Computers and Electronics in Agriculture 144 (2018) 284–296

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Data synthesis methods for semantic segmentation in agriculture: A *Capsicum annuum* dataset

R. Barth[a,c,*], J. IJsselmuiden[b], J. Hemming[a], E.J. Van Henten[b]

[a] Wageningen University & Research, Greenhouse Horticulture, P.O. Box 644, 6700 AP Wageningen, The Netherlands
[b] Wageningen University & Research, Farm Technology Group, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[c] Harvard University, Biorobotics Laboratory, 60 Oxford Street, Cambridge, MA, United States

ARTICLE INFO

Keywords:
Synthetic dataset
Semantic segmentation
3D modelling
Agriculture
Robotics

ABSTRACT

This paper provides synthesis methods for large-scale semantic image segmentation datasets of agricultural scenes with the objective to bridge the gap between state-of-the art computer vision performance and that of computer vision in the agricultural robotics domain. We propose a novel methodology to generate renders of random meshes of plants based on empirical measurements, including the automated generation per-pixel class and depth labels for multiple plant parts. A running example is given of *Capsicum annuum* (sweet or bell pepper) in a high-tech greenhouse. A synthetic dataset of 10,500 images was rendered through Blender, using scenes with 42 procedurally generated plant models with randomised plant parameters. These parameters were based on 21 empirically measured plant properties at 115 positions on 15 plant stems. Fruit models were obtained by 3D scanning and plant part textures were gathered photographically. As reference dataset for modelling and evaluate segmentation performance, 750 empirical images of 50 plants were collected in a greenhouse from multiple angles and distances using image acquisition hardware of a sweet pepper harvest robot prototype. We hypothesised high similarity between synthetic images and empirical images, which we showed by analysing and comparing both sets qualitatively and quantitatively. The sets and models are publicly released with the intention to allow performance comparisons between agricultural computer vision methods, to obtain feedback for modelling improvements and to gain further validations on usability of synthetic bootstrapping and empirical fine-tuning. Finally, we provide a brief perspective on our hypothesis that related synthetic dataset bootstrapping and empirical fine-tuning can be used for improved learning.

1. Introduction

1.1. Research aim

In recent years the need of robotisation in agriculture has been growing notably to keep up with the increasing demand of productivity and quality of food production whilst decreasing the pressure on resources required (Bac et al., 2014). Although mechanisation has been an ongoing human effort for centuries, the next leap forward to achieve these higher goals is by adding a degree of artificial intelligence to harvesting and crop management systems to enable increased selectivity, precision and robustness.

We identified one of the main current bottlenecks for introducing robotics in agriculture as the computer vision performance. In the past decade, the general field of computer vision made significant progress in object localisation and consecutively was successfully applied in many domains. However, this performance has not been matched for sensing solutions in agriculture (Gongal et al., 2015; Nasir et al., 2012). We argue that one of the main reasons is the absence of detailed and large annotated agricultural datasets that current state-of-the-art methods require, but are infeasible to obtain manually.

Accordingly, to contribute to solving this bottleneck and move the field forward, we provide a method for artificial agricultural data synthesis. We hypothesise it is possible with this approach to generate synthetic images highly similar to empirical images. Specifically, this paper introduces a method for the generation of large-scale semantic segmentation datasets on a plant-part level of realistic agriculture scenes, including automated per-pixel class and depth labeling. One purpose of such synthetic dataset would be to bootstrap or pre-train computer vision models, which are fine-tuned thereafter on a smaller empirical image dataset (Dittrich et al., 2014; Kondaveeti, 2016). Our methodology is designed to be extended to other plants, but for reference a running example is given for a *Capsicum annuum* species, also known as sweet (or bell) pepper. An empirical photographic dataset

* Corresponding author at: Wageningen University & Research, Greenhouse Horticulture, P.O. Box 644, 6700 AP Wageningen, The Netherlands.
E-mail addresses: ruud.barth@wur.nl (R. Barth), joris.ijsselmuiden@wur.nl (J. IJsselmuiden), jochen.hemming@wur.nl (J. Hemming), eldert.vanhenten@wur.nl (E.J.V. Henten).

https://doi.org/10.1016/j.compag.2017.12.001
Received 4 May 2017; Received in revised form 20 October 2017; Accepted 2 December 2017
Available online 19 December 2017

was gathered and partially annotated to (i) use as reference for modelling and (ii) to verify the performance of any computer vision method that used the synthetic data for bootstrapping and applied to the empirical images.

In the following sections we report on the methodology for the data synthesis, with the requirement of similarity with the empirical dataset. The method starts with the modelling of a structural plant model using empirical plant parameter measurements and images. This model was used for generating randomised mesh instances of plants, which were imported into render software to mimic scenes of a commercial agricultural environment. To synthesise color and per-pixel label and depth data, these scenes were rendered with similar characteristics as the hardware used in a harvesting robot prototype.

In the results section, examples are given for both the synthetic and empirical datasets. Although subjectively these sets are comparable, the sets were analysed for differences in distributions of colors per class to verify our first hypothesis of similarity. To test our second hypothesis that the datasets can be used for improved learning, 5 experiments were performed using a basic semantic segmentation deep learning network. However, the scope of this paper was primarily on the synthetic image generation methodology. Therefore we discussed the machine learning part briefly to give future perspective on our follow-up companion paper. In that paper we will more extensively discuss the impact of synthetic data for semantic segmentation of plant parts and the requirements. At the end of this paper we discussed the challenges and limitations of this approach for realistic data generation and its potential use for computer vision. We conclude the paper by making the used scripts, models and datasets publicly available. The objective of this release is to (i) enable comparison of state-of-the-art computer vision methods for this domain, (ii) further validate the approach of synthetic modelling and empirical fine-tuning and (iii) gain community feedback on modelling deficiencies and improvements.

1.2. Research context

In the domain of agriculture, progress in image classification performance has been lagging behind state-of-the-art results in other domains (Gongal et al., 2015; Nasir et al., 2012). Although some progress in high level object detection or localisation have lately been accomplished (Sa et al., 2016), lower level detailed object part recognition in scenes reflecting realistic structural object complexity, remains unsolved. The challenges of computer vision in the agricultural domain come from the high amount of variation within object classes and changing environment conditions, throughout the day and seasons (e.g. light, growth stages). To overcome this variability, large annotated and detailed datasets are needed for capture all different situations. Although collecting image data can be automated (van der Heijden et al., 2012), it remains required and time consuming to manually annotate.

Synthetic image dataset generation methods are emerging as an important tool in the computer vision community to automatically create annotated training data for bootstrapping machine learning models (Dittrich et al., 2014; Kondaveeti, 2016). Consecutively, such models can be fine-tuned by and applied to empirical image data. Recent examples showing improved object recognition performance can be found in multiple domains, e.g. urban scene segmentation (Ros et al., 2016), 3D human pose estimation from depth images (Shotton et al., 2013) and multi-modal magnetic resonance imaging for pathological cases (Cordier et al., 2016).

Previous work on methods for plant architecture modelling have been also successful for synthetic plant image generation. For example, OpenAlea (Pradal et al., 2015, 2017) is able to generate anatomical and functional plant models and furthermore can be used to simulate images with a virtual camera. Other approaches such as ElonSim (Benoit et al., 2014) provide a simulator of plant growth, specifically root systems, and a simulator of the image acquisition to generate synthetic images including ground truth. The simulator uses plant and camera parameters. Furthermore, recently a method was created for automatic model based synthetic dataset generation for crop and weeds detection on a per-pixel level (Cicco et al., 2016), though no plant parts could be differentiated.

The required level of labeling detail depends on the task and in turn determines how much annotation effort is needed. One approach is to only label images on a high level using a single class or a few keywords per image in order to classify an image globally or give a shortlist of objects in the image (Everingham et al., 2015). This can be partially automated through combined image and label retrieval using context from search engines (Fergus et al., 2005). For other datasets like ImageNet or PASCAL VOC, manual annotation was performed using crowd sourcing (Everingham et al., 2010; Russell et al., 2008). A second approach is to weakly label the data with bounding boxes around objects or their parts (Papandreou et al., 2015). However, some computer vision tasks require a per pixel level labeling of the image, also known as semantic segmentation. Specifically for agriculture, per-pixel segmentations are required for localisation in robotics for harvesting (Bac et al., 2013), disease detection (Polder et al., 2014) and phenotyping (van der Heijden et al., 2012). For example in harvest robotics, obstacle maps on the plant part level resolution improves successful motion planning (Bac et al., 2014, 2016). Registered depth images can provide an additional dimension for motion control (Barth et al., 2016).

With the advent of state-of-the-art machine learning methods for computer vision, most notably convolutional neural networks for image classification and segmentation, the training dataset size requirement has been further increased (Najafabadi et al., 2015). Such learning models can have up to 10^{11} free parameters (Dean et al., 2012), which depend upon a large number of distinct data samples for the optimisation to converge properly without overfitting to occur (Trask et al., 2015). Without access to large datasets, domains such as agriculture previously used traditional computer vision methods using manual feature crafting (Bolón-Canedo et al., 2013) whilst capturing a limited subset of the variability. Our aim is to facilitate the agricultural computer vision domain with the benefits of state-of-the-art machine learning, e.g. the supervised hierarchical feature representation learning and the performance increase that comes with large datasets (LeCun et al., 2015).

2. Materials and methods

In Fig. 1 our method to obtain the synthetic and empirical datasets is shown in a flowchart. Empirical data was a cornerstone for two objectives. First, it was used as a reference to create both a realistic model and conditions to render the synthetic dataset. Second, to provide fine-tuning data and a verification test set for computer vision methods that use the synthetic dataset for bootstrapping. This section provides some intermediate results as prerequisite for consecutive methods; the final results of the synthetic and empirical datasets are reported on in Section 3.

2.1. Empirical reference dataset and scans

The empirical photographic image dataset was acquired using imaging hardware of a sweet pepper harvest robot prototype, consisting of a uEye SE industrial camera (UI-5250RE-C-HQ PoE Rev.2, GigE, Germany) with resolution of 1600x1200 pixels and a lens with focal length of 4.16 mm (CMFA0420ND, Lensagon, Germany). The scene was illuminated with a matrix of white LEDs, flashed for 50 μs, producing a light level of approximately 200.000 lx at a distance of 50 cm from the crop. The distribution of the light was highly centered with a sharp falloff towards the edges in the field of view of the camera. By using the flash, the global illumination was suppressed, although this resulted in dark images.

From distances ranging from 50 cm to 10 cm (in 10 cm increments) in front of the plant stems, images were captured from −45, 0 and 45

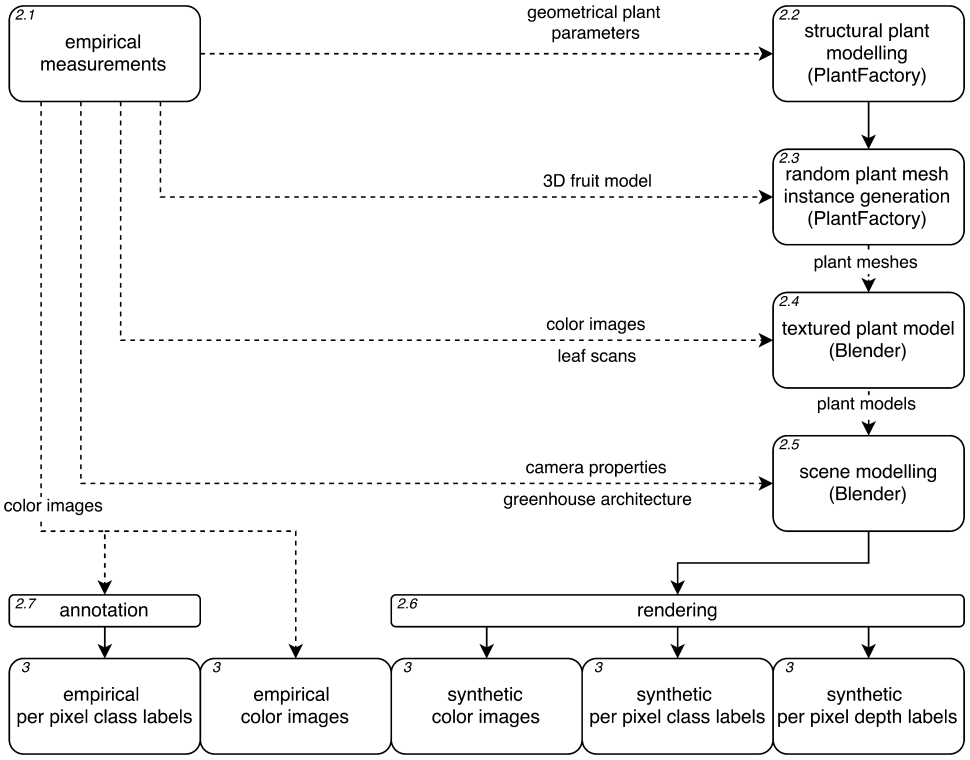

Fig. 1. Methodological flowchart to obtain empirical and synthetic datasets for agricultural plant scenes. Empirical measurements feed information to the plant and scene modelling processes. First a structural model of the plant was created using geometrical plant parameters. The structural model was then used to create instances of polygonal plant meshes that included a 3D mesh scans of fruit. The meshes were imported to Blender, where color images and leaf scans textured the model. Multiple plant models were thereafter included in scenes which mimiced the greenhouse architecture. Camera and illumination properties were added as well. The scenes were then rendered to obtain the sythetic data. A subset of the empirical color images was annotated intended for computer vision fine-tuning and verification test material. Each box is described in a section as referred by its top left number.

degree orientations with the horizontal plane. Furthermore, the camera was angled 20 degrees upwards as previous research suggested this would reduce occlusions (Hemming et al., 2014). In total 50 incremental positions of 20 cm along the row of plants were imaged in a 800×600 binned pixel resolution under two conditions; (i) facing towards and (ii) away from the sun. The increment size of 20 cm along the row was chosen to reflect the plant spacing in the greenhouse in order to approximate one new plant in the field of view per image. Overall weather conditions were clear and sunny with occasional clouds (cumulus humilis). At the position of the camera, the average irradiance was 6,000 lx under a clear sky and 5,000 lx when the sun was occluded by a cloud. 3D meshes were obtained of 3 sweet pepper fruit, cultivar Kaite (E20B.0073, Enza Zaden, the Netherlands) with a Spider 3D scanner (Artec, Luxembourg) with a 3D point accuracy of 0.05 mm. Scanning was performed manually by covering all perspectives using a rotating platform. Bottom occlusions were solved by using multiple poses of the same fruit and merging the resulting meshes automatically with Artec's software package. A set of 10 leaves were flattened and scanned using a consumer flatbed color scanner to obtain their shape, color and texture. At the nodes of the plant, occasionally there were cuts present where fruit were harvested. Frontal photographs were taken to obtain textures for this plant part as well as for the stem.

2.2. Structural plant modelling

A plant can be modelled functionally, structurally, or both (Vos et al., 2007, 2010). Functional plant models represent the interaction of internal and external plant processes. On the other hand, structural plant models focus solely on the physical appearance. When both types of models are combined, the influences of processes on the plant structure are taken into account. The scope of the current dataset was purely structural, as only fixated images irrespective of other influences were modelled.

Structurally, a plant consists of elements of various types and shapes. Based on these elements, a plant architecture can be defined globally or modularly. A global architecture is considered as a single shape and such an architecture inhibits plant variability on a detailed plant part level. In contrast, a modular plant architecture consists of the combination of three types of information; (i) the decomposition information that describes which components a plant consist of, (ii) the topological information that characterises the hierarchy and connection of components with others and (iii) the geometrical information that describes the sizes and poses of the components irrespective of other plant parts (Godin, 2000). With this decomposition either a regular or a multi-scale representation of a plant can be created. In the latter case self-similarity at different levels in the plant hierarchy can occur. In our approach, we created a regular structural modular plant model with a multi-scale representation for side shoots of the plant. This enabled detailed part modelling in which empirical measurements could be included and variability could be expressed.

2.2.1. Decomposition of Capsicum annuum

We decomposed the sweet pepper plant in the following plant parts: main stem section, nodes, sideshoot, leaf stem, leaf, peduncle, fruit and flower. To facilitate robotic harvesting, our model focussed on the generative stage of the crop only. At this stage most flowers have been pollinated and only fruit remain. Therefore the flower was omitted from the model.

2.2.2. Topology of Capsicum annuum

In Fig. 2 a typical section of a sweet pepper plant is shown, with a node in the center. From this empirical situation, we defined our hierarchy of plant components, as shown in Fig. 3a.

2.2.3. Geometry of Capsicum annuum

The geometry of a plant describes its parts in terms of dimensions and poses irrespective of other parts. Similar to other approaches for *Capsicum annuum* (Ballina-Gomez et al., 2013; IPGRI, 1995), 22 relevant plant parameters were identified that capture the geometry between plant parts. These parameters were measured in the early production season (April) on 15 plants of the same cultivar as used for the collection of the empirical image dataset and the scanned fruit and leaves. The collection of the measures included length, width, diameter and angles. Top view angles of plant parts were measured around the

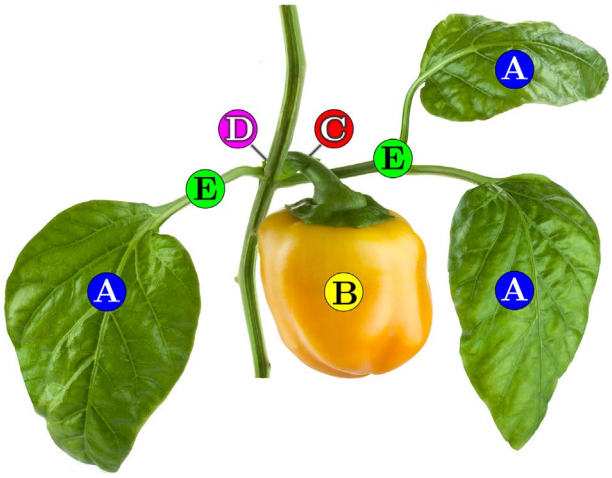

Fig. 2. A section of a *Capsicum Annuum* plant with one node in the center from which all plant parts grow. Plant parts are: (A) leaf, (B) fruit, (C) peduncle, (D) a node at a stem section and (E) sideshoot or leaf stem.

stem, counter-clockwise starting from the anterior side of the plant (hence perpendicular from the aisle towards the plant row) as depicted in Fig. 3b. Side view angles were measured counter-clock-wise, perpendicular with the floor as reference plane as shown in Fig. 3c. Results of the measurements are provided in Table 1 with the angular distributions plotted in Fig. 4.

2.3. Plant instance modelling

To model the plant and create object meshes, the commercial software PlantFactory 2015 Studio (EON Software, 2016) on OSX 10.10 was used. Originally intended for realistic game and video production modelling, it includes functionality to generate randomised plant instances based on plant parameter distributions, which can be exported as mesh models.

Randomised instances of sweet pepper plant meshes were generated by a procedural algorithm within PlantFactory 2015. The structural modular plant topology of Fig. 3a was used as a guideline. In Fig. 5 the procedural structure of the algorithm is shown. Plant part parameters of the obtained geometry from Section 2.2.3 were used. The metric parameters were based on the averages and standard deviations. For the angular values, their distributions were used as shown in Fig. 4, imported to PlantFactory as a curve. Per plant, 40 stem parts were generated and vertically concatenated, resulting in a mesh of approximately 4 m in height. Mesh instances were manually checked for the occasional inconsistency when a fruit intersected with other meshes. Those instances were replaced. Each mesh instance was exported in the open Wavefront OBJ 3D model format. In Fig. 6 example meshes are shown with 5 stem parts.

2.4. Textured modelling from polygon meshes

The randomly generated plant meshes were imported in the open-source software Blender 2.77a (Foundation, 2016; Kent, 2015), which for our purpose supports the composition, simulation and rendering of 3D scenes. To the polygon meshes color, texture, local mesh displacement (bump mapping), glossiness and specularity properties were added.

For the leaves and the cuttings on the stem, where fruit were previously removed, the photocopies were used. For the other plant parts, a color overlay was applied based on average colors from patches of corresponding plant parts in the empirical data set. In order to simulate fruit maturity levels, a color gradient from unripe green to ripe yellow was projected on a noise texture. The gradient and noise parameters

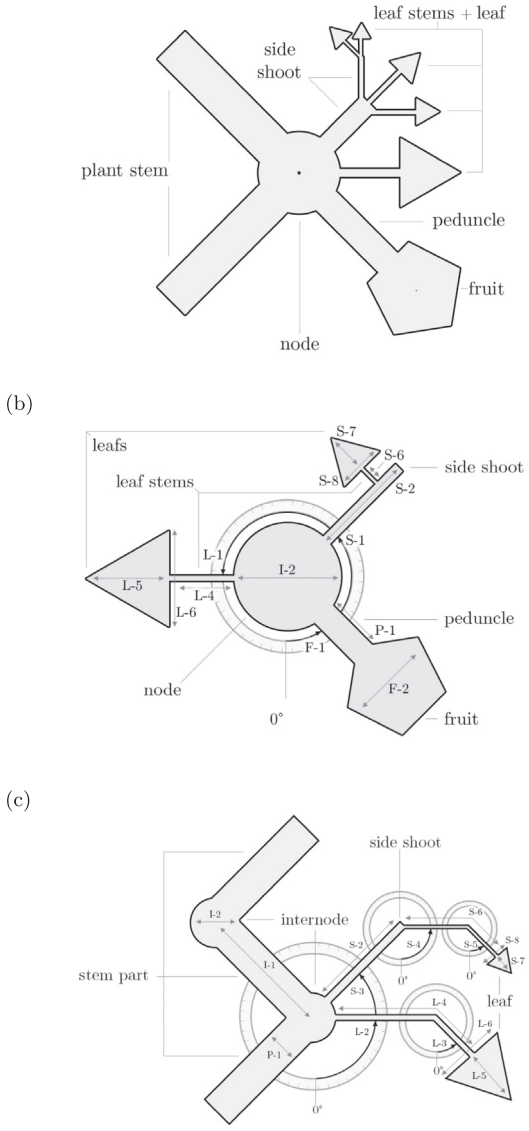

Fig. 3. (a) Structural modular plant topology of *Capsicum annuum*, showing the hierarchy and connection of plant elements. We considered the node as the central part of each section of the plant. The node joins 2 stem sections and connects to (i) sideshoots, (ii) leaf stems and leaves, (iii) peduncle and fruit. A side-shoot can have up to 3 leaves or new side shoots at its end. This topology omits the flower, which has grown into a peduncle and fruit. (b) Schematic top view of *Capsicum annuum*. An intersection is shown at the level of an node of the plant. Parts are connected around the plant: (i) peduncle and fruit, (ii) leaf stem with leaf, (iii) side shoot with leaf stem and leaf. Geometrical plant parameter measures are depicted, as reported in Table 1. Angles F-1, S-1 and L-1 are measured counter-clockwise starting from the anterior side of the plant in respect to the greenhouse aisle. (c) Schematic side view of *Capsicum annuum*. Geometrical plant paramter measures described in Table 1 are depicted. Angles L-2, L-3, S-3, S-4 and S-5 are measured perpendicular to the ground and counter-clockwise.

were manually determined with reference to additional unripe to ripe fruit images taken in the greenhouse. To simulate local leaf deformations (wobbles), a noise texture was used for bump displacement mapping of the leaf meshes. The parameters of this displacement were visually determined in comparison with the empirical image dataset. The polygons of the stem parts were displaced with a flattened 3D scan of the stem, processed with an edge filter to simulate their vertical grooves. For each set of plant parts, light reflection was manually modelled by adding glossiness and specularity modifiers. To add a background, a partial cloudy sky was generated. The sun was modelled as a light emitting sphere and was placed in the background with a lens flare effect. To complete the scene, the vertical wire used in

Table 1
Geometrical *Capsicum annuum* plant parameters per plant part that were measured in 15 plants at 115 node positions. Averages and standard deviations were used for modelling in PlantFactory. Descriptive names are displayed in Figs. 3b and c.

Name	Plant Part	Measure	Average (mm)	SD (mm)
I-1	NODE	Internode length	99	14
I-2	NODE	Node width	16	2
L-4	LEAF	Stem length	99	14
L-5	LEAF	Leaf length	16	2
L-6	LEAF	Leaf width	112	11
S-1	SIDESHOOT	Length	48	30
S-6	SIDESHOOT	Leaf stem length	97	12
S-7	SIDESHOOT	Leaf length	123	16
S-8	SIDESHOOT	Leaf width	102	12
P-1	PEDUNCLE	Length	46	6
F-2	FRUIT	Diameter	84	11
P-1	PLANT	Stem diameter	9	1

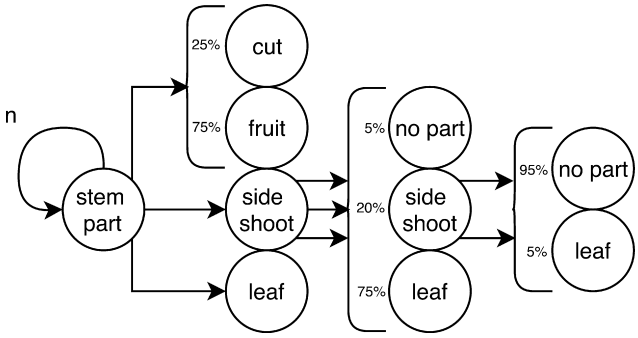

Fig. 5. Procedural structure for plant instance generation as implemented in PlantFactory. Each node represents a plant part for which another plant part is generated for each arrow attached. Brackets imply a random choice was made between plant parts included with indicated probabilities. The stem part grew another stem part *n* times.

horticultural practice to support the plant was modelled by applying a white texture and bump map that curled around the stem.

2.5. Scene modelling

Using Blender, a scene was modelled that represented a part of a Dutch commercial high-tech sweet pepper greenhouse (Bac et al., 2016) by reproducing the plant growing architecture as a double row of plants. In Fig. 7 a frontal and top perspective view of the scene is shown. For each scene, 7 randomised plants were generated, imported and positioned in a row with a 20 cm spacing in between. Similar to horticultural practice, a second row of 6 random plants was added 20 cm behind the first, shifted 10 cm in parallel. Due to memory constraints, up to 13 plants could maximally be instanced in each scene. To virtually collect data, a simulated camera and illumination were added with similar optical properties as the hardware described in Section 2.1. Blender allows to set the focal length and sensor size according to the hardware manufacturer's specifications. The illumination intensity and distribution was empirically matched with the reference color images (see Fig. 8).

The simulated image acquisition hardware followed an arc path upwards at a fixed distance of 40 cm from the center of the 4th plant in the row of 7 plants. Along this path, 250 frame triggers were equally spaced. The camera was placed under an angle of 20 degrees looking

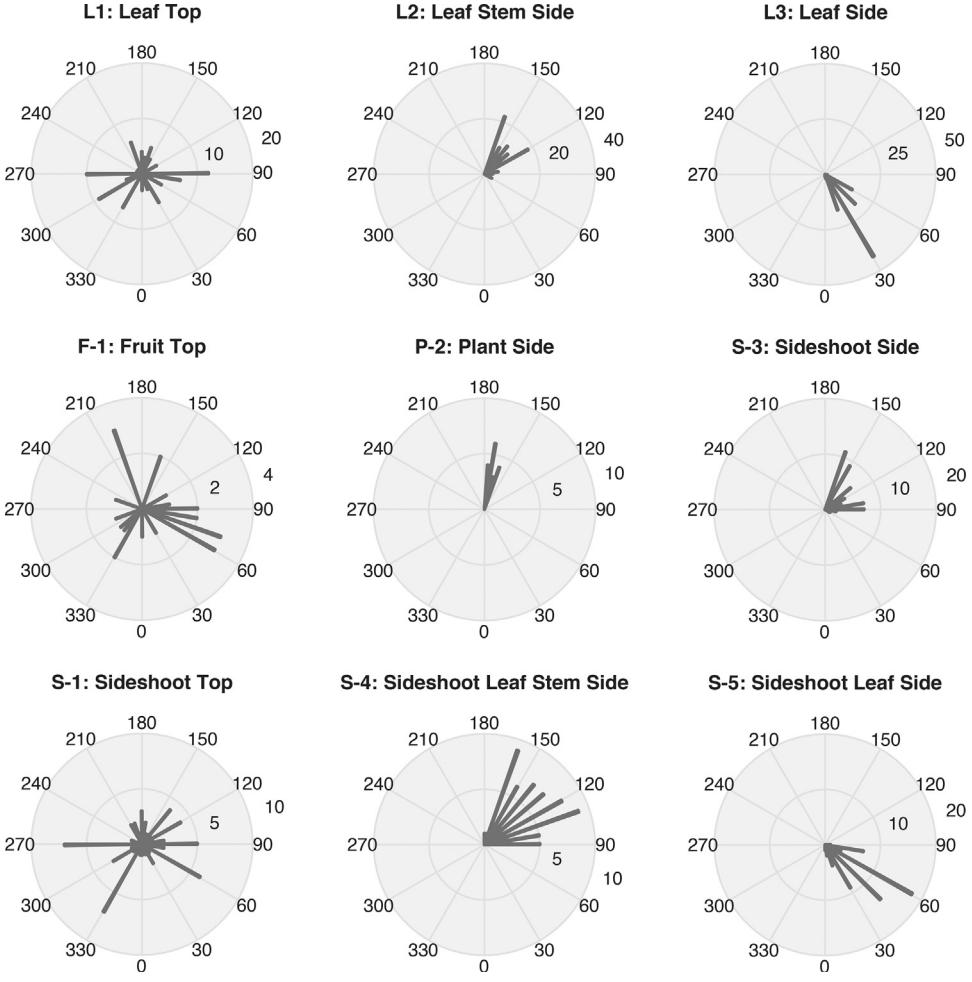

Fig. 4. Angle distributions (number of occurences per 0–360°) for *Capsicum annuum* plant parameters as measured of 15 plants according the schematics in Fig. 3b and c. P-2 measured the angle of the total plant in the plane perpendicular to the plant rows.

Fig. 6. Perspective views of three meshes of randomly generated *Capsicum annuum* instances with 5 stem parts. Color encodes surface normals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

upwards, as also used when obtaining the empirical dataset in the greenhouse.

To create multiple scenes per set of 7 plants, each plant was translated 1 position in the row until all 7 plants had once occupied the center of the row at the location where the simulated data collection occurred. The plants in the second back row also translated along. Furthermore, 6 additional unique scenes were generated with 7 new plants. Hence in total, 42 scenes were created (6 scenes, 7 positions per scene).

2.6. Rendering

For each scene, render computations (color, class label and depth label images) were run on the Odyssey supercomputer cluster supported by the Research Computing Group of the FAS Division of Science at Harvard University. Each frame was assigned to a single computing node with 16 cores, rendering one frame in 10 min on average.

To obtain the per-pixel ground truth label images in which each plant part was represented by a unique color encoding, the scene was duplicated and the color mapping of each plant part was replaced by primary or secondary colors. The background of this duplicate scene was set to black. To avoid interaction of colors in the scene, which would result in more colors than labels, the virtual camera was set to only register a single direct ray of light without bounces. Unfortunately, rendered edges between plant parts still interpolated colors. Therefore the synthetic image labels were post-processed in the commercial image processing software package Halcon 12 (MVTech, 2016) by removing any interpolated color pixels and replacing them with the most frequent neighbour color in a 3×3 patch If this convolution failed in case the majority of the neighbouring pixels also had been interpolated, the window was enlarged until all pixels were equal to one of the class labels. Furthermore, the colors were replaced by grayscale values in a single channel to finally reduce the label image size by 98%.

Ground truth depth images were rendered separately per frame by using the mist environment variable in Blender. For each pixel in the image, the light ray distance between the object and the camera's projection center was obtained. Hence encoded distances were not equal to real world XYZ-coordinates, which could not be obtained due to the absence of corresponding camera poses. The distance was encoded in image grayscale values, ranging from 0 to 255. In this image, the distance in centimetres for each pixel z_p could be recovered by the function $z_p = \left(\frac{255 - I_p}{\left(\frac{2 * 150}{255} \right)} \right)$, where the intensity of a pixel I_p decreases from the maximum intensity 255 with a factor based on the range of the mist in the render, which was set at 150 cm. The depth images were also rendered in a colorscale with high contrasts for intuitive viewing. The resolution in depth of this ground truth is coarse, though an exact representation could be obtained if needed by exporting the scene's Z-buffer in Blender to the OpenEXR (Kainz et al., 2004) linear format.

2.7. Annotation

The empirical image set obtains in the greenhouse contains 750 images at a range of 5 distances. From the 150 images taken at 40 cm, 50 images were annotated using Photoshop CC (Adobe, 2016) by manually outlining and coloring plant part classes. The suction cup of the robot's end-effector occluded the image and was labeled as background. Note that unlike its synthetic counter-part, ground truth in dark areas of the images were hard to manually discern and annotate. Hence only parts were annotated that could be clearly recognised. Average manual annotation time was 30 min per image.

2.8. Semantic segmentation

We gathered evidence for our hypothesis that synthetic bootstrapping and fine-tuning with a small empirical dataset can be effective by running 5 experiments with a semantic segmentation deep learning network, using the DeepLab framework (Papandreou et al., 2015) based on Caffe (Jia et al., 2014).

Specifically we used the Deeplab VGG-16 Vanilla model (Papandreou et al., 2015) with a receptive field of 128 pixels and a stride of 8 pixels. The hyperparameters of the network were manually optimised as suggested by Bengio (2012) and resulted in using Adaptive Moment Estimation (ADAM) (Kingma and Ba, 2015) with $\beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 10^{-8}$ and a base learning rate of 0.00005 for 30,000 iterations with a batch size of 10.

For each experiment we changed the dataset composition (synthetic or empirical images) for learning, fine-tuning or testing. The following compositions were investigated. Brackets indicate image indexes used.

- **A** *Train: synthetic (1–8750). Test: synthetic (8851–8900).*
  This experiment was run to obtain a performance reference point of the model when having access to a large and detailed annotated dataset for this domain.
- **B** *Train: synthetic (1–8750). Test: empirical (41–50).*
  To determine to what extent a synthetically trained model can generalise to a similar set in the same domain without fine-tuning.
- **C** *Train: empirical (1–30). Test: empirical (41–50).*
  As a reference to see if the model can learn using a small dataset, using empirical data.
- **D** *Train: PASCAL VOC. Fine-tune: empirical (1–30). Test: empirical (41–50).*
  To compare the effect of bootstrapping with a non-related dataset.
- **E** *Train: synthetic (1–8750). Fine-tune: empirical (1–30). Test: empirical (41–50).*
  To assess the effect of bootstrapping with a related dataset.

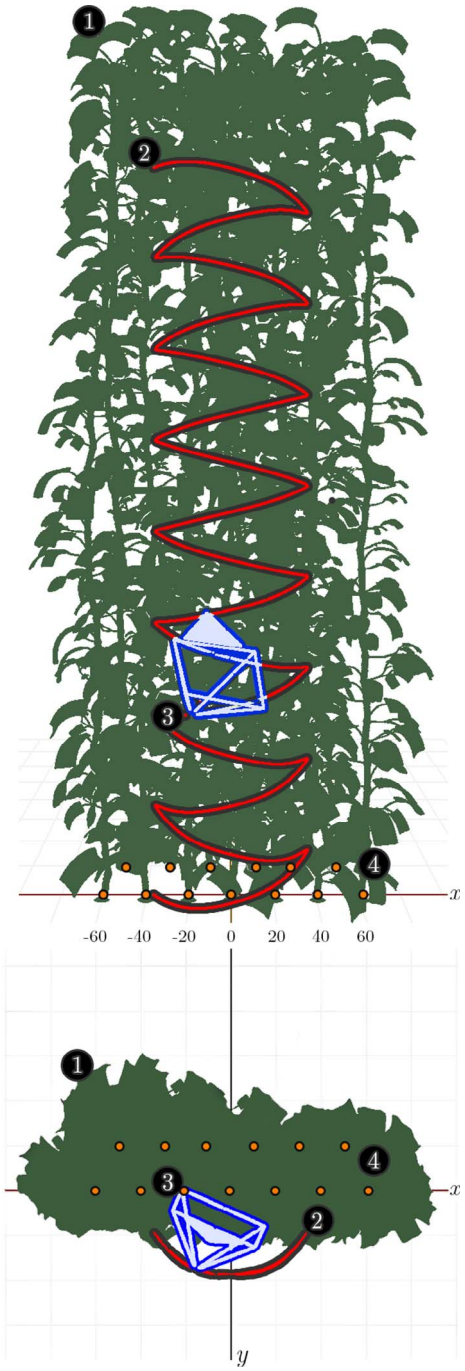

Fig. 7. Front and top perspective views of a sweet pepper crop scene in Blender, without textures. The camera plus illumination (blue, 3) and their path (red, 2) were placed in front of 2 rows of plants (green, 1). The position of each plant on the floor plane is indicated with a dot (orange, 4). Grid spacing is 20 cm. Note that leaves are rectangular; though during render time the shape of the leaf was refined by applying an opacity map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For each experiment, overfitting was prevented by selecting the optimal model by periodically checking the model's performance on an separate validation set. For the synthetic set, these were unique images (8751–8800) from the 6th scene. For the empirical set, these were images of unique plants (31–40).

2.8.1. Performance evaluation

To calculate the performance of our method and to enable equal comparison of future methods, we used the Jaccard Index similarity coefficient as an evaluation measure. This index is also known as the intersection-over-union (IOU) (He and Garcia, 2009) and is widely used for semantic segmentation evaluation (Everingham et al., 2010). The measure is defined in Eq. (2), where the mean IOU per class equals the intersection of the semantic segmentation and the ground truth divided by their union. To derive the measure, a pixel-level confusion matrix C was calculated first for each image I in dataset D:

$$C_{ij} = \sum_{I \in D} |\{p \in I | S^I_{gt}(p) = i \wedge S^I_{ps}(p) = j\}|, \tag{1}$$

where $S^I_{gt}(p)$ is the ground truth label of pixel p in image I and $S^I_{ps}(p)$ is the predicted label. This implies that C_{ij} equal the number of predicted pixels i with label j. The IOU can then be derived as an average for each class L by:

$$\text{IOU} = \frac{1}{L} \sum_{i=1}^{L} \frac{C_{ii}}{G_i + P_i - C_{ii}}, \text{where} \tag{2}$$

$$G_i = \sum_{j=1}^{L} C_{ij} \text{ and } P_j = \sum_{i} C_{ij} \tag{3}$$

Hence G_i denotes the total number of pixels labeled with class i in the ground truth and P_j the total number of pixels with prediction j in the image.

3. Results

A synthetic image dataset of 10,500 images was generated using 6 unique scenes and an empirical dataset of 750 images was obtained. A pixel-level ground truth segmentation of 8 classes was created automatically for all images in the synthetic dataset and manually for 50 images in the empirical dataset. This section first provides example images of both sets after which the sets will be compared on differences in color, class and spatial distribution to verify to what extent the requirement of similarity was met for our first hypothesis. To answer the second hypothesis that such datasets for agriculture are a valid and valuable tool for computer vision learning methods, results of the 5 experiments will be presented.

3.1. Datasets description

In Fig. 9, examples of real and synthetic images are shown with their corresponding ground truths. The datasets and the source material can be found at: https://doi.org/10.4121/uuid:884958f5-b868-46e1-b3d8-a0b5d91b02c0

3.1.1. Datasets comparison

Results of the comparisons between the synthetic and empirical datasets are presented in this section. First, pixel frequencies of classes and their spatial distributions between both sets were compared because such distributions reflect if the structure of the object we intended to model was similar. In Fig. 10 the pixel class frequencies are shown for both datasets. To investigate their spatial distribution, the normalised per class pixel label distributions are shown in Fig. 11.

Property distributions within classes themselves was another comparative perspective, for example color distributions. Some computer vision and learning methods are sensitive to object color, affecting the generalisation of the method to new images with different color distributions. In Fig. 12 the color spectrum for each plant part in both sets are shown.

The spectra were obtained by transforming the color images to hue, saturation and value (HSV) colorspace. The hue channel in this image represented for each pixel which color on the visible spectrum was present, irregardless of illumination and saturation intensity. Due to the heterogeneous illumination distribution in the images, the dark edges of the image were overrepresented with colors in the end of the

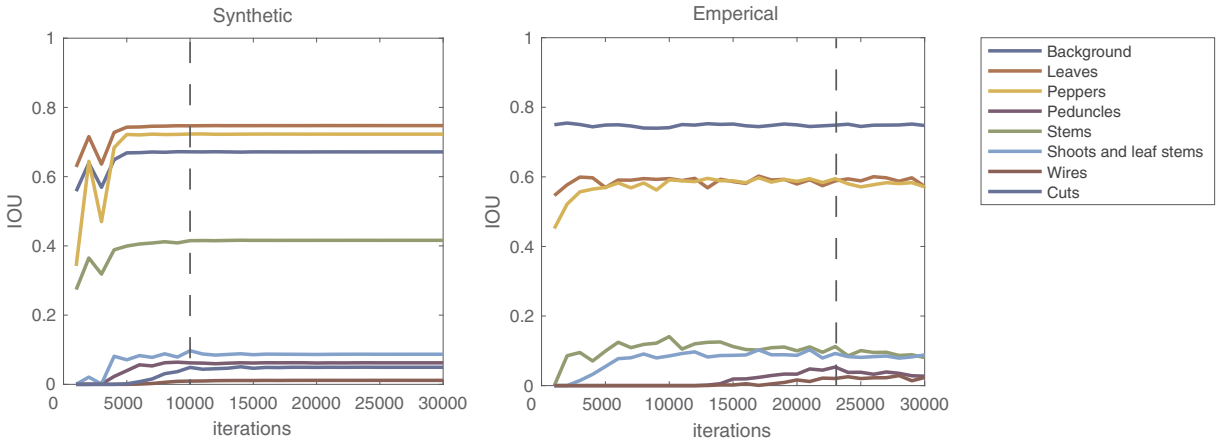

Fig. 8. Average IOU over the validation set per class per iteration in validation set of synthetic bootstrapping (left) and empirical fine-tuning (right). Dashed vertical lines indicate at which iteration the model was fixated before training stabilized or overfitted. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

spectrum. For this reason, we focussed the color analysis on a 300x300 pixel patch in the well illuminated middle of the image.

Intensity was another dimension of interest. With an average intensity of 37 and standard deviation of 12, comparison of differences between plant parts was coarse. Instead we investigated the average spatial intensity distribution over all images in each set as shown in Fig. 13. This enabled us to verify the similarity of the simulated illumination heterogeneity with the empirical set.

3.2. Semantic segmentation results

IOU results for experiment A through E are shown in Fig. 14, separated by class. Segmentation results for the best performing network on synthetic data (A) and empirical data (E) are shown in Fig. 15.

4. Discussion

In this paper we aimed to help the computer vision performance in agriculture towards a state-of-the-art level required for the next generation robotics, e.g. for harvesting, disease detection and phenotyping. Our conjecture of the cause of the low performance in this domain was the unavailability of detailed large annotated label data sets, necessary for most novel machine learning approaches. Recently, generating and training on synthetic image datasets has proven to be a popular and effective solution in other domains.

To extend such an approach to the domain of agriculture, we have described a novel methodology to generate synthetic images of plants. Although the modelling can be time consuming itself, it facilitates the generation of large-scale and more detailed datasets under a broader set of conditions, e.g. different illumination conditions, perspectives or sensors. This would otherwise not be feasible to obtain due to the required large annotation effort. Our approach is generic and applicable to any crop with a consistent modular plant architecture after obtaining an accurate and exhaustive definition of the corresponding plant parameters (Vos et al., 2010, 2007).

Our dataset is an important contribution for the availability of a variety of datasets in the computer vision community so that methods can verify their robustness and generalisation. Currently the focus in the research community is on tuning and validation using type restricted datasets, e.g. human or urban scenes. Our datasets provide a use-case for detailed hierarchical part recognition in bio-related fields.

In this paper we presented a modelling example of a single point in time for a single variety of *Capsicum annuum*. However, our methodology for generating plant models is extendable to include plant parameters for plants under multiple stages of growth. The pipeline allows to interpolate between seasonal plant parameters to generate plants in different growth stages. However, this would require more empirical measurements and there is a trade-off between modelling accuracy over the season and the accuracy that is desired from the application. For our purpose of using machine learning, we assumed a single point in time for modelling would already be sufficient to improve synthetic based learning.

As we did not have depth data in our empirical dataset, it was not possible to test any hypothesis regarding machine learning that made use of the synthetic depth data. We added the methodology for generating depth data for future research and as an example for the community how to obtain such data.

In the following subsections we zoom in on our hypotheses (i) that stated that we can create synthetic images similar to empirical images by discussing to what extent our requirement of similarity was met and (ii) that a synthetically bootstrapped model can be used for improved learning when only fine-tuned on a small annotated empirical dataset.

4.1. Synthetic and empirical set similiarity

At the beginning of this paper we posed that the synthetic images should be similar to the empirical images. We can both qualitatively and qualitatively observe the differences between sets. The former is subjective and it must be noted that human perceptual evaluation of images often employs sensory completion to make up for differences or absences (McNamara, 2001). Nonetheless it is valuable to compare the sets in this regard because it provides clues for objective comparison and possible improvements.

4.1.1. Qualitative comparison

When visually compared we noted slight differences in color. This can be explained by the manual color tuning process in which it was hard to find colors close to the empirical situation, given that only the result of the illumination interaction effects of the materials, the light source and the environment could be observed. Future research should include methods for automated color optimisation. Also obtaining a calibrated color ground truth is recommended in combination with proper camera color calibration.

Another difference was the greater perceived variance of shapes and poses in the empirical set. This was the result of excluding part shape variance, e.g. not taking into account an exhaustive set of leaf curls and poses or local side-shoot deformations. Hence, the plant parameters were not adequately capturing all the variation. In forthcoming research we suggest to include also intra plant part pose variations and deformations.

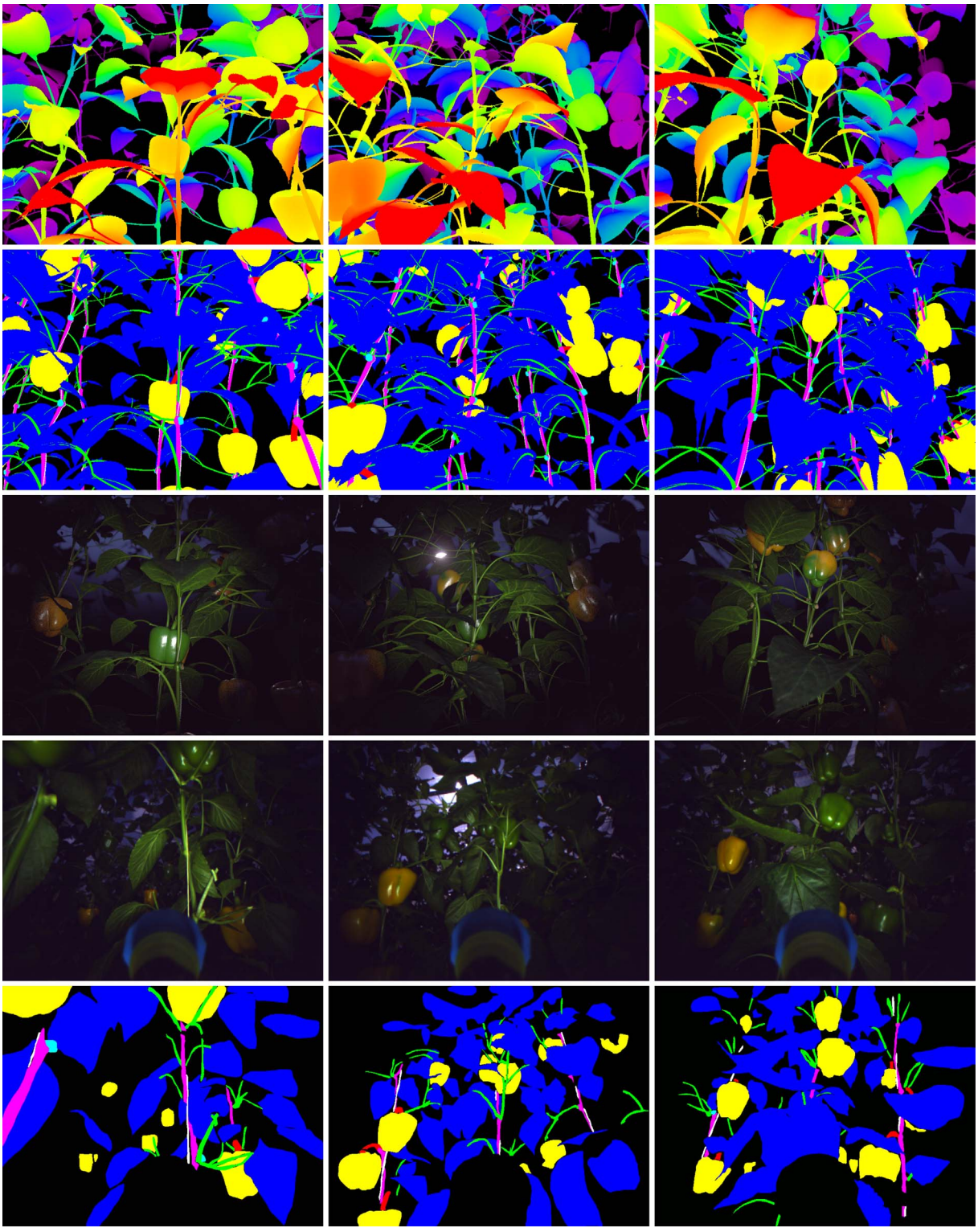

Fig. 9. Three examples of the synthetic and empirical color images and their corresponding ground truth labels. The first three rows contain column pairs of (i) synthetic ground truth depth labels, (ii) class labels and (iii) color images. The last two rows contain column pairs of (i) real images and (ii) ground truth class labels. Note the depth label has an arbitrary colorscale for intuitive viewing. Class labels: ● background, ● leaves, ● peppers, ● peduncles, ● stems, ● shoots and leaf stems, ○ wires and ● cuts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1.2. Quantitative comparison

Superordinate image similarity measures were previously not well defined in literature. However, studies of individual measures like color histogram comparison (Swain and Ballard, 1991) or shape measures (Mehtre et al., 1997) have been performed. In this paper, the similarity requirement was also not quantified in a single measure. Instead we looked at: (i) label set distributions, (ii) label image position distributions, (iii) part color spectra and (iv) illumination intensity distributions to gain a more quantitative insight into the similarity between sets.

i Within a dataset, label frequencies are often highly unbalanced (Caesar et al., 2015), resulting in neglected classes in some type of

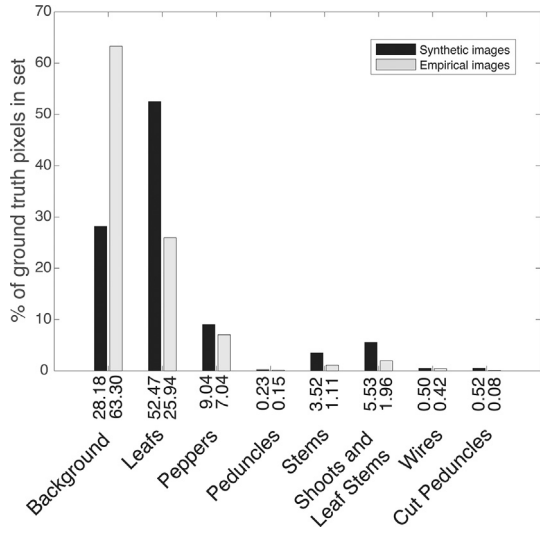

Fig. 10. Percentage of ground truth pixels per class, compared between real and synthetic datasets.

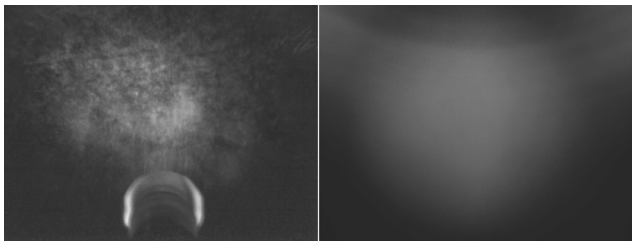

Fig. 11. Per class normalised pixel label distributions for the emperical and synthetic dataset. Each image was obtained by summation of class masks of all images in the set, divided by the maximum resulting pixel value.

computer vision approaches. To counter that effect, object occurrence statistics can be used for normalisation (Chawla, 2010). In Fig. 10 we observe an unbalance within and between the sets. The latter can partially be explained by the methodological difference in obtaining the ground truth in both cases. For the synthetic dataset, the labels of all pixels in the color image were computable irregardless of illumination. For the empirical dataset, a subjective decision on the label in the dark edge areas was often not possible.

ii In Fig. 11 we compared the plant part spatial distribution in the images between sets. Overall, the sparsity in the empirical set is notable and was due to the small size of the annotated set. In the synthetic set, we note 3 vertical hotspots of stem + wire classes whereas in the empirical set the spatial variance of these classes was higher. This can be explained by a more regular plant distance in the synthetic set, due to the fact that our model did not take intra-plant properties into account.

The distributions of the peppers between sets was similar, e.g. a hot spot in the center of both distributions and less to the left and right explained by the particular occlusions resulting from the chosen scanning path. Closely correlated with peppers are the peduncles, however the peduncle distributions were hard to compare due to their sparsity.

Lastly, there was a difference between leaf class distributions. Whilst empirically more centered, in the synthetic set there was a higher occurrence at the top of the image. This can be explained by improper modelling the shapes and poses of the leaves, resulting in a discrepancy of silhouettes when viewed from a 20 degree upward angle.

iii Evaluating Fig. 12, we observe plant part color similarities but also differences in our sets.

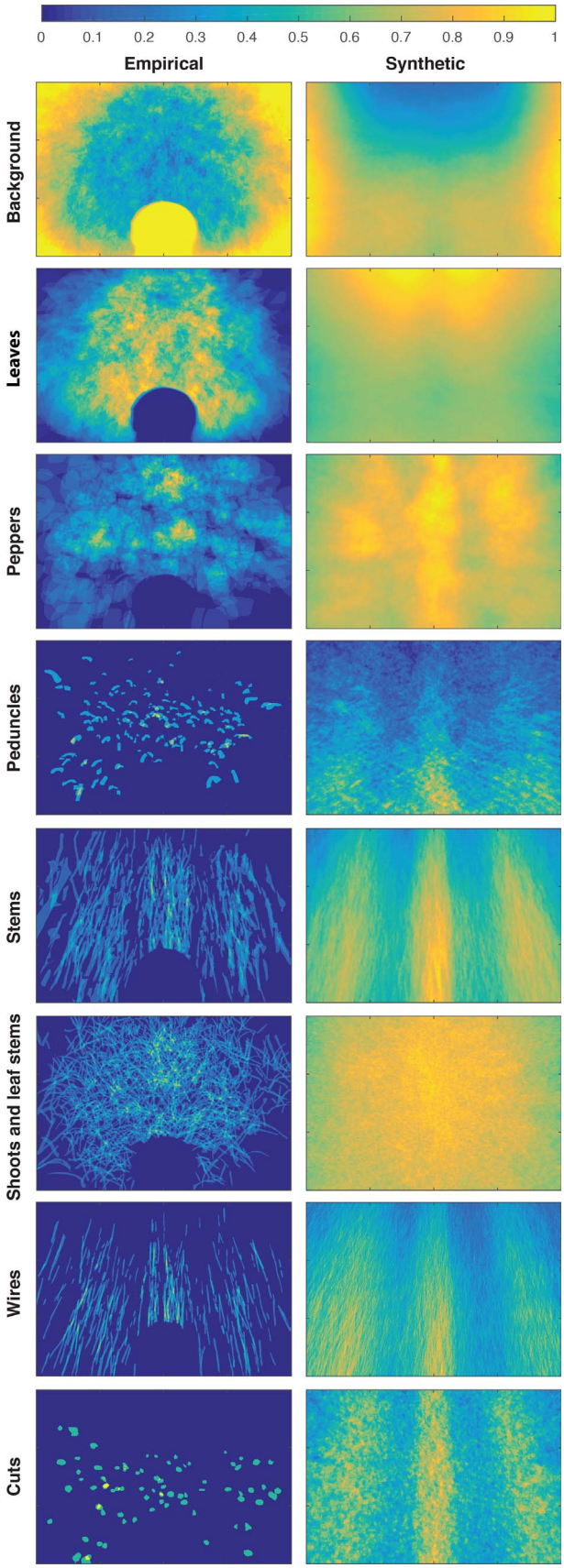

Fig. 12. Color distribution per plant part of synthetic and empirical images. The vertical axis shows the percentage of plant part color averaged over the image set. The horizontal axis shows the hue value in HSV colorspace. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

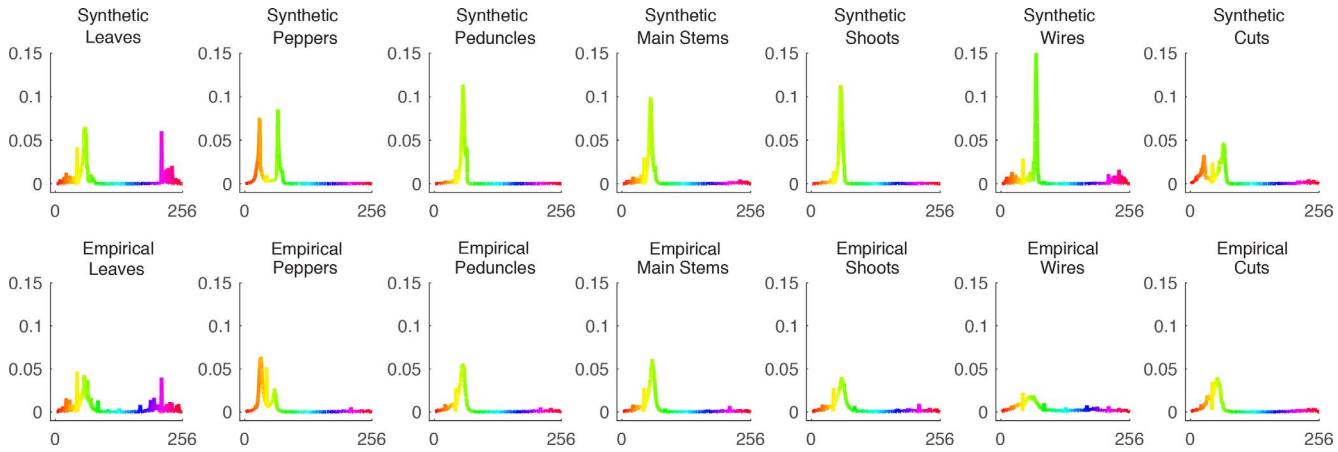

Fig. 13. Average illumination intensity distribution over all images in the emprical (left) and the synthetic (right) sets, with an average pixel intensity of 37 and 38 respectively. In the images shown here, the intensity of both images was doubled to increase contrast for the reader. (For interpretation of the referencens to color, the reader is referred to the web version of this article.)

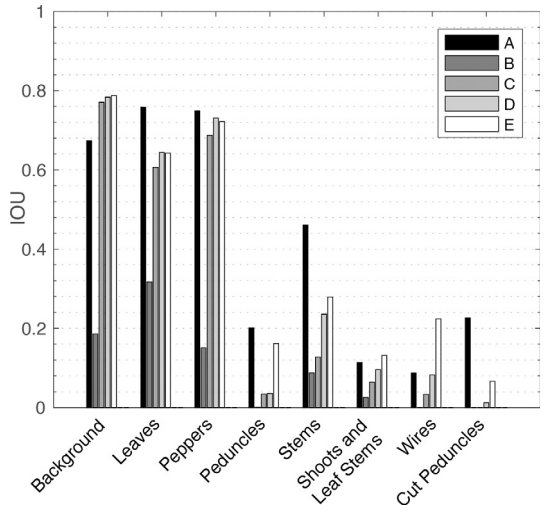

Fig. 14. Average IOU per class over the test set for experiments A through E.

Leaves: The proportion of the left spectrum of the leaf class was similar, though the synthetic leaves should contain relatively less greens. The purple spectrum on the right of this class can be explained by the outward position of some leaves, leading to underexposure and therefore taking on the purple global background illumination color.

Peppers: The pepper colors seem far off at first, though the green peak height might be caused by the improper modelling of the percentage of green (50%) and yellow peppers (50%). In the empirical set, more ripe fruit were present because a part of the crop was selected in the greenhouse with abundance of yellow peppers to increase the spatial density of robot harvest trials. Irregardless of this ripeness imbalance, the color of the ripe peppers in the synthetic set needs to be adjusted towards the yellow end of the spectrum.

Peduncles, stems and shoots: All were modelled too green and lack yellows.

Wires: The synthetic wire color distribution shows a peak in green. By inspecting the data this was likely the result by the rendering a color interpolation at the edges of the wire class and background stem class. Furthermore, a relative large part of the wire pixels were edge pixels and these pixels were included in the ground truth of the wire class. A thicker synthetic wire might increase color similarity.

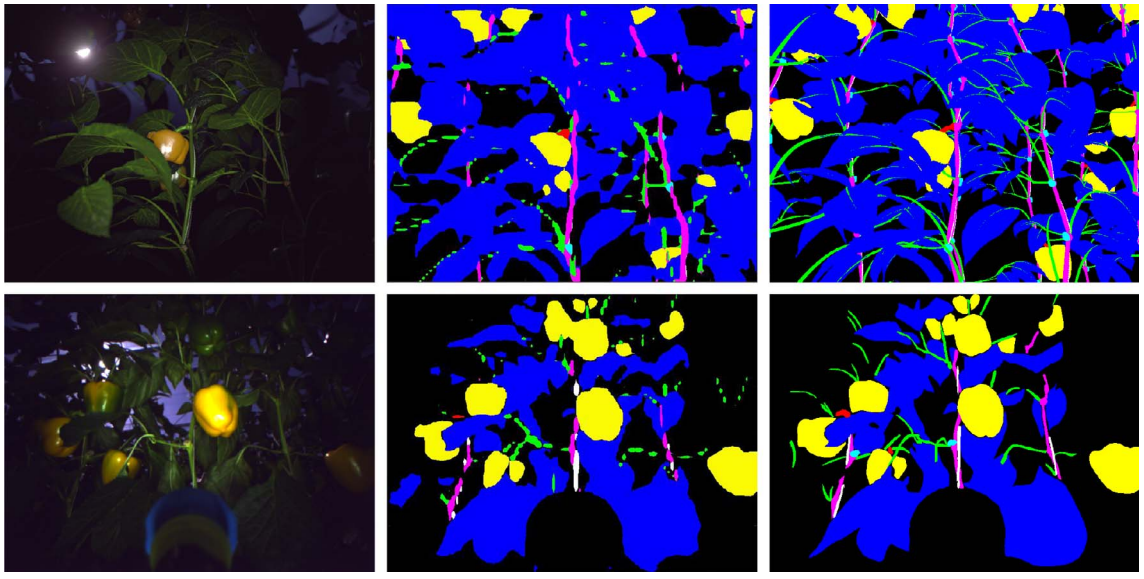

Fig. 15. Segmentation results for synthetic test set from experiment A (top row) and empirical test set from experiment E (bottom row). Color images (left), classification segmentation (middle) and ground truth (right) are shown in each row. Class labels: ● background, ● leaves, ● peppers, ● peduncles, ● stems, ● shoots and leaf stems, ○ wires and ● cuts.

Cuts: Although the cuts were textured with a photograph, the absence of color calibration most likely resulted in the difference we observe in this class.

iv When looking at the average illumination intensity distribution over all images in each set (Fig. 13) we observe a comparable heterogenous illumination distribution with a strong vignetting effect. This was caused by (i) an interaction of illumination hardware that focussed a centered beam on the scene and (ii) the lens type which had a default vignet. The average intensity of these images was similar.

For aspects discussed under both (i) and (ii), these differences might also be caused due to the empirical set measurement only consisting from frontal, −45 and +45 degree views, whereas the synthetic set included also intermediate waypoints that furthermore moved vertically per next frame, as depicted in Fig. 7.

Although the differences discussed in i-iv could be valuable for improving the similarity between image sets, their combined impact can only be evaluated in the perspective of a specific task performance, e.g. machine learning. In the context of computer vision, we can state that image sets are sufficiently similar when they in any manner can be used for improved recognition. Therefore we evaluated the similarity also in the context of the task of segmentation in Section 4.2, where we quantitatively compare the IOU performance of the 5 experiments.

4.2. Bootstrapping and segmentation

Although differences between the image sets exist, experiments A through E indicated that bootstrapping with related synthetic data improved the learning performance compared to solely training on a small empirical dataset or bootstrapping with non-related data. From each experiment, we concluded the following:

A The model trained and tested using the synthetic dataset provided a baseline performance on the task when having access to a large amount of detailed annotated data.

B Without any fine-tuning, the synthetically bootstrapped model did not generalise well to empirical data.

C Using only a small empirical dataset for training, the model learned to differentiate plant parts to a certain extent. However, this primarily holds for the classes background, leaves and peppers, with an average IOU of 0.68. We observe that the model learns the most frequent classes that were also most discriminative in color, e.g. black, dark green and yellow correspondingly. The other classes that were infrequent and overlapped in color with the frequent classes were segmented poorly, with an average IOU of 0.05.

D When bootstrapping with an non-related dataset (PASCAL VOC) and fine-tuning with empirical data, performance on empirical test images was increased over the previous experiments (B,C).

E When bootstrapping with our synthetic related dataset and fine-tuning with empirical data, the best performance was achieved testing on empirical images.

From the results we conclude that the increase in performance using synthetic data bootstrapping compared to the other approaches might be caused by the increased training sample number with high similarity that was made available to the CNN.

In the experiments we observed a correlation between class frequency and class performance, suggesting the model had a bias for class availability. Future efforts should be focussed on coping with this bias, for example using normalisation during the computation of the loss.

Our experiments and their conclusions are indicative because we could not prove there does not exist a different convolutional model architecture, hyper-parameter combination or initialisation per experiment that would have a better performance. However, the results present a starting point to determine how synthetic data can be used to improve segmentation performance.

Although plant part segmentation in reconstructed 3D models has previously been achieved in smaller plants (Golbach et al., 2016; Paproki et al., 2011), segmenting multiple plant parts from single 2D images previously remained unsolved. For plant robotics and phenotyping, the requirement of plant part localisation is currently a bottleneck (Minervini et al., 2015). Our results show a promising method for meeting this requirement when observing the final segmentations qualitatively.

Our work contributes to image segmentation challenges in the plant domain. In agricultural applications, our approach of segmenting individual plant parts in high detail will enable a large range of possibilities. For example, from leaf volume estimations in vineyards to a all kinds of phenotyping applications to determine plant parameters from images.

4.3. Conclusion

A new method for generating synthetic data sets for agricultural computer vision was presented. Based on empirical data, a sweet pepper plant model was created, randomised plant instances were generated and rendered to mimic realistic greenhouse conditions. Our hypothesis that with this approach we can create a synthetic image dataset similar to empirical images holds perceptually and qualitatively, though quantitatively there were differences in class and color distributions. However, we also stated that the requirement of similarity on the task, e.g. pre-training models for image segmentation. Our hypothesis that bootstrapping a convolutional neural network that fine-tunes on a small empirical dataset outperforms other methods of training has been confirmed by our experiments. Segmentation results show a promising next step for semantic part localisation in agriculture. Future efforts should be aimed in further optimising the network architectures, focussing on the performance of the infrequent classes. The datasets and their source material are publicly released and can be found at: https://doi.org/10.4121/uuid:884958f5-b868-46e1-b3d8-a0b5d91b02c0

Acknowledgements

This research was partially funded by the European Commission in the Horizon2020 Programme (SWEEPER GA No. 644313). The authors would like to thank prof.dr. R. D. Howe and dr. D. Perrin for their input of this research and making computing resources available. The authors declare that they have no conflict of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2017.12.001.

References

Adobe, 2016. Photoshop. <http://www.adobe.com/products/photoshop.html>.

Bac, C., Hemming, J., van Henten, E., 2013. Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper. Comput. Electron. Agric. 96, 148–162. http://dx.doi.org/10.1016/j.compag.2013.05.004. <http://www.sciencedirect.com/science/article/pii/S0168169913001099>.

Bac, C.W., van Henten, E.J., Hemming, J., Edan, Y., 2014. Harvesting robots for high-value crops: state-of-the-art review and challenges ahead. J. Field Robot. 31 (6), 888–911. http://dx.doi.org/10.1002/rob.21525.

Bac, C., Hemming, J., van Henten, E., 2014. Stem localization of sweet-pepper plants using the support wire as a visual cue. Comput. Electron. Agric. 105, 111–120. http://dx.doi.org/10.1016/j.compag.2014.04.011. <http://www.sciencedirect.com/science/article/pii/S0168169914000933>.

Bac, C.W., Roorda, T., Reshef, R., Berman, S., Hemming, J., van Henten, E.J., 2016. Analysis of a motion planning problem for sweet-pepper harvesting in a dense obstacle environment. Biosyst. Eng. 146, 85–97. http://dx.doi.org/10.1016/j.biosystemseng.2015.07.004. special Issue: Advances in Robotic Agriculture for Crops. <http://www.sciencedirect.com/science/article/pii/S1537511015001191>.

Ballina-Gomez, H., Latournerie-Moreno, L., Ruiz-Sanchez, E., Perez-Gutierrez, A., Rosado-Lugo, G., 2013. Morphological characterization of *Capsicum annuum* L. accessions

from southern Mexico and their response to the Bemisia tabaci-Begomovirus complex. Chilean J. Agric. Res. 73, 329–338.
Barth, R., Hemming, J., van Henten, E.J., 2016. Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation. Biosyst. Eng. 146, 71–84. http://dx.doi.org/10.1016/j.biosystemseng.2015.12.001. special Issue: Advances in Robotic Agriculture for Crops. <http://www.sciencedirect.com/science/article/pii/S1537511015001816>.
Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures, CoRR abs/1206.5533. <http://arxiv.org/abs/1206.5533>.
Benoit, L., Rousseau, D., Belin, tienne, Demilly, D., Chapeau-Blondeau, F., 2014. Simulation of image acquisition in machine vision dedicated to seedling elongation to validate image processing root segmentation algorithms. Comput. Electron. Agric. 104 (suppl C), 84–92. http://dx.doi.org/10.1016/j.compag.2014.04.001. <http://www.sciencedirect.com/science/article/pii/S0168169914000830>.
Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., 2013. A review of feature selection methods on synthetic data. Knowl. Inf. Syst. 34 (3), 483–519. http://dx.doi.org/10.1007/s10115-012-0487-8.
Caesar, H., Uijlings, J.R.R., Ferrari, V., 2015. Joint calibration for semantic segmentation, CoRR abs/1507.01581. <http://arxiv.org/abs/1507.01581>.
Chawla, N.V., 2010. Data Mining for Imbalanced Datasets: An Overview. Springer US, Boston, MA, pp. 875–876. http://dx.doi.org/10.1007/978-0-387-09823-4_45.
Cicco, M.D., Potena, C., Grisetti, G., Pretto, A., 2016. Automatic model based dataset generation for fast and accurate crop and weeds detection, CoRR abs/1612.03019. <http://arxiv.org/abs/1612.03019>.
Cordier, N., Delingette, H., Le, M., Ayache, N., 2016. Extended modality propagation: image synthesis of pathological cases. IEEE Trans. Med. Imaging(99). http://dx.doi.org/10.1109/TMI.2016.2589760. 1-1.
Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., aurelio Ranzato, M., Senior, A., Tucker, P., Yang, K., Le, Q.V., Ng, A.Y., 2012. Large scale distributed deep networks. In: In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, vol. 25. Curran Associates, Inc., pp. 1223–1231. <http://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks.pdf>.
Dittrich, F., Woern, H., Sharma, V., Yayilgan, S., 2014. Pixelwise object class segmentation based on synthetic data using an optimized training strategy. In: 2014 First International Conference on Networks Soft Computing (ICNSC), 2014, pp. 388–394. doi:http://dx.doi.org/10.1109/CNSC.2014.6906671.
Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. Int. J. Comput. Vision 88 (2), 303–338.
Everingham, M., Eslami, S.M., Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: a retrospective. Int. J. Comput. Vision 111 (1), 98–136. http://dx.doi.org/10.1007/s11263-014-0733-5.
Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A., 2005. Learning object categories from google's image search. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Vol. 2, pp. 1816–1823. doi:http://dx.doi.org/10.1109/ICCV.2005.142.
Blender Foundation, 2016. Blender. <https://www.blender.org>.
Godin, C., 2000. Representing and encoding plant architecture: a review. Ann. For. Sci. 57 (5), 413–438. http://dx.doi.org/10.1051/forest:2000132.
Golbach, F., Kootstra, G., Damjanovic, S., Otten, G., van de Zedde, R., 2016. Validation of plant part measurements using a 3d reconstruction method suitable for high-throughput seedling phenotyping. Mach. Vis. Appl. 27 (5), 663–680. http://dx.doi.org/10.1007/s00138-015-0727-.
Gongal, A., Amatya, S., Karkee, M., Zhang, Q., Lewis, K., 2015. Sensors and systems for fruit detection and localization: a review. Comput. Electron. Agric. 116, 8–19.
He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284. http://dx.doi.org/10.1109/TKDE.2008.239.
Hemming, J., Ruizendaal, J., Hofstee, J.W., van Henten, E.J., 2014. Fruit detectability analysis for different camera positions in sweet-pepper. Sensors (Basel) 14 (4), 6032–6044, 24681670. http://dx.doi.org/10.3390/s140406032. [pmid]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029692/>.
I.P.G.R.I. (IPGRI), 1995. Descriptors for Capsicum (Capsicum spp.).
Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. Available from: arXiv preprint arXiv:<1408.5093>.
Kainz, F., Bogart, R., Hess, D., 2004. The openexr image file format. GPU Gems: Programming Techniques, Tips and Tricks for Real-Time Graphics, R. Fernando, Ed. Pearson Higher Education.
Kent, B.R., 2015. 3D Scientific Visualization with Blender, 2053-2571, Morgan and Claypool Publishers. doi:http://dx.doi.org/10.1088/978-1-6270-5612-0.
Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization, CoRR abs/1412.6980. <http://arxiv.org/abs/1412.6980>.
Kondaveeti, H.K., 2016. Synthetic isar images of aircrafts. doi:http://dx.doi.org/10.5281/zenodo.48002.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. http://dx.doi.org/10.1038/nature14539. insight.
McNamara, A., 2001. Visual perception in realistic image synthesis. Comput. Graphics Forum 20 (4), 211–224. http://dx.doi.org/10.1111/1467-8659.00550.
Mehtre, B.M., Kankanhalli, M.S., Lee, W.F., 1997. Shape measures for content based image retrieval: a comparison. Inform. Process. Manage. 33 (3), 319–337. http://dx.doi.org/10.1016/S0306-4573(96)00069-6. <http://www.sciencedirect.com/science/article/pii/S0306457396000696>.
Minervini, M., Scharr, H., Tsaftaris, S.A., 2015. Image analysis: the new bottleneck in plant phenotyping [applications corner]. IEEE Signal Process. Mag. 32 (4), 126–131. http://dx.doi.org/10.1109/MSP.2015.2405111.
MVTech, 2016. Halcon. <http://www.halcon.com/>.
Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. J. Big Data 2 (1), 1–21. http://dx.doi.org/10.1186/s40537-014-0007-7.
Nasir, A., Rahman, M., Mamat, A., 2012. A study of image processing in agriculture application under high performance computing environment. Int. J. Comput. Sci. Telecommun. 3.
Papandreou, G., Chen, L.-C., Murphy, K., Yuille, A.L., 2015. Weakly- and semi-supervised learning of a DCNN for semantic image segmentation. In: ICCV.
Paproki, A., Fripp, J., Salvado, O., Sirault, X., Berry, S., Furbank, R., 2011. Automated 3d segmentation and analysis of cotton plants. In: 2011 International Conference on Digital Image Computing: Techniques and Applications, pp. 555–560. doi:http://dx.doi.org/10.1109/DICTA.2011.99.
Polder, G., van der Heijden, G.W., van Doorn, J., Baltissen, T.A., 2014. Automatic detection of tulip breaking virus (tbv) intulip fields using machine vision. Biosyst. Eng. 117, 35–42. http://dx.doi.org/10.1016/j.biosystemseng.2013.05.010. image Analysis in Agriculture. <http://www.sciencedirect.com/science/article/pii/S1537511013000883>.
Pradal, C., Fournier, C., Valduriez, P., Cohen-Boulakia, S., OpenAlea: scientific workflows combining data analysis and simulation. In: SSDBM 2015: 27th International Conference on Scientific and Statistical Database Management, San Diego, United States, 2015. doi:http://dx.doi.org/10.1145/2791347.2791365. URL <https://hal.archives-ouvertes.fr/hal-01166298>.
Pradal, C., Artzet, S., Chopard, J., Dupuis, D., Fournier, C., Mielewczik, M., Ngre, V., Neveu, P., Parigot, D., Valduriez, P., Cohen-Boulakia, S., 2017. Infraphenogrid: a scientific workflow infrastructure for plant phenomics on the grid. Fut. Gen. Comput. Syst. 67 (suppl C), 341–353. http://dx.doi.org/10.1016/j.future.2016.06.002. <http://www.sciencedirect.com/science/article/pii/S0167739X16301820>.
Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A., 2016. The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes.
Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T., 2008. Labelme: a database and web-based tool for image annotation. Int. J. Comput. Vision 77 (1), 157–173. http://dx.doi.org/10.1007/s11263-007-0090-8.
Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. Deepfruits: A fruit detection system using deep neural networks. Sensors 16 (8), 1222. http://dx.doi.org/10.3390/s16081222. <http://www.mdpi.com/1424-8220/16/8/1222>.
Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A., Blake, A., 2013. Efficient Human Pose Estimation from Single Depth Images. Springer, London, London.
EON Software, 2016. Plant factory. <http://www.plantfactory-tech.com/>.
Swain, M.J., Ballard, D.H., 1991. Color indexing. Int. J. Comput. Vision 7 (1), 11–32. http://dx.doi.org/10.1007/BF00130487.
Trask, A., Gilmore, D., Russell, M., 2015. Modeling order in neural word embeddings at scale, CoRR abs/1506.02338. <http://arxiv.org/abs/1506.02338>.
van der Heijden, G., Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., van Eeuwijk, F., Glasbey, C., 2012. Spicy: towards automated phenotyping of large pepper plants in the greenhouse. Funct. Plant Biol. 39 (11), 870–877. http://dx.doi.org/10.1071/FP12019.
Vos, J., Marcelis, L.F.M., Visser, P., Struik, P.C., Evers, J., 2007. Functional-Structural Plant Modelling in Crop Production. Springer, Netherlands.
Vos, J., Evers, J.B., Buck-Sorlin, G.H., Andrieu, B., Chelle, M., de Visser, P.H.B., 2010. Functional-structural plant modelling: a new versatile tool in crop science. J. Exp. Bot. 61 (8), 2101–2115. http://dx.doi.org/10.1093/jxb/erp345. arXiv:http://jxb.oxfordjournals.org/content/61/8/2101.full.pdf+htm. <http://jxb.oxfordjournals.org/content/61/8/2101.abstract>.