Effectiveness of Training with Procedurally Generated Synthetic Images of Crop Plants

Nazifa Azam Khan[1], Mikolaj Cieslak[2], Mark Eramian[1], Ian McQuillan[1]
nazifa.khan@usask.ca, msciesla@ucalgary.ca, mark.eramian@usask.ca, mcquillan@cs.usask.ca
[1]Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[2]Department of Computer Science, University of Calgary, Calgary, Alberta, Canada

Abstract

Artificial neural networks are often used to identify features of crop plants. However, training their models requires many annotated images, which can be expensive and time-consuming to acquire. Procedural models of plants, such as those developed with Lindenmayer-systems (L-systems) can be used to create visually realistic images, where annotations are implicitly known. These synthetic images can either augment or be used on their own instead of using real images in training neural networks for phenotyping tasks. The objectives of this paper are two fold. Firstly, we explore the degree to which realism in the synthetic images improves prediction. Secondly, we systematically vary amounts of real images used for training in both maize and canola to better understand situations where only synthetic images generated from L-systems can accurately predict phenotypic properties on real images. We achieved a mean absolute error (MAE) of 1.05 in predicting leaf count in maize and of 1.59 in predicting inflorescence branch count in canola using only synthetic images with U-Net. Furthermore, predictions made with only synthetic images as training data were improved by almost a ten-fold factor (in terms of MAE) by carefully calibrating the procedural model to real images.

1. Introduction

The quality and quantity of agricultural outputs depends on many factors such as crop genotype, disease, growth patterns, nutritional deficiency, and environmental conditions [23, 47]. The assessment and determination of plant growth, morphology, function, composition, disease detection, or any phenotype, is collectively called plant phenotyping. Many of these factors are evident from consistent monitoring, which has been found to be crucial data towards helping with the successful cultivation of new crops. This has traditionally been done for massive numbers of plants in a manual fashion by plant breeders, however this is both time-consuming and often overly reliant on intuition [15]. The possibility of using automated high-throughput phenotyping has long been recognized as an important step forward [49], and is now in the initial stages of being used to help cultivate new crop varieties [8, 45].

Image-based plant phenotyping is gaining in popularity as a method to automatically extract useful information from plant images [48] and to identify phenotypic traits throughout a plant's life [11], with machine learning being an increasingly useful approach to analyze the massive volumes of data generated by phenotyping platforms. Artificial neural networks (ANNs) and deep learning are providing excellent results on many data analysis and image analysis tasks [12], and have successfully been used in image classification, multi-instance detection, and segmentation [25]. ANNs have been used for certain complex image-based plant phenotyping tasks, such as leaf counting, age estimation, mutant classification [52], plant disease detection and diagnosis from images [34], the classification of fruits and other organs [40], and pixel-wise localization of root and shoot tips [41]. A comprehensive summary of deep learning algorithms for identification, regression, and prediction of plant stress phenotypes was presented by Singh et al. 2018 [46]. Some other plant phenotyping tasks where deep learning has been successfully used are counting plant stalks and stalk width [6], rice panicle segmentation [56], cotton bloom detection [57], wheat spike detection [19], maize tassel counting [30], and leaf counting in rosette plants [13, 17, 51]. Infected leaves of maize plants were identified by simultaneously localising wheat spikes and spikelets, demonstrating the power of multi-task deep architectures [42].

One of the most popular and powerful networks for deep learning is the convolutional neural network (CNN). The CNN has layers of neurons representing image pixels, and connections between layers that perform linear filtering. CNNs are capable of learning highly discriminative features during the training phase, and can classify plants without

needing segmentation, or feature extraction. Many works in the literature provide examples of different plants where important phenotypic traits have been extracted with CNNs; e.g. automatic joint feature and classifier learning for temporal phenotype/genotype classification [37], crop lodging detection [31], and leaf counting after segmenting rosette leaves with a deconvolutional network [2]. Lee et al. 2015 [26] proposed the Deep Plant framework that uses CNNs to learn feature representations from leaves using 44 different plant species. Deep Plant Phenomics (DPP) [52] has pre-defined trained CNNs built using TensorFlow for several common phenotyping tasks involving object detection, object counting, and semantic segmentation.

There are different popular networks that have been built using CNNs. U-Net [44] is a predefined model architecture that is widely used for segmentation and regression tasks. U-Net has been used extensively for biomedical image segmentation [4, 20, 21, 44, 58]. For crops, Ullah et al. [53] developed a segmentation model based on U-Net to segment canola plants from weeds and background in canola field images. Najafian et al. [36] customized U-Net for wheat head segmentation. Li et al. [27] used U-Net to extract tobacco plants from UAV images. Naik et al. [35] developed a semantic segmentation method with U-Net for weed segmentation. Ullah et al. [54] used U-Net to segment crop rows. U-Net has been modified for counting tasks from the obtained segmentations, e.g. Jeong et al. [22] applied a 3D cell-counting method using U-Net to identify initial seed cell numbers, and Bhagat et al. [7] used a modified U-Net model for plant leaf segmentation and counting.

Supervised machine learning methods require a training set where correct labels for the phenotype of interest are known, and the model is trained on the labelled dataset. A major challenge for deep learning applications for plant phenotyping tasks is the availability of a large quantity of annotated data for training models. Unfortunately, plant image datasets of the desired species, environment, scale, and size, labelled with the phenotypic properties of interest, are likely not available, and they might be difficult to obtain due to the large cost associated with collecting and annotating this type of data [51]. Hence, a motivation arises for using an alternative source of training data.

Computational systems for modelling and simulating plants have been an important area of research. One of the most common types of models used for this purpose are Lindenmayer systems (L-systems), introduced by Lindenmayer in 1968 [28] as a formalism for simulating the development of multicellular organisms in terms of division, growth, and death of individual cells. They can be used as a mathematical theory of plant development and for creating visually rich simulations of plants [29]. L-systems are a formal grammar system, with a set of rewriting rules that are applied in parallel to all letters of a string, which then iterates to the next string. Each string can describe an image and, a sequence of strings can describe a sequence of images for a temporal process. L-systems can have different plant geometries [18, 29, 38], environmental factors [3, 55], and mechanistic controls [38, 43].

The idea of using synthetic images from these simulations to train ANNs for the purpose of identifying phenotypic properties was initially explored by Ubbens et al. [51]. This idea overcomes many of the obstacles of manual annotations. If the phenotype of interest is built into the L-system model, then the correct annotations are automatically known. For the purposes of identifying plant organs and architecture, they can be precisely incorporated into an L-system model [29, 32]. In comparison to the cost of real data collection, and the precise labelling of them, it is easier to obtain synthetic data from these procedural models of plants. Furthermore, once a model has been created for a species, it is possible to generate arbitrarily large synthetic images datasets at no additional expense. In [51], it was shown that synthetic images from an L-system can be used to augment datasets of real plant images or can even be used alone as a source of training data. Mixing real and synthetic datasets led to better prediction of leaf count in *Arabidopsis thaliana*. A similar approach was also used with maize to predict leaf count [33] but using an existing procedural model of maize within Plant Factory Exporter [1] instead of an L-system to create synthetic images. In their work, results were mixed and there were cases (depending on the number of synthetic images used) where adding synthetic images to a real dataset for training improved prediction versus only training on real images, but often adding synthetic images did not help.

Cieslak et al. [10] demonstrated an approach for creating L-system models to match a set of images and their background, using maize and canola as a case study. They also showed how to calibrate these models so they can accurately capture the growth and visual characteristics in order to generate usable synthetic images. Calibrating these models is an important step for making realistic-looking synthetic images. Indeed, one might expect that the more realistic the synthetic images look, the more replaceable they are for real images for training. But the degree to which this is true is not yet understood. Their approach also demonstrates that it is particularly easy to take an existing L-system model and adapt it to a new environment, which is another big potential advantage of using synthetic images over manually labelling in each new environment.

Here, we continue the study of using synthetic images for training ANNs. We use two complex plants, maize and canola, and start with the same L-systems created in [10].

For maize, we systematically vary different amounts of real images combined together with synthetic images from L-system simulations, and measure success at leaf counting

on real images using both mean absolute error (mean absolute value of correct number minus predicted number) and Pearson's correlation. We find that using synthetic images in addition to real images does not improve success. However, using synthetic images has a large benefit vs. a small number of real images, and the benefit decreasing as the number of real images increases. The MAE when trained on synthetic images is 1.59 times higher than the MAE when trained on real images.

The next experiment involves canola, which is more complex in our dataset with much larger numbers of occluded leaves than in maize. We try to predict the number of inflorescence branches (flowering branches). We again find that in most situations real annotated images worked better for training than a combination of real and synthetic images. ANNs trained with only synthetic images and no real images work well at predicting inflorescence branch numbers in real images. We again found that MAE when trained on synthetic images is 1.5 times higher than MAE trained on real images. While visually these synthetic images look similar to the real plant images, refining and calibrating the L-system for the purposes of creating even more realistic synthetic images had an almost 10-fold decrease in MAE. The ANN prediction results were used to help improve the L-system calibration process, which were in-turn used to create even more accurate L-systems and synthetic images. Hence, this same process can either be used to make phenotype predictions from synthetic images, or to improve the L-systems themselves, and it can be seen as two different sides of the same coin.

To note, unlike past works, on using synthetic data for phenotype prediction in crops, validation datasets were used to avoid overfitting, which has an impact on whether both real and synthetic combined data provided better results than only real images in our tests.

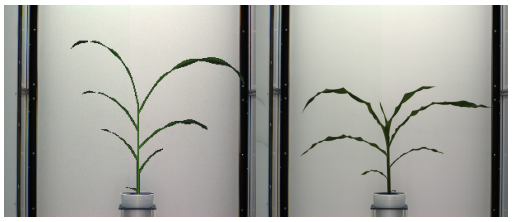
2. Dataset

Synthetic images were generated starting from the maize and canola models implemented in [10] that followed an interactive method for creating and calibrating L-systems.

Maize dataset: Real images of maize were obtained from an open dataset [39] called UNL-CPPD-I of 700 total images taken using the visible light camera at the UNL Lemnatec Scanalyzer 3D High-Throughput Phenotyping Facility [9]. Images were taken of 13 maize plants with different genotypes during 27 consecutive days of development during the vegetative stage starting 2 days after seed planting. Images were taken from two different side views separated by 90 degrees, which we call the view-0 image and view-90 image. The dataset also contains ground-truth annotations where leaves that are visible in the images are marked (leaves that were completely occluded in the images were not annotated). This is evident because the number of annotated leaves from the two views can differ substantially. Additional details regarding the imaging setup, dataset organization, and genotypes are in [9].

Maize follows an alternate phyllotaxy with each leaf developing on the opposite side of the previous leaf, and therefore the leaves form a planar-like surface. As the view-0 image was not always orthogonal to this plane, a pre-processing step was used to determine the best view. The same procedure was performed as described in [5]. Briefly, background subtraction was used to extract the foreground to remove the fixed background of the phenotyping system, Otsu thresholding was used on the grayscale image of the foreground image to obtain the segmented image, and another thresholding was performed by calculating the excess green index of the image. At this point, for each plant and each day, the convex hulls of the binarized plant images of both views were calculated, and the one with the largest convex hull by area was selected as the view for that plant/day. Only the resulting images chosen via this procedure were used for all subsequent analyses. Lastly, only images with at least one leaf were kept, and this dataset is known as the real maize dataset, denoted by M^R. In total, this dataset contains 328 images.

In total, the development of 50 plants was simulated from the maize L-system where they were calibrated to use the same timeline as the real plants. However, only a single view was generated for each day and plant, taken directly orthogonal to the leaves. In total, this created 1500 images. Ground-truth annotations for the number of leaves were automatically determined from the model. Furthermore, these annotations were calibrated to the real images in that it only counted a leaf if it was not occluded (algorithmically determined), and if it was larger than a threshold chosen so that the leaf would be visible by a human. Lastly, only images with at least one leaf were kept, and the resulting dataset is known as the synthetic dataset. This synthetic dataset contained 1350 images and is referred to as M^S. A real image on day 25 from 0 degrees and a synthetic image on the same day is in Figures 1b and 1a.

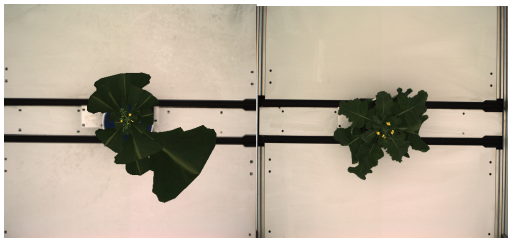


(a) (b)

Figure 1. (a) A synthetic maize image on day 25 generated from maize L-system. (b) A real maize plant image on day 25 from 0 degree view.

Canola dataset: Real images of *Brassica napus* (canola) are available in a dataset called P2IRC Flagship 1 Data that contains images of 50 different spring type genotypes, with six replicates of each, grown under two treatment conditions for over 38 days at the LemnaTec Scanalyzer 3D facility at University of Nebraska, Lincoln. The dataset and its details can be found in [14]. For the inflorescence branch count predictions we selected only top view images and those that had flowers. Also, while this dataset was a case/control dataset (meaning some plants were watered sufficiently, and some were not to study the affect of drought/stress), the 'case' images had technical issues [14]. Since the control images were enough for the purposes of this work, we only use the control images here. In total, the real dataset contained 385 images, referred to as C^R. The inflorescent branch count annotations were scored manually (details about manual scoring in [14]).

The development of 200 plants was simulated from the canola L-system in [10] using the same timeline as the real plants. Only flowering images were considered. As canola is a more complex plant to model in terms of leaf architecture and in terms of determining inflorescence branches than in maize, calibrating the model was challenging. Ultimately, we created 5 different synthetic datasets of canola called $C_1^S, C_2^S, C_3^S, C_4^S, C_5^S$, each contained 1200 images generated from five different variants of the canola plant model. These variants were obtained using a refinement procedure as described below. Figure 2 shows a synthetic image of canola from C_4^S at approximately the same time point together with a real image. This is similar to Klein et al. [24] who used synthetic images to train a classifier to distinguish between healthy and infected tomato plants, where they iteratively generated synthetic images to obtain optimal performance.



(a) (b)

Figure 2. (a) A synthetic image of a canola plant generated from C_4^S, at the same time point. (b) A real canola plant image at approximately the same time.

3. Improving Realism in Synthetic Images

Before using ANNs trained with synthetic images to detect phenotype (explained in Section 4) in this section, we describe a similar preliminary methodology and its results. The goal is to iteratively improve the realism of L-system models and the visual accuracy of its synthetic images by examining how close the mean absolute loss (MAE) is with a CNN trained on its synthetic images and tested on real images vs. one trained with real images and tested on real images. If there is a large discrepancy, then we can visually inspect the images and calibrate the model as done in [10]. We implement a simple CNN inspired from [51]. As canola has a more complex structure than maize, we explain the process with canola. A similar experiment was conducted with Maize, however the results showed that the maize L-system was already well calibrated and results did not improve. Hence, no refinement phase was required.

3.1. Methodology

An inflorescence branch counting procedure and experiment was performed on canola using the Deep Plant Phenomics (DPP) platform [52] using a CNN that was created and trained according to [50]. The model structure contained six convolutional layers with filter dimensions $(5, 5, 3, 32)$, $(5, 5, 32, 64)$, $(5, 5, 32, 64)$, $(3, 3, 64, 64)$, $(3, 3, 64, 64)$, and $(3, 3, 128, 128)$, stride length 1, and the $tanh$ activation function. Each convolutional layer was followed by a pooling layer with kernel size 3 and stride length 2. The model parameters and training hyper-parameters were: batch size 4, image dimensions 256×256, learning rate 0.0001, and number of epochs 500. A data augmentation procedure was done that consisted of cropping, flipping, and adjusting brightness/contrast. The testing network had an additional output layer. We did not further investigate other architectures as this experiment is to only improve the level of realism in synthetic images, and to better calibrate the L-system.

First only real images from all 39 distinct genotypes (285 images) were used for training. Testing was performed using 100 randomly selected real images from all images not used for training. Evaluation was measured by comparing the number of real leaves to the number of predicted leaves using MAE, standard deviation, and the square of Pearson's correlation coefficient (r^2). Next, we trained with only synthetic images and real images for testing. This entire procedure was repeated five additional times by using each of $C_1^S, C_2^S, C_3^S, C_4^S, C_5^S$ respectively.

3.2. Results

Identifying the best synthetic image dataset of canola was done by calculating the inflorescence branch count predictions for canola. Table 1 shows the results when testing on 100 real images for each of $C_1^S, C_2^S, C_3^S, C_4^S, C_5^S$. The refinement of these datasets from one to the next is described in Section 3.3. The MAE improved when transitioning the synthetic training images from C_1^S to C_4^S from 12.48 to 1.45 which is 8.6 times smaller.

Real images for training	Synthetic C_1^S for training	Synthetic C_2^S for training	Synthetic C_3^S for training	Synthetic C_4^S for training	Synthetic C_5^S for training
0.7 (0.87, 0.95)	12.48 (8.78, -7.22)	16.61 (5.63, -9.86)	1.98 (2.25, 0.68)	1.45 (1.76, 0.81)	5.6 (3.63, -0.57)

Table 1. Each table cell contains the MAE (respectively standard deviation and r^2) for the inflorescence branch counting task in canola. This inflorescence branch counting task results are a primary investigation to identify the best synthetic dataset only. Training was done on real images as indexed in the first column, and column 2 through 6 were trained by synthetic data from $C_1^S, C_2^S, C_3^S, C_4^S, C_5^S$ respectively and tested on 100 real images.

Figure 3 shows some synthetic images of canola from the same day generated from four different procedural models of canola plants (Figure 3a from C_1^S, 3b from C_2^S, 3c from C_3^S, and 3d from C_4^S).

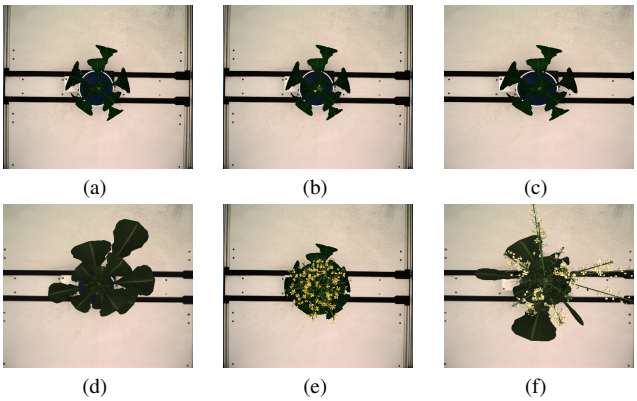


Figure 3. Synthetic images of canola on the same day generated from four different procedural models of canola plants (a – d) and two flowering images of same time point from different models. (a) C_1^S (b) C_2^S (c) C_3^S (d) C_4^S (e) C_1^S (f) C_4^S.

3.3. Refinement of Canola L-systems

Poor prediction results in earlier L-system variants and from visual inspection between real and synthetic images of canola helped to create refined L-system variants. Leaf shape and curvature, inflorescence branches, and flower structures were adjusted using L-system parameters to generate more realistic looking canola leaves, which substantially helped in improving the deep learning results. This observation is evident from Figure 3. A closer look at these figures shows how leaf shape and curvature were adjusted. Another scenario is shown in Figure 3e and 3f, where the inflorescence branches and flowers were improved from the first to the fourth model.

Specifically, the first L-system was the canola model from [10] (modified to count inflorescence branch numbers). The second was created by changing to the branch vigour parameter with the mean value and standard deviation being increased for greater variability of the feature in the synthetic images. This small change to the branching parameter made a difference to plant morphology. From the second to third L-system, the canola model parameters were changed to generate a distribution of inflorescence branch numbers that better reflect the real dataset. Additionally, the mean value and standard deviation were made slightly larger to decrease the number of lateral branches, and to have better control over the apex behaviour.

From the third to fourth L-system, there were a number of adjustments. The means and standard deviations of growth, branching, and leaf parameters were changed. The user-defined functions for growth of all organs (e.g., leaves and internodes) were stretched (so organs grow over a longer period). The function used to determine leaf width was changed to increase the leaf width. The leaf texture was also changed from a simple green color with a white midvein to an image-based color (i.e., the leaf texture was extracted from one of the real canola images). Hence, the fourth model was made substantially more realistic and was more accurately calibrated to the real images, compared to the first or second model. For the last L-system, few changes were made, such as to petal color, although that did not improve prediction.

Hence, the rest of the paper uses C_4^S for synthetic canola images.

4. Phenotyping Methodology

This section describes the phenotype prediction in both maize and canola.

4.1. Leaf counting task in maize

A leaf counting procedure and experiment on maize was performed using the predefined network U-NET with the DPP platform [52]. The model parameters and training hyper-parameters were: batch size 4, image dimensions 256×256, learning rate 0.001. The number of epochs was chosen from the training and validation loss curve to avoid overfitting. A data augmentation procedure was done that consisted of cropping, flipping, and adjusting brightness/contrast. In all of the studies, 10% of the training data was used as validation data.

Two sets of experiments were conducted for leaf counting with maize. In the first set, only real images were used for training. As M^R consisted of 13 plants with one image per day across 27 days, for each i from 1 to 8, training was done by randomly sampling without

replacement i of the 13 plants, and then using all of the images of those i plants for training. This was done in contrast to training with a randomly selected subset of images from the entire real image set, as it could be easier for an experimenter to create ground truth annotations for a small number of plants over time than to find and create ground truth annotations for a large number of different plants and randomly selected time points. For each, 5 of the remaining real plants were used for testing in two ways, first using 100 randomly selected images, and second using all remaining images. Evaluation was measured by comparing the number of real leaves to the number of predicted leaves using MAE, standard deviation, and the square of Pearson's correlation coefficient. In this second set of experiments, the procedure above was repeated by using the synthetic images for training but using real images for testing.

4.2. Inflorescence branch counting task in canola

A similar experiment was performed on canola for inflorescence branch count. The U-Net model and parameters were the same as described above in Section 4.1 for maize. As there are 39 distinct genotypes, we use images from i genotypes, in increments of 3 for training (again, using all images from each selected genotype). For example, the first training set has images of flowering plants from 3 genotypes, and the next training set has images of flowering from 6 genotypes, and so on. Testing was again done both using 100 randomly selected real images and with all remaining real images, and evaluation was done between number of real and predicted inflorescent branch numbers using MAE and standard deviation.

5. Phenotyping Results

The results regarding prediction of maize leaf count are in Table 2. Of note is the first row only synthetic images for training but testing on real data. Finally, we calculated the Root Mean Squared Error (RMSE) while training with all real images, and tested on 100 real maize images, and obtained the value of 1.2. This value increased to 1.61 while the training was done on all the synthetic images.

Inflorescence branch count predictions for canola were calculated on the best synthetic canola dataset C_4^S with U-Net. Table 3 shows the results, with the first row showing results for training on synthetic but testing on real data.

6. Discussion

As has been found consistently in the literature (e.g. [16]) training with real images performed better than training with synthetic images. Hence, the focus is on seeing how close prediction can be when training on synthetic images vs. real images, depending on the number of real images (or real plants over time).

For maize, training with only synthetic images gives better results than training with 2 real plants with 51 images. The MAE for training with 2 real maize plants or less is at least 1.19, whereas synthetic images only improves it to 1.05. The MAE when trained with synthetic images is quite low at, 1.05 vs. 0.66 when trained on real images (testing on 100 real images) respectively. It should be noted that it was relatively easy to produce visually realistic synthetic images of maize due to its simpler architecture, and indeed using these synthetic images was very good at leaf prediction. The difference in MAE between training with all synthetic versus training with all real is only 0.39.

In maize, while training with all real images, the RMSE was 1.2, and it increased to 1.61 while training on all synthetic images. In contrast, in [33] the RMSE obtained after training 720 real images was 1.33, and their RMSE for training with synthetic images was $\geq$ 2. While their real images of maize were taken at the same imaging facility, we caution against any firm conclusions as their real images were different and included up until 66 days after planting, versus 29 days in our dataset which could affect predictions. However, their often worse prediction results when using synthetic images were likely caused by less realistic synthetic images. Indeed, they indicated that despite maize following alternating phyllotaxy, their procedurally generated images had successive leaves that emerged from the stalk at random angles [33]. Our L-system was built to properly follow alternating phyllotaxy. This also shows the importance of biological realism in creating synthetic images. Figure 4 shows distributions of relative leaf count difference histograms for training on all real (4a, blue), and training on all synthetic (4a, orange). In addition, Figure 4 plots real leaf count versus predicted count, which shows the frequency of the data points (4b for training on all real, and 4c for training on all synthetic).

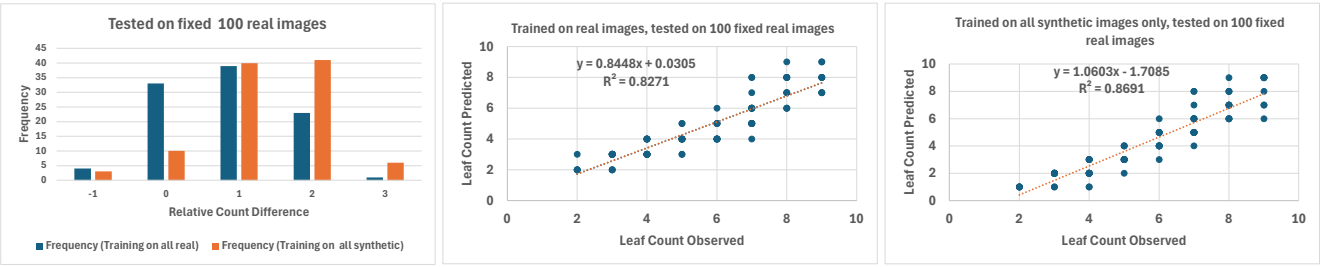
For canola, there were later stage images with more overlapping and occluded components with flowering images. In addition, canola has a more a complex branching pattern, so the ANN has to distinguish different orders of branches. The fourth L-system variant showed by far the most promising results. For canola, the MAE for training with only synthetic images gives better results than training with 3 genotypes of 9 plants and 32 images (1.59 and 2.17 respectively). Similarly to maize, the results with synthetic training only, vs. trained on all real images is quite close; 1.59 vs. 1.06 (testing on 100 fixed real images) respectively. There are however several points of interest. Figure 5 presents the distribution of relative count difference for inflorescence branch count in canola, while testing on 100 real canola images and trained on all the remaining real images (5a, blue), and all the synthetic images (5a, orange). Figures 5b and 5c plots real inflorescence branch count versus predicted count for training on all real images and

Numbers of real plants used in Training (number of real images)	**MAE (Absolute Loss Standard Deviation, r^2), tested with remaining real images**	**MAE (Absolute Loss Standard Deviation, r^2), tested with 100 real images**
Synthetic images used for training; 1350	0.82 (0.64, 0.77)	1.05 (0.66, 0.63)
1 real plant (26)	1.21 (0.84, 0.54)	1.44 (0.92, 0.31)
2 real plant (51)	0.94 (0.71, 0.69)	1.19 (0.71, 0.54)
3 real plant (77)	0.55 (0.64, 0.83)	0.67 (0.73, 0.76)
4 real plant (103)	0.75 (0.77, 0.74)	0.74 (0.76, 0.73)
5 real plant (126)	0.96 (0.82, 0.65)	1.04 (0.82, 0.58)
6 real plant (150)	0.87 (0.92, 0.66)	0.79 (0.77, 0.70)
7 real plant (175)	0.71 (0.66, 0.79)	0.7 (0.65, 0.78)
8 real plant (201)	0.65 (0.64, 0.84)	0.66 (0.65, 0.86)

Table 2. The MAE (respectively standard deviation, r^2) of leaf number prediction in maize appears in each table cell. The first column indexes the number of real plants (total number of images in parentheses) used for training. Results when testing with all remaining images and 100 real images are in the second and third column respectively.

Numbers of real plant genotypes (Number of real plants, Number of real images) used in Training	**MAE (Absolute Loss Standard Deviation, r^2), tested with remaining real images**	**MAE (Absolute Loss Standard Deviation, r^2), tested with 100 real images**
Synthetic images used for training; 1200	1.68 (1.96, 0.84)	1.59 (2.05, 0.76)
3 genotypes (9, 32)	2.30 (2.84, 0.65)	2.17 (2.61, 0.59)
6 genotypes (17, 63)	1.05 (1.66, 0.89)	0.78 (0.93, 0.94)
9 genotypes (26, 80)	1.14 (1.90, 0.87)	0.9 (1.01, 0.93)
12 genotypes (34, 101)	1.15 (2.19, 0.84)	0.8 (0.97, 0.94)
15 genotypes (43, 123)	1.15 (2.04, 0.86)	1.18 (1.98, 0.81)
18 genotypes (50, 146)	1.20 (1.80, 0.86)	1.19 (1.56, 0.86)
21 genotypes (59, 165)	1.32 (2.05, 0.83)	1.18 (1.79, 0.83)
24 genotypes (66, 184)	1.11 (1.83, 0.87)	1.03 (1.43, 0.88)
27 genotypes (73, 200)	0.90 (1.69, 0.90)	0.88 (1.30, 0.91)
30 genotypes (80, 215)	1.05 (1.79, 0.89)	0.97 (1.26, 0.90)
33 genotypes (89, 236)	1.12 (1.89, 0.89)	0.96 (1.37, 0.90)
36 genotypes (96, 254)	1.16 (2.08, 0.88)	1 (1.6, 0.87)
39 genotypes (104, 285)	NA	1.06 (1.69, 0.85)

Table 3. The MAE (respectively standard deviation, r^2) of inflorescence branch counting task in canola. The first column indexes the number of real plant genotypes (total number of real plants and images in parentheses) used for training. Results when testing with all remaining images and 100 real images are in the second and third column respectively.

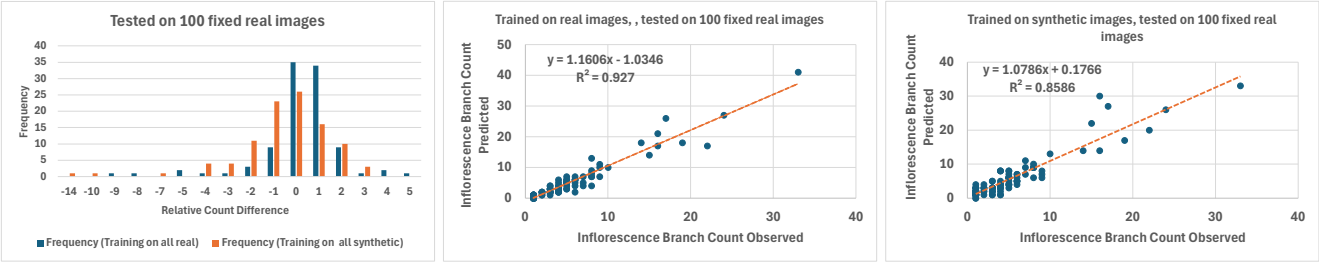


(a) (b) (c)

Figure 4. For leaf count in maize (a) Distributions of relative count difference while trained on all real images (in blue) and distributions of relative count difference while trained on all synthetic images (in orange) (b) Scatter plots of actual and predicted leaf counts while trained on all real images (c) Scatter plots of actual and predicted leaf counts while trained on all synthetic images. In all cases, test set had fixed 100 real images.

all synthetic images respectively.

The columns of Table 3 demonstrates how improvements were made to the the L-systems, thereby leading to improved deep learning results. Pearson's r^2 values show that the synthetic images from the fourth canola L-system were quite similar to the real canola images as the Pearson's correlation coefficients indicates the strength of associations between variables. The $r^2 = 0.85$ when trained on 285 real images and tested on 100 real images, and $r^2 = 0.76$ when trained on the 1200 synthetic images



(a) (b) (c)

Figure 5. For inflorescence branch count in canola (a) Distributions of relative branch count difference while trained on real images (in blue) and distributions of relative branch count difference while trained on synthetic images (in orange) (b) Scatter plots of actual and predicted branch counts while trained on real images (c) Scatter plots of actual and predicted leaf counts while trained on synthetic images.

generated from C_4^S and tested on the same 100 real images. And, $r^2 = 0.59$ when trained on 32 real images and tested on 100 real images. The RMSE we obtained for inflorescence branch count in canola is 1.94 while trained on all the real images and tested on 100 fixed images, and 2.43 while trained on all the synthetic images and tested on the same test set.

It is evident that these results were quite sensitive to the precise L-system construction. For example, when trained on only C_2^S, the MAE was exceptionally large at 16.6, but it was quite good when trained only on C_4^S at 1.45. This again shows that producing realistic synthetic images is extremely important. This is an important contribution as there is a large spectrum from training on simple geometry (e.g., ball-and-stick) images to those that accurately reflect phenotype captured in the described improvements from C_2^S to C_4^S. It also shows the simulated images alone even in canola can perform quite well, while requiring no groundtruth labelling.

One interesting contrast between the maize and canola experiments is that when both training and testing on only synthetic images (a separate experiment, not described), the MAE in maize was extremely low. This seems to indicate that the canola L-system is considerably more complicated and more difficult for the ANN to make accurate predictions. This is important as this was one of the important points of discussion in [51] regarding the *A. thaliana* model.

Finally, we chose to always train with all the available synthetic images when training rather than smaller numbers (such as the same number of real images). This was done because it is easy to generate arbitrarily large sets of synthetic images from the same L-system. Also, we ran a separate experiment to test if the amount of synthetic images played a role in prediction accuracy, and it showed that it was better to always use all the synthetic images rather than gradually adding them in, but further experimentation would be useful.

7. Conclusions

Machine learning and procedural plant models are two broad and impactful fields in plant phenotyping research. Here, we showed how each can help the other. This study conducted computational experiments in maize and canola regarding the use of synthetic images of plants for training artificial neural networks. For the task of leaf counting in maize, we calculated the mean absolute value of the difference between the predicted and correct number of leaves. The L-system used to create synthetic images was quite realistic, and this realism may have been a key contributor to its improved success versus others in the literature. Furthermore, it was more accurate to only use synthetic images with no groundtruth labelling than using 51 real images. The task of inflorescence branch counting in canola was conducted and results were evaluated using a similar procedure. This study again showed the importance of realism for some phenotypes and species. In some cases, the L-systems required calibration, with changes leading to remarkable improvements in prediction.

Here, we also saw that the deep learning results served to quantify how replaceable the synthetic images were for real images rather than using only visual inspections. The results show that using only synthetic images worked reasonably well at predicting on real images, while requiring no groundtruth labelling. After adjusting for overfitting, we achieve 1.05 MAE in predicting leaf count in maize and 1.59 MAE in predicting inflorescence branch count in canola using only synthetic images with U-Net.

8. Acknowledgements

This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund (CFREF), from the National Sciences and Engineering Research Council of Canada (NSERC), and a scholarship from SaskCanola. We thank Dr. Przemyslaw Prusinkiewicz for helpful discussion.

References

[1] Plant factory exporter. e-on Software. 2

[2] Shubhra Aich and Ian Stavness. Leaf counting with deep convolutional and deconvolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2080–2089, 2017. 2

[3] MT Allen, Przemyslaw Prusinkiewicz, and TM DeJong. Using L-systems for modeling source–sink interactions, architecture and physiology of growing trees: The L-PEACH model. *New Phytologist*, 166(3):869–880, 2005. 2

[4] Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[5] Nazifa Azam Khan, Oliver AS Lyon, Mark Eramian, and Ian McQuillan. A novel technique combining image processing, plant development properties, and the Hungarian algorithm, to improve leaf detection in maize. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 74–75, 2020. 3

[6] Harjatin Singh Baweja, Tanvir Parhar, Omeed Mirbod, and Stephen Nuske. Stalknet: A deep learning pipeline for high-throughput measurement of plant stalk count and stalk width. In *Field and Service Robotics*, pages 271–284. Springer, 2018. 1

[7] Sandesh Bhagat, Manesh Kokare, Vineet Haswani, Praful Hambarde, and Ravi Kamble. Eff-unet++: A novel architecture for plant leaf segmentation and counting. *Ecological Informatics*, 68:101583, 2022. 2

[8] Aakash Chawade, Joost van Ham, Hanna Blomquist, Oscar Bagge, Erik Alexandersson, and Rodomiro Ortiz. High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy*, 9(5):258, 2019. 1

[9] Sruti Das Choudhury, Srinidhi Bashyam, Yumou Qiu, Ashok Samal, and Tala Awada. Holistic and component plant phenotyping using temporal image sequence. *Plant Methods*, 14(1):35, 2018. 3

[10] Mikolaj Cieslak, Nazifa Khan, Pascal Ferraro, Raju Soolanayakanahally, Stephen J Robinson, Isobel Parkin, Ian McQuillan, and Przemyslaw Prusinkiewicz. L-system models for image-based phenomics: case studies of maize and canola. *in silico Plants*, 4(1):diab039, 2022. 2, 3, 4, 5

[11] Peter A Crisp, Diep Ganguly, Steven R Eichten, Justin O Borevitz, and Barry J Pogson. Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. *Science Advances*, 2(2):e1501340, 2016. 1

[12] Sruti Das Choudhury, Ashok Samal, and Tala Awada. Leveraging image analysis for high-throughput plant phenotyping. *Frontiers in Plant Science*, 10:508, 2019. 1

[13] Andrei Dobrescu, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Leveraging multiple datasets for deep leaf counting. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2072–2079, 2017. 1

[14] Jana Ebersbach, Nazifa Azam Khan, Ian McQuillan, Erin E Higgins, Kyla Horner, Venkat Bandi, Carl Gutwin, Sally Lynne Vail, Steve J Robinson, and Isobel AP Parkin. Exploiting high-throughput indoor phenotyping to characterize the founders of a structured *B. napus* breeding population. *Frontiers in Plant Science*, 12:780250–780250, 2021. 4

[15] Dionysia A Fasoula, Ioannis M Ioannides, and Michalis Omirou. Phenotyping and plant breeding: overcoming the barriers. *Frontiers in Plant Science*, 10:1713, 2020. 1

[16] Scott Geng, Ranjay Krishna, and Pang Wei Koh. Training with real instead of synthetic generated images still performs better. In *Synthetic Data for Computer Vision Workshop @ CVPR 2024*, 2024. 6

[17] Mario Valerio Giuffrida, Peter Doerner, and Sotirios A Tsaftaris. Pheno-Deep counter: a unified and versatile deep learning architecture for leaf counting. *The Plant Journal*, 96(4):880–890, 2018. 1

[18] Christophe Godin and Pascal Ferraro. Quantifying the degree of self-nestedness of trees: application to the structural analysis of plants. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7(4):688–703, 2009. 2

[19] Md Mehedi Hasan, Joshua P Chopin, Hamid Laga, and Stanley J Miklavcic. Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods*, 14(1): 100, 2018. 1

[20] Nabil Ibtehaz and M Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks*, 121:74–87, 2020. 2

[21] Ahmed Iqbal, Muhammad Sharif, Muhammad Attique Khan, Wasif Nisar, and Majed Alhaisoni. Ff-unet: a u-shaped deep convolutional neural network for multimodal biomedical image segmentation. *Cognitive Computation*, 14 (4):1287–1302, 2022. 2

[22] Eun Ji Jeong, Donghyuk Choi, and Dong Woo Lee. U-net deep-learning-based 3d cell counter for the quality control of 3d cell-based assays through seed cell measurement. *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, 26 (5):468–476, 2021. 2

[23] James W Jones, John M Antle, Bruno Basso, Kenneth J Boote, Richard T Conant, Ian Foster, H Charles J Godfray, Mario Herrero, Richard E Howitt, Sander Janssen, et al. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural systems*, 155:269–288, 2017. 1

[24] Jonathan Klein, Rebekah E Waller, Sören Pirk, Wojtek Pałubicki, Mark Tester, and Dominik Michels. Synthetic data at scale: A paradigm to efficiently leverage machine learning in agriculture. *Available at SSRN 4314564*, 2023. 4

[25] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 1

[26] Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino. Deep-plant: Plant identification with convolutional neural networks. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 452–456. IEEE, 2015. 2

[27] Qianxia Li, Lihui Yan, Zhongfa Zhou, Denghong Huang, Dongna Xiao, and Youyan Huang. Study on tobacco plant cross-level recognition in complex habitats in karst mountainous areas based on the u-net model. *Journal of the Indian Society of Remote Sensing*, 52(9):2099–2114, 2024. 2

[28] Aristid Lindenmayer. Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. *Journal of Theoretical Biology*, 18(3):280–299, 1968. 2

[29] Aristid Lindenmayer and Przemyslaw Prusinkiewicz. *The Algorithmic Beauty of Plants*. New York: Springer-Verlag, 1990. 2

[30] Yuzhen Lu, Yuping Huang, and Renfu Lu. Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: A review. *Applied Sciences*, 7(2):189, 2017. 1

[31] Sara Mardanisamani, Farhad Maleki, Sara Hosseinzadeh Kassani, Sajith Rajapaksa, Hema Duddu, Menglu Wang, Steve Shirtliffe, Seungbum Ryu, Anique Josuttes, Ti Zhang, Sally Vail, Curtis Wozniak, Isobel Parkin, Ian Stavness, and Mark Eramian. Crop lodging prediction from UAV-acquired images of wheat and canola using a DCNN augmented with handcrafted texture features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 2

[32] Radomir Mech. *Modeling and Simulation of the Interaction of Plants with the Environment using L-systems and their Extensions*. Calgary, 1997. 2

[33] Chenyong Miao, Thomas P Hoban, Alejandro Pages, Zheng Xu, Eric Rodene, Jordan Ubbens, Ian Stavness, Jinliang Yang, and James C Schnable. Simulated plant images improve maize leaf counting accuracy. *BioRxiv*, page 706994, 2019. 2, 6

[34] Sharada P Mohanty, David P Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419, 2016. 1

[35] K Jairam Naik. Deep learning based segmentation of weed images in crop fields using u-net. In *2024 3rd International Conference for Advancement in Technology (ICONAT)*, pages 1–6. IEEE, 2024. 2

[36] Keyhan Najafian, Alireza Ghanbari, Mahdi Sabet Kish, Mark Eramian, Gholam Hassan Shirdel, Ian Stavness, Lingling Jin, and Farhad Maleki. Semi-self-supervised learning for semantic segmentation in images with dense patterns. *Plant Phenomics*, 5:0025, 2023. 2

[37] Sarah Taghavi Namin, Mohammad Esmaeilzadeh, Mohammad Najafi, Tim B Brown, and Justin O Borevitz. Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods*, 14(1): 66, 2018. 2

[38] Taishin Nishida. KOL-system simulating almost but not exactly the same development: case of japanese cypress. *Memoirs of the Faculty of Science, Kyoto University, Series B*, 8(1):97–122, 1980. 2

[39] University of Nebraska-Lincoln Plant Vision Initiative. UNL plant phenotyping datasets, 2018. https://plantvision.unl.edu/dataset. 3

[40] Pornntiwa Pawara, Emmanuel Okafor, Olarik Surinta, Lambert Schomaker, and Marco Wiering. Comparing local descriptors and bags of visual words to deep convolutional neural networks for plant recognition. In *International Conference on Pattern Recognition Applications and Methods*, pages 479–486. SciTePress, 2017. 1

[41] Michael P Pound, Jonathan A Atkinson, Alexandra J Townsend, Michael H Wilson, Marcus Griffiths, Aaron S Jackson, Adrian Bulat, Georgios Tzimiropoulos, Darren M Wells, Erik H Murchie, et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, 6(10):gix083, 2017. 1

[42] Michael P Pound, Jonathan A Atkinson, Darren M Wells, Tony P Pridmore, and Andrew P French. Deep learning for multi-task plant phenotyping. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2055–2063, 2017. 1

[43] Przemyslaw Prusinkiewicz, Scott Crawford, Richard S Smith, Karin Ljung, Tom Bennett, Veronica Ongaro, and Ottoline Leyser. Control of bud activation by an auxin transport switch. *Proceedings of the National Academy of Sciences*, 106(41):17431–17436, 2009. 2

[44] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 2

[45] Nadia Shakoor, Scott Lee, and Todd Mockler. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Current Opinion in Plant Biology*, 38:184–192, 2017. 1

[46] Asheesh Kumar Singh, Baskar Ganapathysubramanian, Soumik Sarkar, and Arti Singh. Deep learning for plant stress phenotyping: trends and future perspectives. *Trends in Plant Science*, 23(10):883–898, 2018. 1

[47] José C Soares, Carla S Santos, Susana MP Carvalho, Manuela M Pintado, and Marta W Vasconcelos. Preserving the nutritional quality of crop plants under a changing climate: importance and strategies. *Plant and Soil*, 443: 1–26, 2019. 1

[48] François Tardieu, Llorenc Cabrera-Bosquet, Tony Pridmore, and Malcolm Bennett. Plant phenomics, from sensors to knowledge. *Current Biology*, 27(15):R770–R783, 2017. 1

[49] RD Tillett. Image analysis for agricultural processes: a review of potential opportunities. *Journal of Agricultural Engineering Research*, 50:247–258, 1991. 1

[50] Jordan Ubbens. Deep plant phenomics documentation. github, 2017. https://deep-plant-phenomics.readthedocs.io/en/latest/. 4

[51] Jordan Ubbens, Mikolaj Cieslak, Przemyslaw Prusinkiewicz, and Ian Stavness. The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods*, 14(1):6, 2018. 1, 2, 4, 8

[52] Jordan R Ubbens and Ian Stavness. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science*, 8:1190, 2017. 1, 2, 4, 5

[53] Hafiz Sami Ullah, Muhammad Hamza Asad, and Abdul Bais. End to end segmentation of canola field images using dilated u-net. *IEEE Access*, 9:59741–59753, 2021. 2

[54] Muhib Ullah, Fatimah Islam, and Abdul Bais. Quantifying consistency of crop establishment using a lightweight u-net deep learning architecture and image processing techniques. *Computers and Electronics in Agriculture*, 217:108617, 2024. 2

[55] Tomonari Watanabe, Jim S Hanan, Peter M Room, Toshihiro Hasegawa, Hiroshi Nakagawa, and Wataru Takahashi. Rice morphogenesis and plant architecture: measurement, specification and the reconstruction of structural development by 3D architectural modelling. *Annals of Botany*, 95(7):1131–1143, 2005. 2

[56] Xiong Xiong, Lingfeng Duan, Lingbo Liu, Haifu Tu, Peng Yang, Dan Wu, Guoxing Chen, Lizhong Xiong, Wanneng Yang, and Qian Liu. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods*, 13(1):104, 2017. 1

[57] Rui Xu, Changying Li, Andrew H Paterson, Yu Jiang, Shangpeng Sun, and Jon S Robertson. Aerial images and convolutional neural network for cotton bloom detection. *Frontiers in Plant Science*, 8:2235, 2018. 1

[58] Hasib Zunair and A Ben Hamza. Sharp u-net: Depthwise convolutional network for biomedical image segmentation. *Computers in biology and medicine*, 136:104699, 2021. 2