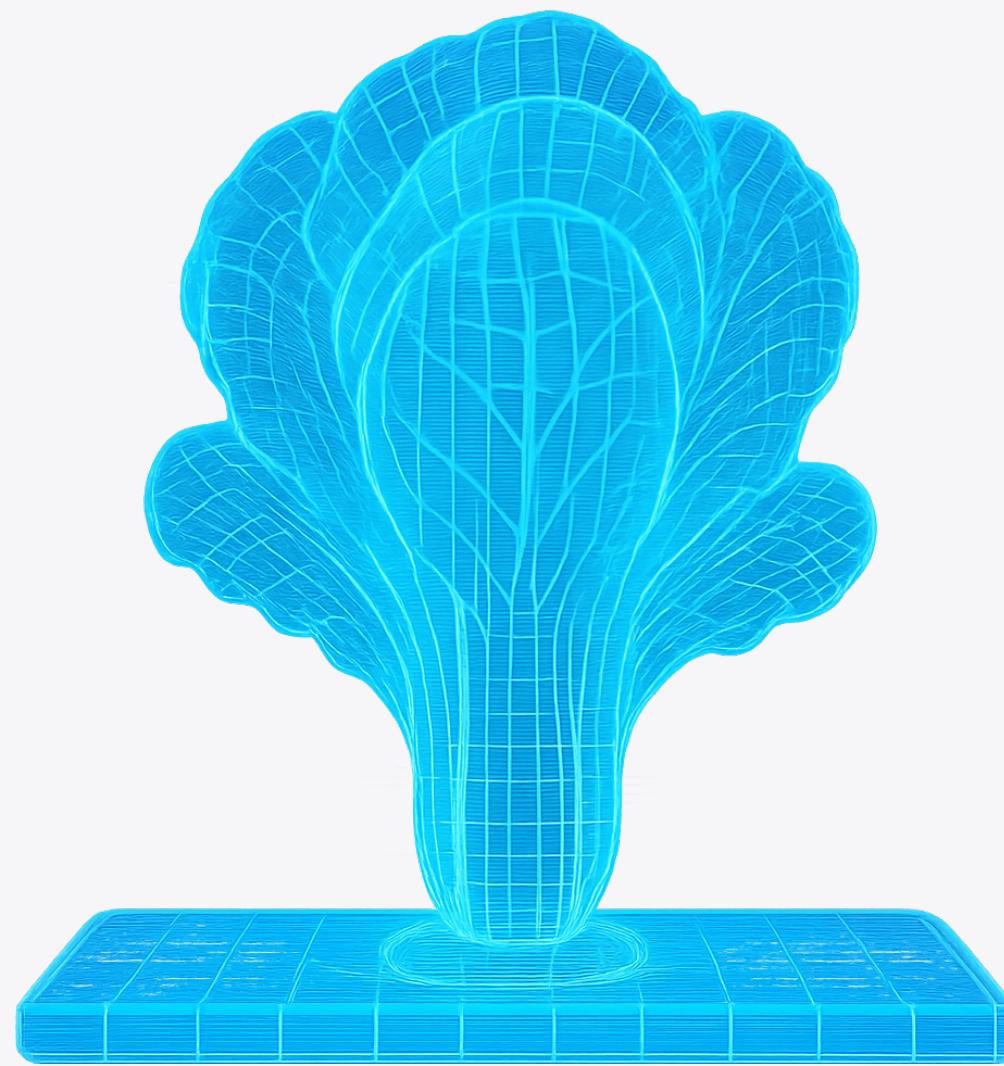




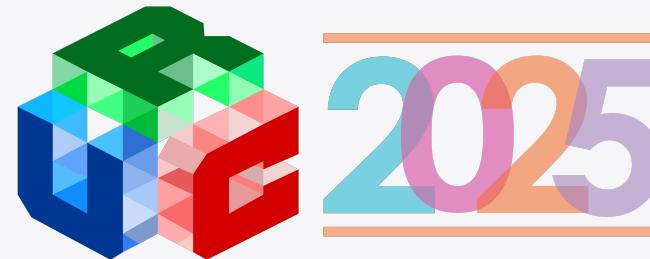


Actual crop

Digital
replica

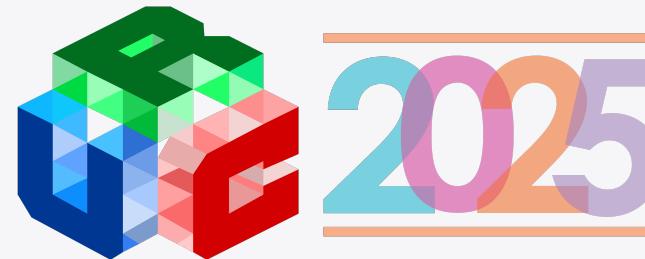


Digital Twin



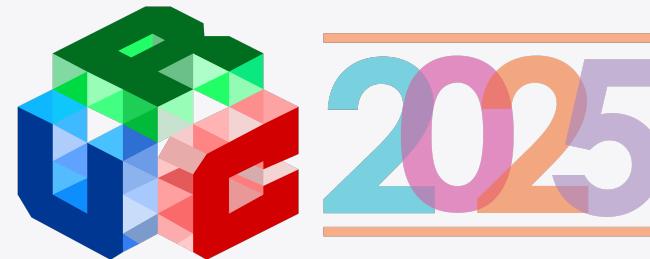
UNIVERSITY *of* SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

How would my plant respond if I change the temperature to X?



UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Will growth stagnate if I change
the pH level to Y?



Current farming practice



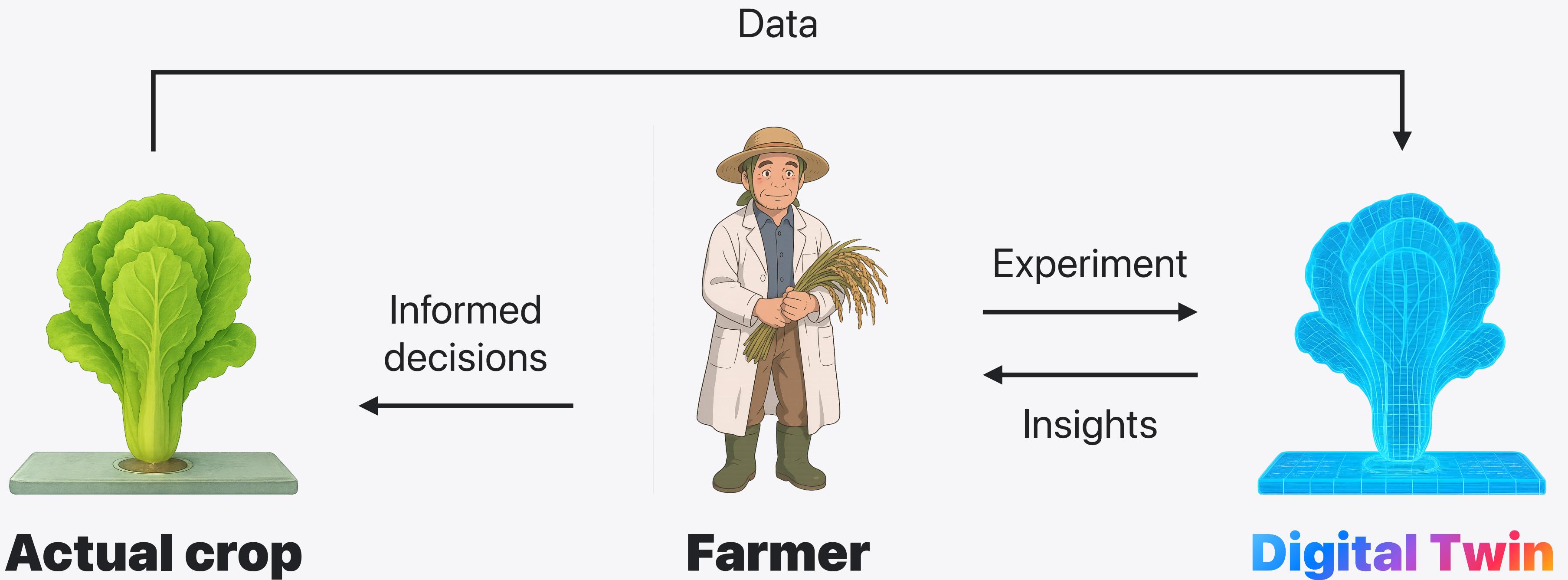
Actual crop

Direct decisions



Farmer

Our vision



Technological impacts

Farmers can make informed agricultural decisions



More Sustainable and Efficient farming

Healthier crops

Higher yields

SUSTAINABLE DEVELOPMENT GOALS

2 ZERO HUNGER



9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

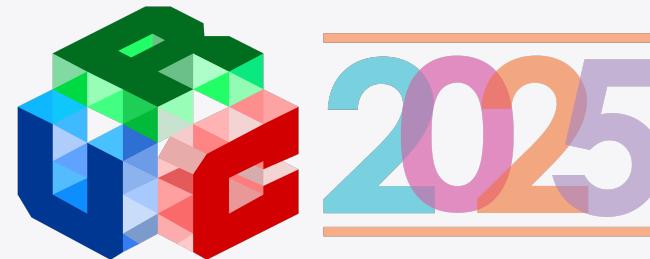


12 RESPONSIBLE CONSUMPTION AND PRODUCTION

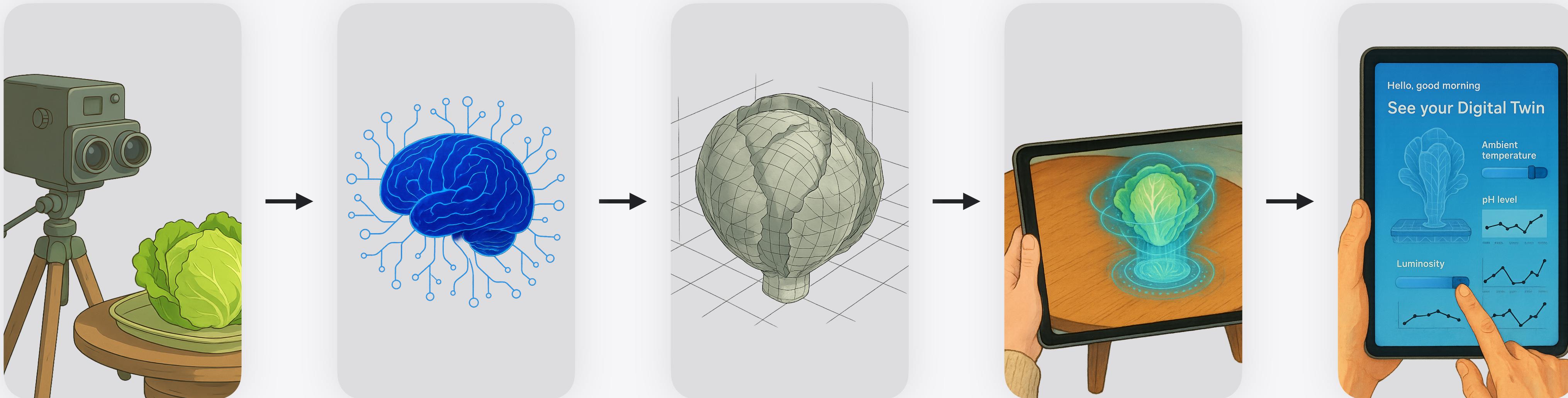


Entrepreneurial opportunities for developers

Laying the groundwork



R&D roadmap



**Data
Collection**

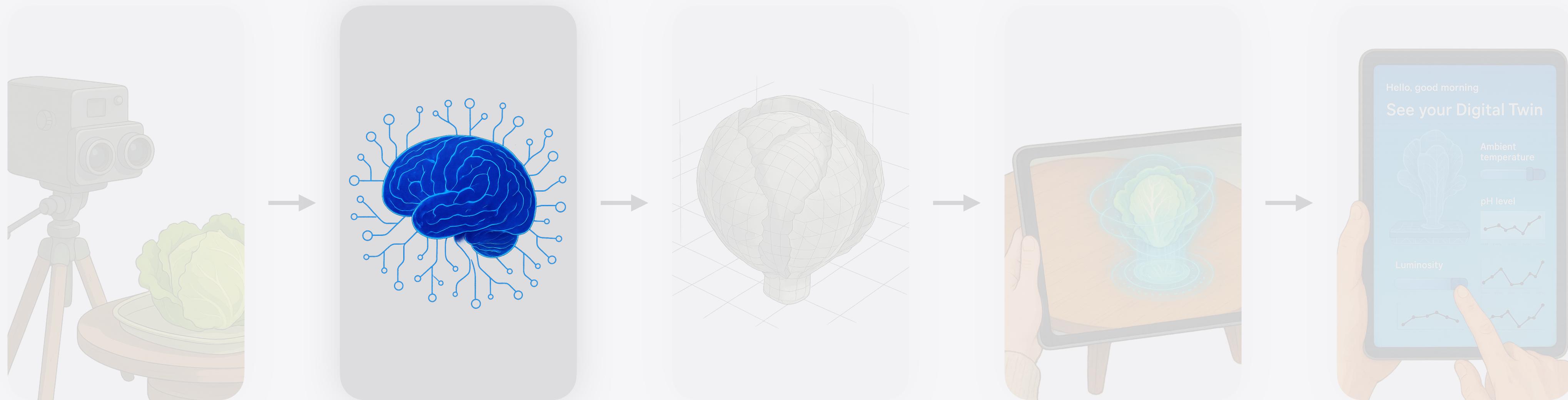
**Artificial
Intelligence**

3D Modeling

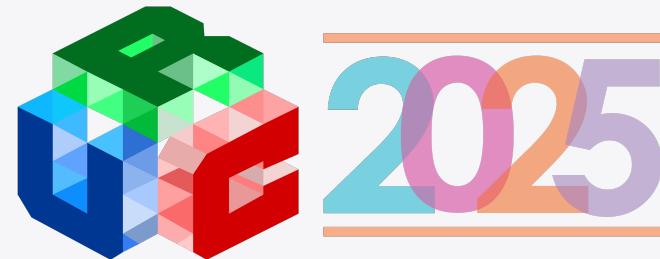
**Augmented
Reality**

**User
Interface**

In this presentation

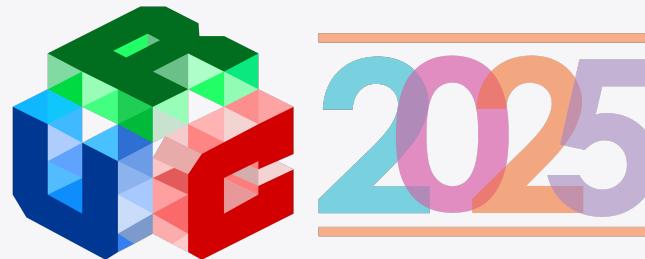


Artificial
Intelligence



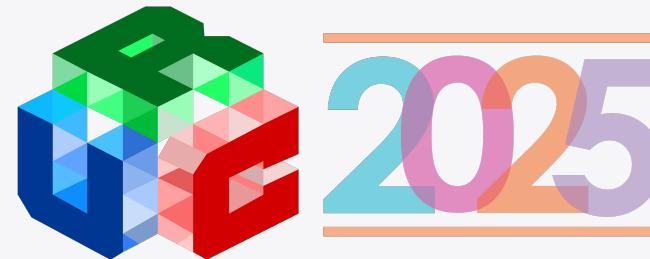
In this presentation

- ① Data collection process
- ② Model architecture
- ③ Training and testing results
- ④ Unified loss function
- ⑤ Conclusions and recommendations

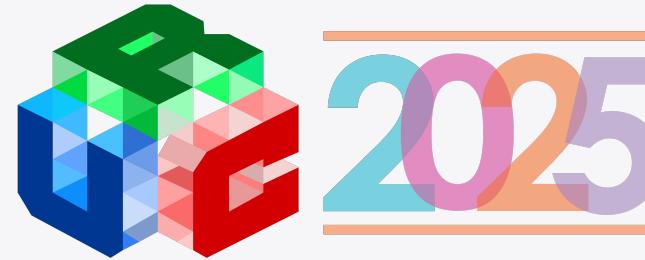


In this presentation

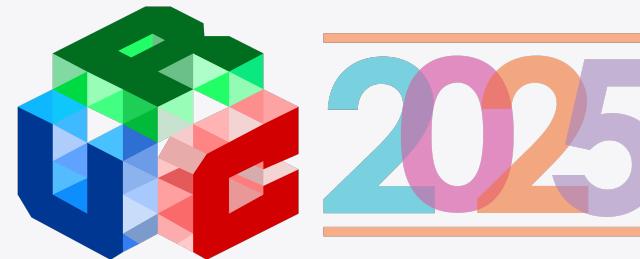
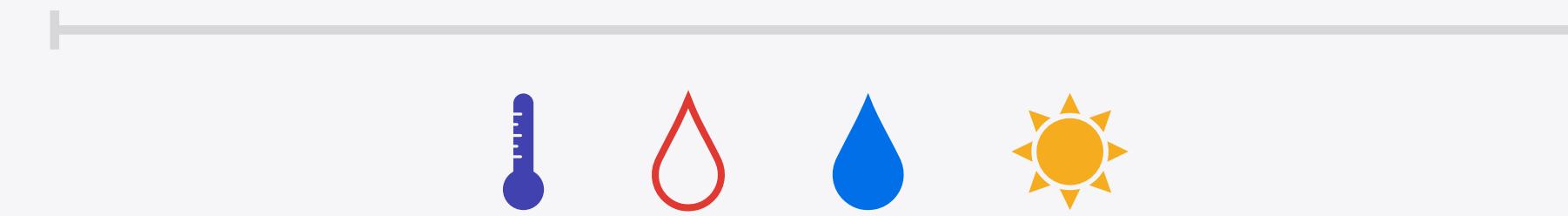
- ① Data collection process



A deep learning model that predicts
future appearances of crops with respect
to external conditions around them



A deep learning model that predicts
future appearances of crops with respect
to external conditions around them



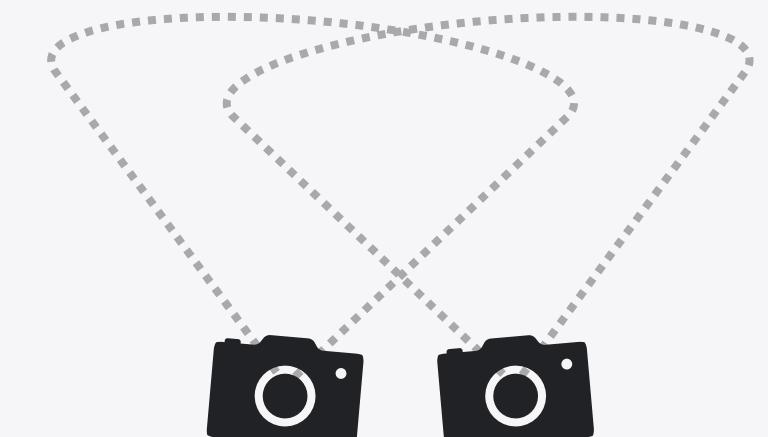
Collecting data

Images



Numerical data

	Ambient temperature	25°C
	pH level	7.5
	Soil moisture	50%
	Luminosity	700 lux



Stereo cameras



Sensors

Collecting data

Frame 1



12 hours

Frame 2

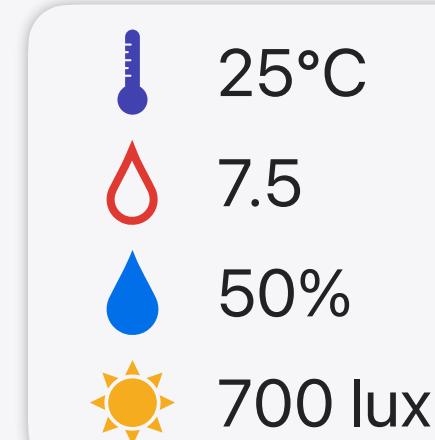
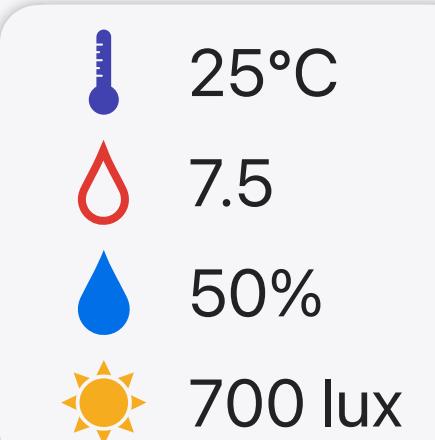


12 hours

Frame 3



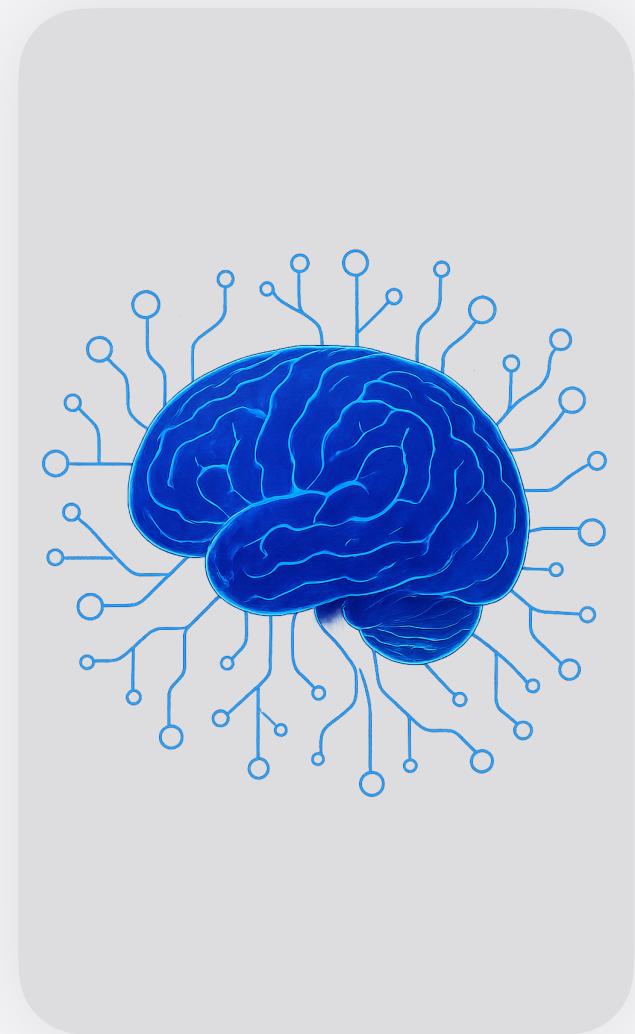
...



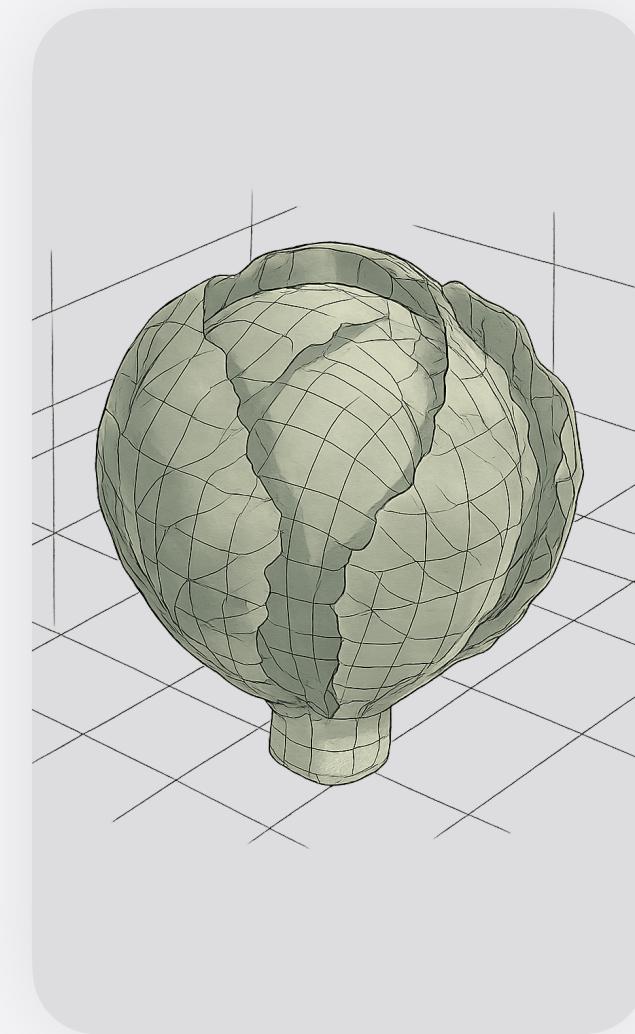
R&D roadmap



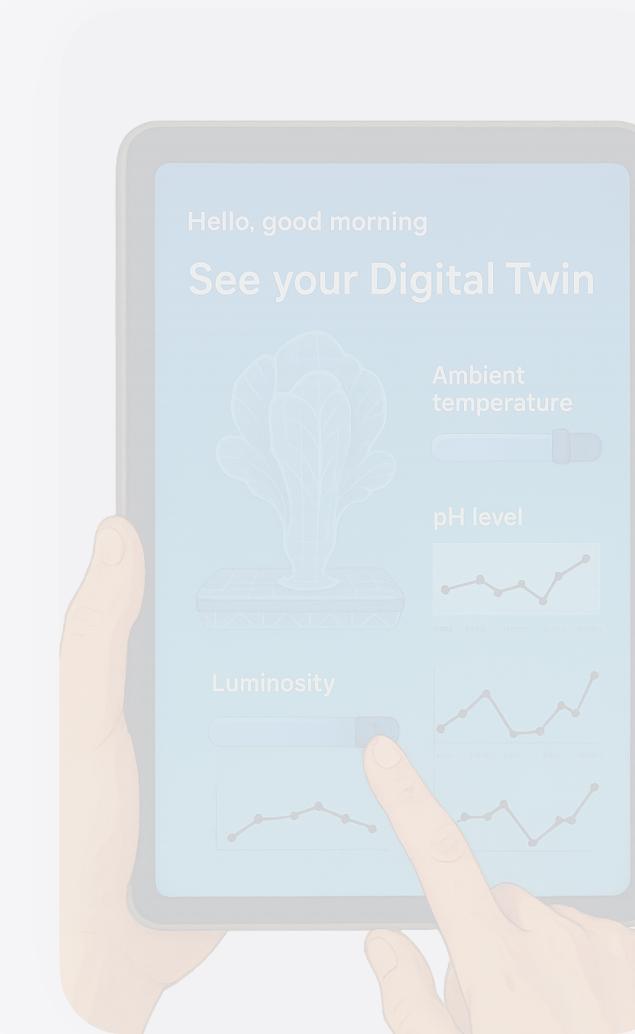
**Data
Collection**



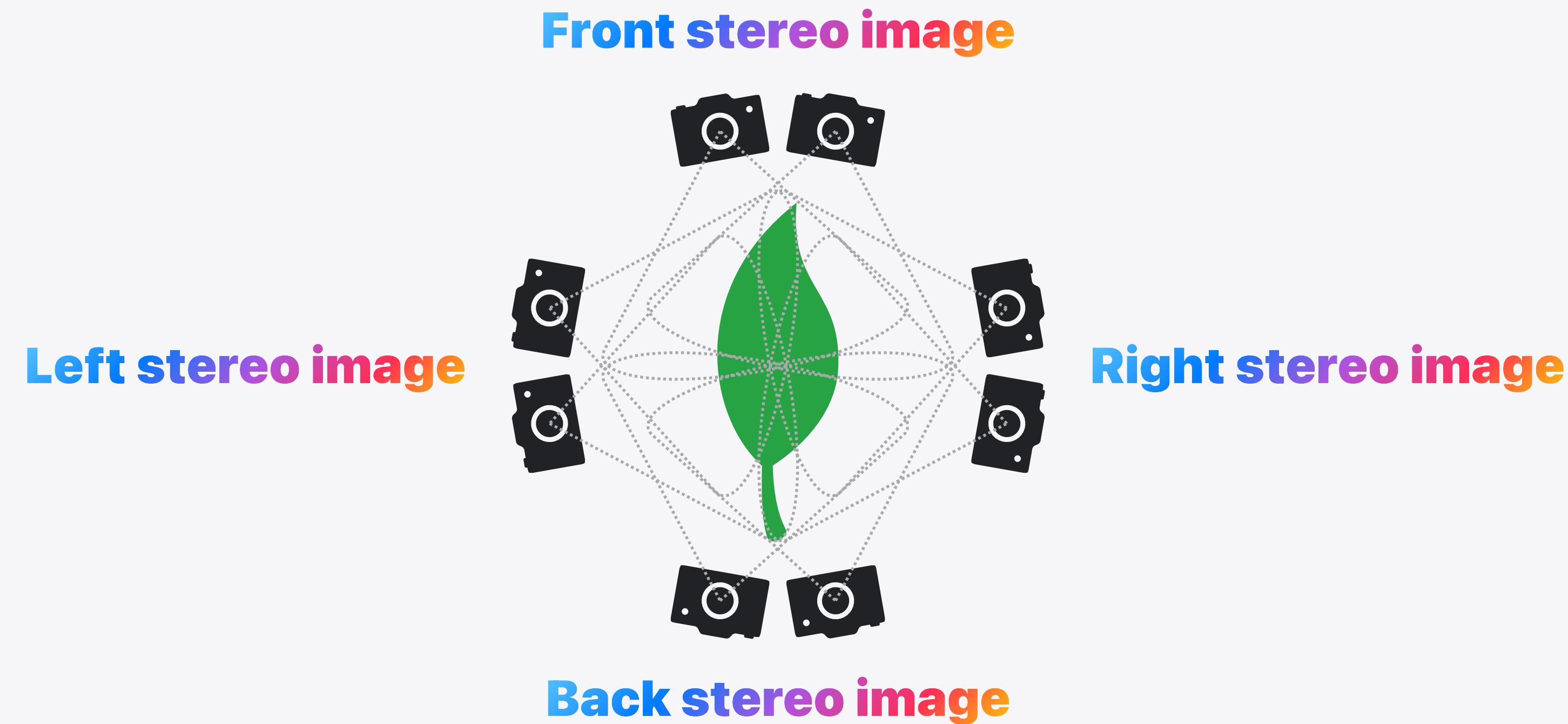
**Artificial
Intelligence**



3D Modeling



Imaging setup



UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

150th SVD 2025

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-on, 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. Current farming practices like vertical farming allow crops to grow in controlled environments, enabling farmers to manipulate external conditions such as pH levels. However, these decisions might not always be favorable to the crops. Applying too little or too much water to a plant can lead to water-stressed, unhealthy crops. Crops can become damaged or experience stunted 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 representation of a physical object or system that is used to monitor and control 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 gain insights and use them to make informed decisions that will benefit their crops. This leads to better water usage for their crops, and higher yields. It redefines modern farming practices to be more sustainable and efficient, ultimately benefiting mankind and the planet.

Promotes SUSTAINABLE GOALS
2 ZERO HUNGER, 9 INDUSTRY INNOVATION, AND INFRASTRUCTURE, 12 RESPONSIBLE CONSUMPTION AND PRODUCTION

Farmers can make informed agricultural decisions

The Technological Impacts

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.

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

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.

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.

Investigating Segmentation Architectures Trained Solely for the Plant Class

Correct segmentation of plant objects positively affects succeeding phases such as 3D modeling. YOLOv6 and Detectron2 are the initial architectures chosen for comparison.

Sample of whole process

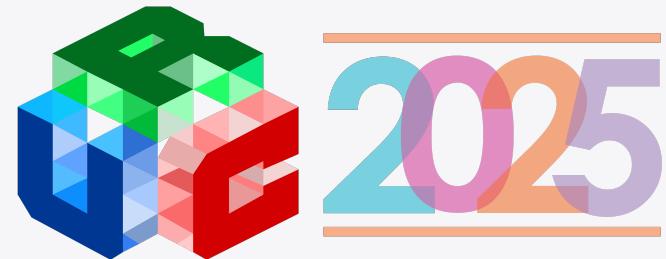
Plant's past appearances Images

External conditions associated Numerical data

Day 1, Day 1.5, Day 15, Day 15.5, Day 15.5 map, Day 15.5 cloud, Day 15.5 point cloud, Day 15.5 3D model, Day 15.5 future appearance

Recommendations

- Continue collecting data. Grow more plants, expand their 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.
- 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.



UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Collecting data

Frame 1

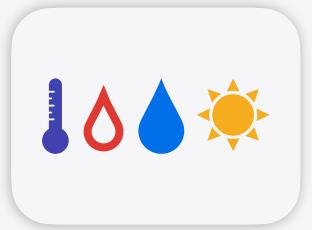
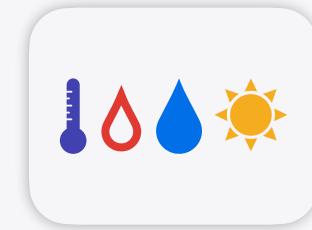
Frame 2

Frame 3

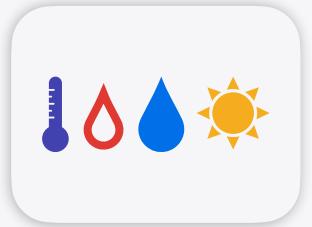
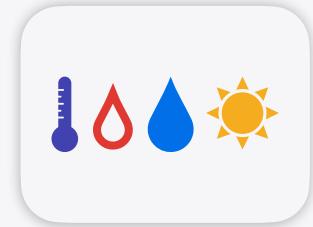
Frame 4

...

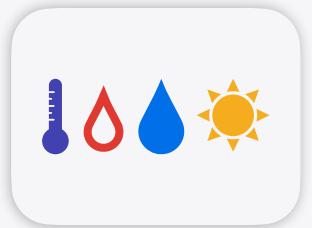
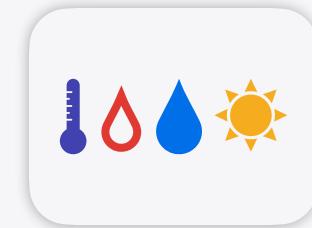
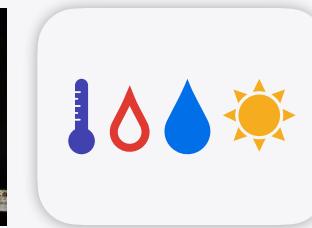
Sequence 1



Sequence 2

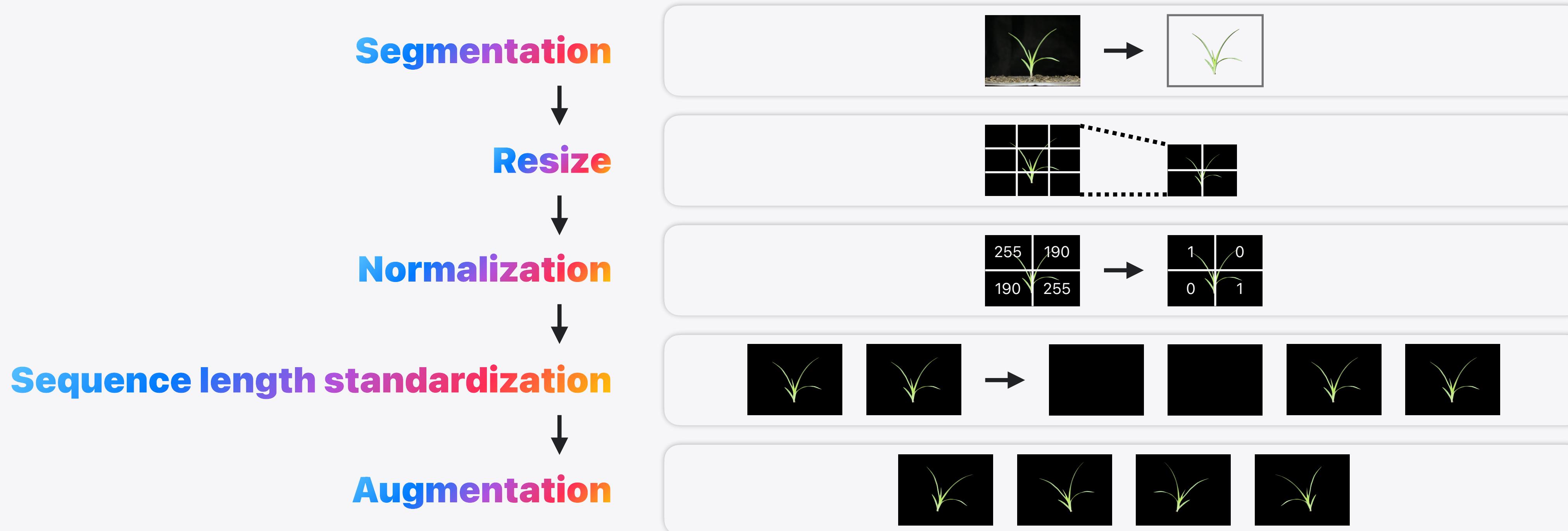


Sequence 3



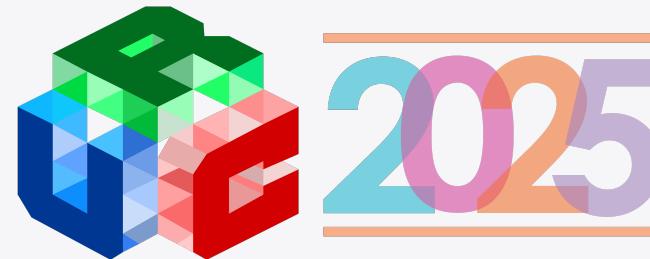
...

Preprocessing pipeline



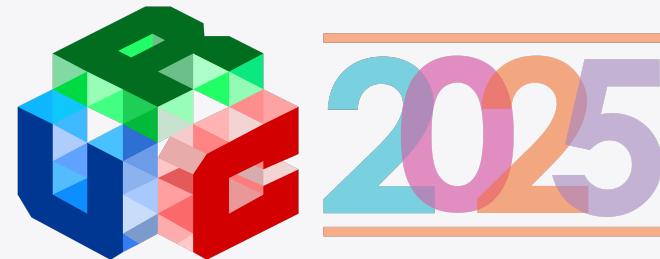
In this presentation

- ① Data collection process

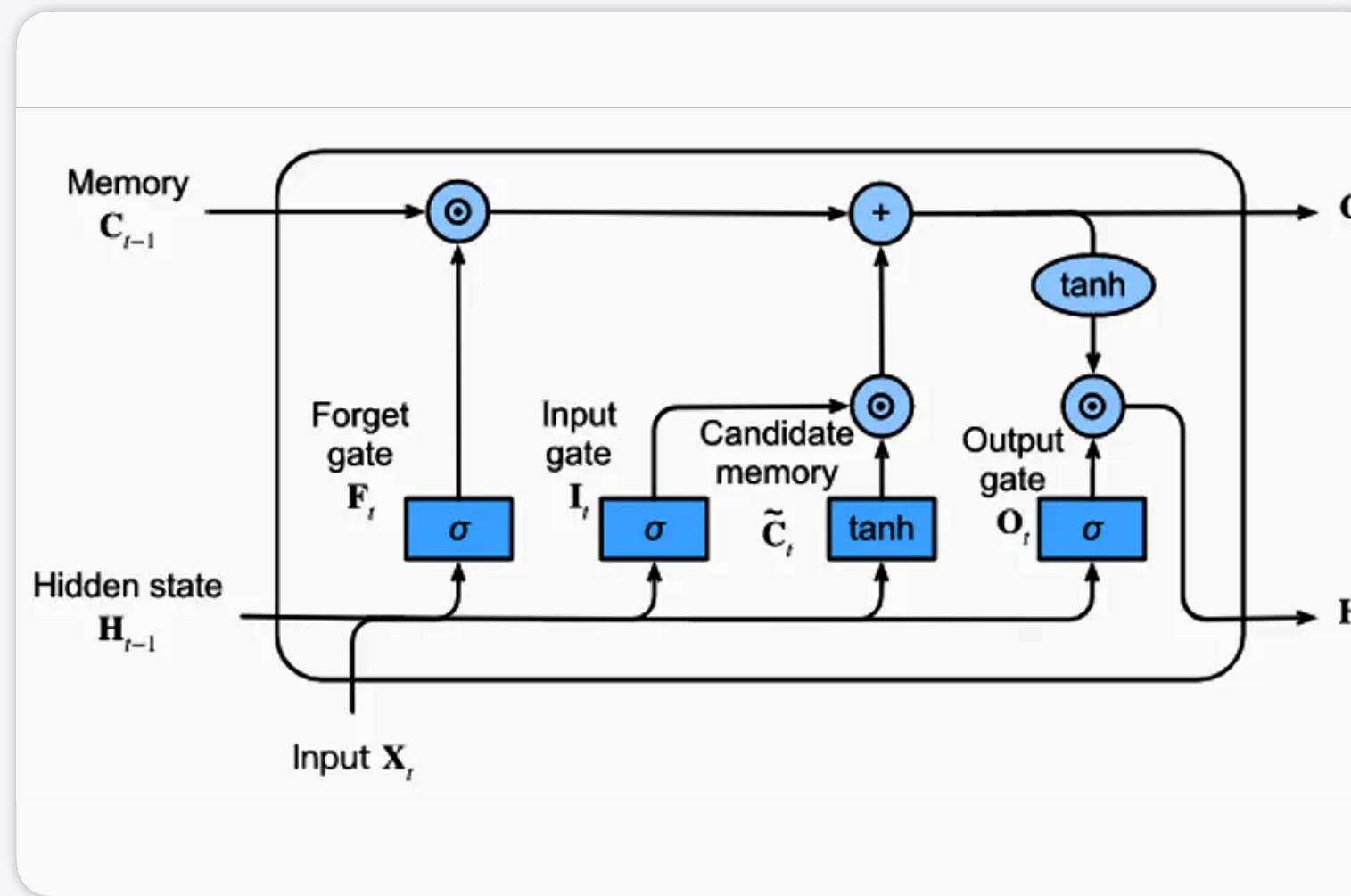


In this presentation

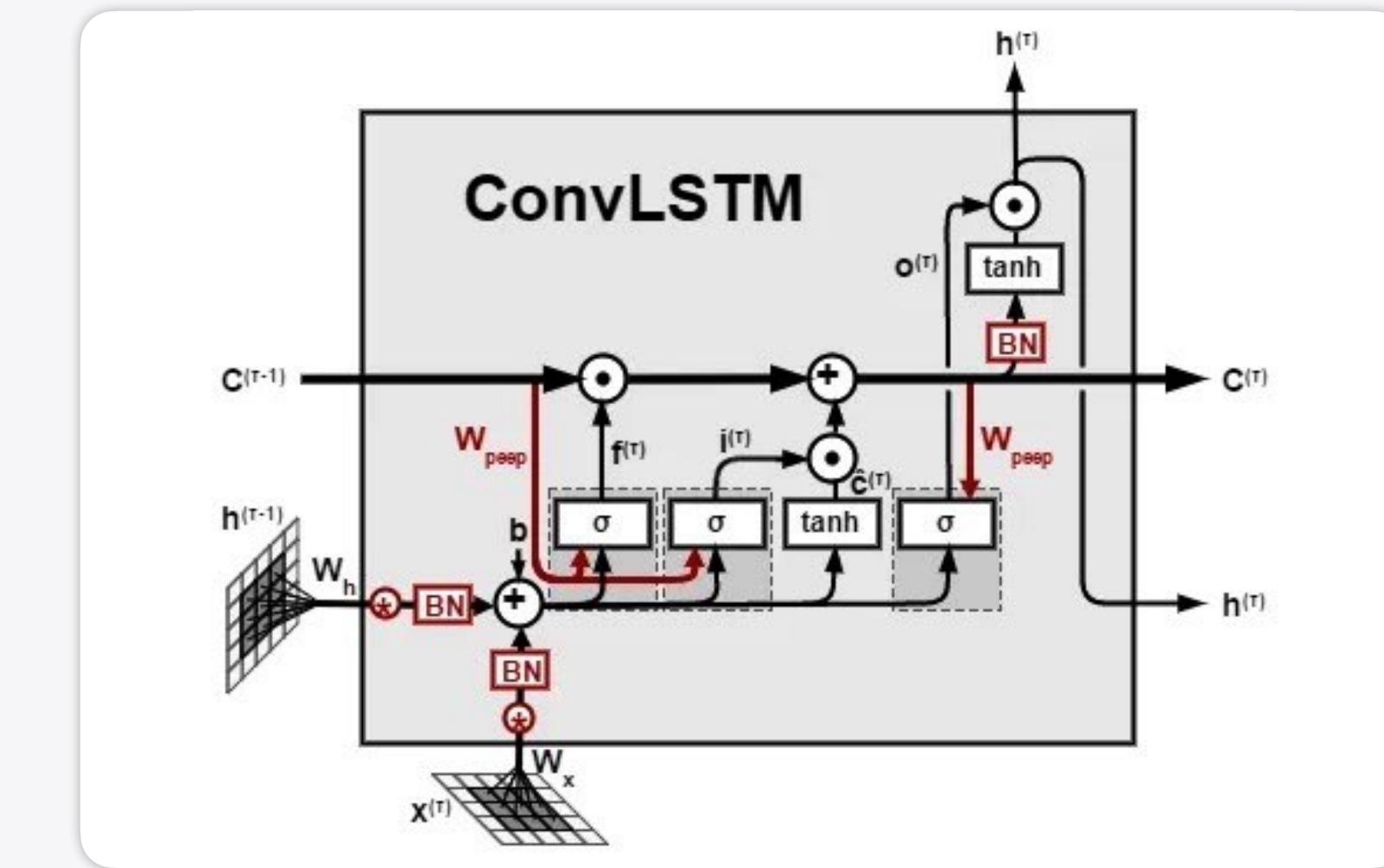
② Model architecture



Recent deep learning advancements

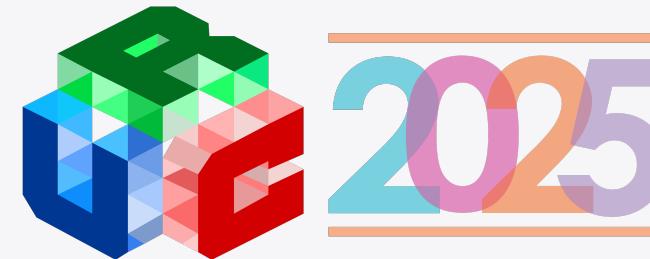


Traditional LSTM



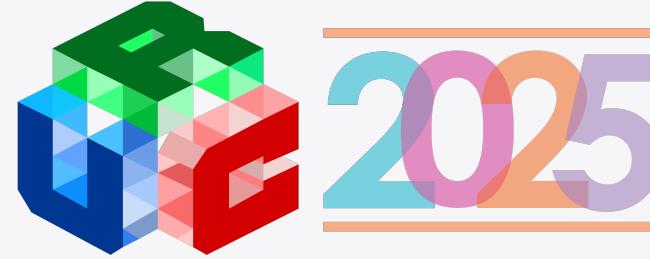
Convolutional LSTM

Convolutional LSTM



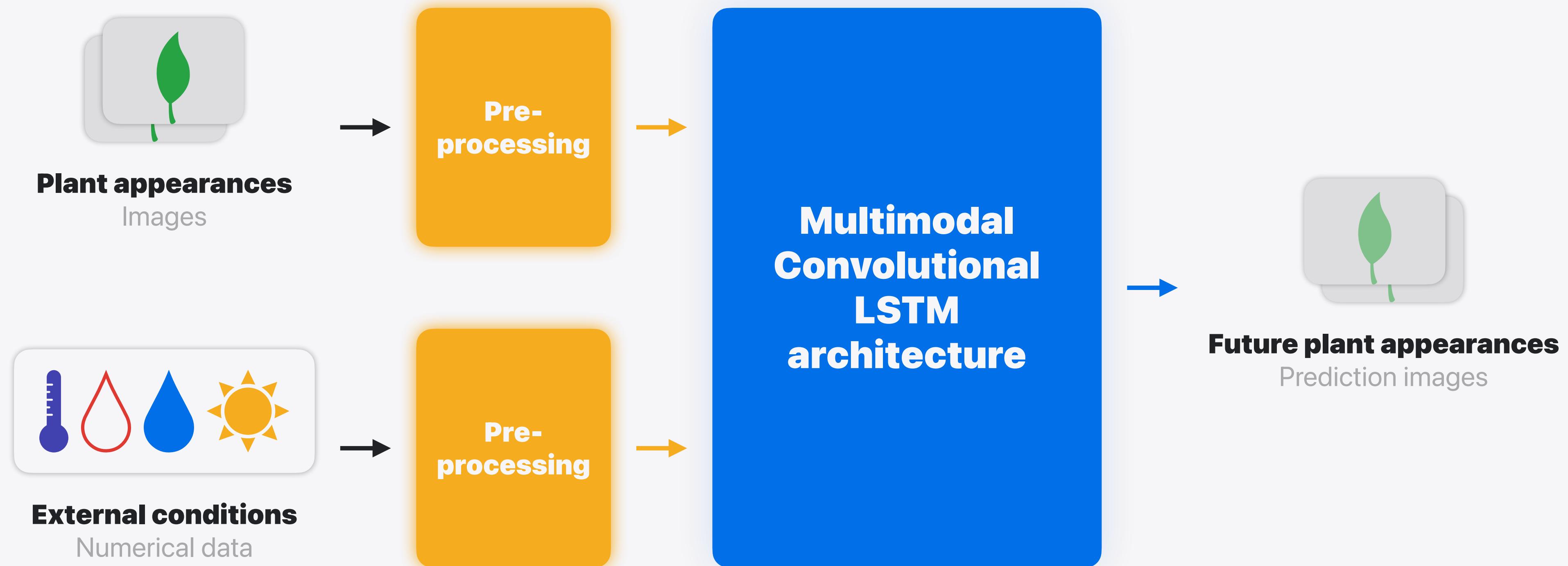
UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Multimodal Convolutional LSTM

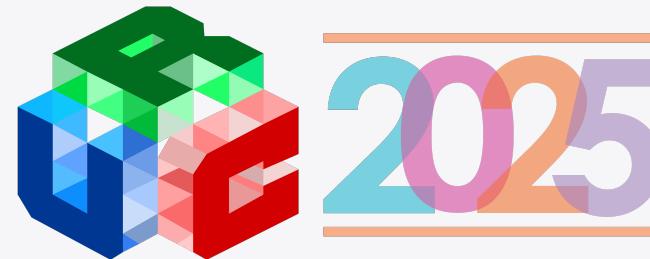
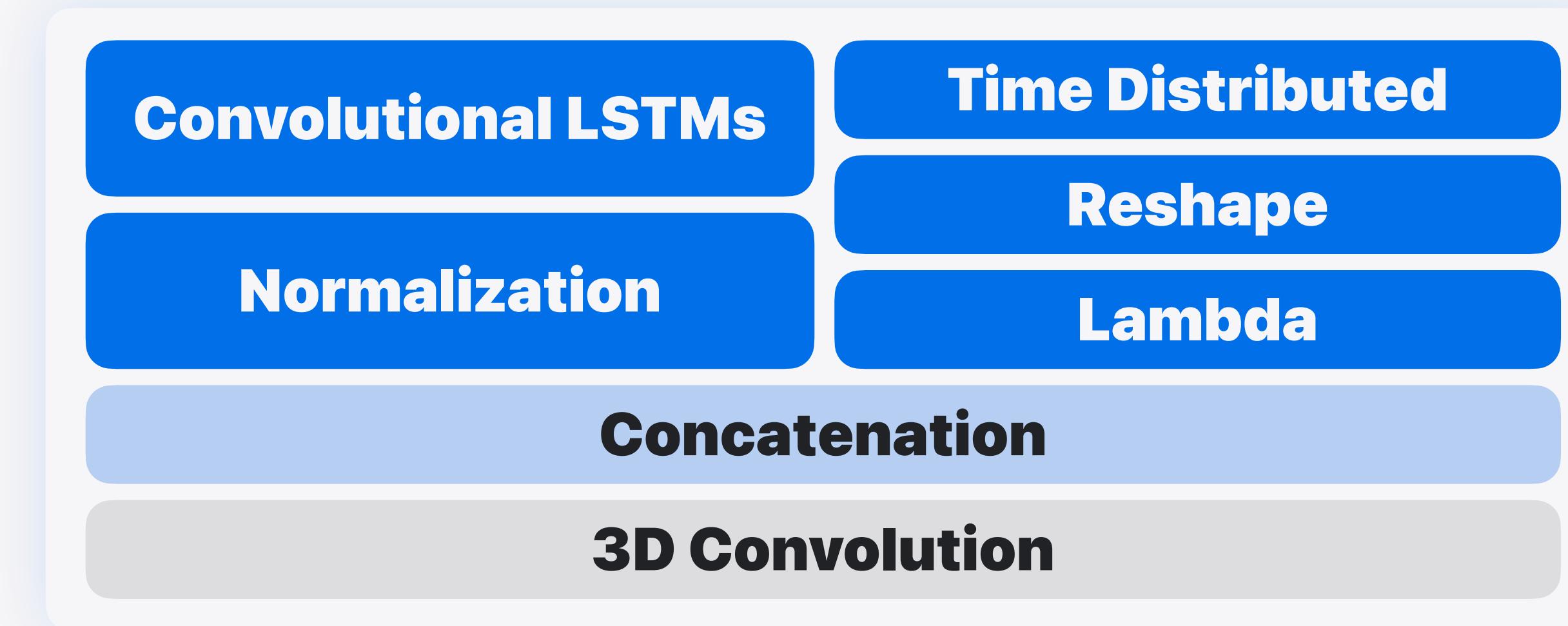


UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Multimodal learning

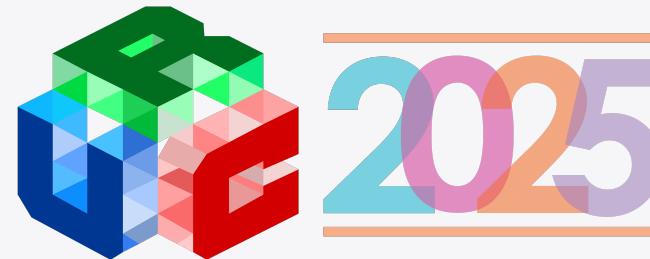


Model architecture



In this presentation

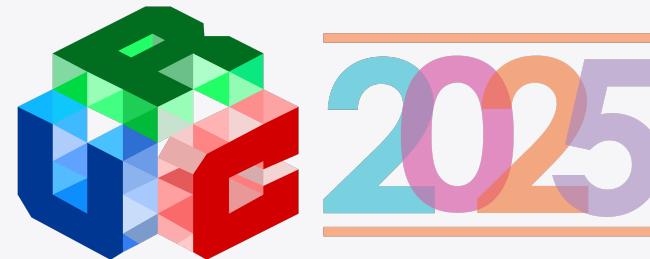
② Model architecture



In this presentation

③

Training and testing results



UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

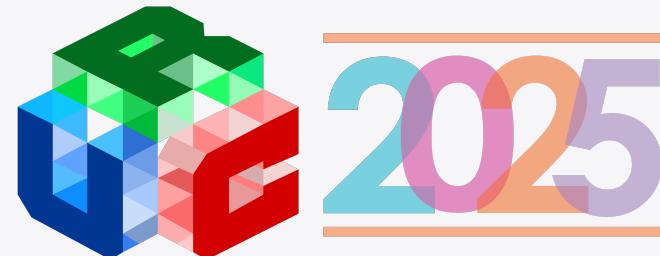
Evaluation

MSE
model

Model trained with
Mean Squared Error
loss function

TC
model

Model trained with
Temporal Consistency
loss function



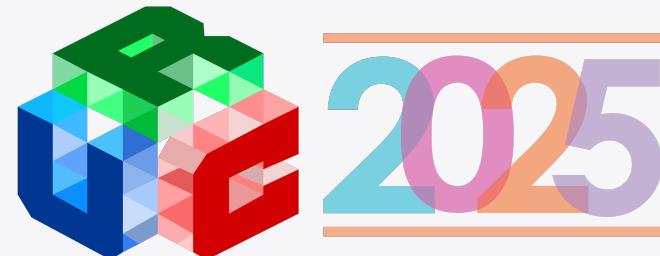
Evaluation

MSE
model

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (x_{s,L,i} - \hat{x}_{s,L,i})^2$$

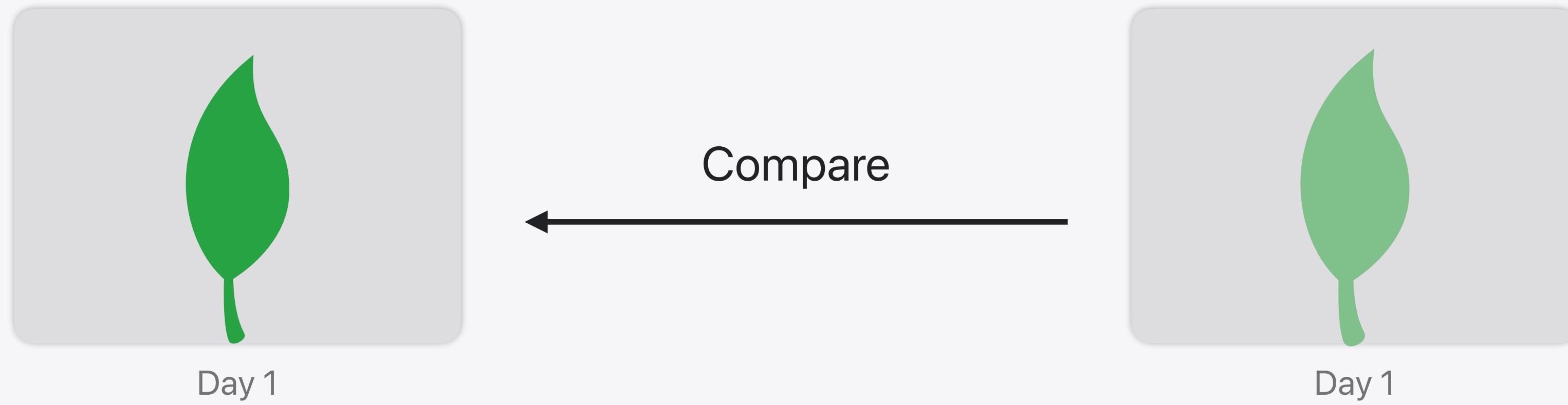
TC
model

$$L_{\text{TC}} = \frac{1}{N(T_S - 1)} \sum_{t=2}^{T_S} \sum_{i=1}^N [(x_{s,t,i} - x_{s,t-1,i}) - (\hat{x}_{s,t,i} - \hat{x}_{s,t-1,i})]^2$$



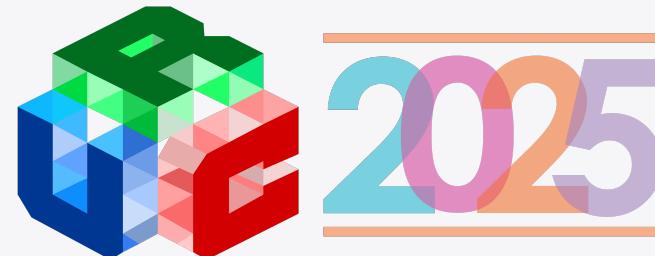
UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

Evaluation



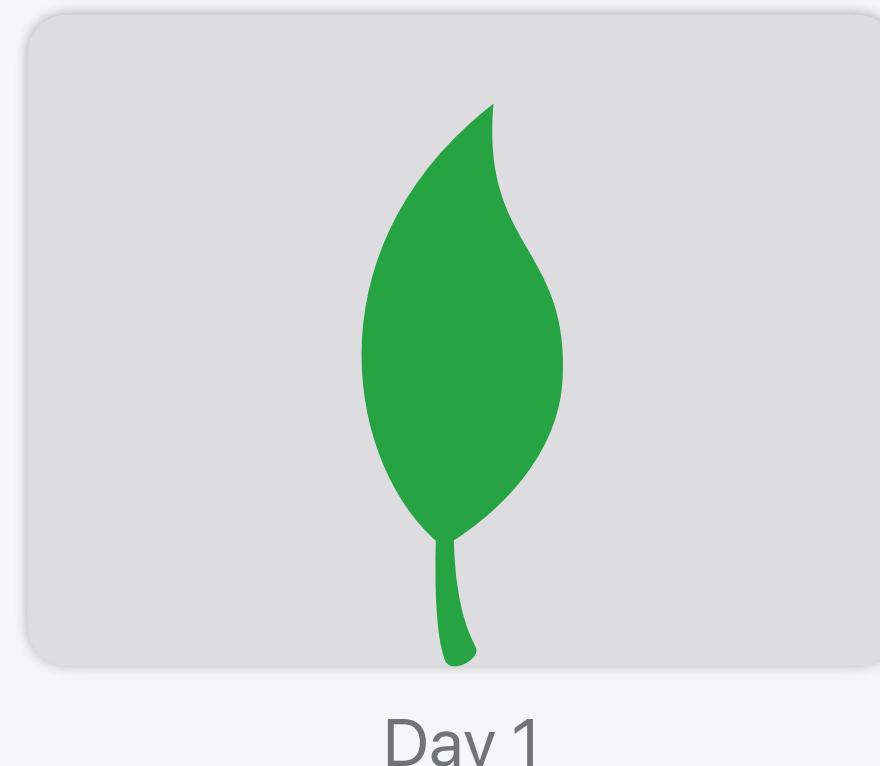
Actual image
from the dataset

Predicted image
by the model



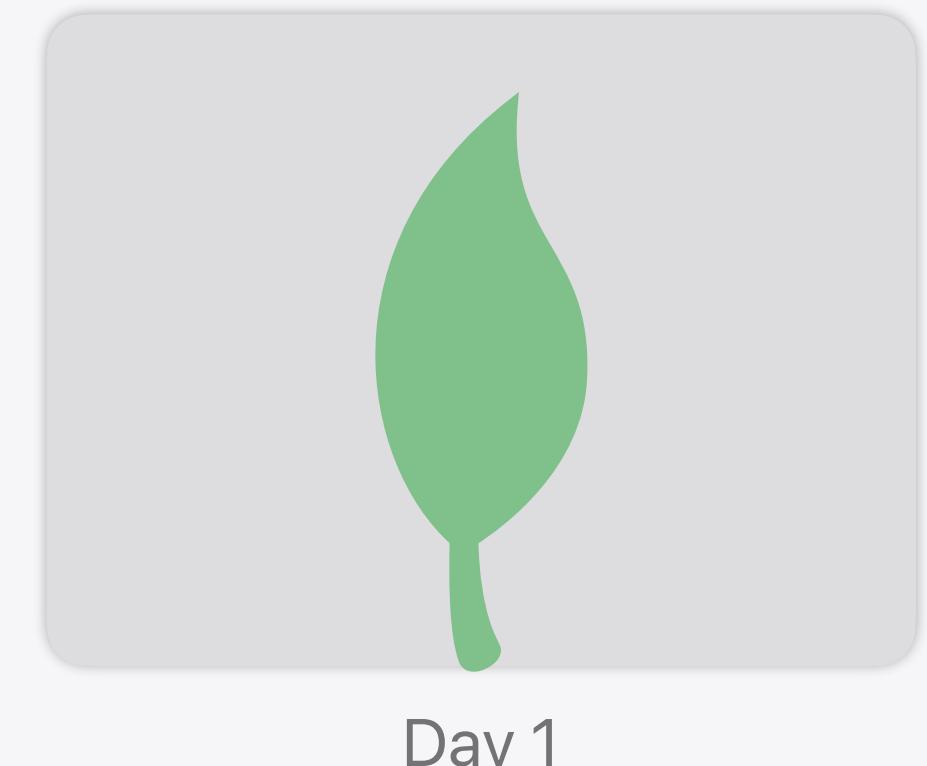
Evaluation

Metrics



Actual image
from the dataset

Mean Squared Error
Peak Signal-to-Noise Ratio
Structured Similarity Index
Temporal Consistency
Total Variation
Visual inspection

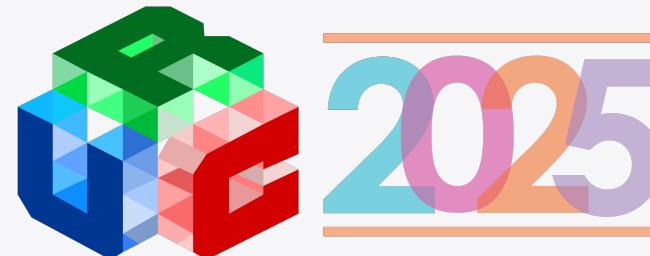


Predicted image
by the model

Evaluation

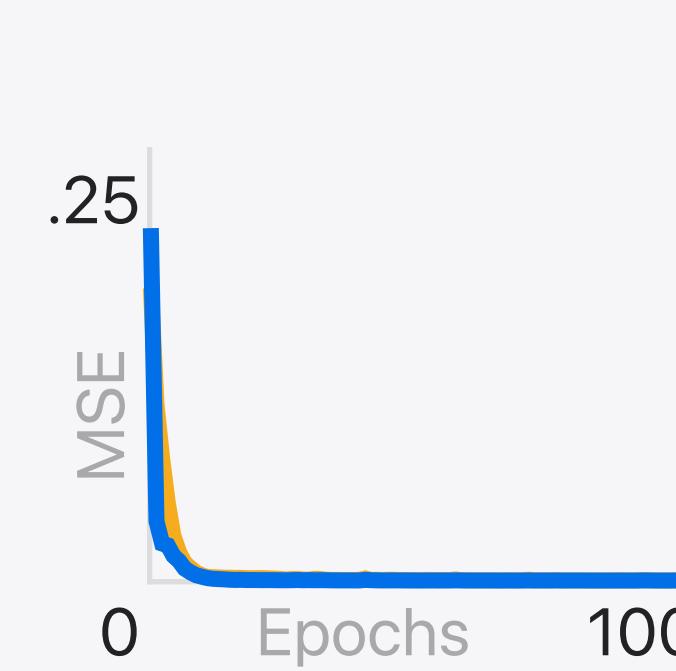
Quantitative metrics

Bad		Good
Greater than 0	MSE	0
Below 40 dB or above 50 dB	PSNR	Within 40-50 dB
Lesser than 1	SSIM	1
Greater than 0	TC	0
Greater than 0	TV	0



Training results

Metric curves of **MSE model**



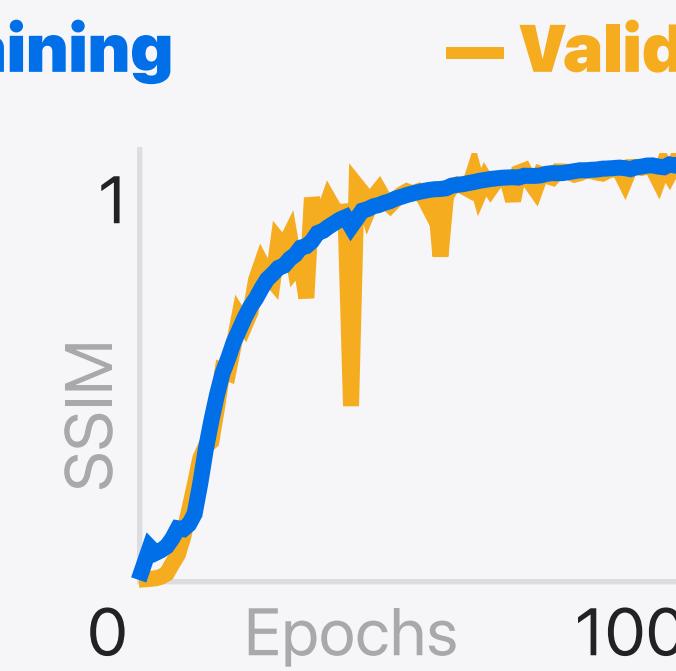
MSE

✓ Met good benchmarks



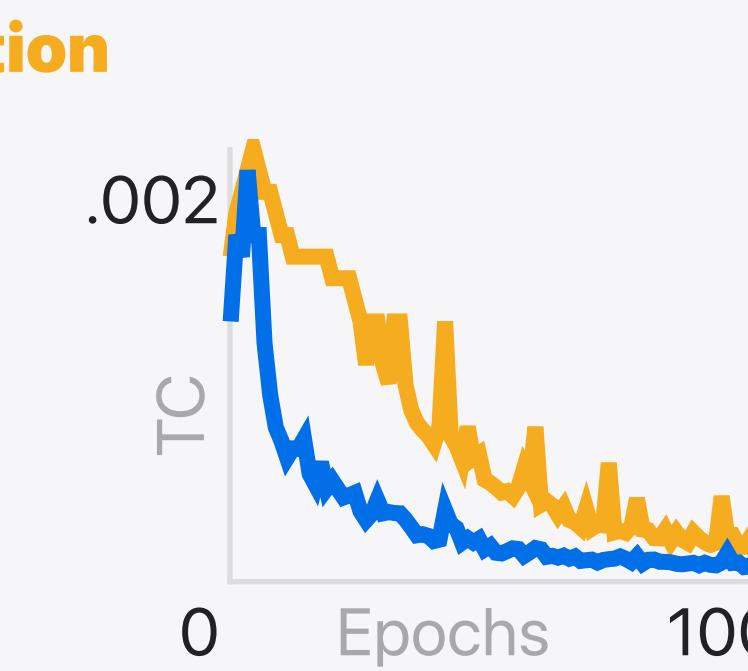
PSNR

✓ Met good benchmarks



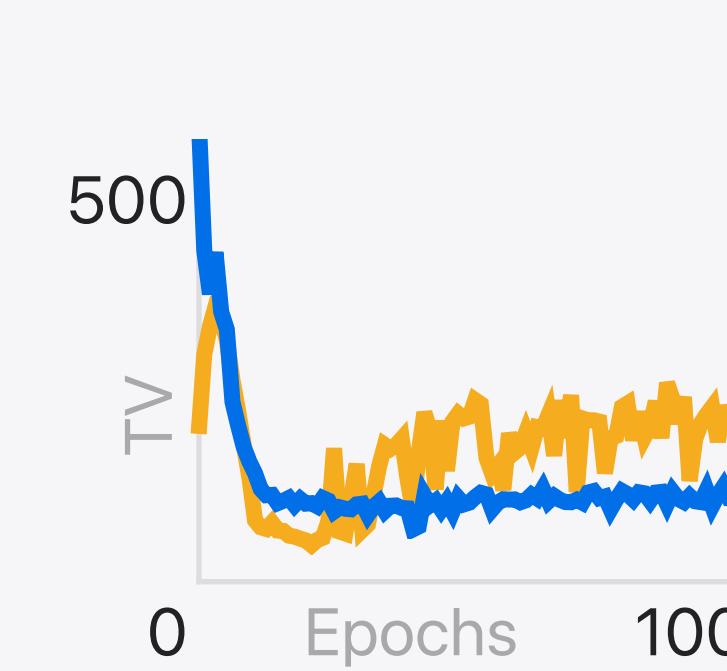
SSIM

✓ Met good benchmarks



TC

✓ Met benchmarks but slower than TC model

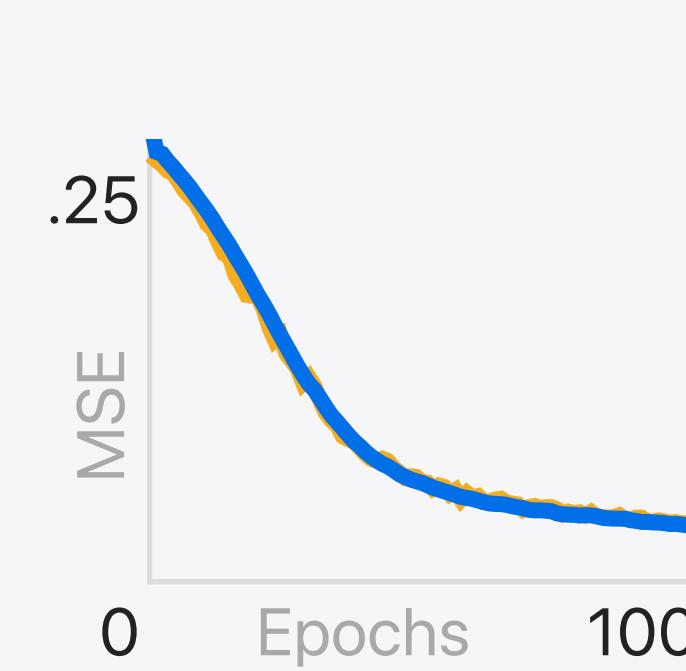


TV

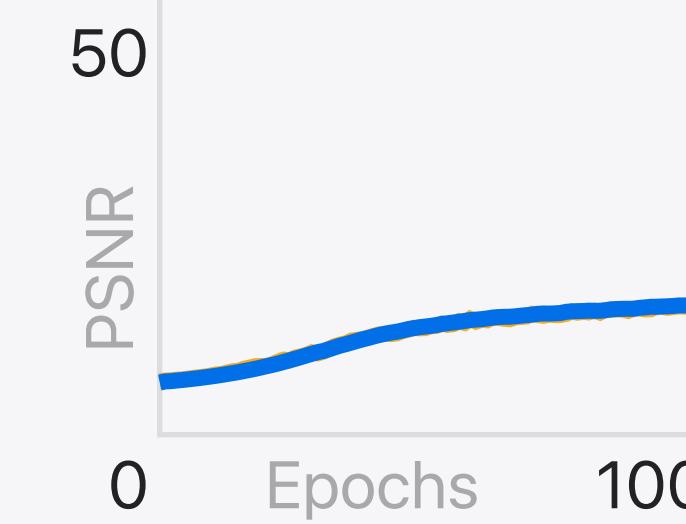
✓ Met good benchmarks

Training results

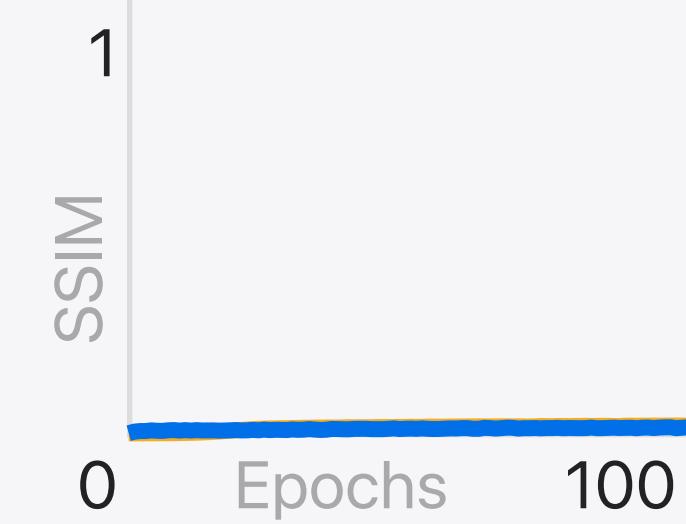
Metric curves of **TC** model

**MSE**

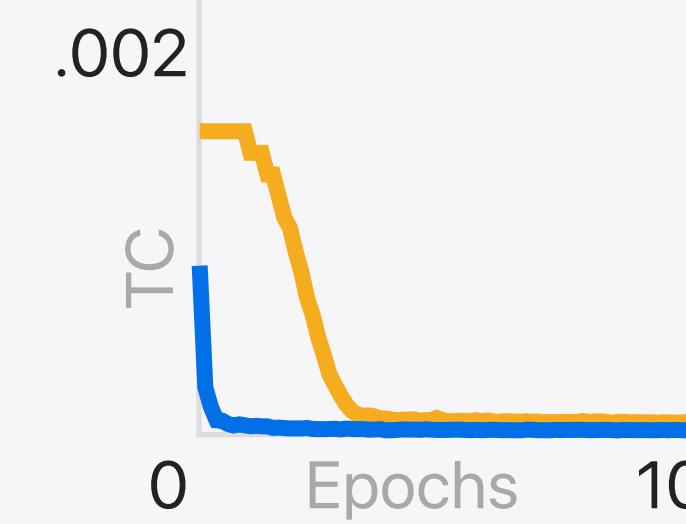
✖ Did not meet good benchmarks

**PSNR**

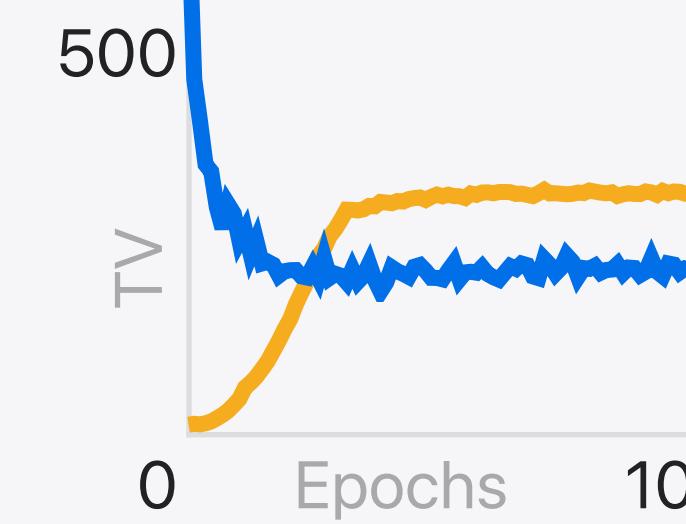
✖ Did not meet good benchmarks

**SSIM**

✖ Did not meet good benchmarks

**TC**

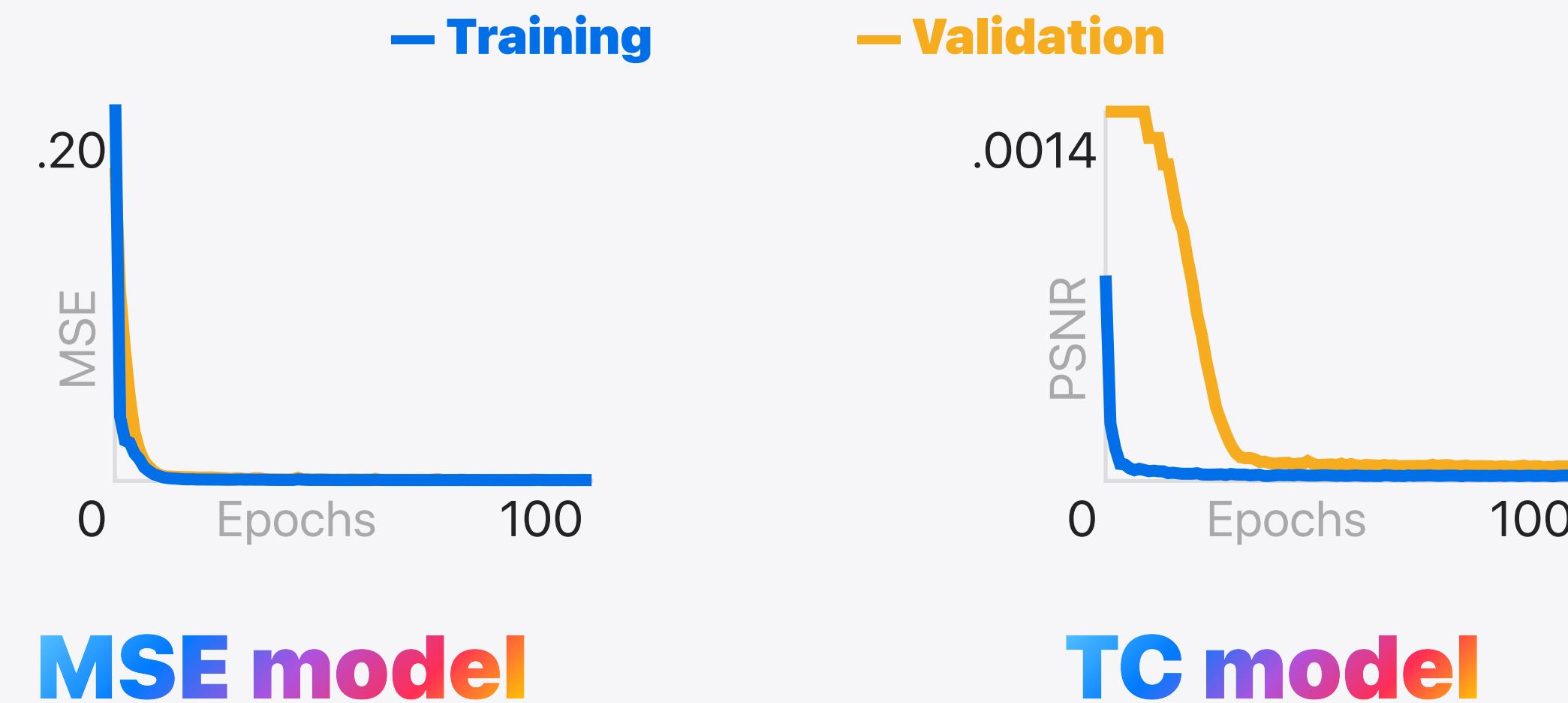
✓ Met good benchmarks

**TV**

✖ Slightly worse than MSE model

Training results

Loss curves



⚠ Early convergence

Test results

Quantitative metrics

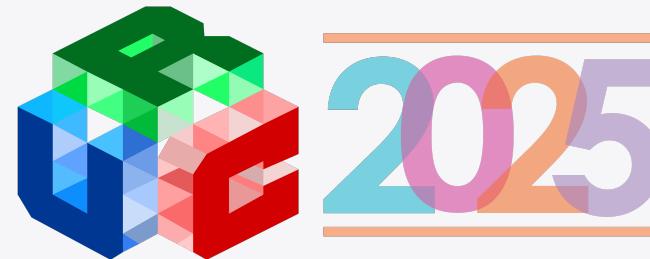
	MSE model	TC model
MSE ★ Lower is better	0.0001 ★	0.0326
PSNR ★ Higher is better	45.1848 dB ★	14.8740 dB
SSIM ★ Higher is better	0.9633 ★	0.0162
TC ★ Lower is better	0.0001	0.0000 ★
TV ★ Lower is better	148.1984 ★	239.4765

MSE model

Model trained with
Mean Squared Error
loss function

TC model

Model trained with
Temporal Consistency
loss function



Unified loss function

MSE
model

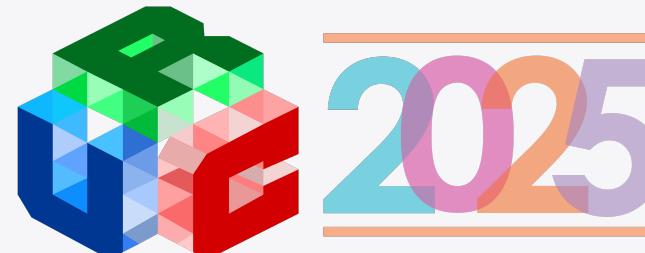
Model trained with
Mean Squared Error
loss function

TC
model

Model trained with
Temporal Consistency
loss function

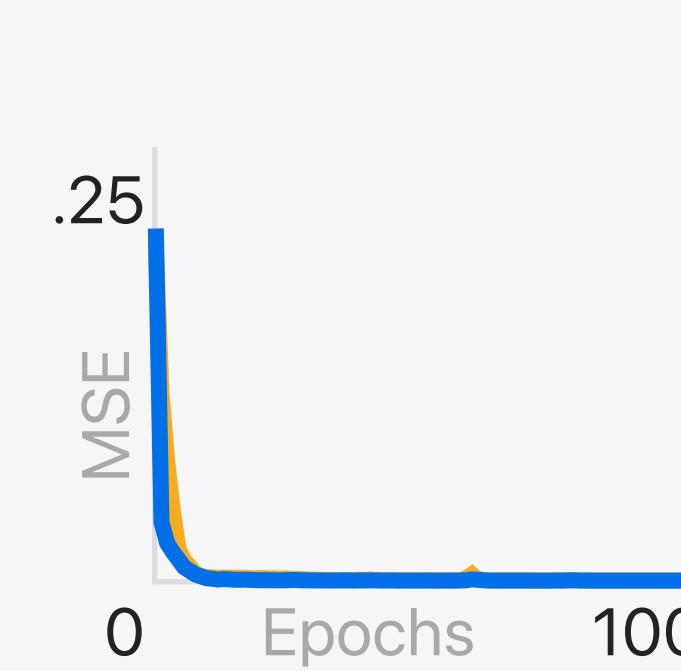
MSE+TC
model

Model trained with
unified MSE and TC
loss functions



Training results

Metric curves of **Unified model**



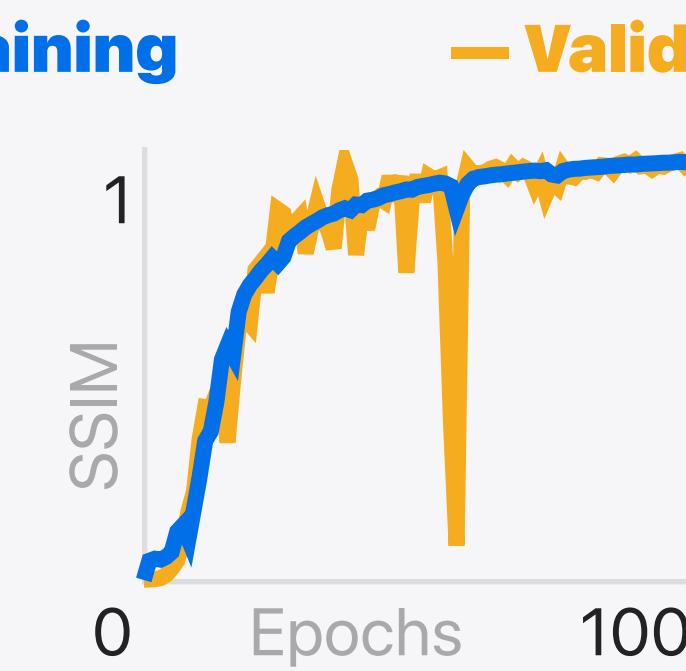
MSE

✓ Similar to MSE model



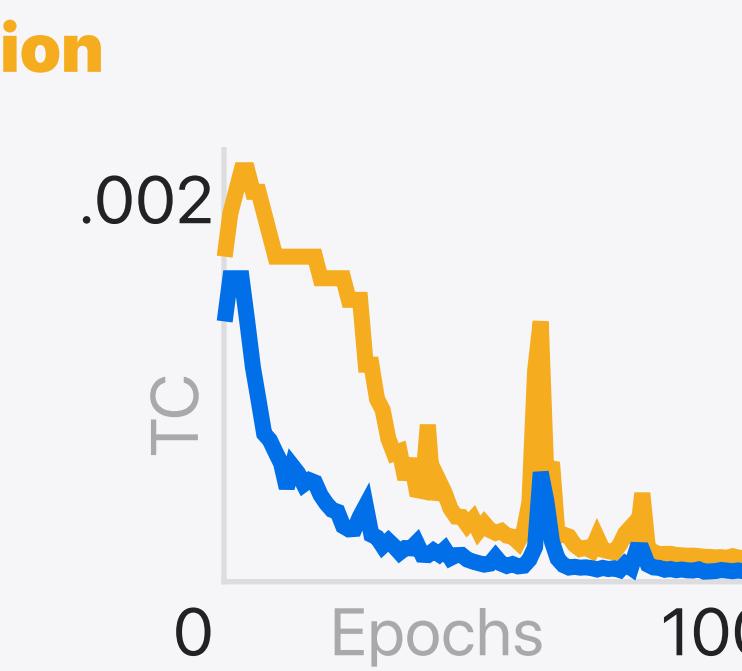
PSNR

✓ Slightly faster than MSE model



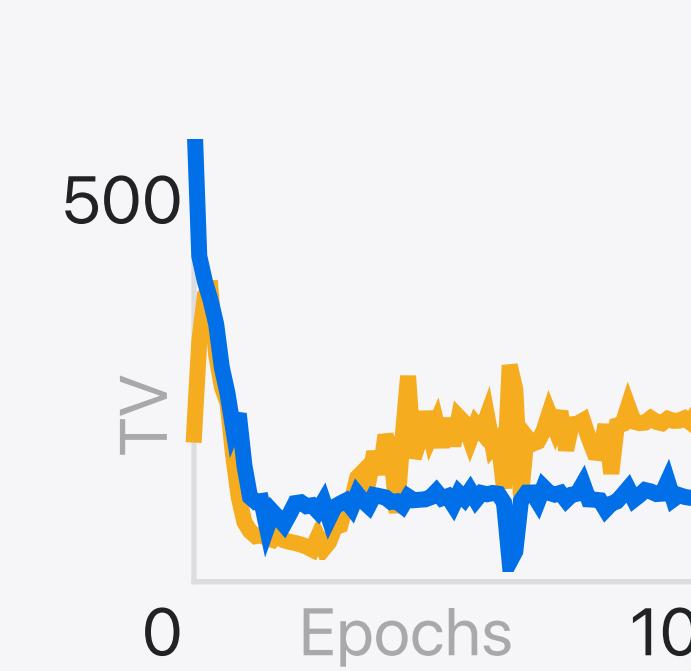
SSIM

✓ Similar to MSE model



TC

✓ Similar to MSE model



TV

✓ Similar to MSE model

Test results

Quantitative metrics

	MSE model	TC model	Unified model
MSE ★ Lower is better	0.0001 ★	0.0326	0.0001 ★
PSNR ★ Higher is better	45.1848 dB	14.8740 dB	46.9794 dB ★
SSIM ★ Higher is better	0.9633	0.0162	0.9692 ★
TC ★ Lower is better	0.0001	0.0000 ★	0.0001
TV ★ Lower is better	148.1984	239.4765	141.8375 ★

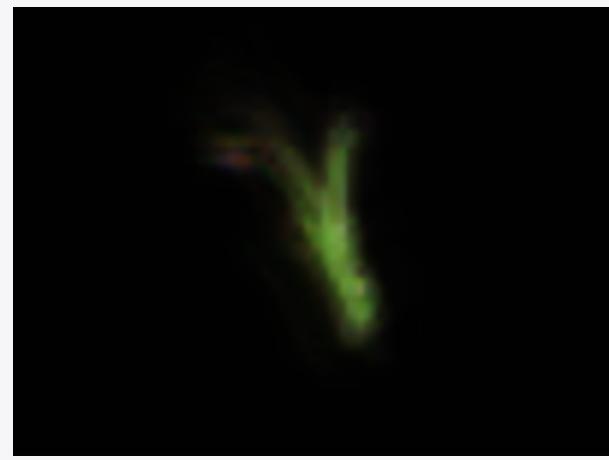
Test results

Qualitative inspection

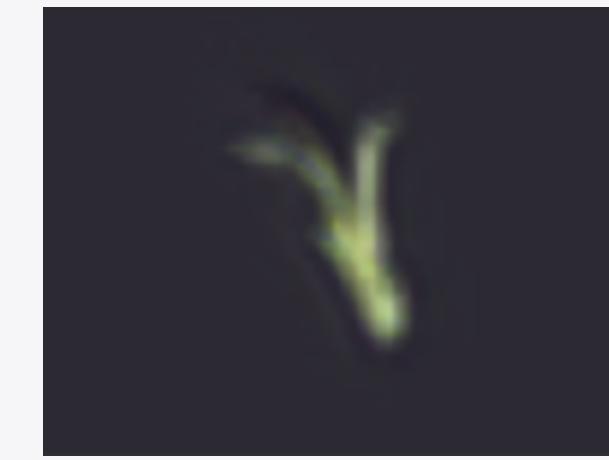
Predicting one future time step
Based on past 35 images



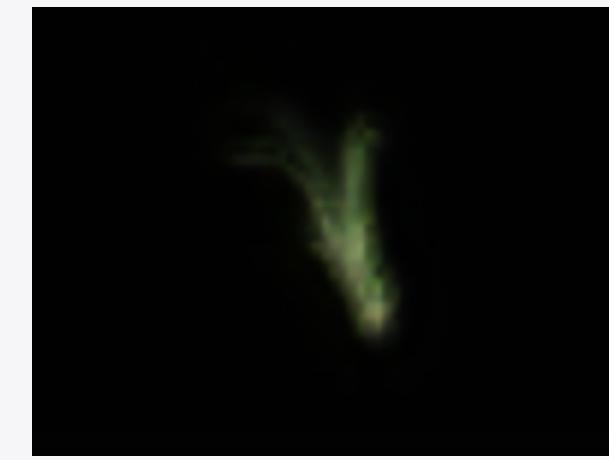
Ground truth



MSE model



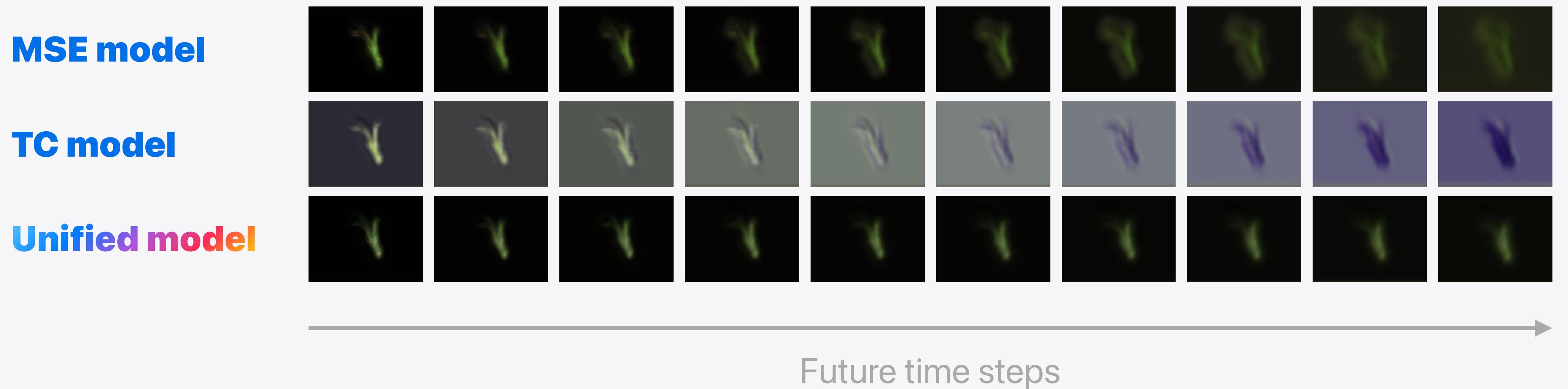
TC model



Unified model

Test results

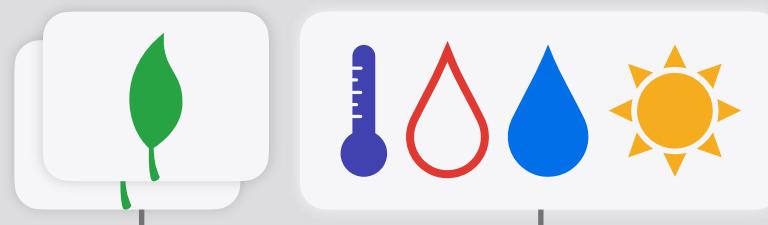
Qualitative inspection



Recap

Predicting future plant growth appearances with Multimodal Convolutional Long Short-Term Memory

Multimodal learning



Preprocessing

Multimodal architecture



Model architecture

ConvLSTM

Normalization

Concatenation

3D Convolution

Time Distributed

Reshape

Lambda

MSE

Mean Squared Error

Excels image quality

TC

Temporal Consistency

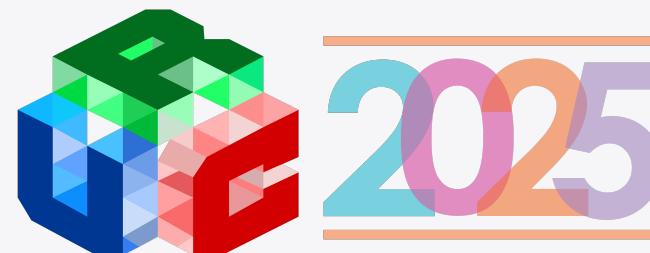
Excels smoothness in changing frames of plant appearance

MSE
+TC

Unified MSE + TC

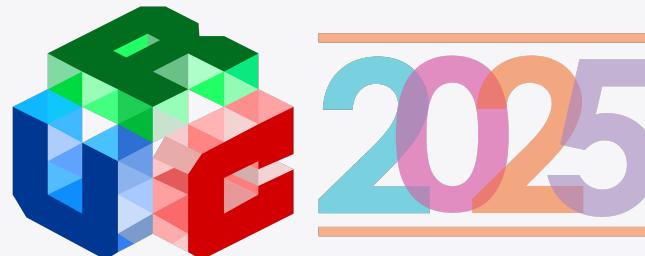
Surpasses other models with slightly better train & test results

Unified Loss Function
A unique approach by blending MSE and TC as a single loss function



Limitations

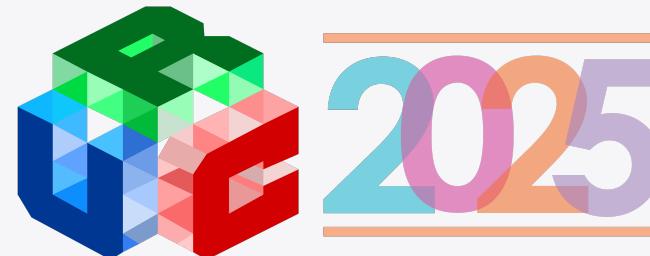
- ! Dataset lacks diversity in plant images
The causal of early convergence during model training
- ! Dataset lacks plant images grown in different external conditions
The causal why multimodality feature remain untested



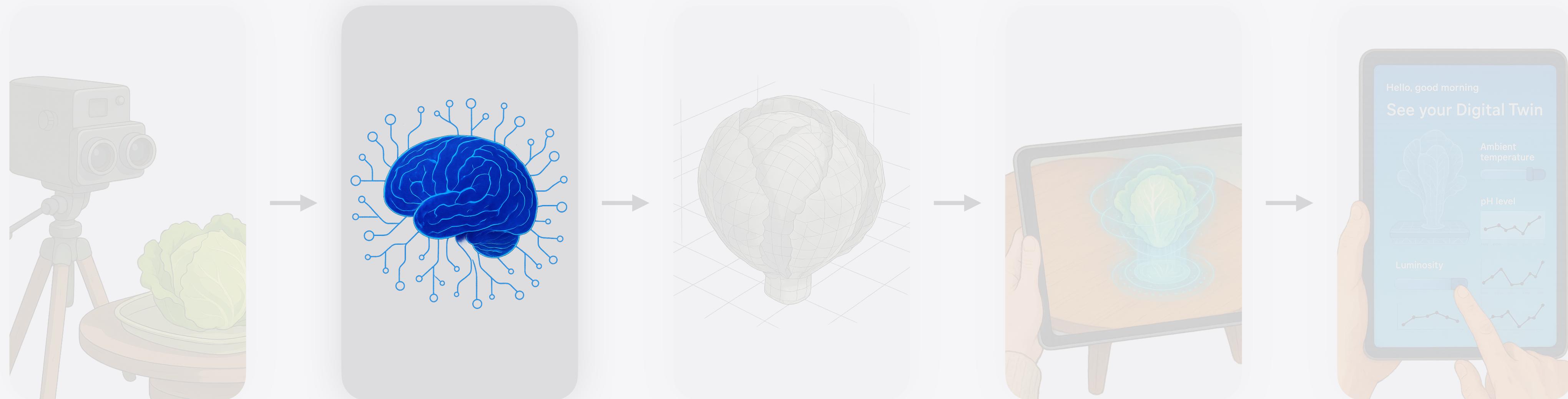
Recommendations

For succeeding researchers,

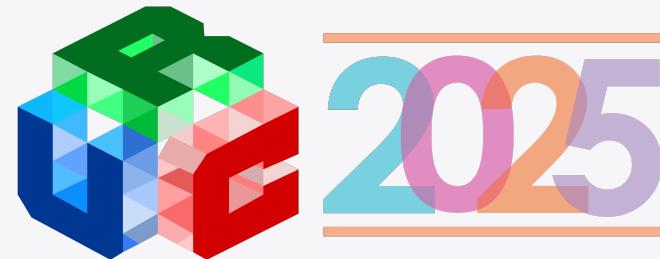
- ✓ Continue collecting data (i.e., add more sequences)
- ✓ Expand the dataset with other plant types (i.e., beyond lettuce)
- ✓ Grow them in different external conditions (e.g., cooler temperatures)
- ✓ Explore other techniques that help solve the weaknesses found in current models (e.g., blurry predictions)



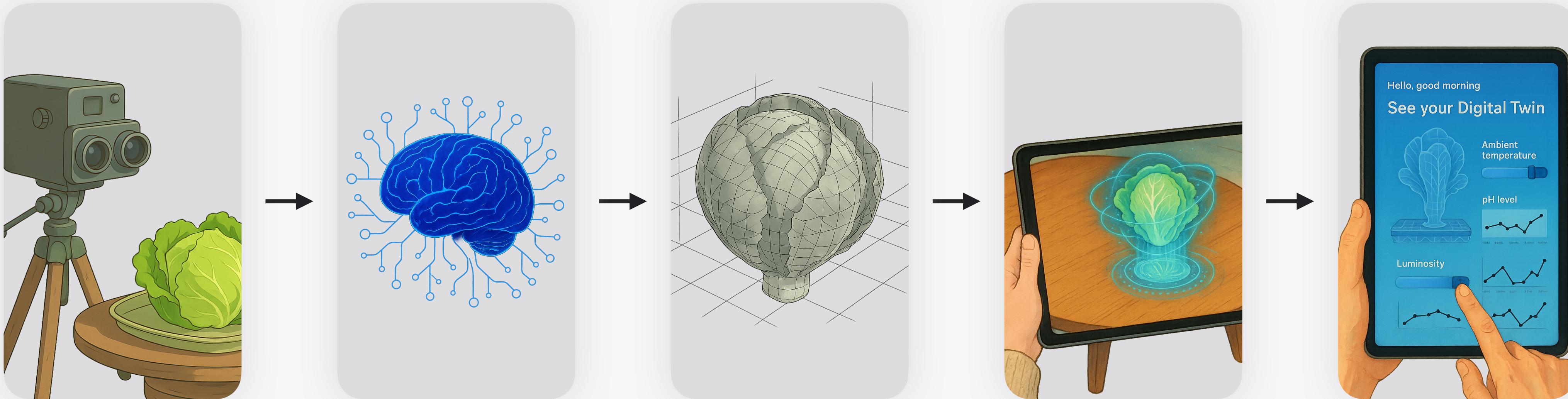
R&D roadmap



**Artificial
Intelligence**



R&D roadmap



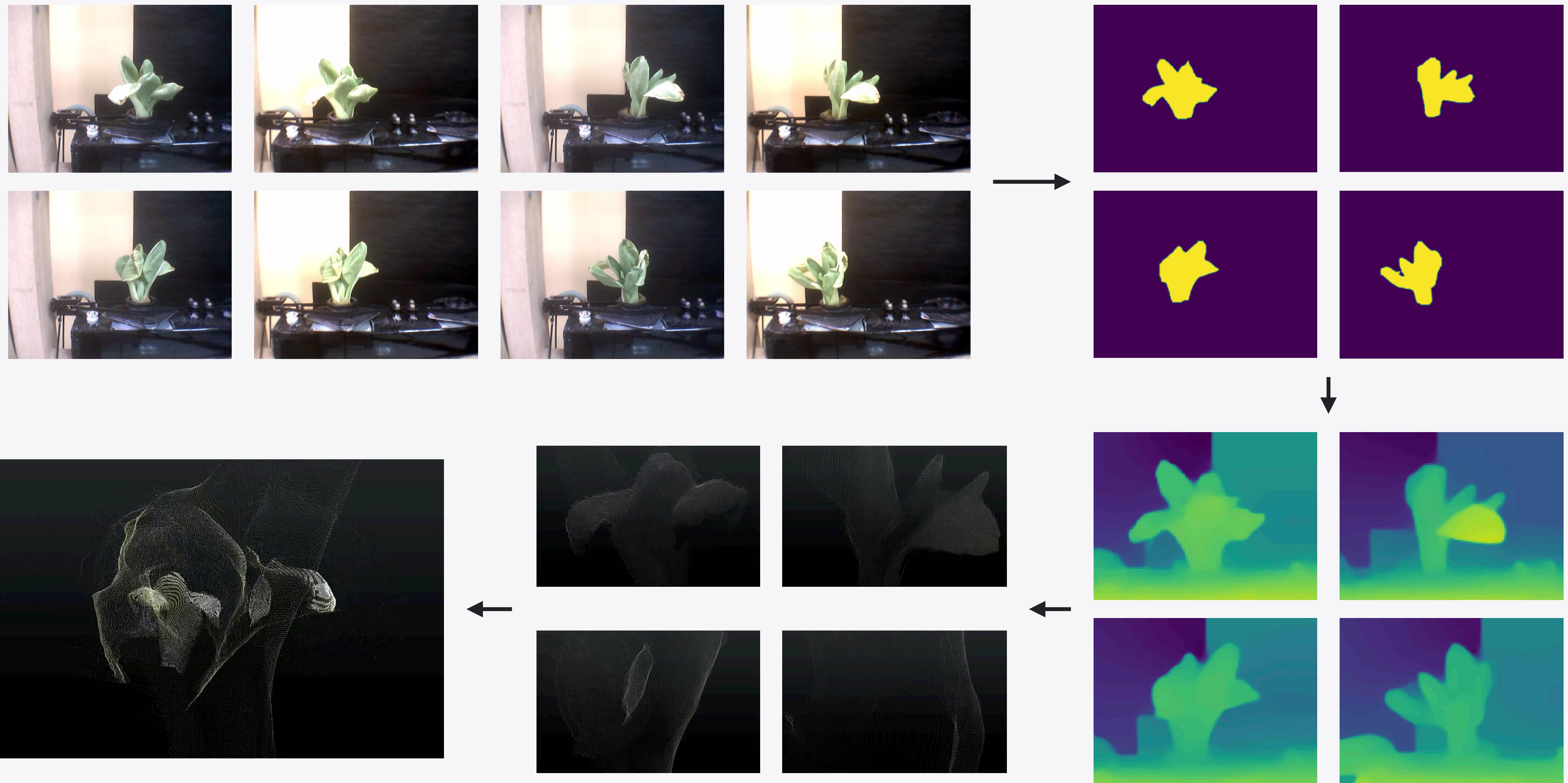
**Data
Collection**

**Artificial
Intelligence**

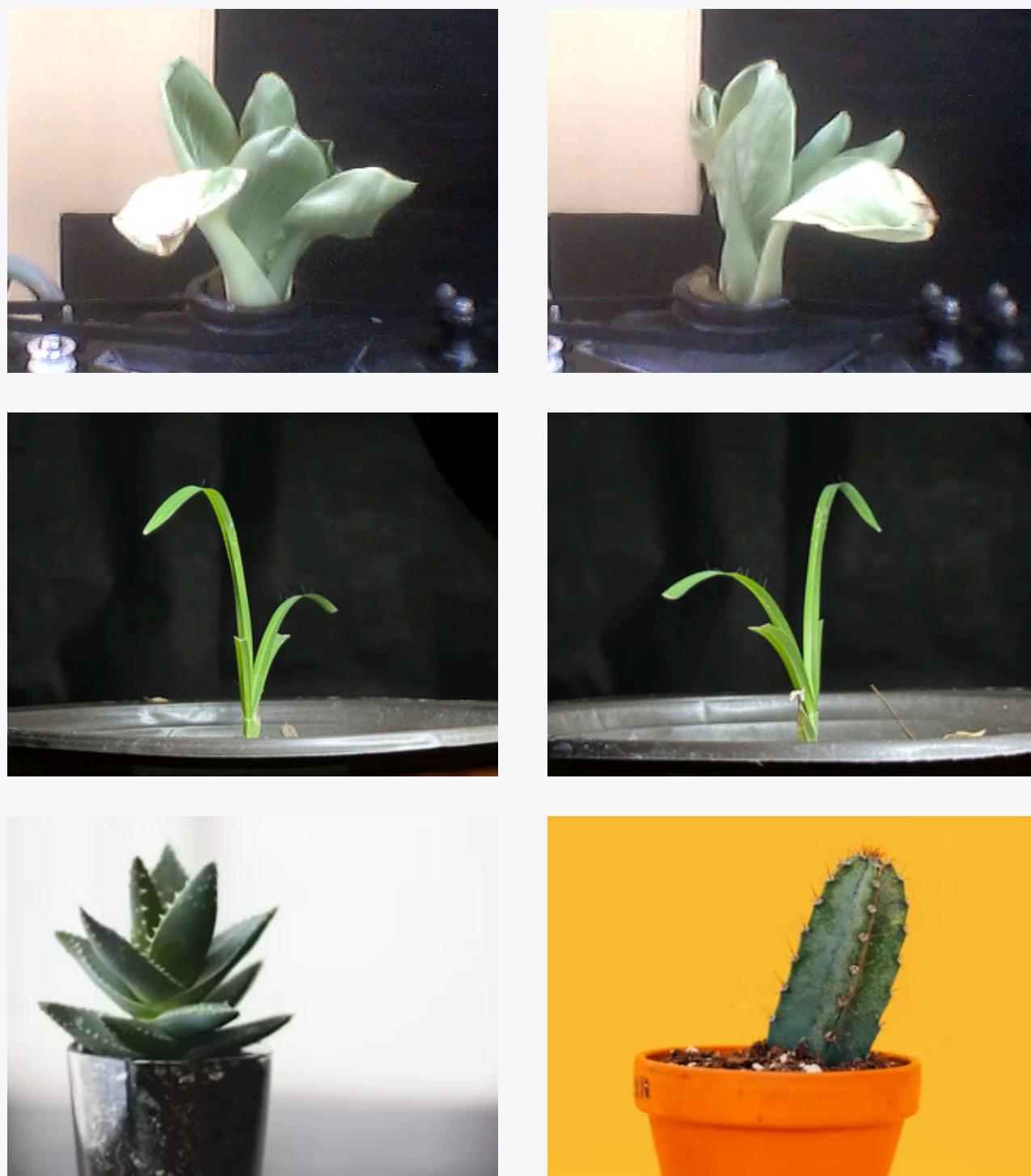
3D Modeling

**Augmented
Reality**

**User
Interface**



Training images



YOLOv8



Detectron2



UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

150th SVD 2025

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-on, 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. Current farming practices like vertical farming allow crops to grow in controlled environments, enabling farmers to manipulate external conditions such as pH levels. However, these decisions might not always be favorable to the crops. Applying too little or too much water, for example, can lead to unhealthy crops. Crops can become damaged or experience stunted 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 representation of a physical object or system that is used to monitor and control 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 gain insights and use them to make informed decisions that will benefit their crops. This leads to better yields, healthier crops, and higher yields. It redefines modern farming practices to be more sustainable and efficient, ultimately benefiting mankind and the planet.

Promotes SUSTAINABLE GOALS
2 ZERO HUNGER, 9 INDUSTRY INNOVATION AND INFRASTRUCTURE, 12 RESPONSIBLE CONSUMPTION AND PRODUCTION

Farmers can make informed agricultural decisions

The Technological Impacts
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.

Developing a Multimodal Convolutional Long Short-Term Memory Architecture for Predicting Future Plant Appearance
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.

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.

Investigating Segmentation Architectures Trained Solely for the Plant Class
Correct segmentation of plant objects positively affects succeeding phases such as 3D modeling. YOLOv6 and Detectron2 are the initial architectures chosen for comparison.

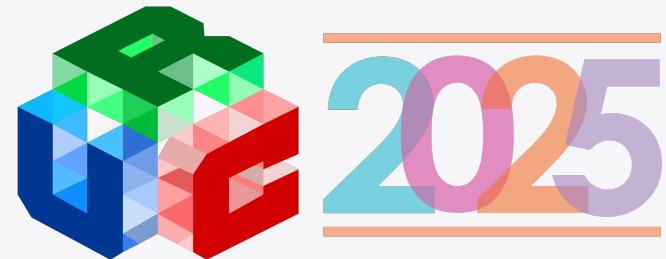
Sample of whole process
Plant's past appearances Images, External conditions associated Numerical data

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
Most General: Point Cloud Registration, Point Cloud Refinement, Point Cloud Generation. Point Cloud generation uses 3D registration. Point cloud refinement uses the boundary mesh with colored and colored color.

Recommendations

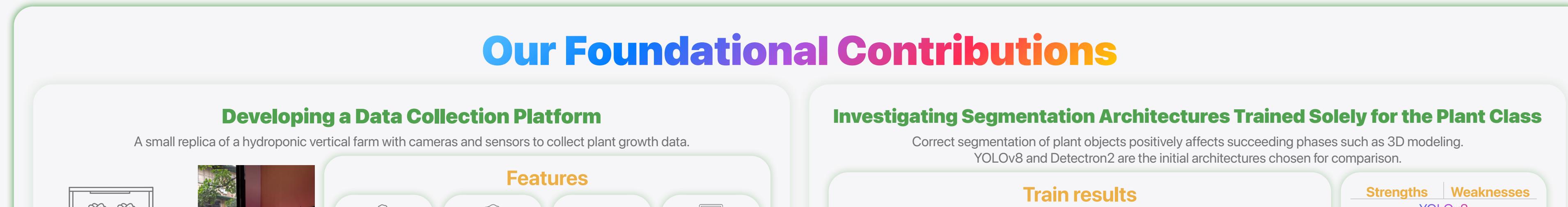
- Continue collecting data. Grow more plants, expand their plant types, and grow them under different external conditions.
- Explore data collection techniques that capture top and bottom plant views and yield denser point clouds.
- Explore deep learning techniques that can address weaknesses found in current models, such as blurriness in predictions.
- 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.

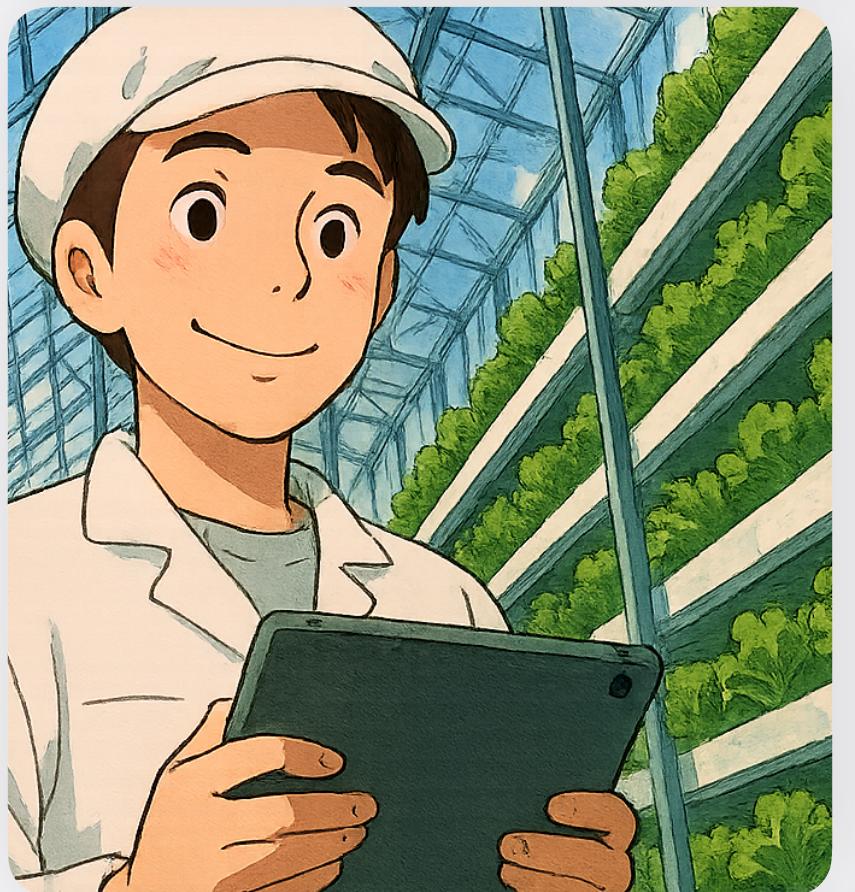
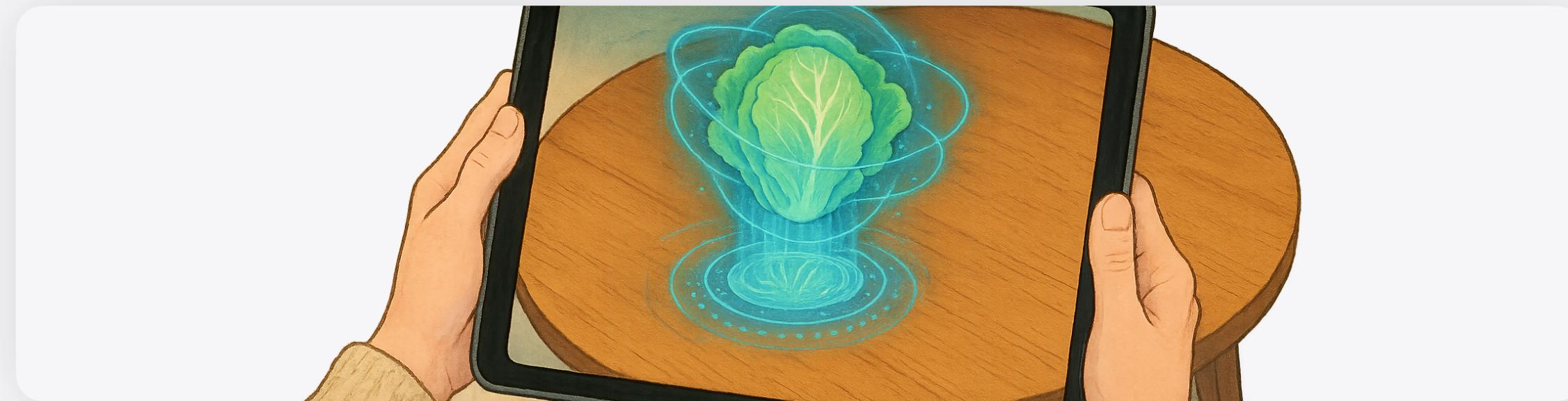


UNIVERSITY of SAN CARLOS
SCIENTIA • VIRTUS • DEVOTIO

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

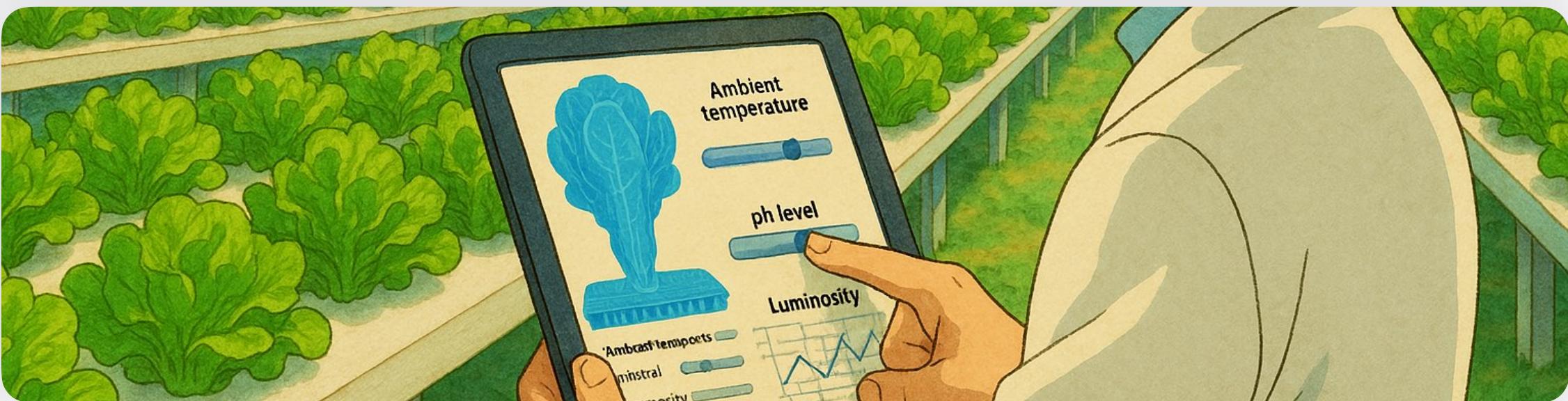
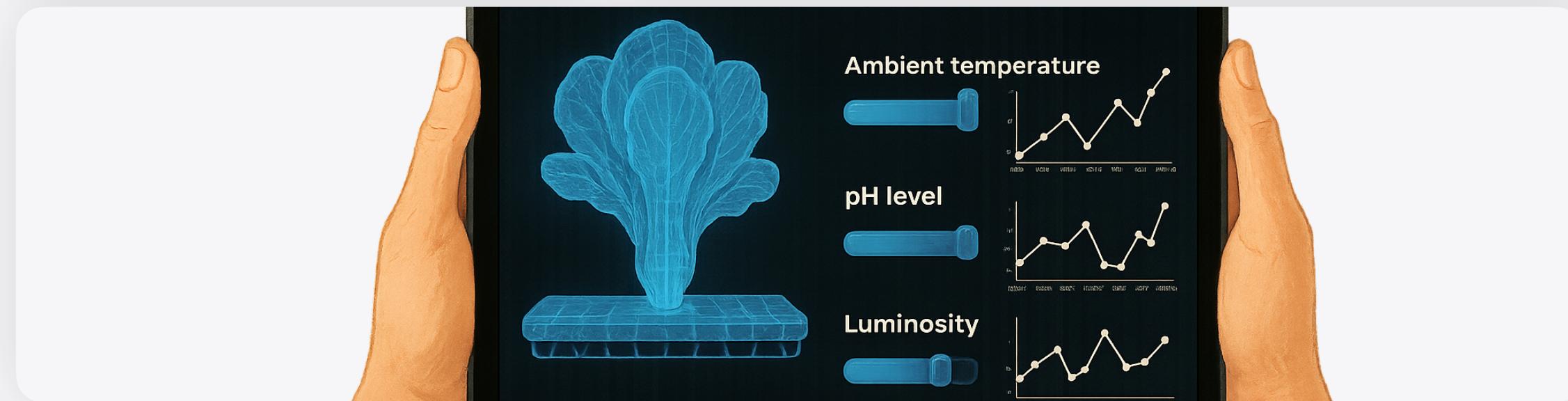
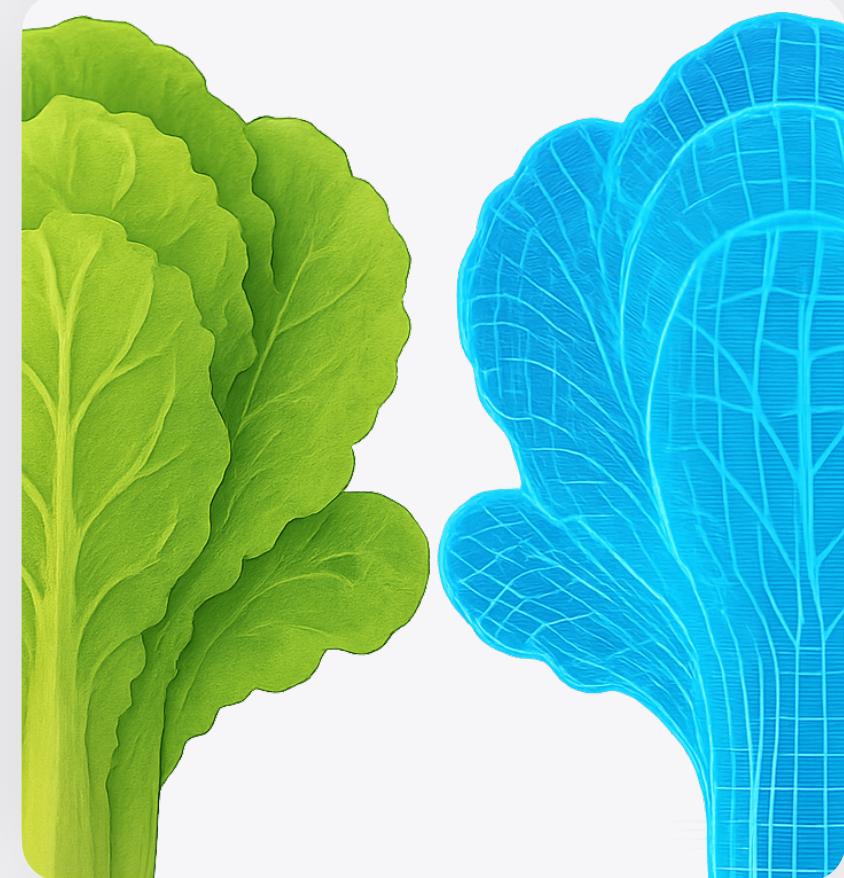




Our vision

Digital twins in agriculture

Empowering farmers with informed decisions.
Promoting sustainability and efficiency.
For humanity. For the planet.



Unified Temporal Consistency and Mean Squared Error Loss Function for Predicting Plant Growth Structures through Multimodal Convolutional Long Short-Term Memory: 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, School of Engineering