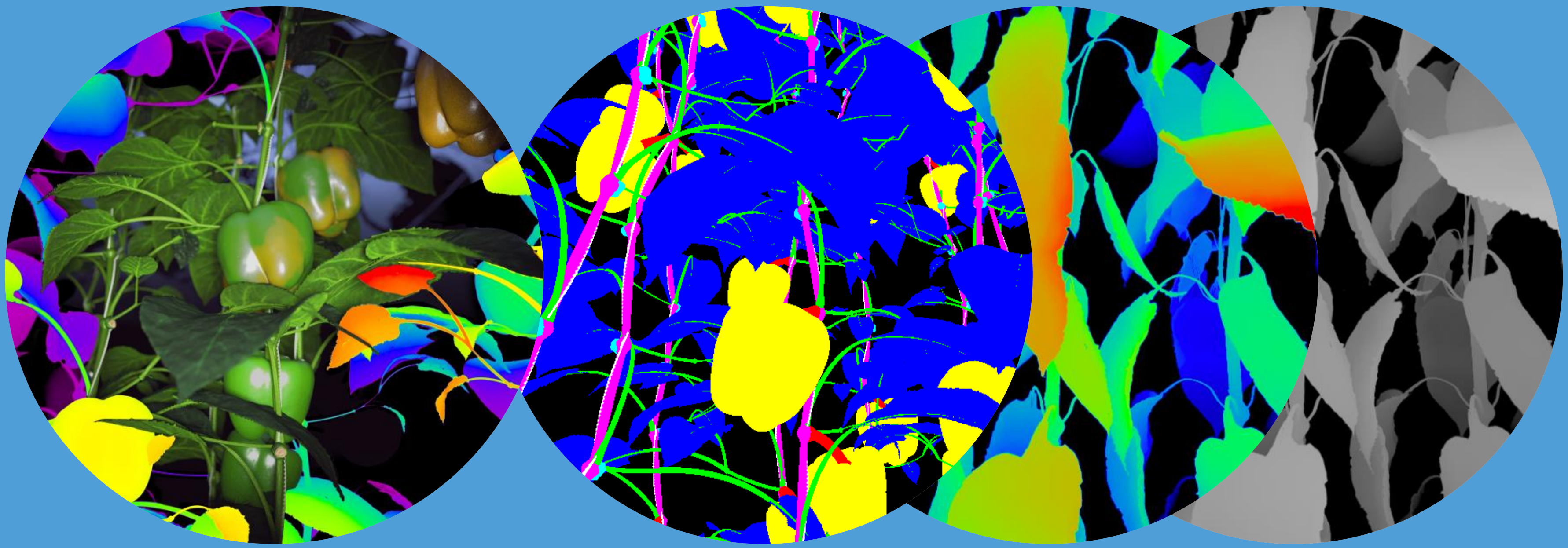
Deep Learning for Agro-Food Robotics

Manya Afonso, Ruud Barth, Aneesh Chauhan, Ron Wehrens



WAGENINGEN UR
For quality of life


Sweeper
FANUC Robot
LR Mate 200iD

Obstacle detection for a sweet-pepper harvest robot.

↓

Plant part localization.

↓

Using state-of-the-art machine learning: Deep Learning.

↓

Large annotated datasets on a per-pixel level.

↓

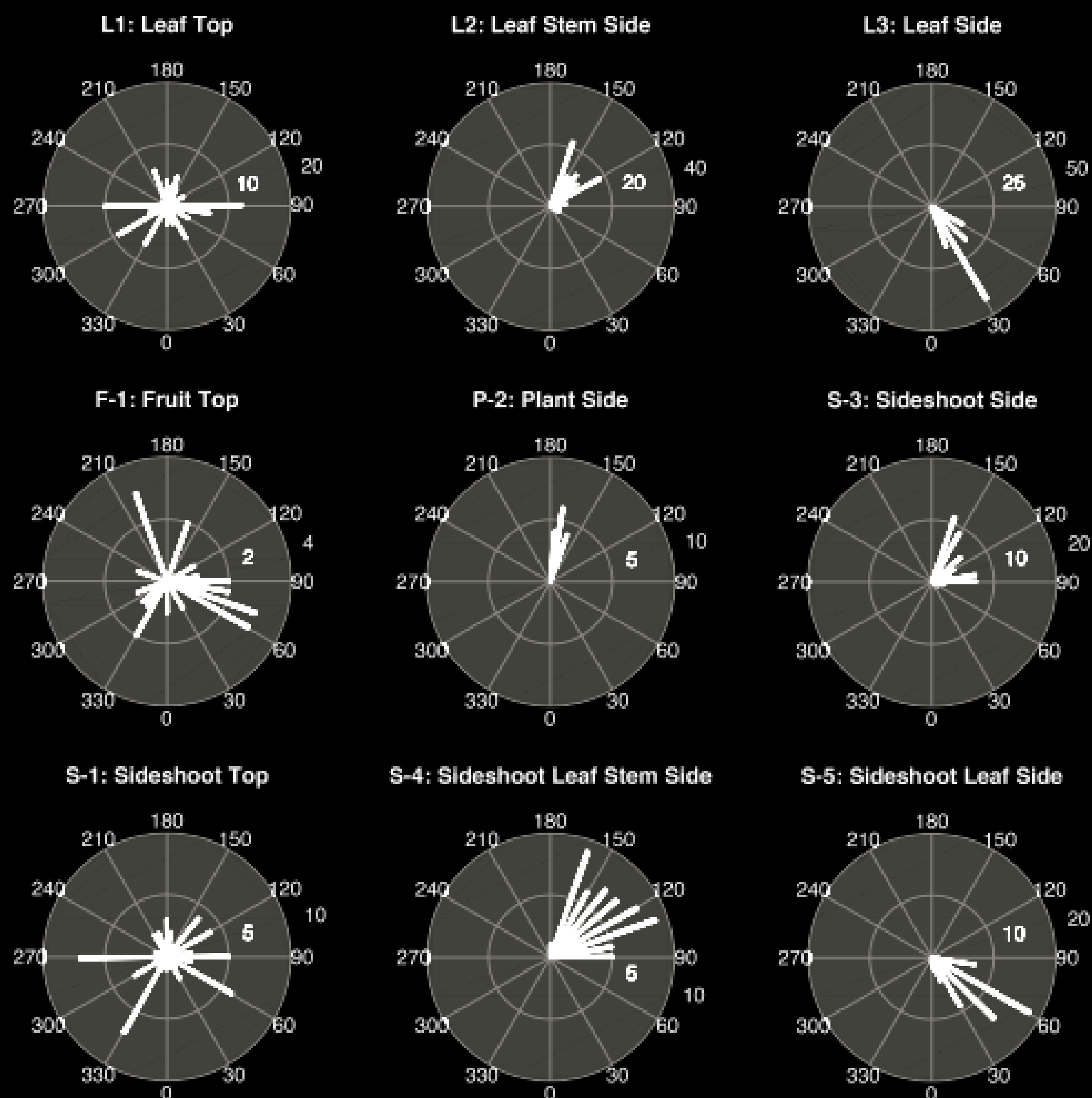
Synthetic dataset to bootstrap the model.

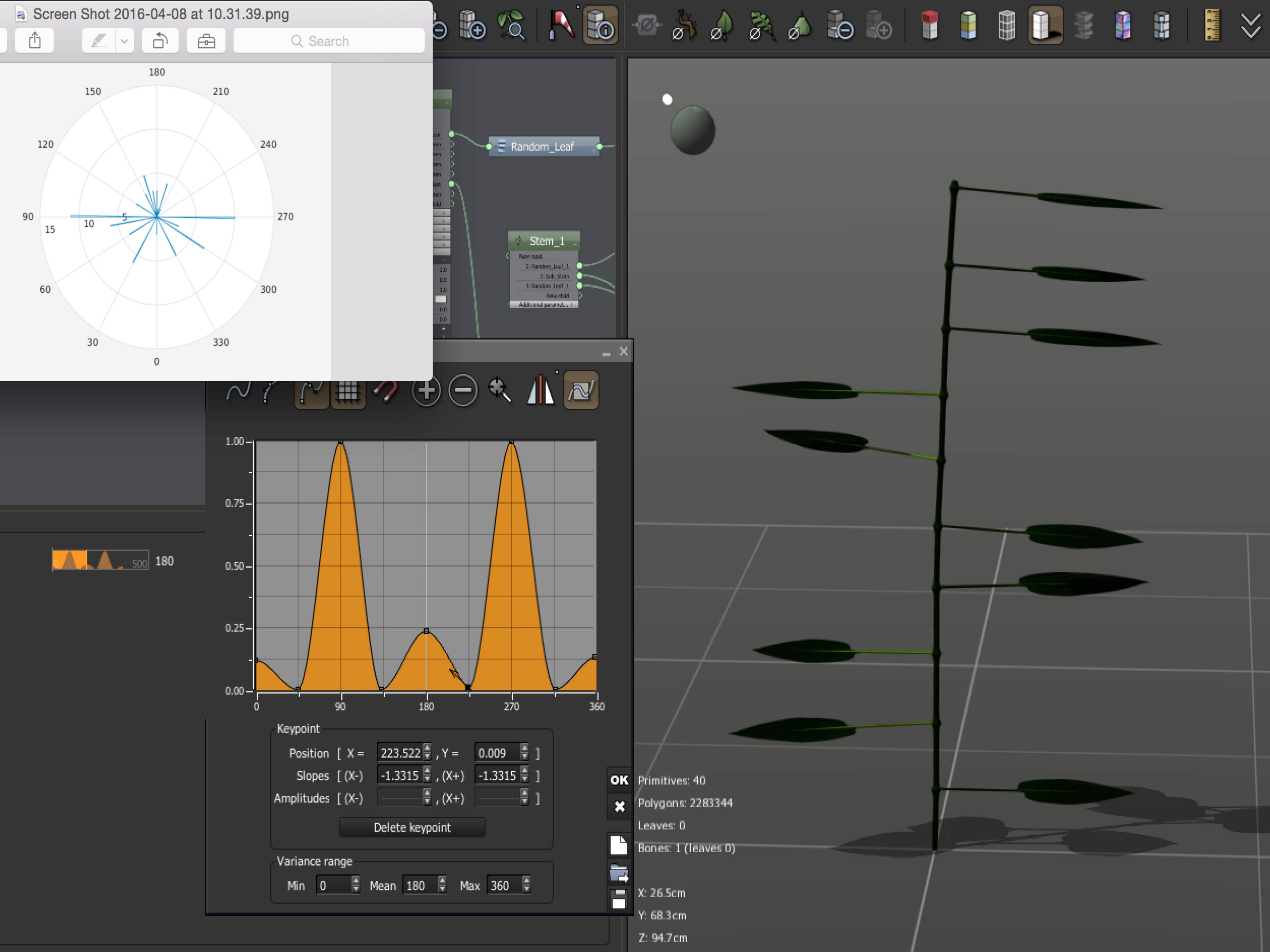

Real Image

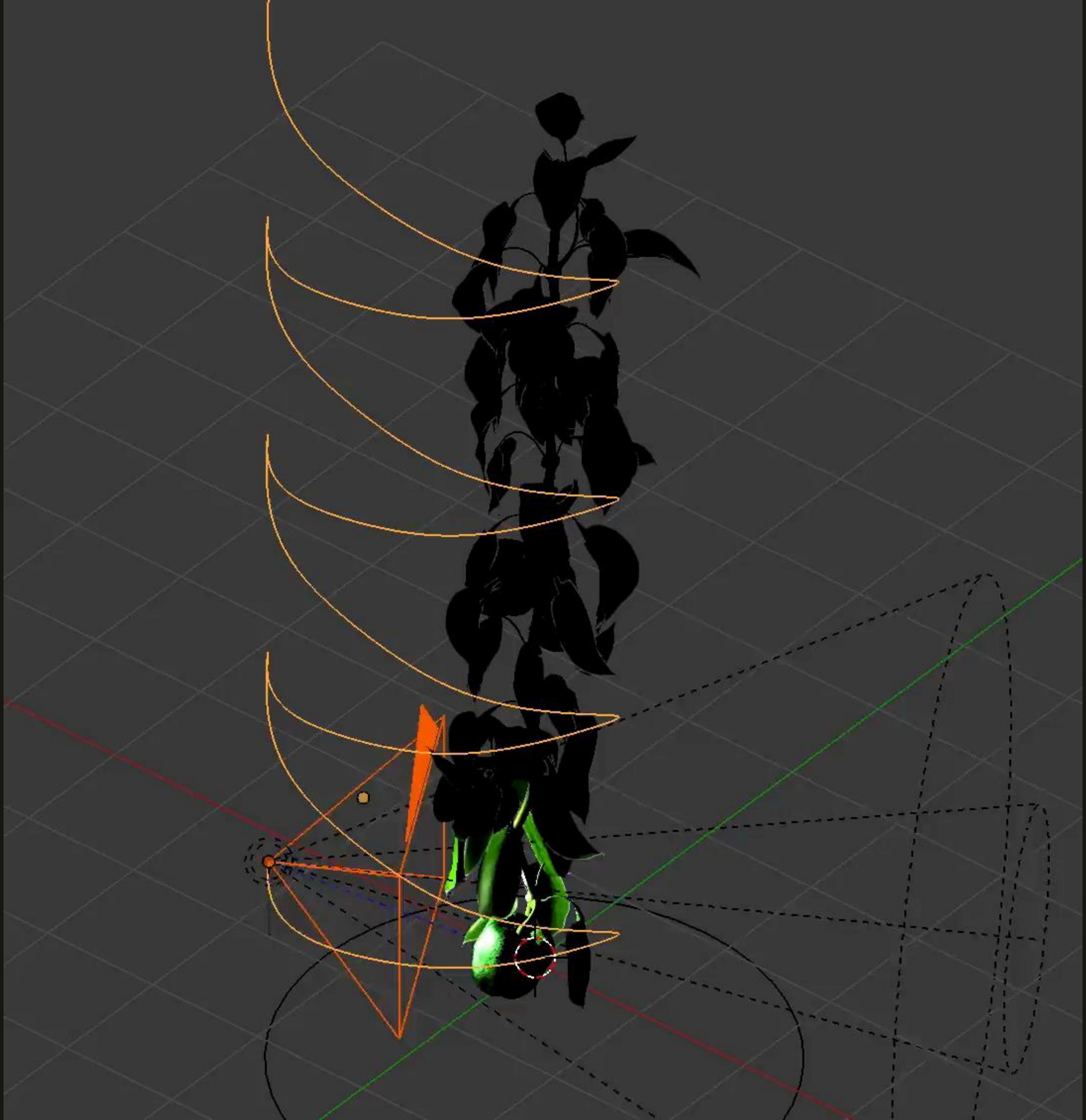

Ground
Truth

WAGENINGEN UR
For quality of life

The Good Dinosaur, Pixar, 2015


L1: Leaf Top
L2: Leaf Stem Side
L3: Leaf Side
F-1: Fruit Top
P-2: Plant Side
S-3: Sideshoot Side
S-1: Sideshoot Top
S-4: Sideshoot Leaf Stem Side
S-5: Sideshoot Leaf Side
180
150
120
90
60
30
0
330
300
270
240
210


Screen Shot 2016-04-08 at 10.31.39.png
Search
180
150
210
120
240
90
270
60
300
30
330
0
15
10
5
Random_Leaf
Stem_1
1.00
0.75
0.50
0.25
0.00
0
90
180
270
360
Keypoint
Position [X = 223.522 , Y = 0.009]
Slopes [(X-) -1.3315 , (X+) -1.3315]
Amplitudes [(X-) , (X+)]
Delete keypoint
Variance range
Min 0
Mean 180
Max 360
OK
Primitives: 40
Polygons: 2283344
Leaves: 0
Bones: 1 (leaves 0)
X: 26.5cm
Y: 68.3cm
Z: 94.7cm




Real Image


Synthetic Image


Ground
Truth

Generated 10k images on a supercomputer.

Annotated 50 empirical photographs manually.

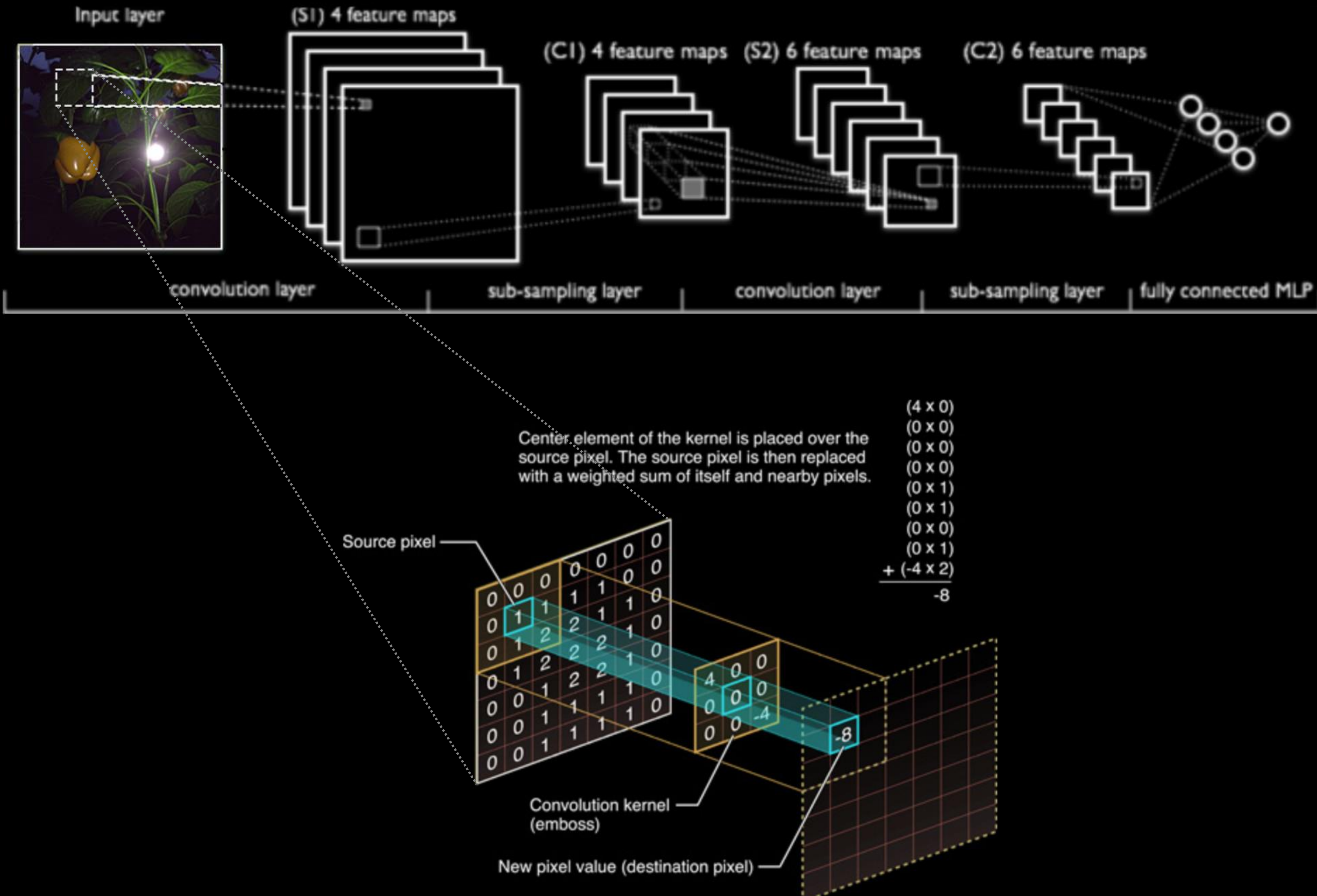
Create a semantic segmentation deep learning pipeline!

↓

Train on synthetic data

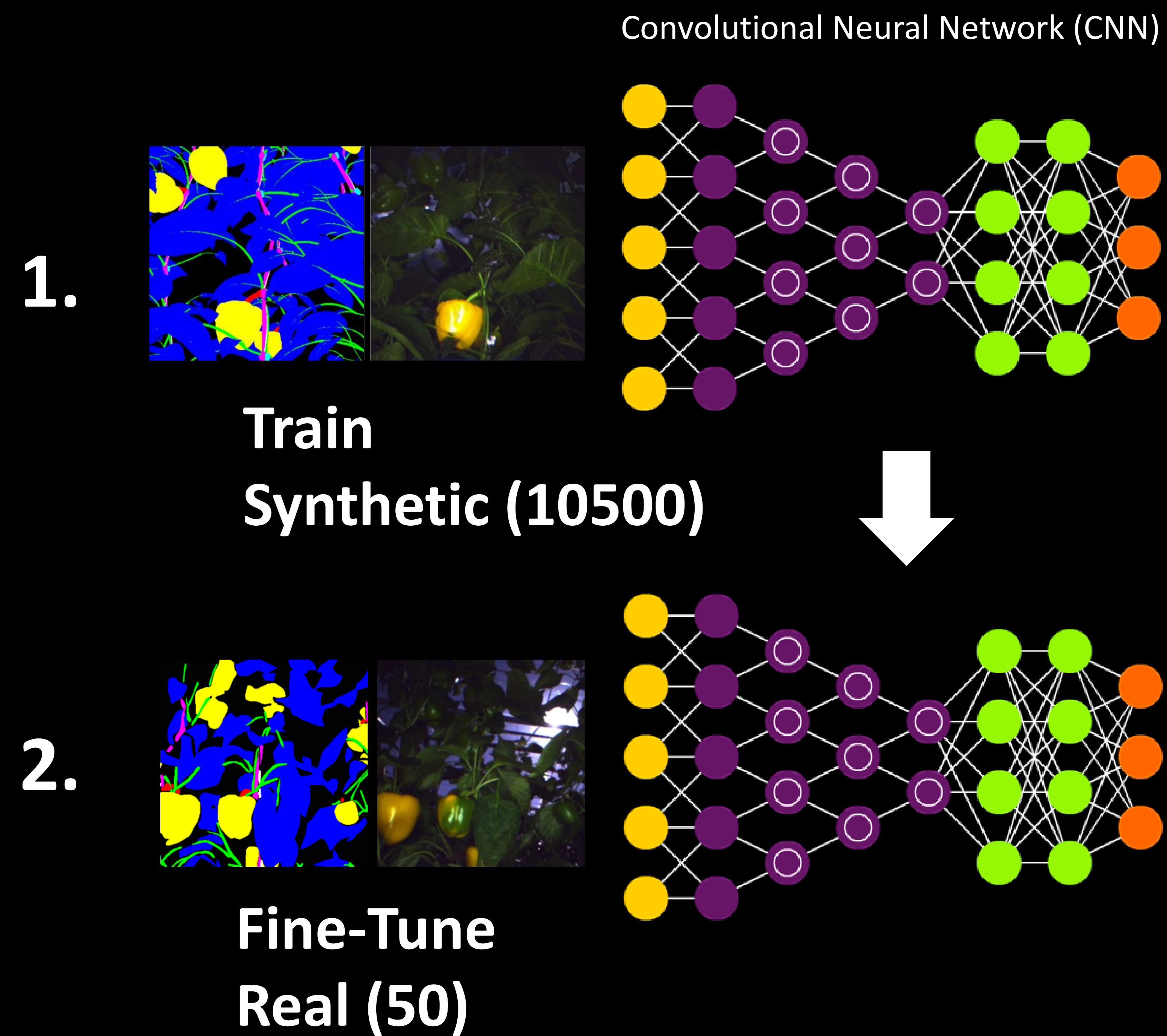
↓

Fine-tune on empirical data



Input layer
(S1) 4 feature maps
(C1) 4 feature maps
(S2) 6 feature maps
(C2) 6 feature maps
convolution layer
sub-sampling layer
convolution layer
sub-sampling layer
fully connected MLP
Center element of the kernel is placed over the source pixel. The source pixel is then replaced with a weighted sum of itself and nearby pixels.
(4 x 0)
(0 x 0)
(0 x 0)
(0 x 0)
(0 x 1)
(0 x 1)
(0 x 0)
(0 x 1)
+ (-4 x 2)
-8
Source pixel
Convolution kernel (emboss)
New pixel value (destination pixel)

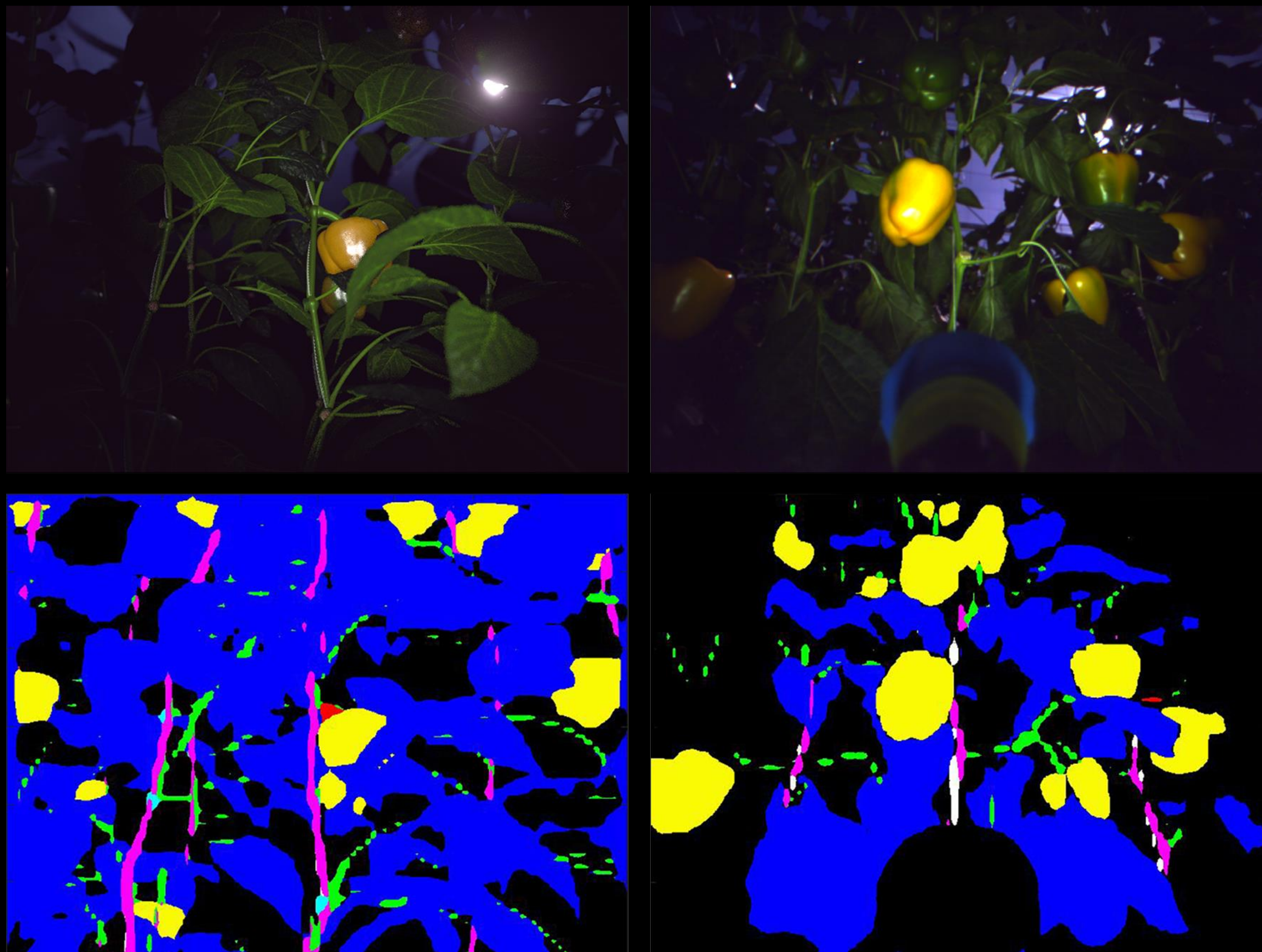
Image as published in “Conditional Random Fields as Recurrent Neural Netw


Convolutional Neural Network (CNN)
1.
Train
Synthetic (10500)
2.
Fine-Tune
Real (50)

synthetic | real

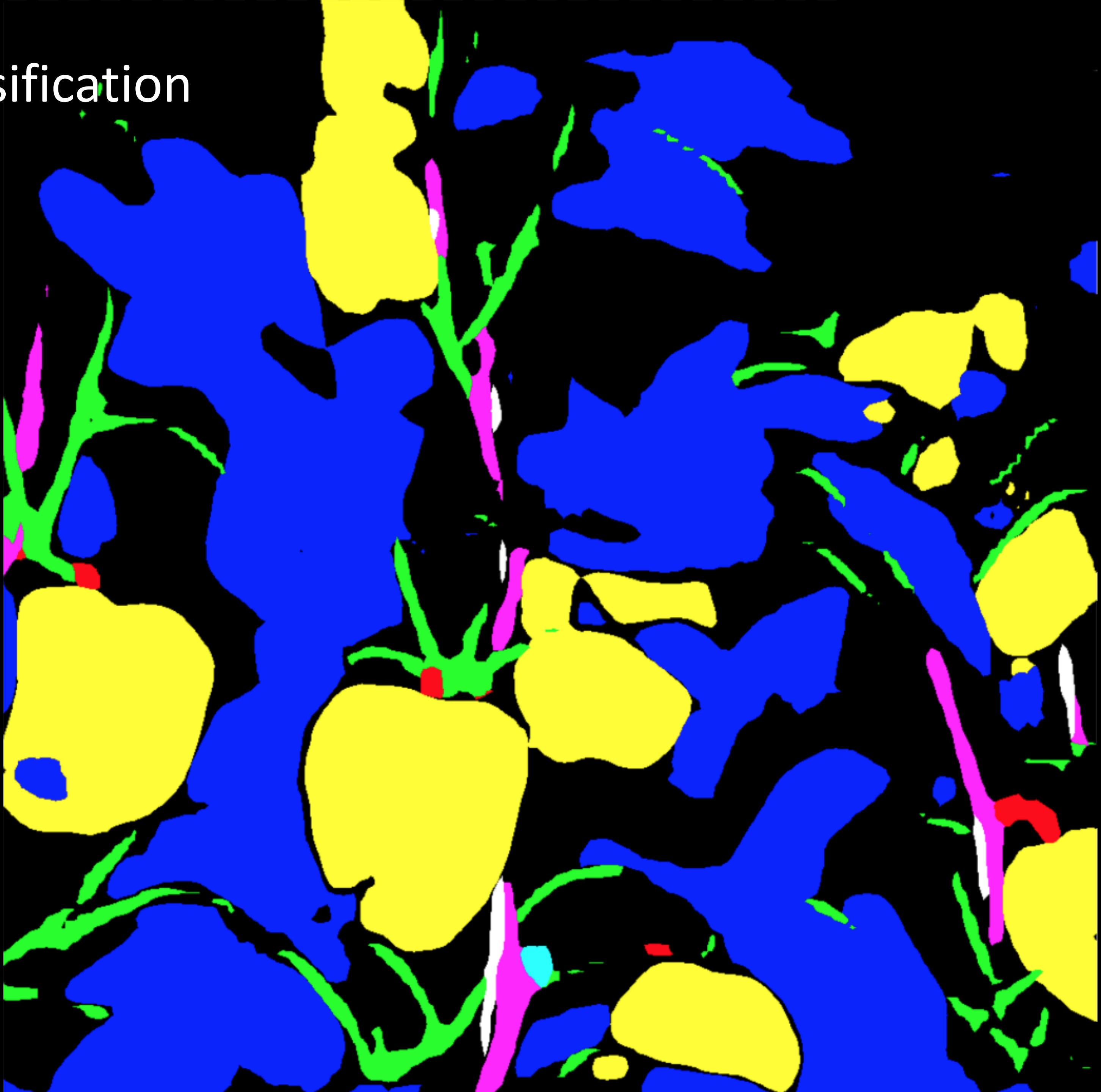
color image

computer prediction



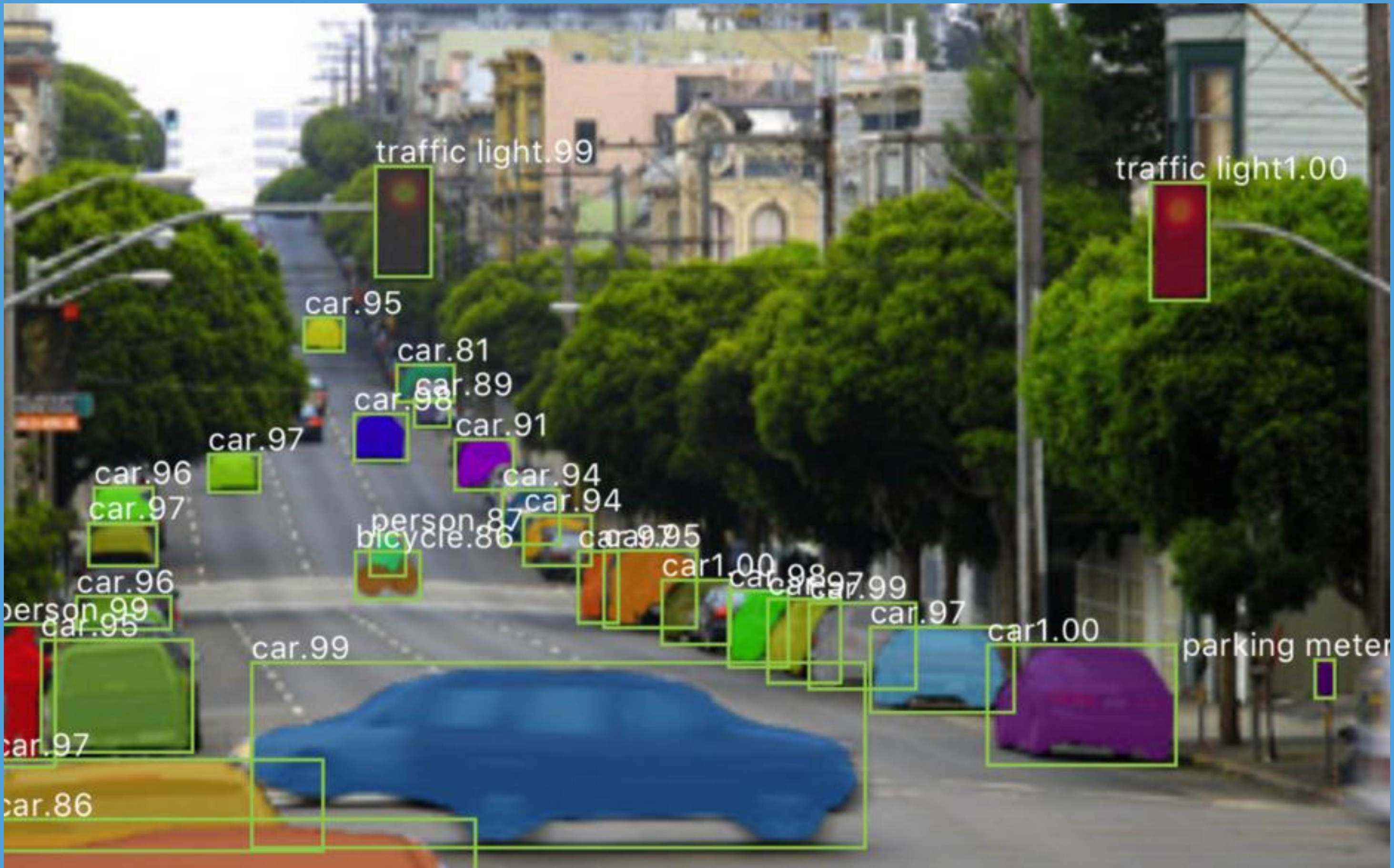
Real Image

Classification



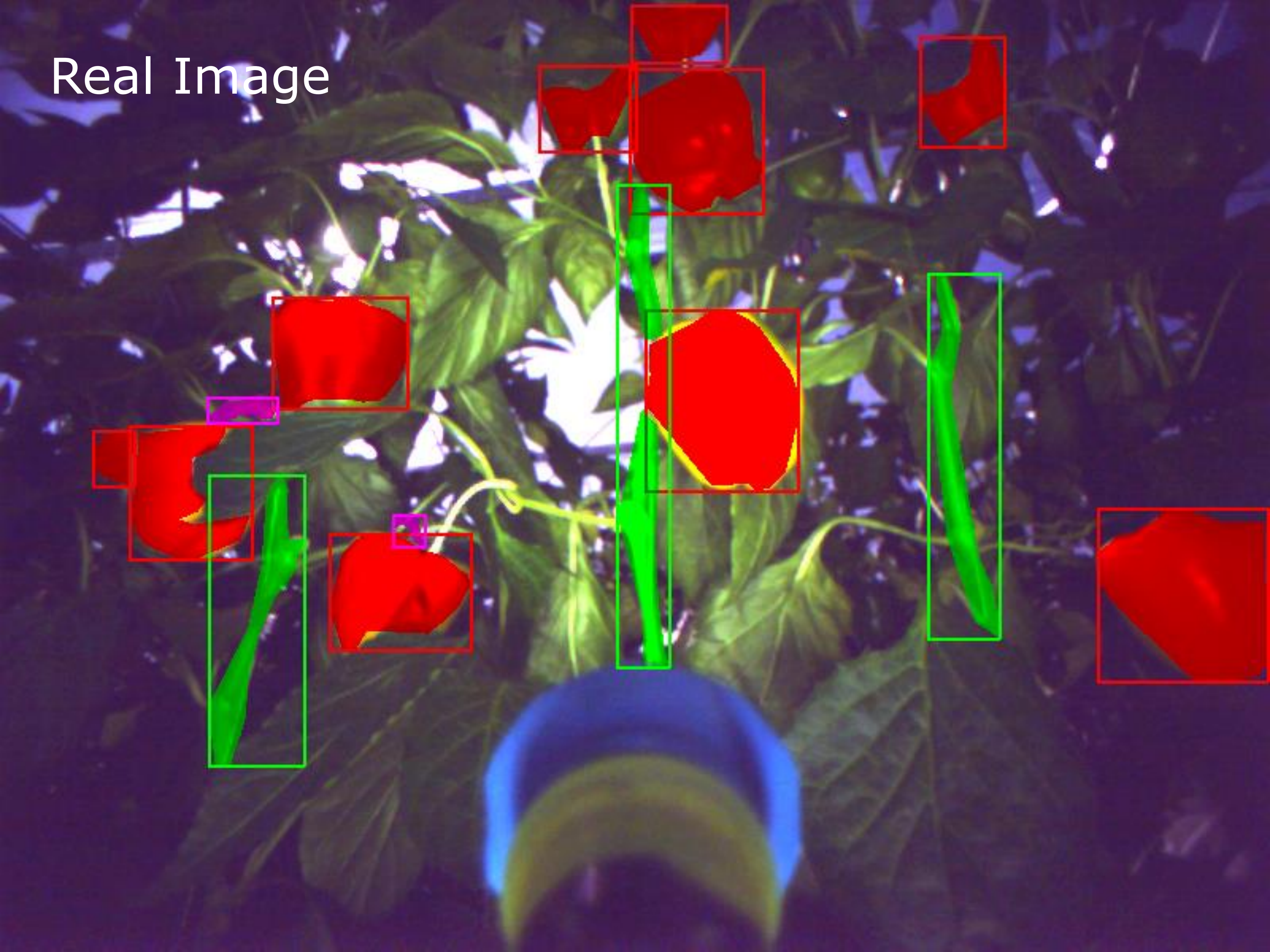
However per pixel segmentation cannot distinguish individual parts!

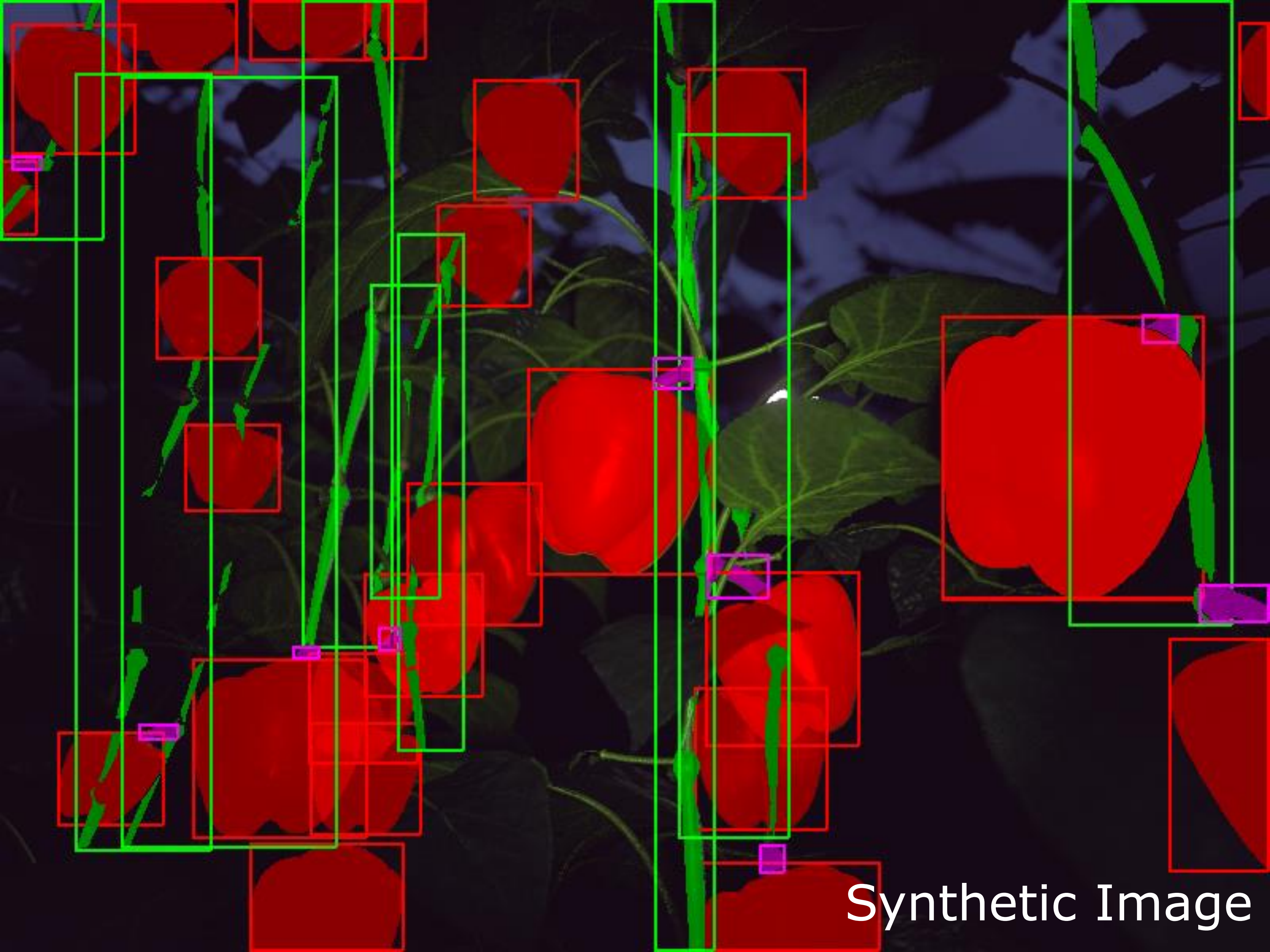
Instance Segmentation



We manually labelled 50 empirical images for the classes fruit, peduncle and stem (30 training, 20 validation).

For the synthetic dataset, 250 images were re-rendered with separated instances (200 training, 50 validation).

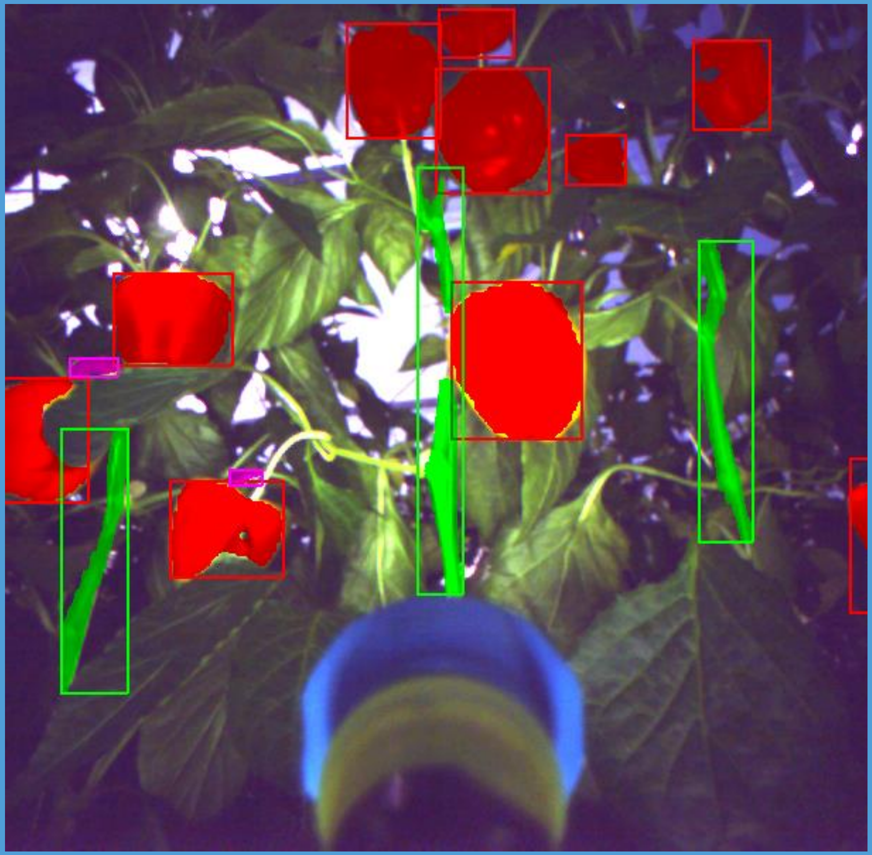

Real Image


Synthetic Image

Results Synthetic detection

Results Empirical detection

Results Empirical detection (fine-tuned)



WAGENINGEN UR
For quality of life

Results

- Fruit detection over empirical test set

Network	Precision	Recall	IoU
Pixelwise	--	--	0.76
MaskRCNN over R101 trained on synthetic	0.54	0.94	0.63
MaskRCNN with finetuning	0.89	0.96	0.79

Conclusion

- We have succesfully trained a model to tell us where plant parts are in the image.
- Synthetic data helps to improve the performance.
- Mask-RCNN moreover provides the recognition of instances.
- This is of extreme improtance for agricultural robotics.
- The computer vision can now cope with a lot of the variation it can encounter. We now have a method to exactly tell where the robot must go.

Questions