

Root Length Estimation: Automatizing Image Analysis without Segmentation

Short title: Root Length Estimation - no Segmentation

Faina Khoroshevsky^{1*#}, Kaining Zhou^{2,6#}, Sharon Chemweno^{3,6}, Yael Edan¹, Aharon Bar-Hillel¹, Ofer Hadar⁴, Boris Rewald⁵, Pavel Baykalov⁵, Jhonathan E. Ephrath⁶, and Naftali Lazarovitch⁶

*Address correspondence to: bordezki@post.bgu.ac.il

These authors contributed equally to this work.

¹ Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer Sheva, Israel.

² The Jacob Blaustein Center for Scientific Cooperation, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer, Israel.

³ The Albert Katz International School for Desert Studies, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer, Israel.

⁴ Department of Communication Systems Engineering, School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel.

⁵ Department of Forest and Soil Sciences, Institute of Forest Ecology, University of Natural Resources and Life Sciences, Vienna, Austria.

⁶ French Associates Institute for Agriculture and Biotechnology of Drylands, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer, Israel.

Abstract

Image-based root phenotyping technologies, including the minirhizotron (MR), have expanded our understanding of the *in-situ* root responses to changing environmental conditions. The conventional manual methods used to analyze MR images are time-consuming and therefore limit their implementation. In this study, previous convolutional neural network (CNN)-based models we developed were adapted to automatically estimate the total root length (TRL) from MR images, obviating the need for image segmentation, where TRL is the sum of length of all the roots of a given image. Training data were derived from manual annotations in Rootfly, a commonly used software for MR image analysis. TRL estimation by two models (a regression-based model and a detection-based model which detects the coordinates of points along the roots) was compared. This can assist in examining human annotations by providing a visual inspection of roots in MR images. The models were trained and tested with 4015 images acquired using two MR system types (manual and automated) and from four crop species (corn, pepper, melon, and tomato) grown under various abiotic stresses. These datasets are made publicly available as part of this publication. The coefficients of determination, R^2 , between the measurements made using Rootfly and the suggested models when estimating TRL were 0.929–0.986 for the main datasets, demonstrating that this tool is accurate and robust. Additional analyses were conducted to examine the effects of 1) the data acquisition system and thus the image quality on the models' performance, 2) automated differentiation between images with and without roots, and (3) the use of the transfer learning technique. These approaches can support precision agriculture by providing real-time root growth information and thus facilitating decision-making.

1. Introduction

Prevailing stresses caused by climate change, such as drought and salinity, put significant constraints on crop yields and thus pose threats to food security and economic development (Molotoks et al., 2021). Root traits (e.g., root length and rooting depth) related to nutrient and water acquisition play critical roles in stress tolerance, given their high levels of plasticity (Koevoets et al., 2016; Lambers et al., 2006; Lynch, 2018, Lynch et al., 2021). Therefore, trait-based root phenotyping has been proposed as a promising approach for crop selection and improvement (Ajmera et al., 2022; Amtmann et al., 2022; Ephrath et al., 2020; McGrail et al., 2020; Zhao et al., 2018). The spatial distribution of the roots and their growth is very sensitive to various physical, chemical, and biological factors, as well as to the hydraulic properties of the soil that affect the availability of water, nutrients, and oxygen for the plants (Hartmann et al., 2018; Kuppe et al., 2022;

1 Lak et al., 2020). Therefore, it is important to describe root growth under the influence of diverse
2 environmental conditions to accurately understand agricultural systems and to develop modeling
3 capabilities for decision-making.

4 Investigation of root dynamics in response to changing environments has been achieved with
5 image-based root phenotyping techniques (Busener et al., 2020; Hui et al., 2022). Among these, the
6 minirhizotron (MR) technique has been widely used for the nondestructive *in situ* observation of
7 roots (Rewald & Ephrath, 2013). The MR consists of a transparent observation tube embedded in
8 the root zone and an image acquisition component that allows images to be collected repeatedly,
9 allowing the fates of individual roots to be followed through time (Rewald & Ephrath, 2013). Roots
10 information (e.g., length, diameter, mortality) can be extracted from MR images using image
11 analysis programs such as Rootfly (Wells and Birchfield, Clemson University, South Carolina,
12 USA), RootTracker (Duke University, Durham, North Carolina, USA), RootPainter (Smith et al.,
13 2022), and WinRHIZO (Régent Instruments, Quebec, Canada). Roots length is especially important
14 since it allows to calculate various root parameters based on it, such as rooting depth, root length
15 density (RLD), and root growth rate.

16 A major drawback of the MR technique is that the conventional image collection and analysis
17 processes are performed manually and are therefore time-consuming, which considerably limits the
18 size and number of experiments that can be reasonably conducted (Danilevicz et al., 2021).
19 Moreover, the outcomes of image analyses are subjective and thus dependent upon the knowledge
20 and experience of the annotator (Bauer et al., 2022; Zeng et al., 2008). Automated imaging systems
21 and analysis tools are required to overcome these challenges.

22 In recent years, deep learning (DL) algorithms (LeCun et al., 2015), a subset of machine learning
23 tools, have emerged as a driving force to provide state-of-the-art performance in image-based plant
24 phenotyping, especially in plant stress phenotyping (Pound et al., 2017; Singh et al., 2018). DL
25 tools are increasingly used by plant scientists to process large datasets of images collected using
26 high-throughput phenotyping platforms, facilitating recent developments in the automation of the
27 agricultural domain (Jiang & Li, 2020; Saleem et al., 2021; Fajardo & Whelan, 2021). More
28 specifically, convolutional neural network (CNN)-based architectures have become very popular
29 due to their excellent performance in complex computer vision tasks (Castro-Valdecantos et al.,
30 2022; Farjon et al., 2020; Kalantar et al., 2019; Koirala et al., 2019). The ability of CNNs to learn
31 features directly from data without prior knowledge has allowed them to overcome the limitations
32 of traditional machine-learning approaches.

33 Although CNN for root phenotyping is evolving rapidly, some existing CNN tools were used for
34 root images taken from artificial plant cultivation systems such as glass tubes filled with transparent

gellan gum (Han & Kuo, 2018), or growth pouches in a controlled environment chamber with individual pouches transferred to a copy stand for imaging (Yasrab et al., 2019). Analyzing root systems on images taken *in situ* is more difficult for three reasons: (1) the low contrast between the roots and the heterogeneous rhizosphere; (2) the inconsistent scene illumination caused by the light source, soil water content, and soil type; and (3) the various artifacts, such as scratches or water bubbles on the wall of observation tube. These issues have limited the automated analysis of roots in MR images using the available CNN models (Zeng et al., 2008).

Current CNNs used for root image analysis, predominantly rely on segmentation-based architectures (Atanbori et al., 2019; Bauer et al., 2022; Han & Kuo, 2018; Nair et al., 2022; Smith et al., 2020; Smith et al., 2022; Wang et al., 2019; Yasrab et al., 2019). Segmentation requires that each input image is paired with a set of labels corresponding to each of the pixels in the input image [Smith et al., 2022]. These models require significant human effort to annotate the segmentation masks for model training, i.e., to mark the contour of each root in the annotation tool to obtain all the relevant pixels of the roots in the image. Creating such dense per-pixel annotations for training is a time-consuming process [Smith et al., 2022].

In this paper we propose two fully automated CNN-based MR image analysis models to estimate the total root length (TRL) in each image, without requiring root segmentation annotation for training. We compare our models to two recent studies (Bauer et al., 2022; Smith et al., 2022) that also perform automated root length estimation which rely on models that learn to perform root segmentation as a preliminary step to root length estimation. Smith et al. (2022) presented RootPainter, an open-source graphical user interface-based software for creating a dataset and training a fully convolutional network, a modified version of U-Net (Ronneberger et al., 2015), to perform image segmentation. They use this tool to train a model to perform root segmentations of rhizotron-based root images and to obtain root length estimates by doing skeletonization of generated segmentations and pixel counting. Bauer et al. (2022) incorporated a pipeline for TRL estimation by combining open-source software tools that included root segmentation with the RootPainter (Smith et al., 2022) followed by extracting TRL estimates from the resulting segments. This was achieved with RhizoVision Explorer (Seethepalli et al., 2021) by the sum of the Euclidean distances between the connected skeletal pixels of the root topology.

The suggested CNN-based models have been adapted for the TRL estimation task and are based on our previous CNN architectures that were initially developed to count the number of leaves on a single plant in an image (Itzhaky et al., 2018). The “ground truth” (GT) values for training these models were acquired using Rootfly, a manual MR image analysis software (Fig. S1). Rootfly

allows to estimate the TRL values based on the points coordinates that the annotator marks along the entire length of each root during the annotation process (Fig. 1).

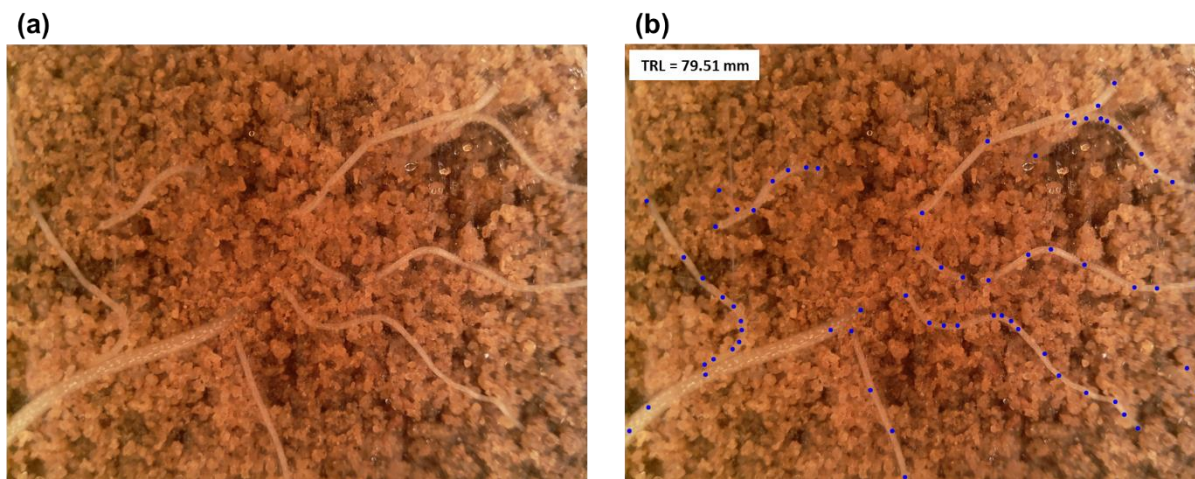


Fig 1. Example of (a) a raw root image and (b) the same image with added GT points annotations drawn on top and the resulting TRL value from Rootfly.

Both the TRL values and the coordinates of these points in each MR image are referred to as the GT values in this study. Our suggested models for automated TRL estimation only require the inherited information provided by Rootfly for training, without additional annotations from other software.

The first suggested model is a regression-based network, which implements a function that maps the entire image to a single value that represents a specific attribute of the image (here, the TRL value). Only the GT value of TRL per image is required to train this model. The second model is based on explicit object detection which requires the GT values of both the TRL and the coordinates of the points per image for training. In general, a detection task is based on the detection of a bounding box around each object in the image (Lin et al., 2017b; Tan et al., 2020) or the detection of an object's central point (Sindagi & Patel, 2017); however, in the suggested model, it refers to the detection of the points along the roots. The detection of these points also allows visual inspection of the roots in the image—once there is a trained model that automatically outputs its estimates of the coordinates of the root points, the annotator can use them to re-check the current annotations. Since human annotations are inconsistent and subjective (Bauer et al., 2022; Kume et al., 2018), the proposed visual inspection feature could reduce errors in annotations leading to improved annotation quality. In addition, the performance of current semi-automated and automated software used for root image analysis has been measured on small datasets (Atanbori et al., 2019; Han & Kuo, 2018; Smith et al., 2022; Yasrab et al., 2019) and on a limited range of species each time (e.g., corn, wheat, and rice). These limitations have led to overfitting issues when new root images are

introduced (Paez-Garcia et al., 2015). An important method for handling this kind of difficulty is transfer learning, in which a base network is trained using a large dataset and then used for another task that has a smaller dataset (Yosinski et al., 2014). A trained network can be used for feature extraction based on one dataset and then used for another dataset; e.g., when the features of a CNN that were trained using a simulated X-ray image of a soil–root system are used for another model applied to real images (Douarre et al., 2018). Alternatively, transfer learning can be done by using the first trained layers of the base (trained) network as part of a new target network, while the rest of the layers are randomly initialized and trained toward the target task (Yosinski et al., 2014). The copied layers can then be fine-tuned or left “frozen” (i.e., not altered during training on the new task). We examine the contribution of the transfer learning technique using the later approach with an entire trained CNN being fine-tuned for a new dataset. We also contribute by making publicly available the datasets that were acquired in this study. This includes 4015 images with the corresponding TRLs and points’ coordinate annotations (Khoroshevsky et al., 2022).

The objectives of this study were to (1) adapt two existing CNN models to the TRL estimation task (a completely different task without segmentation annotation); (2) examine the influence of sample size and image quality (which depend on the image acquisition system) on the models’ performance; (3) suggest a method for distinguishing between images with and without roots; (4) demonstrate the models’ contributions to root research by comparing RLD plots generated based on the models’ TRL estimates and the GT annotations from Rootfly; (5) examine the transfer learning opportunities when testing a trained model on a new dataset with different properties; and (6) make publicly available a new MR dataset that includes two types of MR systems (manual and automated) with high-quality annotations.

The rest of the paper is organized as follows: the Materials and Methods, including the models’ architectures, data acquisition, datasets, and evaluation, are described in Section 2; the Results and Discussion are presented in Section 3; and the Conclusions and Outline are presented in Section 4.

2. Materials and Methods

2.1. Architectures of the models

The two suggested CNN models (Fig. S2) were developed for the task of part counting (Itzhaky et al., 2018) and implemented to count the number of leaves on a single plant in an image (Farjon et al., 2021; Itzhaky et al., 2018). In later studies, they were used as alternative options in a two-stage network for counting parts-per-object in an image with multiple objects (Khoroshevsky et al., 2020; Khoroshevsky et al., 2021), in which the objects were first detected using the RetinaNet architecture

(Lin et al., 2017b) and for each detection, these models were used for part counting of the detected object. In the present study, these two models were adapted for the completely different task of TRL estimation, with multiple objects (roots) in an image.

The first model is a multiple-scale regression (MSR) model, which is based on direct regression and is designated the *regression model*. It outputs the total length of roots without using additional annotations for training. Instead of receiving annotations of part count per image, it now learns to output the TRL values when there are multiple roots or no roots at all.

The second model is a detection and regression (D+R) model, designated the *points model*. It learns to detect the coordinates of points along the roots in the image and uses them to output the TRL estimates. Unlike the original detection-based counting model, the detected points here are not the roots' centers but points annotated along each root and, instead of parts counting, the model estimates the TRL. Based on the GT coordinates of these points, a two-dimensional "density estimation" heat map is generated for each image. This is done with a Gaussian kernel placed around the coordinates of each point, to obtain a heat map with Gaussian distribution around each point, from which the model learns to output the TRL estimates. This was done instead of generating a binary map of the points' coordinates, for two reasons. The first is that the specific coordinates of the points depend on the subjective annotations of the user. Given the nature of the task, the exact coordinates can vary and cannot be defined with pixel-level precision. Second, it is more difficult to train a network to output a binary map, where even a deviation of a single pixel in the root's estimated location implies a full error. When small deviations are considered to be partial successes, gradual learning is promoted.

The generation of GT heat maps is part of the preparation of the total GT annotations for each image. They are used in training when the heat map generated by the model (i.e., detecting the roots) is compared with the GT heat map, and the relevant loss is calculated (Itzhaky et al., 2018). A trained model can generate this map by itself in test time, which can then be used for the final estimation of TRL. Fig. 2b shows an example of a GT map that was generated according to the points' annotation data from Rootfly on a raw MR image (Fig. 2a). The predicted map, which was generated in test time by a trained model for this image, is shown in Fig. 2c. In this example, the middle root in the image was not annotated because of an error of the annotator (Fig. 2b) but was detected by the model (Fig. 2c).

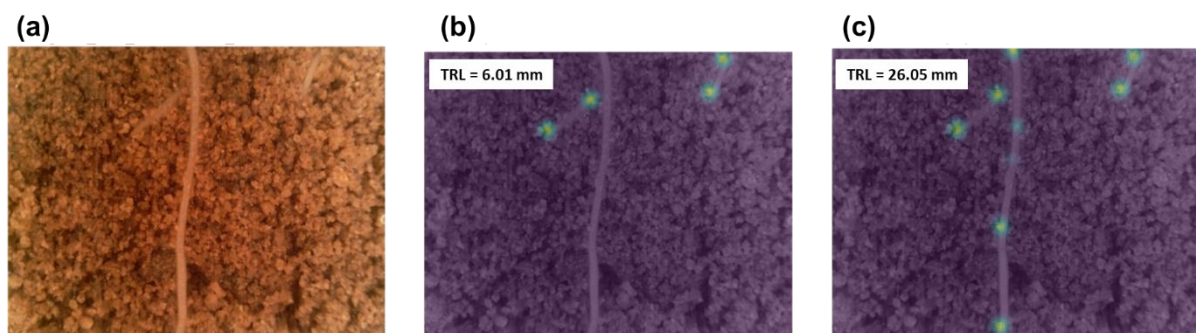


Fig. 1. Comparison of a GT heat map of a root's points (human annotations) and the heat map generated by the points model. (a) raw MR image. (b) the GT heat map of the points coordinates on top of the raw image and its corresponding TRL value. The heat map was generated based on the points manually annotated in Rootfly. (c) the predicted map with predicted points and the corresponding TRL value.

2.2. Data acquisition

Data were acquired with two systems, manual and automated (Fig. 3), which collectively record images that display the distribution of the root length across a range of depths in the soil profile. The automated system was acquired using MR cameras with higher image resolution, larger observation areas, and better illumination, than the manual system resulting in significant differences in image quality.

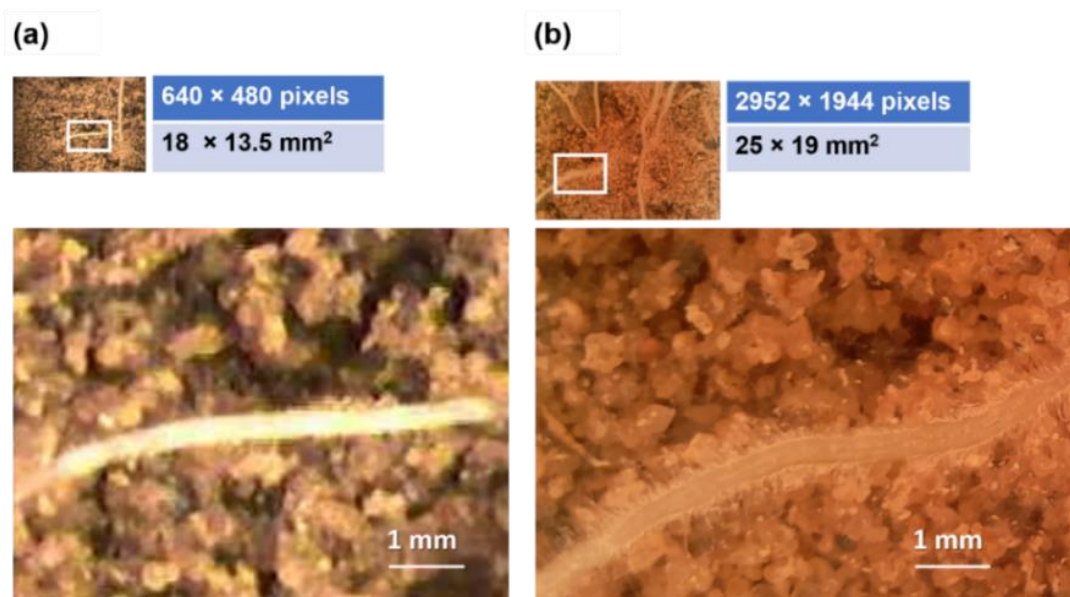


Fig. 3 Comparison of root images taken using (a) a conventional MR system or (b) an automated MR system. Images in the second row are enlarged versions of the marked parts (with a white border) in the original root images in the first row.

Manual image acquisition was performed using a Bartz Technology Co. (Carpinteria, California, USA) MR camera system with its default settings. The camera was attached to an indexing handle and controlled using a laptop-computer-based image capture system (I-CAP, Bartz Technology

1 Co., Carpinteria, California, USA). Lighting was supplied by four small incandescent light bulbs
2 surrounding the camera. The camera was manually lowered into an observation tube installed 10–
3 15 cm in front of the measured plant. One operator moved the camera from the bottom of the tube
4 upward, in increments of 13.5 mm, over the entire length of the tube, while another operator
5 monitored the quality of the image on the laptop and captured it simultaneously. After all images
6 were taken from the measured tube, the camera and laptop were moved to the next tube for image
7 collection. Each image viewed an area of $18 \times 13.5 \text{ mm}^2$ at a resolution of 640×480 pixels.
8 Acquired images had a redundant frame that was cropped, resulting in images with a size of
9 624×450 pixels.

10 **Automated image acquisition** was performed using the integrated system RootCam©
11 (CrystalVision, Samar, Israel; 2952×1944 resolution) with its default settings. RootCam includes
12 the camera itself and software designed to move the camera along a rail to acquire images every
13 18.75 mm along a plant root. Lighting was supplied by LED strips. The images were saved to a
14 “Raspberry Pi” device, which was accessible via a network cable and allowed the image acquisition
15 time intervals to be set with remote control. Each image viewed an area of $25 \times 19 \text{ mm}^2$.

16 Analysis of the root images captured by the MR systems was challenging since they may have been
17 blurry or had poor contrast between the roots and the background. Moreover, the background of
18 some images contained artifacts such as scratches, stains, and water bubbles on the observation
19 tubes. However, these challenges are significantly more prominent in the manual MR system
20 because of its inconsistent illumination and lower resolution which caused significantly poorer
21 quality of the obtained images. These had more roots ‘blended’ with the soil than images obtained
22 with the automated system. Therefore, the datasets acquired with the manual MR system were more
23 challenging even for a human annotator. Due to the significant differences in image quality, the
24 datasets are distinguished based on the acquisition system and are referred to as different *dataset*
25 *types*. Examples of these challenges, from both the manual and automated acquisition systems, can
26 be seen in Fig. 4, in which the annotations of the GT points are given on top of each root.

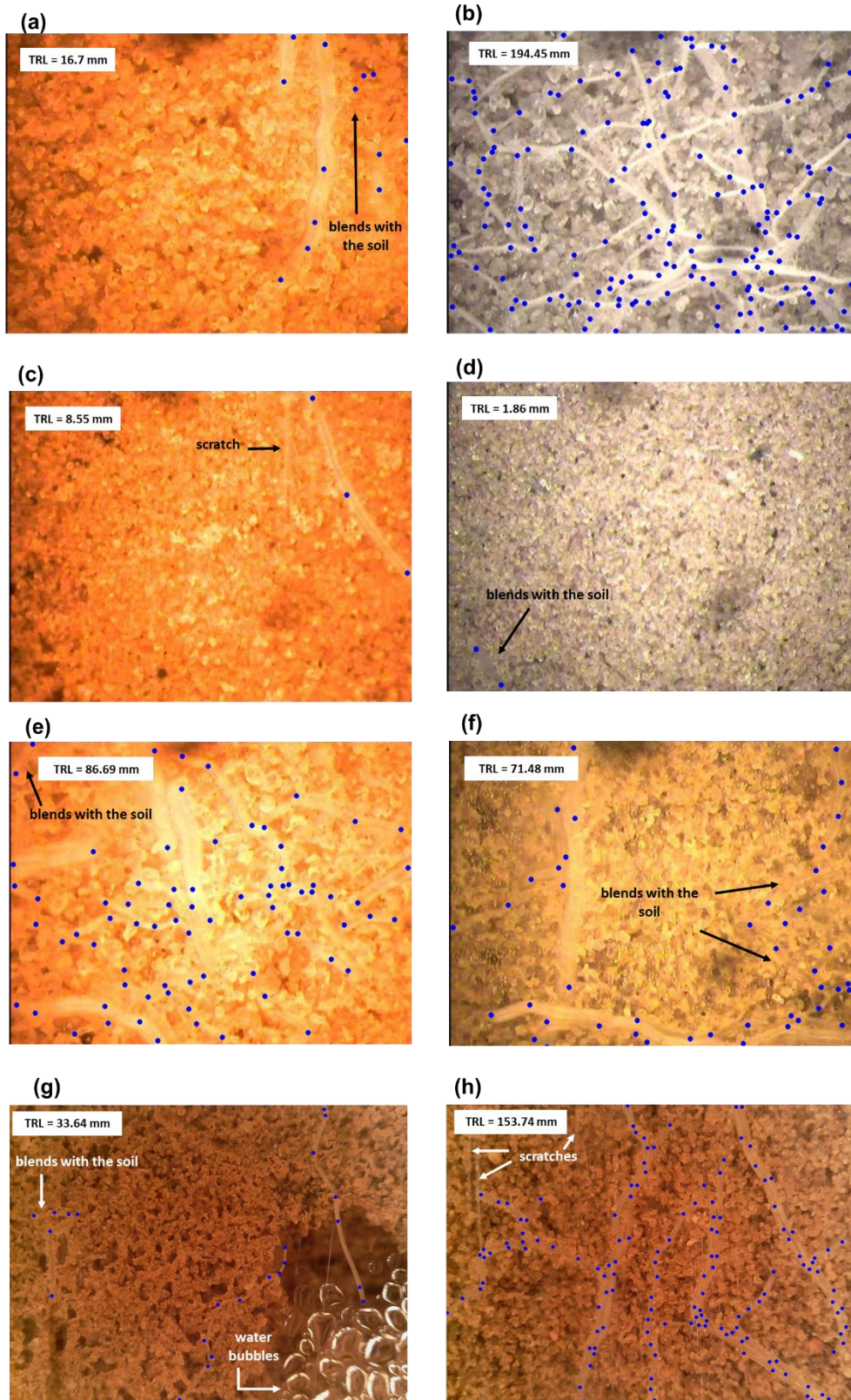


Fig. 4. Examples of images with GT point annotations (blue points, drawn on top) and the GT TRL values (in mm). The images demonstrate the different difficulties in handling such images and are of different crops and different acquisition systems: **(a and b)** corn, **(c and d)** melon, **(e and f)** tomato, and **(g and h)** pepper, where images a–f were acquired with the manual system, and images g and h with the automated system. These difficulties include the differences when **(c and d)** a single root or **(b)** multiple roots are present in an image; **(a–h)** inconsistent scene illumination between images, even of the same crop; **(a, d, e, f, and g)** low contrast between roots and soil; and artifacts, such as **(c and h)** scratches that may look like roots or **(g)** water bubbles.

2.3. Datasets

Using the two acquisition systems, four datasets (denoted datasets 1, 2, 3, and 4) were obtained from six different experiments (Table S1). The experiments included four different crop species (corn, pepper, melon, and tomato) grown under stressed conditions (drought, nitrogen deficiency, low temperature, and salinity combined with container-induced root restriction, respectively) in different years (designated “Melon 2018”, “Melon 2019”, “Tomato 2019”, “Tomato 2020”, “Corn 2020”, and “Pepper 2021”). The soil used for “Pepper 2021” was loamy sand and that for the other experiments was sand (Table S2).

The images of *Dataset 1* and *Dataset 2* were randomly chosen from different tubes, locations (soil depths), and dates of measurement and were used for both training and testing the models. Additional datasets, *Datasets 3* and *Dataset 4*, were obtained to further test the trained models and to generate RLD curves. To do so, each of the datasets used for the estimation of RLD included time-series images acquired from the same tube at different locations. Since the acquisition protocol for these additional datasets was not randomized by crop, tube, location, or time, they were not used for training.

Dataset 1. Included images from all six experiments, initially with 200 images per experiment, which were randomly selected. After 57 extremely corrupted images were deleted¹, the data were randomly divided into training, validation, and test sets, with 73%, 9%, and 18% of the data in each set, respectively. This resulted in 836 images for training, 104 for validation, and 203 for testing.

Dataset 2. Included 420 images acquired using the automated MR system, randomly chosen from the “Pepper 2021” experiment. The set was randomly split such that 72% was used for training, 10% for validation, and 18% for testing. *Datasets 3* and *4* were only used for testing.

Dataset 3. Included 832 images from the “Pepper 2021” experiment, acquired using the automated MR system in a time series, at intervals of 5 days.

Dataset 4. Included a total of 2452 images acquired using the manual MR system from all six experiments. The time series of root images in each experiment was acquired using the manual MR system, at intervals of 14 days after planting.

Ground truth. GT was obtained with manual annotations using the Rootfly software, in which the user must draw a root by adding points along the selected root. These points usually correspond to the coordinates at the start and the end of the root, the splitting points between the main root and other roots, and points to describe the curvature of the root. These points are then connected in a

¹ This was manually determined.

line, the length of which reflects the real length of the selected root. This is done for all roots within an image and for all images in a dataset.

Sensitivity analysis—creating subsets of *Dataset 1*

To examine the influence of the dataset size and MR system type (manual or automated), an additional subset of images was randomly chosen from *Dataset 1*, with a similar size to *Dataset 2* (including the same numbers of images with and without roots), for training, validation, and testing. This allowed us to compare the two different MR systems without the effect of sample size.

2.4. Evaluation

2.4.1. Metrics

To evaluate the metrics of the models, y_i was defined as observation i (true TRL value of image i), \hat{y}_i as the automated estimation of TRL for image i , n as the sample size (number of images), and \bar{y} as the mean of all n (true) observations.

The results were evaluated using the coefficient of determination, denoted R^2 , between the GT manual annotation (dependent variable) and model's estimated TRL values and the following metrics:

1. $|\Delta RL|$: the mean of the absolute difference between the GT value and model's estimate for TRL per image (Eq. 2).

$$(2) |\Delta RL| = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{n}$$

2. *Mean relative deviation (MRD)*: the average relative error of the model's estimates—defined as the absolute difference between the true and estimated values, divided by the true value (Eq. 3). This is only relevant to images with roots.

$$(3) MRD = \frac{\sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}}{n}$$

3. *Normalized root mean squared error (NRMSE)* (Eq. 4).

$$(4) NRMSE = \frac{\sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}}}{y_{max} - y_{min}}$$

4. *Fraction of explained variance, 1 - fraction of variance unexplained (FVU)* (Eq. 5).

$$(5) 1 - FVU = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

This statistic is used to examine whether the suggested models' performance (its *mean squared error is the nominator*) is significantly preferable than always estimating TRL with the trivial value of mean y_i (*the mean squared error of this estimator is the denominator*).

2.4.2. Analysis

Each model was trained with 300 epochs, using the relevant training set. The relevant validation set was used to choose the best epoch based on the MRD values obtained for it. The trained model with chosen weights was then tested on the test set of the relevant dataset (the results are presented in Section 3). The following cases were evaluated:

1. **The suggested models' performance and comparison to previous segmentation-based models.** For each model, training and testing were conducted on the same dataset type (using the training and validation sets of Dataset 1 and Dataset 2). The presented results are compared with previously reported published studies (Smith et al., 2022; Bauer et al., 2022) that performed automated TRL estimation with a segmentation-based model (Section 3.1.1).
2. **Image quality and dataset size – sensitivity analysis results.** The effects of dataset size and image quality were examined by comparing the results obtained by using Dataset 1 and Dataset 2 with these obtained using the subset of Dataset 1 that was the same size as Dataset 2 (Section 3.1.2).
3. **Binary classification.** Automatic distinguishment between images with roots from those without roots (Section 3.1.3). The classification was based on the two models after they were trained using *Dataset 1* (the largest set). To output the images that had no roots in the test set, a threshold was chosen for the TRL estimate of the model. For images in the test set with estimated TRLs smaller than this threshold, the binary output was “no roots”; otherwise, the output was “has roots”. This threshold was chosen as the average estimated TRL value for images without roots in the validation set of *Dataset 1*.
4. **Comparing root traits with RLD calculations.** The trained models were tested on *Datasets 3* and *4*, and RLD plots were generated to compare the outputs based on the manual annotations vs. the models' estimates (Section 3.1.4). Since the MR technique forms images in two dimensions, the RLD was calculated from the TRL for the specific imaging area (cm/cm^2). Data preparation and graphics were conducted with the packages “dplyr”, “reshape2”, and “ggplot2” in RStudio Desktop (version 1.4.1725) software with R (version 4.1.1).
5. **Comparison with a joint model—using both dataset types for training.** The *points model* was trained and tested by joining both types of images, those that were obtained with

the manual MR system and those obtained with the automated MR system. This was done by combining the training, validation, and test sets of *Datasets 1* and 2 (Section 3.2).

6. **Examining the transfer learning possibilities:** The effect of the dataset type used for training a model on the performance of that model when tested on a different type of data was evaluated (Section 3.3). This was examined by first presenting the test results for *Dataset 2* to a model trained exclusively on *Dataset 1*, and vice versa. Second, evaluating how fine-tuning the model by introducing additional training images from the other data type improved the results. This was demonstrated using a model originally trained on *Dataset 2*, which was then fine-tuned by additional training with a subset of images randomly chosen from the training set of *Dataset 1*. The effect was examined as a function of the additional subset size by varying the number of additional training images in the range of 10–200. In the preliminary examination, it was observed that a model could perform well on another dataset type for images without roots. Therefore, the focus here was specifically on adding images with roots for the additional training.

2.4.3. Computation

All experiments were conducted using AMD Ryzen 2920X CPU, NVIDIA GeForce RTX 2080 Ti GPU, CUDA 11.3, and PyTorch 1.2. TRL estimation took <1 s per image using a graphics processing unit (GPU) in test time.

3. Results and Discussion

3.1. Training and testing with the same dataset type

3.1.1. The suggested models' performance and comparison to previous segmentation-based models

Results revealed that both suggested models (the regression model and points model) achieve higher linear correlation (R^2) values than Smith et al. (2022) and Bauer et al. (2022), whereas the NRMSE values of Bauer et al. (2022) were in the same range (Table 1).

Results (Table 1) show that the *points model* outperformed the *regression model* in all cases on all metrics (except the MRD values for the subset of *Dataset 1*, which differed by 0.16% in favor of the *regression model*). When tested on *Dataset 2*, the differences between the models were very small (e.g., a difference of 0.1% in MRD and 0.005 in the R^2 value). In contrast, for *Dataset 1*, which was the more challenging dataset, the *points model* showed a significant advantage in terms of the average estimation error per image, with an MRD value 7.3% lower than that of the

regression model. This implies that having high-quality images (*Dataset 2*) allowed the automated estimation of TRL values with MRD values of ~10%, even without information on the point coordinates, when the *regression model* was used. However, with a more challenging dataset such as *Dataset 1*, the *points model*, which incorporates additional information on the root (point) coordinates, produced significantly better results, reducing the MRD results from ~28% to ~21%.

Table 1. Evaluation metrics for the estimation of TRL using data from both the manual and automated MR cameras, when trained and tested on the same dataset type using the *regression* and *points models*. The results compared to previous works of automated root length estimations. As part of the sensitivity analysis, the results for the subset of *Dataset 1* that was the same size as *Dataset 2* are presented.

Model	dataset	$ \Delta RL $	MRD (GT>0)	NRMSE	R^2	$1 - FVU$
Smith et al. (2022)		-	-	-	0.89-0.92	-
Bauer et al. (2022)		-	-	0.039 - 0.077	0.59 - 0.81	-
<i>regression model</i>	<i>Dataset 1</i>	4.5	28.2%	0.056	0.929	0.9
	<i>Dataset 2</i>	4.902	10.7%	0.0343	0.981	0.974
	<i>Dataset 1</i> - subset	5.95	29.04%	0.071	0.875	0.86
<i>points' model</i>	<i>Dataset 1</i>	3.19	20.9%	0.041	0.958	0.947
	<i>Dataset 2</i>	4.88	10.6%	0.0304	0.986	0.980
	<i>Dataset 1</i> - subset	5.04	29.2%	0.053	0.923	0.921

3.1.2. Image quality and dataset size – sensitivity analysis results

A comparison of the results for the same model on different dataset types (*Dataset 1* vs *Dataset 2*) revealed that for both models, better results were obtained with the automated MR system images (*Dataset 2*) on all metrics except NRMSE. The MRD values were lower by 17.5% and 10.3% for *Dataset 2* than for *Dataset 1* with the *regression model* and *points model*, respectively, even though *Dataset 2* was more than two-fold smaller. This implies that image quality is a major influence on model performance. To determine the influence of the dataset type (and thus the type of image acquisition system), the results of *Dataset 2* were compared with those obtained for the subset of *Dataset 1* (which was the same size as *Dataset 2*). Increasing the dataset size (using the full *Dataset 1*) improved the results compared with those achieved using its subset, but the results for *Dataset 2* remained significantly better than those for *Dataset 1*.

Although the results for *Dataset 2* remained better with both models, the combination of the larger dataset (full *Dataset 1*) with additional information on the root point coordinates (when using the

points model) yielded improved results for the subset of *Dataset 1*. The improvement due to this increase in dataset size was greater in the *points model*, in which the MRD value decreased from 29.2% to 20.9%, whereas in the *regression model*, the MRD value decreased by only 0.84%. Therefore, with more-challenging data, using the *points model* and increasing the dataset size improved the TRL estimation. Since the *points model* proves to be more robust to changes in image quality, the results of the next sections are based on this model.

3.1.3. Binary classification - distinguishing between images with and without roots

The $|\Delta RL|$ values are of specific interest for images without roots. This will be demonstrated for *Dataset 1* (the largest dataset). Each image in this dataset was classified in either the “no roots” or the “has roots” category based on a preset threshold for $|\Delta RL|$. For the *points model*, the threshold value for deciding if an image had roots based on the validation set was 0.47 mm, resulting in a selected threshold of 0.5 mm. This threshold resulted in a 3% error for this binary task. For the *regression model*, the threshold was set to 0.7 mm and the resulting error on the test set was 5%. Fig. 5 presents the $|\Delta RL|$ values for the test set of *Dataset 1* when the *points model* was used, which included 144 images with roots and 59 images without roots (Table S1). It can be seen from these results that 57 of the 59 images (~97%) without roots had an estimated TRL of <1 mm, indicating that the *points model* is robust in detecting images without roots.

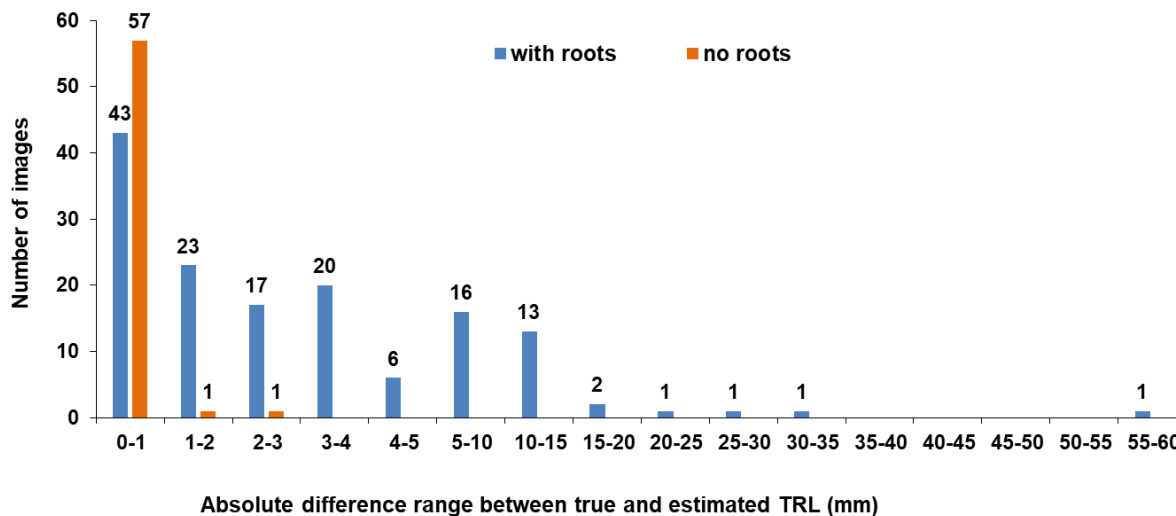


Fig. 5. The number of images in each range of absolute differences between manual annotations of TRL values and the *points model* estimates for the test set of *Dataset 1*. Blue bars show the results for images with roots, and orange bars show those for images without roots.

3.1.4. Comparing root traits with RLD calculations

One specific parameter addressed in this work, which is based on TRL estimates, is the total length of roots per unit of the observed area (RLD). It is an indicator of root distribution in the soil, which is important in understanding the extraction of water and nutrients from the soil (Walter et al., 2009). Therefore, RLD can be used to evaluate the response of roots to various environmental conditions (Machado et al., 2005; Sharma et al., 2018; Soda et al., 2017).

The comparison of the RLD calculations based on data generated from manual annotations (GT values) and those generated from the *points model* TRL estimates for *Dataset 3* (Fig. 6a–b), and *Dataset 4* (Fig. 7) demonstrates similar patterns for the GT values and the model's estimates.

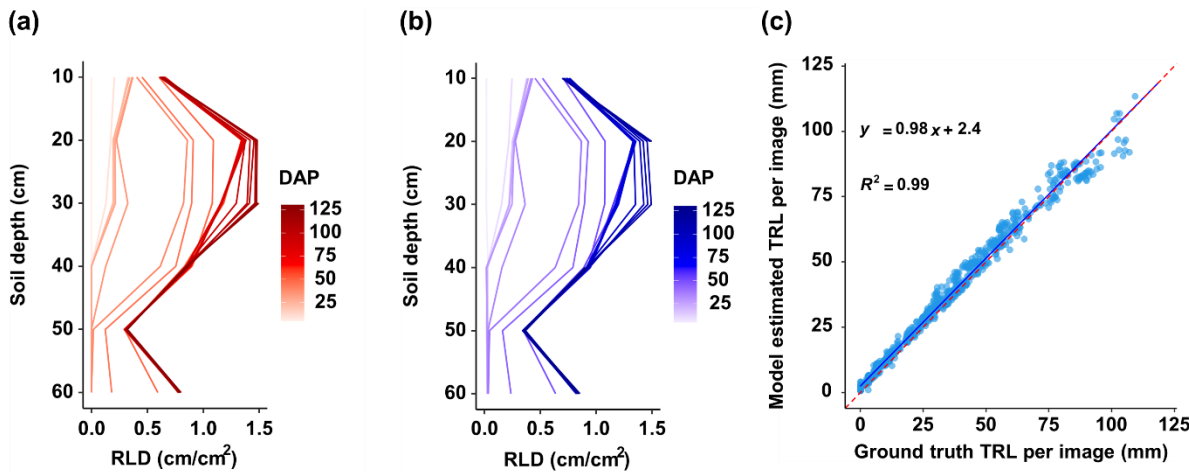


Fig. 6. Comparison of RLD with depth for *Dataset 3* analyzed using (a) manual annotation and (b) model estimation. (c) correlation plot for the *points model* when tested on *Dataset 3*. DAP, days after planting.

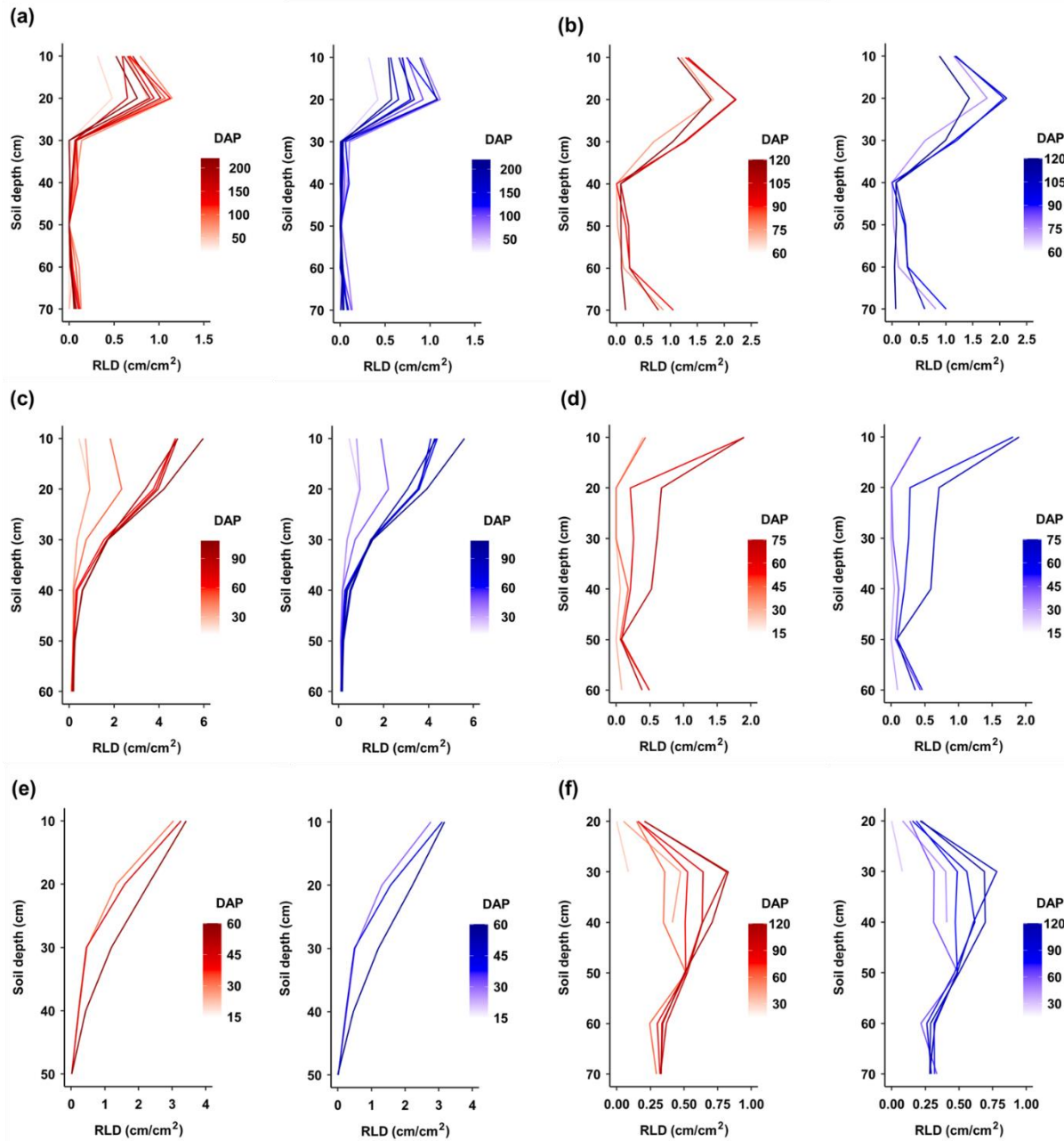


Fig. 7. Comparison of RLD calculated from images captured on different days by manual annotation using Rootfly (red line) or by model estimation (blue line) in six experiments: (a) Melon 2018; (b) Melon 2019; (c) Tomato 2019; (d) Tomato 2020; (e) Corn 2020; and (f) Pepper 2021. DAP, days after planting.

In each case, the model's estimates were generated by a model that was trained on the same data type (manual or automated system), meaning that for *Dataset 3*, a model trained on *Dataset 2* was used, and for *Dataset 4*, a model trained on *Dataset 1* was used. For *Dataset 3*, the model's estimation and manual annotation of TRL values correlated strongly with the 1:1 line (Fig. 6c), demonstrating that the model is robust and reliable. The proposed model outperformed previously reported works that showed proportional bias between the model estimation and manual annotation using RootPainter, an established CNN method (Bauer et al., 2022; Nair et al., 2022; Smith et al., 2022). Therefore, conversion between measurements can be avoided when interpreting the model

results. The root traits results in Fig. 7 and Fig. 8 are important to farmers when deciding the amounts of water and fertilizer to apply based on root exploration and real-time root dynamics. Furthermore, the reduced RLD at later measurement dates on the “Melon 2019” data (Fig. 7b) shows that the *points model* successfully identified “root disappearance”. This is also demonstrated by the points heat map generated by the model (Fig. 8) and indicates that the proposed model can track the appearance, growth, and disappearance of individual roots over time, allowing the accurate estimation of root production and turnover. This is important when studying the permanent disappearance of fine roots, which is a critical criterion for the transition from dead root to soil organic matter via decomposition (Guo et al., 2005; Iversen et al., 2018; Johnson et al., 2001; Primka et al., 2022). Specifically, this model can detect either the loss of length from existing roots or the complete disappearance of a root(s), which indicates that the model can achieve human-expert-level performance.

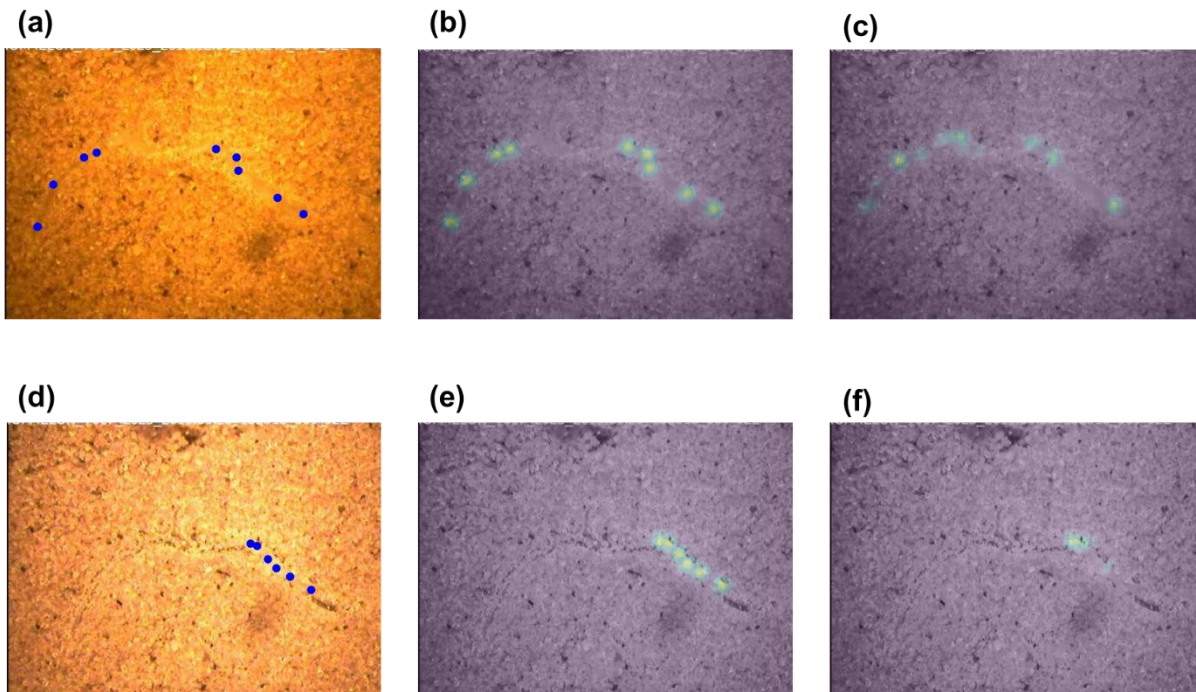


Fig. 8. Examples displaying the models' ability to recognize root disappearance. In the first row, there are (a) one raw image acquired from the "Melon 2018" experiment, (b) GT heat map of this image points on top of the raw image, and (c) model estimation of the points heat map for this image. In the second row, there are (d) the raw image from the same location taken after 50 days, (e) GT heat map of this image points', and (f) model estimation.

Table 2 presents the metrics results when the *points model* was tested on *Datasets 3* and *4*. Although the error values (15.2%) for *Dataset 3* are larger than those for the test set of *Dataset 2* (10.6%; Table 1), the correlation between the GT values and the model's estimates remains very high ($R^2 = 0.989$; Fig. 7c, Table 2).

Table 2. Results of the *points model* when tested on *Dataset 3* (automated MR camera) and *Dataset 4* (manual MR camera).

Test set	Experiment	$ \Delta RL $	MRD (GT>0)	$NRMSE$	R^2	$1 - FVU$
<i>Dataset 3</i>	Pepper 2021	2.613	15.2%	0.031	0.989	0.979
<i>Dataset 4</i>	Melon 2018	1.63	35.7%	0.062	0.908	0.905
	Melon 2019	2.37	17.37%	0.049	0.967	0.958
	Tomato 2019	3.81	21.44%	0.036	0.987	0.976
	Tomato 2020	1.23	18.31%	0.025	0.977	0.977
	Corn 2020	2.87	18.93%	0.030	0.986	0.981
	Pepper 2021	1.49	23.73%	0.046	0.961	0.956

Table 2 also presents the results when the *points model* was tested on *Dataset 4*. For five of the six experiments in *Dataset 4* (all except “Melon 2018”), the results were close to the MRD obtained for the test set of *Dataset 1* (20.9%, Table 1), with higher correlation values. Three of these experiments had better results than those obtained for *Dataset 1* (“Melon 2019”, “Tomato 2020”, and “Corn 2020”), but the results of one experiment (“Melon 2018”) were an exception, with a ~15% increase in MRD relative to the *Dataset 1* results. These experiments' images were especially challenging because of the illumination issue and the extreme blending of the roots with the color of the soil.

3.2. Comparison with a joint model—using both dataset types for training

The “joint model” refers to the *points model* when trained and evaluated using both *Dataset 1* and *Dataset 2*. The results (Table 3) reveal that the model can handle a variety of root images while training, and in fact, it performed as well as it did when trained and tested on the same type of images. It even performed slightly better with a mix of image types during training, with MRD values decreasing by 0.7% for the test set of *Dataset 1* and by 1.1% for *Dataset 2*, and the R^2 values increasing by 0.008 and 0.003 (i.e., remaining almost the same), respectively. The fact that training with more images and different types of images (from both manual and automated systems) provides an enhanced result implies that if multiple dataset types are available for training, it is better to join the datasets for training instead of training each dataset individually. The larger and more varied the training set, the better the model will perform in testing.

Table 3. Test results for the *points model* trained on both dataset types, manual and automated MR camera (Datasets 1 and 2). The results of the model when trained only on the same dataset type as the test set, taken from Table 1, are given in parentheses.

Training and validation set	Test set	$ \Delta RL $	MRD (GT>0)	<i>NRMSE</i>	R^2	$1 - FVU$
joint model (<i>Datasets 1+ 2</i>)	joined model (<i>Datasets 1+ 2</i>)	3.53	16.7%	0.0294	0.980	0.972
	<i>Dataset 1</i>	3.00 (3.19)	20.2% (20.9%)	0.0355 (0.0408)	0.966 (0.958)	0.960 (0.947)
	<i>Dataset 2</i>	4.92 (4.88)	9.7% (10.6%)	0.034 (0.0304)	0.989 (0.986)	0.975 (0.980)

3.3. Examining the transfer learning possibilities

Transfer learning is examined by testing a trained model on another data type and fine-tuning by allowing additional training. The results of testing the *points model* on different data types than the type upon which it was trained (manual vs automated system) revealed that the greatest decline in performance occurred when a model that was trained with automated MR system data (*Dataset 2*) was tested on new data from a manual MR system (*Dataset 1*) (Table 4, Fig. 9). This resulted in an MRD of 39% and R^2 of 0.839 (Table 4) compared with 20.9% and 0.958, respectively, when trained on *Dataset 1* (Table 1).

Table 4. Results of TRL estimation using the points model when tested on a dataset that differed from the training dataset.

Training and validation set	Testing set	$ \Delta RL $	MRD (GT>0)	<i>NRMSE</i>	R^2	$1 - FVU$
<i>Dataset 1</i> (Manual MR system)	<i>Dataset 2</i> (Automated MR system)	21.43	38.6%	0.137	0.956	0.584
<i>Dataset 2</i> (Automated MR system)	<i>Dataset 1</i> (Manual MR system)	7.45	39%	0.083	0.839	0.785

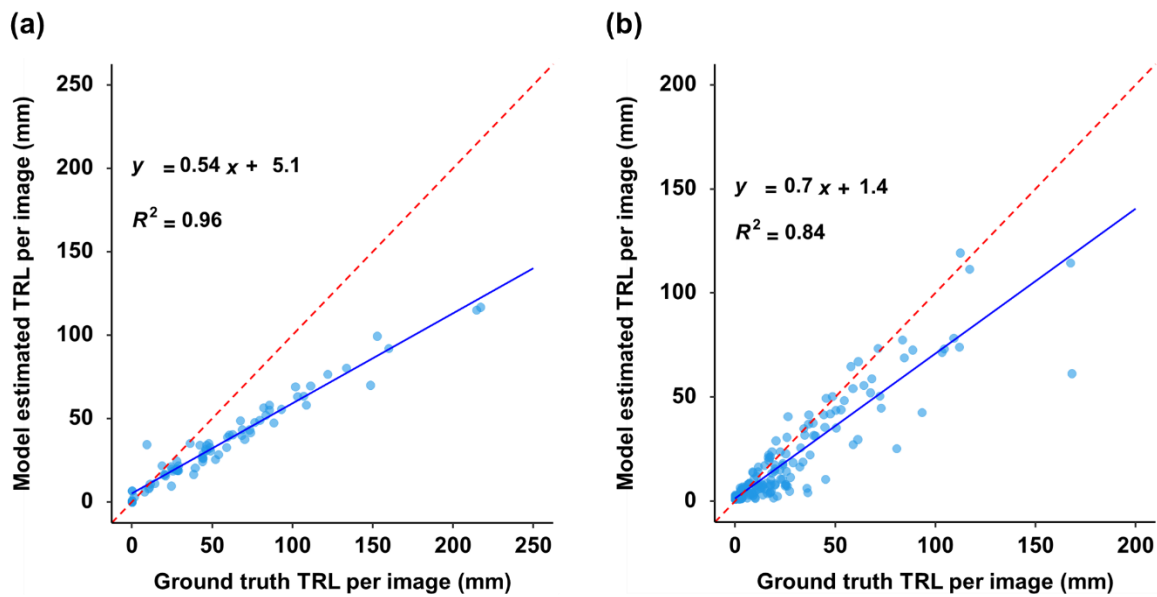


Fig.9. Correlation plots between ground truth TRL per image generated by manual annotation and the estimated TRL per image generated using the *points model* when (a) trained on *Dataset 1* and tested on *Dataset 2*; and (b) trained on *Dataset 2* and tested on *Dataset 1*.

This means that training on lower-quality data (*Dataset 1*) and testing on high-quality data (*Dataset 2*) gave better results than the opposite situation. This is probably attributable to three factors. The first is that *Dataset 1* is almost as twice as large as *Dataset 2*, and the bigger the training set, the better the generalizability of the model to new types of images. Second, the images in *Dataset 1* are more diverse due to their poorer quality than the images from the automated system (*Dataset 2*), which also allows better generalizability. Third, testing a model on more-challenging images when the model was trained on “easier” images also contributed to the reduction in performance on more-difficult data (e.g., in *Dataset 1*, the roots are more difficult to see because their color, more often, blends with the soil).

Fig. 9 also shows a consistent and proportional bias between model estimation and manual annotation, indicating a systematic underestimation of TRL with model estimation. Greater bias was observed (Fig. 9a) when the model was trained on images taken with the manual MR system (*Dataset 1*) and tested on images with the automated MR camera (*Dataset 2*). Furthermore, the more roots present in the images, the more likely was this error (Fig. 9). Although deterioration in performance was observed when the model was tested on images taken from a different MR system, the bias in the method was systematic and very predictable given the high correlation value.

In the next step, an attempt was made to improve results for the case that had a greater performance decline (model trained on *Dataset 2* and tested on *Dataset 1*). The transfer learning options were examined using this model by performing additional training of the given model with datasets of

increasing size from *Dataset 1*. This approach also quantified the additional annotations that were required when an existing trained model was used on a new type of data. The results of these additional training options (Table S3) revealed that by adding only 10 additional training images from another dataset, the model's performance was significantly improved: MRD decreased from 39% to ~31% and R^2 improved from ~0.839 to ~0.927. However, adding further training images improves the results at a slower rate as an increasing number of images are required to get closer to the original performance of a model that was trained upon the same data type as that upon which it was tested.

4. Conclusions and Outlook

Two CNN models that were previously developed for a part counting task were used to automatically analyze root images providing a method for accurately analyzing large numbers of MR images. The architecture of the models was the same as the previously developed models (Itzhaky et al., 2018), however, the training data is different in concept. The points model learns to find another type of points - not the center point of parts (leaves) of a single object (a plant) in an image, but points along each object when having multiple objects in an image.

Results revealed MRD values of 20.9% and 10.6% for the manual and automated MR systems, respectively. The specific contributions of this work are as follows:

1. Automated tools for the estimation of TRL from MR images are presented, the results of which correlate strongly with manually obtained values: $R^2 = 0.929\text{--}0.958$ for images from the manual MR system (Dataset 1) and $R^2 = 0.981\text{--}0.989$ for images from the automated MR system (Dataset 2).
2. A dataset of MR images with thoroughly examined root annotations is made publicly available.
3. The *points model*, which identifies the coordinates of roots in an image (outputs an estimate of the heat map of points coordinates along the roots), allows the visual inspection of the roots in each image and the examination of the human annotations.
4. The ability of the suggested *regression model* to estimate TRL values per MR image, without the need for any annotation for training other than the TRL value, is demonstrated.
5. The models can be used to distinguish between images with and without roots. This was done by setting a threshold to the TRLs estimate, and in the case of an estimate smaller than this threshold, the binary output was “no roots”; otherwise, the output was “has roots”. This resulted in 3% and 5% errors for this binary task, for the *points model* and the *regression model* respectively.

6. The ability of a CNN model trained to estimate TRL values on a specific dataset to analyze new datasets (new distributions) is demonstrated. The change in performance is quantified and the ability of the *points model* to improve as a function of the number of additional training images from the new dataset is demonstrated. It was shown that as few as 10 additional training images can present a significant improvement. This implies that using a small sample from a new dataset (even if it has more “difficult” images) for additional training can significantly enhance the model’s performance, confirming the generalizability of the model. Dozens of additional training images, rather than thousands, are required. This is important, especially for root studies, where datasets tend to be relatively small, and annotation is time-consuming.

This study provides a method for the rapid and reliable estimation of root growth patterns, allowing root phenotyping and analysis with high temporal resolution. This will help growers to make sound decisions about water and nutrient supply based on up-to-date root growth information and to predict plant stress before visible stress symptoms appear. This has important implications for the timely and cost-effective control of stress in precision agriculture. The proposed approach can potentially maximize the utilization of MR-based root phenotyping platforms.

CRedit authorship contribution statement

Faina Khoroshevsky: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft, Visualization. **Kaining Zhou:** Conceptualization, Formal analysis, Data Curation, Writing - Original Draft, Visualization. **Sharon Chemweno:** Data Curation, Writing - Original Draft. **Yael Edan:** Supervision, Conceptualization, Methodology, Writing – review & editing. **Aharon Bar-Hillel:** Conceptualization, Methodology, Writing – review & editing. **Ofer Hadar:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Boris Rewald:** Writing – review & editing. **Pavel Baykalov:** Writing – review & editing. **Jhonathan E. Ephrath:** Supervision, Methodology, Writing – review & editing. **Naftali Lazarovitch:** Conceptualization, Funding acquisition, Supervision, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1 **Acknowledgments**

2 This research was supported by the Israeli Ministry of Agriculture and Rural Development (Eugene
3 Kandel Knowledge Centers) as part of the Root of the Matter—The Root Zone Knowledge Center
4 for Leveraging Modern Agriculture, and grant number 16-38-0044 of the Israeli Ministry of
5 Agriculture and Rural Development, and the European Union’s Horizon 2020 Research and
6 Innovation Program (grant agreement no. 777222) (ATTRACT project “NextMR-IAA”). Partial
7 support was provided by the Ben-Gurion University of the Negev and the W. Gunther Plaut Chair
8 in Manufacturing Engineering.

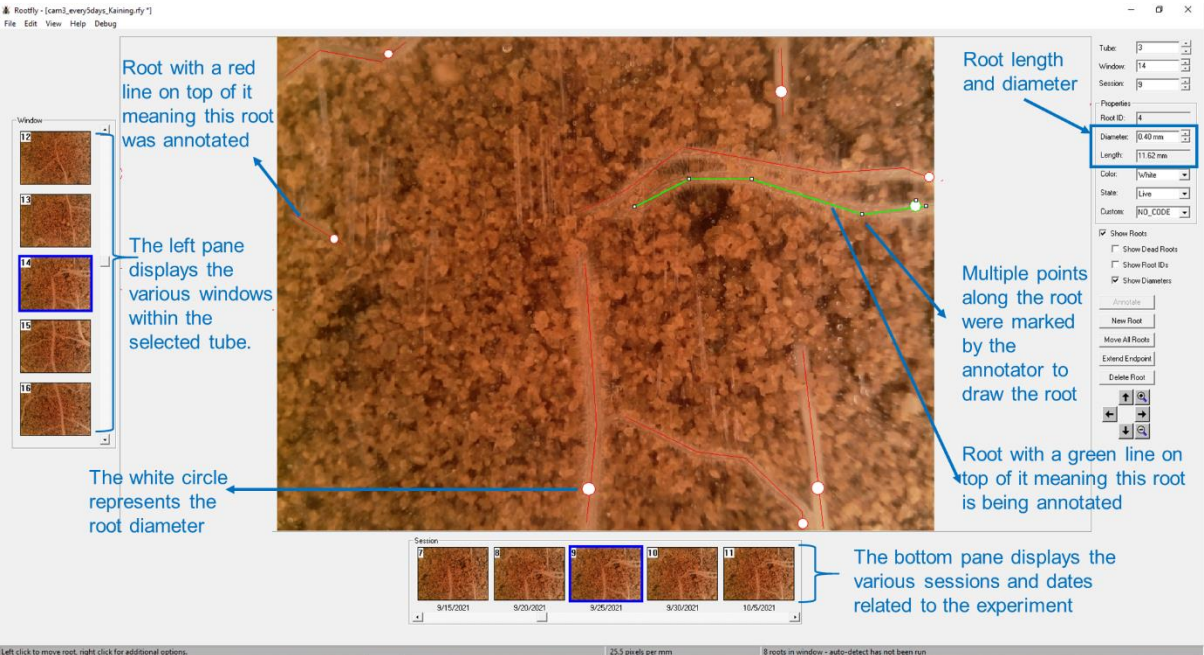
9

10 **Data availability**

11 The datasets generated and analyzed during the current study are available in the Zenodo repository,
12 <https://doi.org/10.5281/zenodo.7482146>.

13

1 **Supplementary Materials**



2
3 **Fig. 1.** Annotated snapshot of the Rootfly graphical user interface (GUI). The image window located in the middle of
4 the workspace displays the current image being interpreted and annotated. The window pane located at the left of the
5 workspace displays the various windows within the selected tube. The session pane located at the bottom of the
6 workspace displays the various sessions and dates related to the experiment. The properties pane located on the right
7 of the workspace displays the properties associated with the roots of a current session.
8

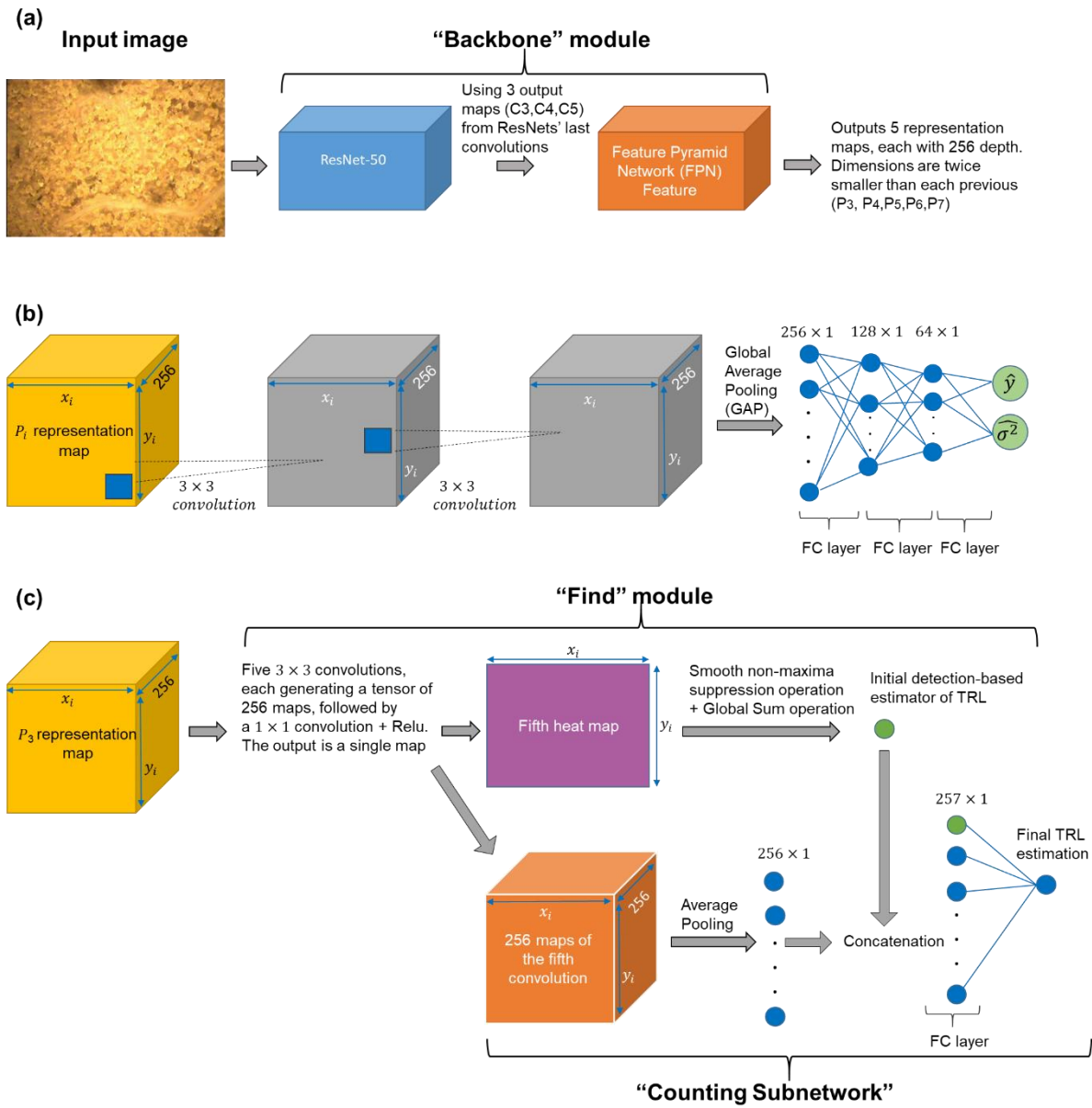


Fig. 2. Schema of the network modules (Itzhaky et al., 2018). (a) modules for generating a multiscale image representation with the “Backbone” module. (b) architecture of the direct regression module of the regression model—the subnetwork that is followed by the “Backbone”. (c) architecture of the modules that follow the “Backbone” in the *points model*.

1 **Table 1.** Description of root image datasets, referring to the MR system type, crop, dataset size, and information on
2 data partitioning to the training, validation, and testing sets.

MR system	Dataset	Experiment name	Total number of images	Images without roots	Training images	Validation images	Test images
Manual	<i>Dataset 1</i>	Melon 2018	175	74	126	18	31
		Melon 2019	199	84	145	18	36
		Tomato 2019	200	82	146	18	36
		Tomato 2020	199	7	146	17	36
		Corn 2020	181	44	135	16	30
		Pepper 2021	189	53	138	17	34
		Total for Dataset 1	1143	413	836 (318 without roots)	104 (36 without roots)	203 (59 without roots)
Automated	<i>Dataset 2</i>	Pepper 2021	420	37	302 (26 without roots)	41 (4 without roots)	77 (7 without roots)
Automated	<i>Dataset 3</i>	Pepper 2021	832	184	-	-	-
Manual	<i>Dataset 4</i>	Melon 2018	521	300	-	-	-
		Melon 2019	217	77	-	-	-
		Tomato 2019	364	92	-	-	-
		Tomato 2020	178	91	-	-	-
		Corn 2020	116	31	-	-	-
		Pepper 2021	224	39	-	-	-

1 **Table 2.** Description of root image datasets—manual and automated cameras.

Camera type	Dataset	Total number of images	Images without roots	Experiment name	Crop type	Stress type	Soil Type
Manual	Dataset 1	175	74	Melon 2018	Melon	Low temperature	Sand
		199	84	Melon 2019	Melon	Low temperature	Sand
		200	82	Tomato 2019	Tomato	Salt stress	Sand
		199	7	Tomato 2020	Tomato	Salt stress and root restriction	Sand
		181	44	Corn 2020	Corn	Drought	Sand
		189	53	Pepper 2021	Pepper	Nitrogen deficiency	Loamy sand
Automated	Dataset 2	420	37	Pepper 2021	Pepper	Nitrogen deficiency	Loamy sand
Automated	Dataset 3	832	184	Pepper 2021	Pepper	Nitrogen deficiency	Loamy sand
Manual	Dataset 4	217	77	Melon 2019	Melon	Low temperature	Sand
		116	31	Corn 2020	Corn	Drought	Sand

2
3

1 **Table 3.** Results of testing the *points model* on the test set of *Dataset 1* when it was initially trained on images from
2 *Dataset 2*, but with additional training using randomly chosen images from *Dataset 1*. This was compared with the
3 original results when the model was both trained and tested on the same data type (*Dataset 1*).

$1 - FVU$	R^2	$NRMSE$	MRD (GT>0)	$ \Delta RL $	
0.947	0.958	0.0408	20.9%	3.19	Original results – training on Dataset 1
0.785	0.839	0.083	39%	7.45	Training on Dataset 2
0.903	0.927	0.056	31.3%	5.38	Adding 10 images
0.923	0.941	0.049	30.6%	4.94	Adding 20 images
0.939	0.947	0.044	30.0%	4.49	Adding 40 images
0.937	0.946	0.045	28.1%	4.26	Adding 60 images
0.934	0.946	0.046	27.8%	4.32	Adding 80 images
0.915	0.938	0.052	26.3%	4.71	Adding 100 images
0.931	0.941	0.047	25.8%	4.41	Adding 120 images
0.940	0.947	0.043	25.5%	4.16	Adding 140 images
0.925	0.953	0.049	25.5%	4.57	Adding 160 images
0.945	0.951	0.042	24.1%	4.07	Adding 180 images
0.953	0.955	0.039	24.03%	4.02	Adding 200 images

4

References

- Ajmera, I., Henry, A., Radanielson, A. M., Klein, S. P., Ianevski, A., Bennett, M. J., Band, L.R., & Lynch, J. P. (2022). Integrated root phenotypes for improved rice performance under low nitrogen availability. *Plant Cell Environ.* 45(3), 805-822.
<https://doi.org/10.1111/pce.14284>
- Amtmann, A., Bennett, M. J., & Henry, A. (2022). Root phenotypes for the future. *Plant Cell Environ.* 45(3), 595-601.
<https://doi.org/10.1111/pce.14269>
- Atanbori, J., Montoya-P, M. E., Selvaraj, M. G., French, A. P., & Pridmore, T. P. (2019). Convolutional neural net-based cassava storage root counting using real and synthetic images. *Front. Plant Sci.* 10, 1516.
<https://doi.org/10.3389/fpls.2019.01516>
- Bauer, F. M., Lärm, L., Morandage, S., Lobet, G., Vanderborght, J., Vereecken, H., & Schnepf, A. (2022). Development and validation of a deep learning based automated minirhizotron image analysis pipeline. *Plant Phenomics*. 2022:9758532.
<http://doi.org/10.34133/2022/9758532>
- Busener, N., Kengkanna, J., Saengwilai, P. J., & Bucksch, A. (2020). Image-based root phenotyping links root architecture to micronutrient concentration in cassava. *Plants, People, Planet.* 2(6), 678-687.
<https://doi.org/10.1002/ppp3.10130>
- Castro-Valdecantos, P., Apolo-Apolo, O. E., Pérez-Ruiz, M., & Egea, G. (2022). Leaf area index estimations by deep learning models using RGB images and data fusion in maize. *Precis. Agric.* 1-18.
<https://doi.org/10.1007/s11119-022-09940-0>
- Danilevicz, M. F., Bayer, P. E., Nestor, B. J., Bennamoun, M., & Edwards, D. (2021). Resources for image-based high-throughput phenotyping in crops and data sharing challenges. *Plant Physiol.* 187(2), 699-715.
<https://doi.org/10.1093/plphys/kiab301>
- Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Li, F. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255).
<http://doi.org/10.1109/CVPR.2009.5206848>

1 Douarre, C., Schielein, R., Frindel, C., Gerth, S., & Rousseau, D. (2018). Transfer learning from
 2 synthetic data applied to soil–root segmentation in x-ray tomography images. *J. Imaging*.
 3 4(5), 65.
 4 <https://doi.org/10.3390/jimaging4050065>

5 Ephrath, J. E., Klein, T., Sharp, R. E., & Lazarovitch, N. (2020). Exposing the hidden half: Root
 6 research at the forefront of science. *Plant Soil*. 447(1), 1-5.
 7 <https://doi.org/10.1007/s11104-019-04417-y>

8 Farjon, G., Itzhaky, Y., Khoroshevsky, F., & Bar-Hillel, A. (2021). Leaf counting: Fusing network
 9 components for improved accuracy. *Front. Plant Sci.* 12, 575751.
 10 <http://doi.org/10.3389/fpls.2021.575751>

11 Farjon, G., Krikeb, O., Hillel, A. B., & Alchanatis, V. (2020). Detection and counting of
 12 flowers on apple trees for better chemical thinning decisions. *Precis. Agric.* 21(3), 503-
 13 521.
 14 <https://doi.org/10.1007/s11119-019-09679-1>

15 Fajardo, M., & Whelan, B. M. (2021). Within-farm wheat yield forecasting incorporating off-farm
 16 information. *Precis. Agric.* 22(2), 569-585.
 17 <https://doi.org/10.1007/s11119-020-09779-3>

18 Guo, L. B., Halliday, M. J., Siakimotu, S. J. M., & Gifford, R. M. (2005). Fine root production and
 19 litter input: its effects on soil carbon. *Plant Soil*. 272(1), 1-10.
 20 <https://doi.org/10.1007/s11104-004-3611-z>

21 Han, T. H., & Kuo, Y. F. (2018). Developing a system for three-dimensional quantification of root
 22 traits of rice seedlings. *Comput. Electron. Agric.* 152, 90-100.
 23 <https://doi.org/10.1016/j.compag.2018.07.001>

24 Hartmann, A., Šimůnek, J., Aidoo, M.K., Seidel, S.J., & Lazarovitch, N. (2018). Implementation
 25 and application of a root growth module in HYDRUS. *Vadose Zone J.* 17(1), 1-16.
 26 <https://doi.org/10.2136/vzj2017.02.0040>

27 He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition.
 28 In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp.
 29 770-778).
 30 <https://doi.org/10.48550/arXiv.1512.03385>

31 Itzhaky, Y., Farjon, G., Khoroshevsky, F., Shpigler, A., & Bar-Hillel, A. (2018, September). Leaf
 32 counting: Multiple scale regression and detection using deep CNNs. In *BMVC* (p. 328).

Iversen, C. M., Childs, J., Norby, R. J., Ontl, T. A., Kolka, R. K., Brice, D. J., McFarlane, K., J., & Hanson, P. J. (2018). Fine-root growth in a forested bog is seasonally dynamic, but shallowly distributed in nutrient-poor peat. *Plant Soil*. 424(1), 123-143.
<https://doi.org/10.1007/s11104-017-3231-z>

Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics*. 2020.
<http://doi.org/10.34133/2020/4152816>

Johnson, M. G., Tingey, D. T., Phillips, D. L., & Storm, M. J. (2001). Advancing fine root research with minirhizotrons. *Environ. Exp. Bot.* 45(3), 263-289.
[https://doi.org/10.1016/S0098-8472\(01\)00077-6](https://doi.org/10.1016/S0098-8472(01)00077-6)

Kalantar, A., Dashuta, A., Edan, Y., Dafna, A., Gur, A., & Klapp, I. (2019). Estimating melon yield for breeding processes by machine-vision processing of UAV images. *Precision Agriculture*'19 (pp. 1386-1393). Wageningen Academic Publishers.
https://doi.org/10.3920/978-90-8686-888-9_47

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precision Agriculture*. 20(6), 1107-1135.
<https://doi.org/10.1007/s11119-019-09642-0>

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*. 30.
<https://doi.org/10.48550/arXiv.1703.04977>

Khoroshevsky, F., Khoroshevsky, S., Markovich, O., Granitz, O., & Bar-Hillel, A. (2020). Phenotyping problems of parts-per-object count. In *European Conference on Computer Vision* (pp. 261-278). Springer, Cham.
https://doi.org/10.1007/978-3-030-65414-6_19

Khoroshevsky, F., Khoroshevsky, S., & Bar-Hillel, A. (2021). Parts-per-object count in agricultural images: Solving phenotyping problems via a single deep neural network. *Remote Sens.* 13(13), 2496.
<https://doi.org/10.3390/rs13132496>

Khoroshevsky, F., Zhou, K., & Lazarovitch, N. (2022). Dataset for "Root Length Estimation: Automated Minirhizotron Image Analysis with Convolutional Networks without Segmentation" (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7482146>

1 Koevoets, I. T., Venema, J. H., Elzenga, J. T. M., & Testerink, C. (2016). Roots withstanding their
 2 environment: Exploiting root system architecture responses to abiotic stress to improve crop
 3 tolerance. *Front. Plant Sci.* 7, 1335.
 4 <https://doi.org/10.3389/fpls.2016.01335>

5 Kume, T., Ohashi, M., Makita, N., Kho, L. K., Katayama, A., Endo, I., Matsumoto, K., & Ikeno,
 6 H. (2018). Image analysis procedure for the optical scanning of fine-root dynamics: errors
 7 depending on the observer and root-viewing window size. *Tree Physiol.* 38(12), 1927-1938.
 8 <https://doi.org/10.1093/treephys/tpy124>

9 Kuppe, C. W., Kirk, G. J. D., Wissuwa, M., & Postma, J. A. (2022). Rice increases phosphorus
 10 uptake in strongly sorbing soils by intra-root facilitation. *Plant Cell Environ.* 45(3), 884-899.
 11 <https://doi.org/10.1111/pce.14285>

12 Lak, Z. A., Sandén, H., Mayer, M., Godbold, D. L., & Rewald, B. (2020). Plasticity of root
 13 traits under competition for a nutrient-rich patch depends on tree species and
 14 possesses a large congruency between intra-and interspecific situations. *Forests.*
 15 11(5), 528.
 16 <https://doi.org/10.3390/f11050528>

17

18 Lambers, H., Shane, M. W., Cramer, M. D., Pearse, S. J., & Veneklaas, E. J. (2006). Root structure
 19 and functioning for efficient acquisition of phosphorus: Matching morphological and
 20 physiological traits. *Ann. Bot.* 98(4), 693-713.
 21 <https://doi.org/10.1093/aob/mcl114>

22 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature.* 521(7553), 436-444.
 23 <https://doi.org/10.1038/nature14539>

24 Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017a). Feature pyramid
 25 networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision*
 26 *and Pattern Recognition* (pp. 2117-2125).
 27 <http://doi.org/10.1109/CVPR.2017.106>

28 Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017b). Focal loss for dense object
 29 detection. In *Proceedings of the IEEE International Conference on Computer Vision and*
 30 *Pattern Recognition* (pp. 2980-2988).
 31 <http://doi.org/10.1109/ICCV.2017.324>
 32 <https://doi.org/10.1016/j.media.2017.07.005>

33 Lynch, J. P. (2018). Rightsizing root phenotypes for drought resistance. *J. Exp. Bot.* 69(13), 3279-
 34 3292.

<https://doi.org/10.1093/jxb/ery048>

Lynch, J. P., Strock, C. F., Schneider, H. M., Sidhu, J. S., Ajmera, I., Galindo-Castañeda, T., Klein, S., P., & Hanlon, M. T. (2021). Root anatomy and soil resource capture. *Plant Soil.* 466(1), 21-63.

<https://doi.org/10.1007/s11104-021-05010-y>

Machado, R., & Oliveira, M. D. R. G. (2005). Tomato root distribution, yield and fruit quality under different subsurface drip irrigation regimes and depths. *Irrig Sci.* 24(1), 15-24.

<http://doi.org/10.1007/s00271-005-0002-z>

McGrail, R. K., Van Sanford, D. A., & McNear Jr, D. H. (2020). Trait-based root phenotyping as a necessary tool for crop selection and improvement. *Agronomy.* 10(9), 1328.

<https://doi.org/10.3390/agronomy10091328>

Molotoks, A., Smith, P., & Dawson, T. P. (2021). Impacts of land use, population, and climate change on global food security. *Food Energy Secur.* 10(1), e261.

<https://doi.org/10.1002/fes3.261>

Nair, R., Strube, M., Hertel, M., Kolle, O., Rolo, V., & Migliavacca, M. (2022). High frequency root dynamics: Sampling and interpretation using replicated robotic minirhizotrons. *J. Exp. Bot.* erac427.

<http://doi.org/10.1093/jxb/erac427>

Paez-Garcia, A., Motes, C. M., Scheible, W. R., Chen, R., Blancaflor, E. B., & Monteros, M. J. (2015). Root traits and phenotyping strategies for plant improvement. *Plants.* 4(2), 334-355.

<https://doi.org/10.3390/plants4020334>

Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., Bulat, A., Tzimiropoulos, G., Wells, D. M., Murchie, E. H., Pridmore, T. P., & French, A. P. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience.* 6(10), gix083

<https://doi.org/10.1093/gigascience/gix083>

Primka IV, E. J., Adams, T. S., Buck, A. S., & Eissenstat, D. M. (2022). Shifts in root dynamics along a hillslope in a mixed, mesic temperate forest. *Plant Soil.* 1-17.

<https://doi.org/10.1007/s11104-022-05469-3>

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One.* 8(6), e66428.

<https://doi.org/10.1371/journal.pone.0066428>

Rewald, B., & Ephrath, J. E. (2013). Minirhizotron techniques. In *Plant Roots: The Hidden Half*, Fourth Edition (pp. 735-750). CRC press.

1 Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical
2 image segmentation. In Medical Image Computing and Computer-Assisted Intervention–
3 MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015,
4 Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.
5 https://doi.org/10.1007/978-3-319-24574-4_28

6 Saleem, M. H., Potgieter, J., & Arif, K. M. (2021). Automation in agriculture by machine and deep
7 learning techniques: a review of recent developments. *Precis. Agric.* 22(6), 2053-2091.
8 <https://doi.org/10.1007/s11119-021-09806-x>

9 Seethepalli, A., Dhakal, K., Griffiths, M., Guo, H., Freschet, G. T., & York, L. M. (2021).
10 RhizoVision Explorer: open-source software for root image analysis and measurement
11 standardization. *AoB Plants*. 13(6), plab056.
12 <https://doi.org/10.1093/aobpla/plab056>

13 Sindagi, V. A., & Patel, V. M. (2017). Generating high-quality crowd density maps using contextual
14 pyramid CNNs. In Proceedings of the IEEE International Conference on Computer Vision
15 (pp. 1861-1870).
16 <http://doi.org/10.1109/ICCV.2017.206>

17 Smith, A. G., Han, E., Petersen, J., Olsen, N. A. F., Giese, C., Athmann, M., Dresbøll, D. B., &
18 Thorup-Kristensen, K. (2022). RootPainter: deep learning segmentation of biological
19 images with corrective annotation. *New Phytol.* 236(2), 774-791.
20 <https://doi.org/10.1111/nph.18387>

21 Smith, A. G., Petersen, J., Selvan, R., & Rasmussen, C. R. (2020). Segmentation of roots in soil
22 with U-Net. *Plant Methods*. 16(1), 1-15.
23 <https://doi.org/10.1186/s13007-020-0563-0>

24 Sharma, S. P., Leskovar, D. I., Volder, A., Crosby, K. M., & Ibrahim, A. M. H. (2018). Root
25 distribution patterns of reticulatus and inodorus melon (*Cucumis melo L.*) under subsurface
26 deficit irrigation. *Irrig Sci.* 36(6), 301-317.
27 <https://doi.org/10.1007/s00271-018-0587-7>

28 Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant
29 stress phenotyping: Trends and future perspectives. *Trends Plant Sci.* 23(10), 883-898.
30 <https://doi.org/10.1016/j.tplants.2018.07.004> 883

31 Soda, N., Ephrath, J. E., Dag, A., Beiersdorf, I., Presnov, E., Yermiyahu, U., & Ben-Gal, A. (2017).
32 Root growth dynamics of olive (*Olea europaea L.*) affected by irrigation induced salinity.
33 *Plant Soil*. 411(1), 305-318.
34 <https://doi.org/10.1007/s11104-016-3032-9>

1 Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In
2 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
3 (pp. 10781-10790).
4 <https://doi.org/10.48550/arXiv.1911.09070>

5 UN–United Nations, Department of Economic and Social Affairs, Population Division. 2022.
6 World Population Prospects 2022, Online Edition. Available from
7 <https://population.un.org/wpp> (Accessed 4 December 2022)

8 Walter, A., Silk, W. K., & Schurr, U. (2009). Environmental effects on spatial and temporal patterns
9 of leaf and root growth. *Annu Rev Plant Biol.* 60(1), 279-304.
10 <http://dx.doi.org/10.1146/annurev.arplant.59.032607.092819>

11 Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A. S., Qiu, Z., & Jin,
12 J. (2019). SegRoot: A high throughput segmentation method for root image
13 analysis. *Comput Electron Agric.* 162, 845-854.
14 <https://doi.org/10.1016/j.compag.2019.05.017>

15 Yasrab, R., Atkinson, J. A., Wells, D. M., French, A. P., Pridmore, T. P., & Pound, M. P. (2019).
16 RootNav 2.0: Deep learning for automatic navigation of complex plant root
17 architectures. *GigaScience.* 8(11), giz123.
18 <http://doi.org/10.1093/gigascience/giz123>

19 Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep
20 neural networks? *Advances in neural information processing systems* 27(NIPS '14), NIPS
21 Foundation.

22 Zeng, G., Birchfield, S. T., & Wells, C. E. (2008). Automatic discrimination of fine roots in
23 minirhizotron images. *New Phytol.* 177(2), 549-557. [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-8137.2007.02271.x)
24 [8137.2007.02271.x](https://doi.org/10.1111/j.1469-8137.2007.02271.x)

25 Zhao, J., Sykacek, P., Bodner, G., & Rewald, B. (2018). Root traits of European *Vicia faba*
26 cultivars—using machine learning to explore adaptations to agroclimatic conditions. *Plant Cell*
27 *Environ.* 41(9), 1984-1996. <https://doi.org/10.1111/pce.13062>