As good as human experts in detecting plant roots in minirhizotron images but efficient and reproducible: The Convolutional Neural Network “RootDetector”

Bo Peters (bo.peters@uni-greifswald.de)
University of Greifswald
Gesche Blume-Werry
Umeå University
Alexander Gillert
Fraunhofer Institute for Computer Graphics Research
Sarah Schwieger
Umeå University
Uwe Freiherr von Lukas
Fraunhofer Institute for Computer Graphics Research
Juergen Kreyling
University of Greifswald

Article

Keywords: plant root dynamics, minirhizotron technique, Computer Vision, Convolutional Neural Network, segmentation, belowground phenology

Posted Date: March 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1434431/v1

As good as human experts in detecting plant roots in minirhizotron images but efficient and reproducible: The Convolutional Neural Network "RootDetector"

Bo Peters[1*], Gesche Blume-Werry[1,2], Alexander Gillert[3], Sarah Schwieger[1,2], Uwe Freiherr von Lukas[3,4], Juergen Kreyling[1]

1: Experimental Plant Ecology, Institute of Botany and Landscape Ecology, University of Greifswald, Soldmannstraße 15, D-19489 Greifswald, Germany

2: Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden

3: Fraunhofer Institute for Computer Graphics Research IGD, Rostock

4: Institute for Visual & Analytic Computing, University of Rostock

* corresponding author: bo.peters@uni-greifswald.de

Running headline: RootDetector – a CNN for minirhizotron images

Abstract

Plant roots influence many ecological and biogeochemical processes, such as carbon, water and nutrient cycling. Because of difficult accessibility, knowledge on plant root dynamics in field conditions, however, is fragmentary at best. Minirhizotrons, i.e. transparent tubes placed in the substrate into which specialized cameras are inserted, facilitate the capture of high-resolution images of root dynamics at the soil-tube interface with little to no disturbance after the initial installation. Their use, especially in field studies with multiple species and heterogeneous substrates, though, is limited by the amount of work that subsequent manual tracing of roots in the images requires. Furthermore, the reproducibility and objectivity of manual root detection is questionable. Here, we use a Convolutional Neural Network (CNN) for the automatic detection of roots in minirhizotron images and compare the performance of our RootDetector with human analysts with different levels of expertise. The minirhizotron data stem from various wetland types on organic soils. RootDetector showed a high capability to correctly segmenting root pixels in minirhizotron images from field observations (F1 = 0.6044; r^2 compared to a human expert = 0.99). Reproducibility among humans, however, depended strongly on expertise level, with novices showing drastic variation among individual analysts and annotating on average almost 3-times higher root length/cm^2 per image compared to expert analysts. Analyses with RootDetector save resources, are reproducible and objective, and are as accurate as manual analyses performed by human experts.

keywords:

plant root dynamics, minirhizotron technique, Computer Vision, Convolutional Neural Network, segmentation, belowground phenology

1. Introduction

Quantifying and monitoring biomass accumulation from plants is of growing interest for many scientific fields as it provides a valuable metric for complex ecosystem dynamics. Around 30-95% of plant biomass is located belowground across biomes in form of roots (Mokany, Raison, & Prokushkin, 2006), and roots mediate the carbon input into the soil through rhizodeposition (Lynch & Whipps, 1990). Thus, knowledge about root dynamics is crucial for the understanding of carbon stocks and fluxes in ecosystems, and their representation is essential in coupled biosphere-atmosphere models (Warren et al., 2015, McCormack et al., 2015). However, even basic data such as root length, density, seasonal activity or growth rates are fragmentary at best due to difficult accessibility and high susceptibility of roots to disturbance (Blume-Werry, 2021; Mooney et al., 2012). Many methods for root biomass surveys under field conditions are not very accurate and highly destructive. The washing out of soil samples, for example, has been shown to record only 60% of the biomass, as fine roots, functionally the most important root type, are commonly lost (Robinson, 2004). Destructive methods also do not allow for insights into root dynamics over time, as they provide a mere snapshot. Therefore, by far the most important tool for recording root dynamics in the field has become the so-called minirhizotron technique (Hansson, Steen, & Andren, 1992). Minirhizotrons are transparent tubes, which, once inserted into the soil, enable regular recordings of root growth at the tube-soil interface via imaging by tube scanners or cameras. As this method of sampling is non-destructive and minimally invasive, it can be conducted as often as required and over any length of time, thus enabling precise measurement and visualization of important root parameters. Therefore, the use of minirhizotrons is a highly effective method for detailed investigation even of the finest and most short-lived root types (lifespans of days to weeks) in high temporal and spatial resolution, allowing for investigations of ever more apparent decoupled belowground and aboveground seasonal growth patterns (‘phenology’) (Blume-Werry, Wilson, Kreyling, & Milbau, 2016; Liu et al., 2021; Schwieger, Blume-Werry, Peters, Smiljanić, & Kreyling, 2019). Until now, the detection of roots in the minirhizotron images is done manually by human

66 analysts, at least in field studies dealing with heterogeneous substrates and multiple species. Depending on image quality and root abundance, the processing of a single image can take several
68 hours. This is one reason why there are very few long-term measurement series in fine temporal resolution on root dynamics. The high time demand does up to now also not allow for the
70 quantification of spatial variation in root traits and root dynamics (Träger, Wilson, & Kudo, 2017). This shows that the biggest obstacle for providing sound, temporally and spatially highly resolved
72 data on root dynamics is the arduous detection of roots in complex minirhizotron images.

The recent development of Convolutional Neural Networks (CNN) has sparked interest due to their
74 capacity to automatically extract relevant features directly from images without the need for human feature design. CNNs have been shown to outperform traditional algorithms in most computer
76 vision tasks (Zheng, Yang, & Tian, 2018) and as such provide a powerful, inexpensive and time-saving new method for (semi-)automatic analysis of minirhizotron images. Indeed, first attempts
78 under ideal conditions with relatively homogeneous substrate and young roots of single species are promising (Narisetti et al., 2021; Smith, Petersen, Selvan, & Rasmussen, 2020; Wang et al., 2019).
80 Another advantage of using CNNs over manual segmentation is their improved objectivity and repeatability in comparison to human analysts. As long as the conditions and image quality are
82 relatively constant (e.g., illumination, contrast), the accuracy of automatic feature recognition using CNNs is constant as well. In contrast, accuracy of human analysts may vary greatly depending on
84 state of mind (e.g., fatigue, time pressure) and even more so between individuals.

Here, we introduce RootDetector, a new Convolutional Neural Network for classifying roots and
86 extracting metrics for root length in minrhizotron images from field studies. We trained RootDetector with data from a mesocosm experiment and a field experiment that included different
88 organic soils and a variety of plant species. We compared RootDetector's performance with that of human analysts and investigated how differences in experience with plant physiology and digital
90 root measuring tools between groups of human analysts (novice, advanced, expert) affect theaccuracy of manual root segmentation. We furthermore validated RootDetector's ability to

classify root pixels and to quantify root length on randomly selected minirhizotron images from the field.

2. Material and Methods

2.1 Datasets

2.1.1. Image Acquisition

For this study, we assembled three datasets: one for training of the RootDetector Convolutional Neural Network (Training-Set), one for a performance comparison between humans and RootDetector in segmenting roots on minirhizotron images (Comparison-Set), and one for the validation of the algorithm (Validation-Set). The Training-Set contained 129 images comprised of 17 randomly selected minirhizotron images sampled in a mesocosm experiment (see section 2.1.2), 47 randomly selected minirhizotron images sampled in a field study (see section 2.1.3) as well as the 65 minirhizotron images of soy roots published by Wang et al. (2019). The Comparison-Set contained 25 randomly selected minirhizotron images from the field-study which all were not part of the images included in the Training- and Validation-Sets. The Validation-Set contained 10 randomly selected minirhizotron images from the same field study, which had not been used in the Training-Set. All images were recorded with 2550 x 2273 pixels at 300 dpi with a CI-600 In-Situ Root Imager (CID Bio-Science Inc., Camas, WA, USA) and stored as .tiff files to reduce compression loss.

2.1.2. Mesocosm Sampling

The mesocosm experiment was established in 2018 on the premises of the Institute for Botany and Landscape Ecology of the University of Greifswald (Fig. S1). It features 108 heavy duty plastic buckets of 100 l each, filled to two thirds of their height with moderately decomposed sedge fen peat. Each mesocosm contained one minirhizotron (inner diameter: 64 mm, outer diameter: 70 mm, length: 650 mm) installed at a 45°angle and capped in order to avoid penetration by light. The

mesocosms were planted with varying compositions of plant species that typically occur in north-east German sedge fens (*Carex rostrata, Carex acutiformis, Glyceria maxima, Equisetum fluviatile, Juncus inflexus, Mentha aquatica, Acorus calamus* and *Lycopus europaeus*). The mesocosms were subjected to three different water table regimes: stable at soil surface level, stable at 20 cm below soil surface and fluctuating between the two levels every 2 weeks. The minirhizotrons were scanned weekly at two depth levels between April 2019 and December 2021, resulting in roughly 9500 minirhizotron images of 216 cm × 196 mm. Manual quantification of root length would, based on own experience, take approx. 3 hours per image, resulting in approximately 28500 h of manual processing for the complete dataset.

2.1.3. Field Sampling

The field study was established as part of the Wetscapes project in 2017 (Jurasinski et al., 2020). The study sites were located in Mecklenburg-Vorpommern, Germany, in three of the most common wetland types of the region: alder forest, percolation fen and coastal fen (Fig. S2). For each wetland type, a pair of drained versus rewetted study sites was established. A detailed description of the study sites and the experimental setup can be found in Jurasinski et al. (2020). At each site, 15 minirhizotrons (same diameter as above, length: 1500 mm) were installed at 45° angle along a central boardwalk. The minirhizotrons have been scanned biweekly since April 2018, then monthly since January 2019 at two to four depth levels, resulting in roughly 12000 minirhizotron images of 216 cm × 196 cm, i.e. an estimated 36000 h of manual processing for the complete dataset.

2.2 The CNN RootDetector

2.2.1. Image Annotation

For the generation of training data for the CNN, human analysts manually masked all root pixels in the 74 images of the Training-Set using GIMP 2.10.12. The resulting ground truth data are binary, black-and-white images in Portable Network Graphics (.png) format, where white pixels represent

144 root structures and black pixels represent non-root objects and soil (Fig. 2). All training data were checked and, if required, corrected by an expert (see 2.3.1 for definition). The Validation-Set was
146 created in the same way but exclusively by experts.

148 **2.2.2. Architecture**

RootDetector's core consists of a Deep Neural Network (DNN) based on the U-Net image
150 segmentation architecture (Ronneberger, Fischer, & Brox, 2015) and is implemented in TensorFlow and Keras frameworks (Abadi et al., 2016). Although U-Net was originally developed for
152 biomedical applications, it has since been successfully applied to other domains due to its generic design.

154 RootDetector is built up of four down-sampling blocks, four up-sampling blocks and a final output block (Fig. 1). Every block contains two 3x3 convolutional layers, each followed by rectified linear
156 units (ReLU). The last output layer instead utilizes Sigmoid activation. Starting from initial 64 feature channels, this number is doubled in every down-block and the resolution is halved via 2x2
158 max-pooling. Every up-block again doubles the resolution via bilinear interpolation and a 1x1 convolution which halves the number of channels. Importantly, after each up-sampling step, the
160 feature map is concatenated with the corresponding feature map from the down-sampling path. This is crucial to preserve fine spatial details.

162 Our modifications from the original architecture include BatchNormalization (Ioffe & Szegedy, 2015) after each convolutional layer which significantly helps to speed up the training process and
164 zero-padding instead of cropping as suggested by Ronneberger et al. (2015) to preserve the original image size.

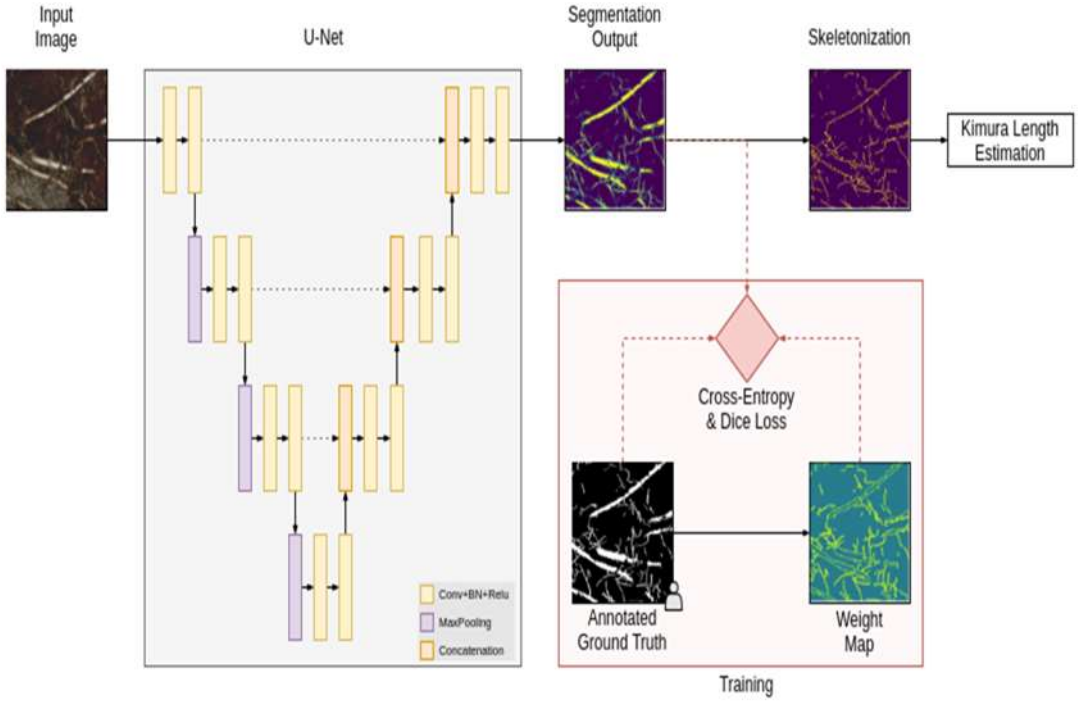


Fig. 1. Overview of the RootDetector system. The main component is a semantic segmentation network based on the U-Net architecture. The root length is estimated by skeletonizing the segmentation output and applying the formula introduced by Kimura et al. (1999). During training only, a weight map puts more emphasis on fine roots.

In addition to the root segmentation network, we trained a second network to detect foreign objects, specifically the adhesive tape that is used as a light barrier on the aboveground part of the minirhizotrons. We used the same network architecture as above and trained in a supervised fashion with the binary cross-entropy loss. During inference, the result is thresholded and used without post-processing.

2.2.3. Training

We pre-trained RootDetector on the COCO dataset (Lin et al., 2014) to generate a starting point. Although the COCO dataset contains a wide variety of image types and classes not specifically related to minirhizotron images, Majurski et al. (2019) showed, that for small annotation counts,

transfer-learning even from unrelated datasets may improve a CNNs performance by up to 20%. We fine-tuned for our dataset with the Adam optimizer (Kingma & Ba, 2015) for 15 epochs and trained on a total of 129 images from the Training-Set (17 mesocosm images, 47 field-experiment images, 65 soy root images). To enhance the dataset size and reduce over-fitting effects, we performed a series of augmentation operations as described by Shorten and Khoshgoftaar (2019). In many images, coarser roots occupied a major part of the positive (white) pixel space, which might have caused RootDetector to underestimate fine root details overall. Similarly, negative space (black pixels) between tightly packed, parallel roots was often very small and might have impacted the training process to a lesser extent when compared to large areas with few or no roots (Fig 2). To mitigate both effects, we multiplied the result of the cross-entropy loss map with a weight map which emphasizes positive-negative transitions. This weight map is generated by applying the following formula to the annotated ground truth images:

Eq. 1 $$w(x) = 1 - \left(\tanh\,(2\tilde{x} - 1)\right)^2$$

where ω(x) is the average pixel value of the annotated weight map in a 5x5 neighborhood around pixel x.

For the loss function we applied a combination of cross-entropy and Dice loss (Milletari, Navab, & Ahmadi, 2016):

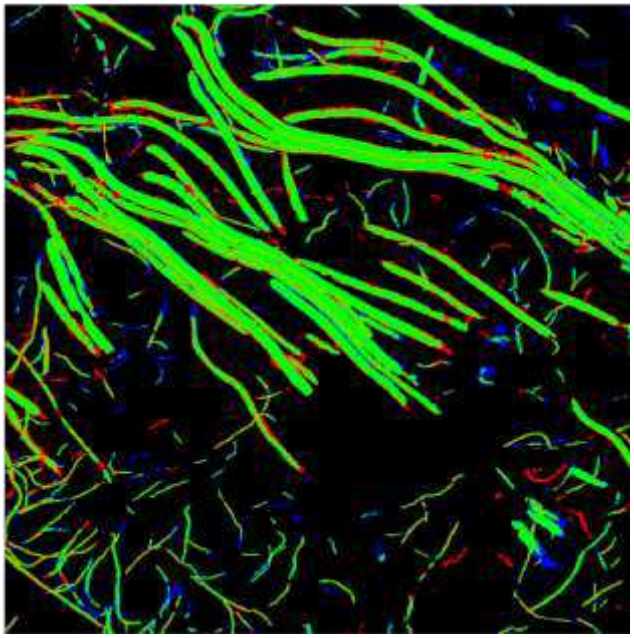
Eq. 2 $$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{Dice} = -\frac{1}{N}\Sigma_i w(x_i) y_i \log(x_i) + \lambda \frac{2\Sigma_i x_i y_i}{\Sigma_i x_i^2\ \Sigma_i y_i^2}$$

where x are the predicted pixels, y the corresponding ground truth labels, N the number of pixels in an image and λ a balancing factor which we set to 0.01. The Dice loss is applied per-image to counteract the usually high positive-to-negative pixel imbalance. Since this may produce overly confident outputs and restrict the application of weight maps, we used a relatively low value for λ.

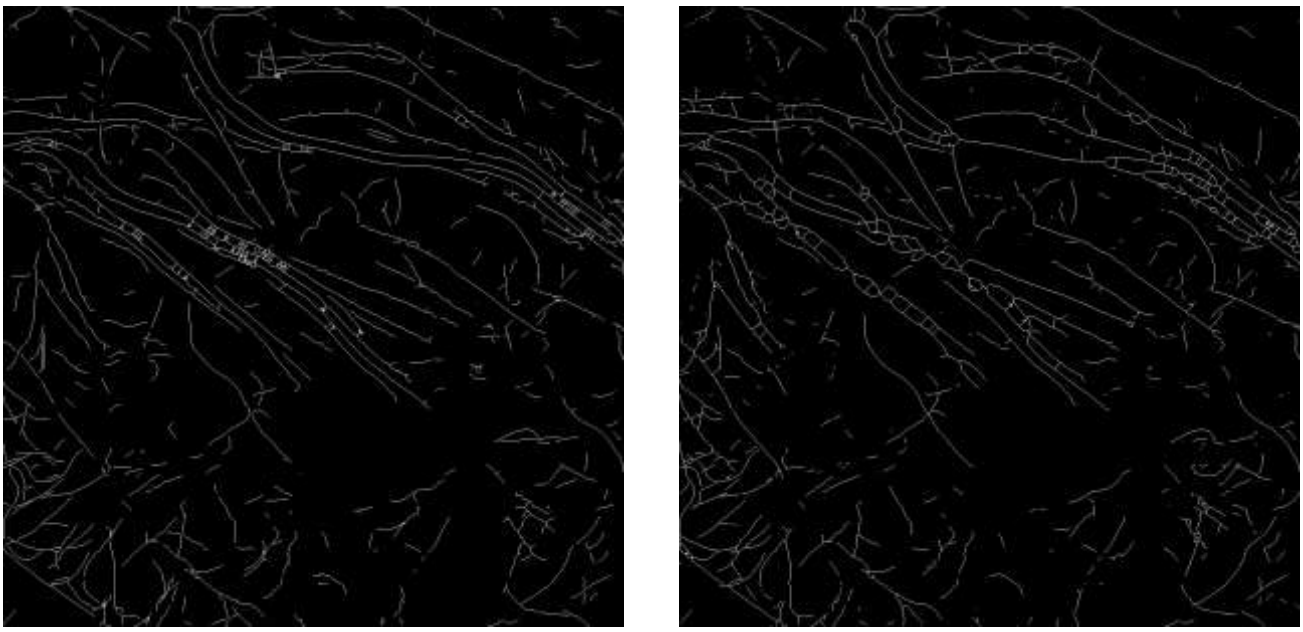
a)

b)

c)



d)



e) f)

Fig. 2: Example of segmentation and result of skeletonization. A 1000 by 1000 pixel input image (a), the manually annotated ground truth image (b), the RootDetector estimation image (c), the combined representation image (error map, d with green indicating true positives, red indicating false positive, blue indicating false negatives), the skeletonized RootDetector estimation image (e), and the skeletonized ground truth image (f).

2.2.4. Output and Post-processing

RootDetector generates two types of output. The first type of output are greyscale .png files in which white pixels represent pixels associated with root structures and black pixels represent non-root structures and soil (Fig. 2c). The advantage of .png images is their lossless ad artifact-free compression at relatively small file sizes. RootDetector further skeletonizes the output images and reduces root-structures to single-pixel representations using the skeletonize function of scikit-image v. 0.17.1 (Zhang & Suen, 1984; Fig. 2e&f). This helps to reduce the impact of large diameter roots or root-like structures such as rhizomes in subsequent analyses and is directly comparable to estimations of root length. The second type of output is a Comma-separated values (.csv) file, with numerical values indicating the number of identified root pixels, the number of root pixels after

skeletonization, the number of orthogonal and diagonal connections between pixels after skeletonization and an estimation of the physical combined length of all roots for each processed image. The latter is a metric commonly used in root research as in many species, fine roots provide most vital functions such as nutrient and water transport (McCormack et al., 2015). Therefore, the combined length of all roots in a given space puts an emphasis on fine roots as they typically occupy a relatively smaller fraction of the area in a 2D image compared to often much thicker coarse roots. To derive physical length estimates from skeletonized images, RootDetector counts orthogonal- and diagonal connections between pixels of skeletonized images and employs the formula proposed by Kimura et al. (1999) (Eq. 3).

Eq. 3 $$L = \sqrt{[{N_d}^2 + (N_d + \frac{N_o}{2})^2]} + \frac{N_o}{2}$$

where N_d is the number of diagonally connected and N_o the number of orthogonally connected skeleton pixels. To compute N_d we convolve the skeletonized image with two 2x2 binary kernels, one for top-left-to-bottom-right connections and another for bottom-left-to-top-right connections and count the number of pixels with maximum response in the convolution result. Similarly, N_o is computed with a 1x2 and a 2x1 convolutional kernels.

2.3 Performance Comparison

2.3.1 Selection of participants

For the performance comparison, we selected 10 human analysts and divided them into three groups of different expertise levels in plant physiology and with the usage of digital root measuring tools. The novice group consisted of 3 ecology students (2 bachelor's, 1 master's) who had taken or were taking courses in plant physiology but had no prior experience with minirhizotron images or digital root measuring tools. This group represents undergraduate students producing data for a Bachelor

thesis or student assistants employed to process data. The advanced group consisted of 3 ecology students (1 bachelor's, 2 master's) who had already taken courses in plant physiology and had at least 100 hours of experience with minirhizotron images and digital root measuring tools. The expert group consisted of 4 scientists (2 PhD, 2 PhD candidates) who had extensive experience in root science and at least 250 hours of experience with digital root measuring tools.

2.3.2. Instruction and Root Tracing

All three groups were instructed by showing them a 60 minute live demo of an expert tracing roots in minirhizotron images, during which commonly encountered challenges and pitfalls were thoroughly discussed. Additionally, all participants were provided with a previously generated, in-depth manual containing guidelines on the identification of root structures, the correct operation of the root tracing program and examples of often encountered challenges and suggested solutions. Before working on the Comparison-Set, all participants traced roots in one smaller-size sample image and received feedback from one expert.

2.3.3. Image Preparation and Root Tracing

Because the minirhizotron images acquired in the field covered a variety of different substrates, roots of different plant species, variance in image quality, and because tracing roots is very time consuming, we decided to maximize the number of images by tracing roots only in small sections, in order to cover the largest number of cases possible. To do this, we placed a box of 1000x1000 pixels at a random location in each of the images in the Comparison-Set and instructed participants to trace only roots within that box. Similarly, we provided RootDetector images where the parts of the image outside the rectangle were occluded. All groups used RootSnap! 1.3.2.25 (CID Bio-Science Inc., Camas, WA, USA) to manually trace roots in each of the 25 images in the comparison set. The combined length of all roots was then exported as a .csv file for each person and image and compared to RootDetector’s output of the Kimura root length.

2.4. Validation

We tested the accuracy of RootDetector on a set of 10 image segments of 1000 by 1000 pixels cropped from random locations of the 10 images of the Validation-Set. These images were annotated by a human expert without knowledge of the estimations by the algorithm and were exempted from the training process. As commonly applied in binary classification, we use the F1 score as a metric to evaluate the performance RootDetector. F1 is calculated from precision (Eq. 4) and recall (Eq. 5) and represents their harmonic mean (Eq. 6). Ranging from 0 to 1, higher values indicate high classification (segmentation) performance. As one of the 10 image sections contained no roots and thus no F1 Score was calculable, it was excluded from the validation. We calculated the F1 score for each of the nine remaining image sections and averaged the values as a metric for overall segmentation performance.

Eq. 4 $$Precision = \frac{truepositive}{truepositive + falsepositive}$$

Eq. 5 $$Recall = \frac{truepositive}{truepositive + falsenegative}$$

Eq. 6 $$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

2.5. Statistical Analysis

We used R Version 4.1.2 (R Core Team, 2021) for all statistical analyses and R package ggplot2 Version 3.2.1 (Wickham, 2016) for visualizations. Correlation analyses were based on least-squares fit and the Pearson method. Performance comparisons between groups of human analysts (novice, advanced, expert) and RootDetector were based on the respective estimates of total root length plotted over the minirhizotron images in increasing order of total root length. Linear models were

296 calculated using the lm function for each group of analysts. To determine significant differences between the groups and the algorithm, 95% CIs as well as 83% CIs were displayed and
298 RootDetector root length outside the 95% CI were considered significantly different from the group estimate at $\alpha = 0.05$ (Gelman & Hill, 2006). The groups of human analysts were considered
300 significantly different if their 83% CIs did not overlap, as the comparison of two 83% CIs approximates an alpha level of 5% (Austin & Hux, 2002; Payton, Greenstone, & Schenker, 2003).

302

3 Results

304 3.1 Performance Comparison

Human analysts differed strongly in total root length annotated per minirhizotron image section (Fig.
306 3). Novices generally estimated highest root length, while experts found the lowest total root length. Novices annotated on average 279 % of root pixels as compared to experts. Variation in total root
308 length estimation, as expressed by the 95% CI in Fig. 3, was highest between the three novices and lowest between the four experts. Root length quantification by RootDetector was indifferent from
310 the expert group, but lower than the novice and advanced groups.

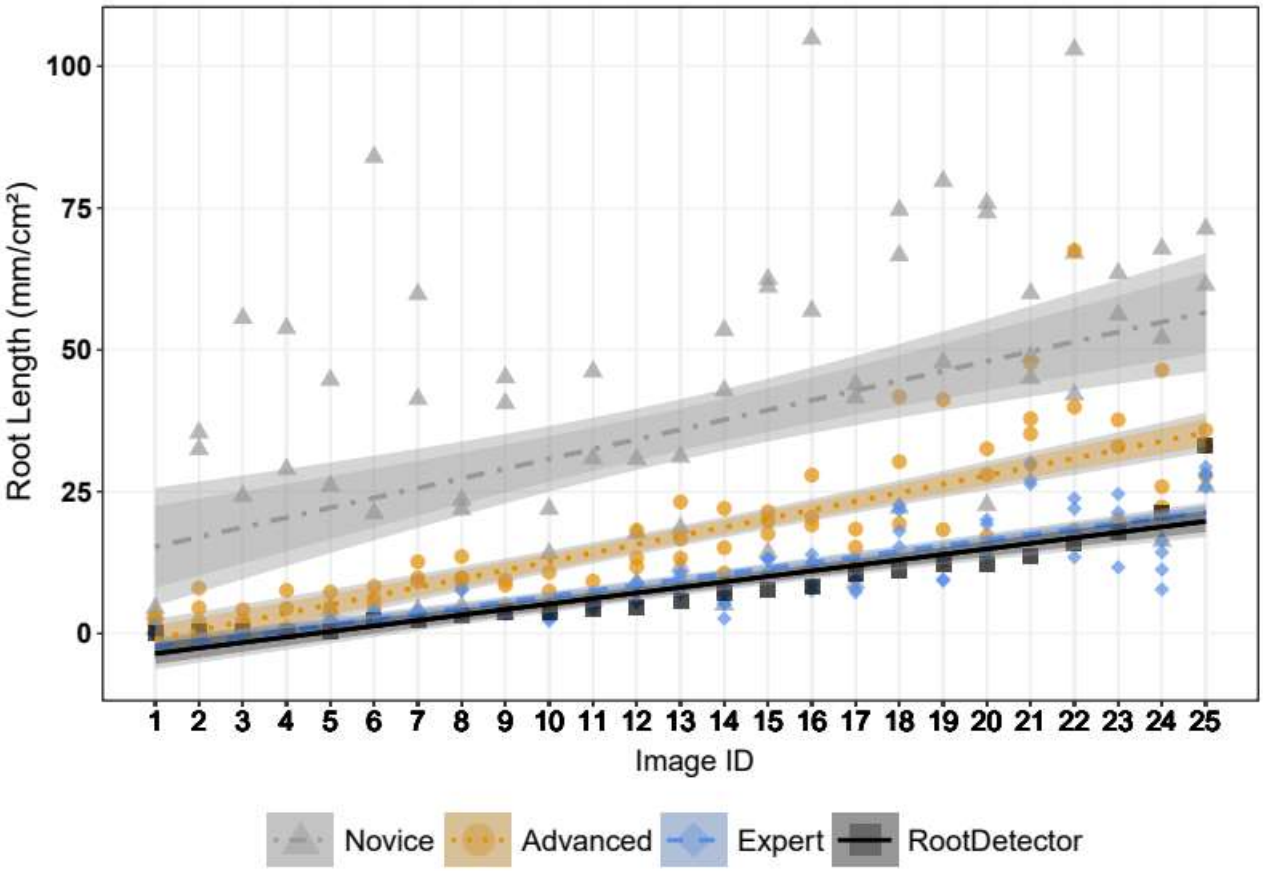


Fig. 3: The CNN RootDetector performs as good as human experts (analysts with scientific background in root ecology and >250 h experience in annotating roots), as indicated by the linear regression of RootDetector lying within the 95% CI (darker shaded ribbons) of the expert group. Groups of human analysts differ in their estimation of root lengths, as indicated by non-overlapping 83% CIs (lighter shaded ribbons). Novices had theoretical input on root ecology but no experience with minirhizotron images, advanced analysts had theoretical input on root ecology and >100 h of experience with annotating minirhizotron images. Images are sorted by increasing root length according to RootDetector along the x-axis.

3.2 Overall Performance of RootDetector

RootDetector showed a high capability to correctly segment roots in the Validation-Set of minirhizotron images not used during the training phase. Total root pixels in the images were detected with a F1 score of 0.6044. The correlation between total number of root pixels detected by

RootDetector and, independently, by one human expert was very high ($r^2 = 0.99$) and with uniform residuals across the data range (Fig. 4a). The correlation was also high after skeletonization ($r^2 = 0.96$; Fig. 4b).

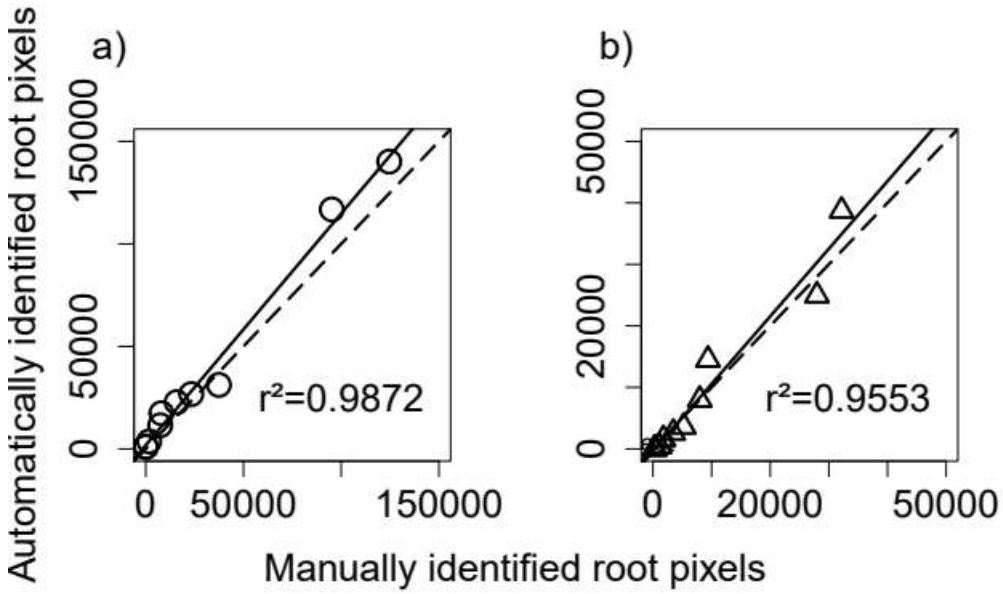


Fig. 4: Correlation between (a) total root pixels and (b) total root pixels after skeletonization as detected by the CNN RootDetector and expert human analysis based on ten 1000 x 1000 pixel image segments cropped from randomly selected minirhizotron images from the field study (Validation-Set). Dotted lines represent the 1:1 line, solid lines the least-squares correlation.

3.3 Discussion

RootDetector provides the general advantage of perfect reproducibility and objectivity, two points that are questionable at best for human analysts (Freschet et al., 2021). Our study clearly showed that there was large variation in root length estimates with novices annotating almost 3-times higher root length/cm² per image compared to expert analysts (Fig. 3). Interestingly, this variation shrank for more experienced analysts, i.e., the more time analysts have already spent on analysing root images, the more similar their estimates become. No matter the level of experience though, the annotation of minirhizotron images clearly is not objective if done by humans, which hinders

comparisons between studies or even between years within long-term studies. For the training of RootDetector, we invested roughly 1300 (±200) h annotating training images. Compared to an estimated duration of >60000 h for manual annotation of the roots in the two experimental setups used here (see methods section), this is an enormous step forward in efficiency.

Up to now, the manual annotation of roots in the minirhizotron images has been the bottleneck for studying root dynamics in high spatial and temporal resolution. Automated minirhizotrons for field studies exist, but manual analysis of the resulting images have so far prevented tapping their full potential concerning temporal resolution and replication (Allen & Kitajima, 2013; Defrenne et al., 2021). Once trained to the given ecosystem (soil type, root morphology, etc.), the algorithm solves this limitation. While we assume that additional training is needed for high-quality analysis of images from other ecosystems, this re-training of the algorithm to other conditions should require less training data than the initial training (Majurski et al., 2019). According to our experience, we estimate that training the algorithm for data from a new experiment will roughly require 25-60 training images of 2550x2273 pixels, which may take approximately 150-300 hours of manual annotation. This would mean that training and using the algorithm becomes less work than analyzing images by hand already after 70 images – which would be reached for 6 minirhizotron tubes with 3 image levels after only one month of weekly sampling. As minirhizotron tubes are often the least expensive part of a respective experimental set-up and as roots are highly variable in space, a high number of replicate tubes is clearly desirable. Similarly, it is known that, especially in highly productive ecosystems, fine root lifespan can be a few days or weeks only (Eissenstat & Yanai, 1997) which illustrates the need for a high temporal resolution in addition to the spatial one. This goal can only be achieved if the resulting images are annotated automatically.

The RootDetector CNN reached a F1 score of 0.6044 for our field study of various wetlands. This is lower than previously published algorithms for root segmentation in rhizotron images achieved (Wang et al. (2019): F1 = 0.6479; Smith et al. (2020): F1 = 0.7; Narisetti et al. (2021): F1 = 0.87). However, those studies were conducted under highly controlled conditions, often with single plant

species and homogeneous mineral soils resulting in much more uniform soil and root appearance and therefore higher quality images than what can commonly be achieved under field conditions. Thus, to increase the understanding of root dynamics and their influence on ecosystem processes, we aimed to develop a tool that gives consistently accurate measures of root pixels and root length on minirhizotron images from natural plant communities, even when those are growing in organic soils consisting of plant material in varying degrees of decomposition. The poor performance of traditional automatic image analysis tools has left field ecologists with little choice but to continue analysing minirhizotron images by hand, limiting the amount of data that can be processed and ultimately our understanding of root ecology. Here, we show a very high correlation between automated annotation by RootDetector and traditional annotation by human experts. There was also no sign of changing variance in the residuals across a wide range of root lengths (Fig. 3 & 4), which further supports the conclusion that this algorithm provides a promising solution for the annotation of roots in minirhizotron images in ecological studies.

Conclusion

CNNs such as RootDetector provide a reliable and efficient method for the detection of roots in minirhizotron images. In comparison with human analysts, whose ability to detect roots varies widely, RootDetector saves resources, is objective and reproducible, and performs as well as human experts. RootDetector furthermore provides not only root pixel estimations but also root length data, which is the most commonly used metric in root ecological research and is not regularly delivered by existing CNNs. RootDetector is supplied as readily usable code on GitHub, enabling easy use by ecologists without the need of advanced programming skills. Transfer to other ecosystems or technical setups of the minirhizotrons will require re-training of the algorithm, but this is an initial and one-time investment which likely will pay off already after around 70 images, i.e. during common study length. In particular coupled with automated minirhizotrons, this tool for automatic analysis of minirhizotron images will allow for unprecedented detail and comprehensiveness in

studies of root dynamics, thereby answering globally important ecological and biogeochemical questions.

Acknowledgements:

We thank Hannah Marx, Kinley, and Robert Mahara on behalf of numerous student workers for annotating minirhizotron images. This joint research project "*DIG-IT!*" is supported by the European Social Fund (ESF), reference: ESF/14-BM-A55-0013/19 and ESF/14-BM-A55-0015/19, and the Ministry of Education, Science and Culture of Mecklenburg-Vorpommern, Germany. Field data stems from the project "Wetscapes" supported by the European Social Fund (ESF) and the Ministry of Education, Science and Culture of Mecklenburg-Western Pomerania [ESF/14-BM-A55-0035/16].

Author Contributions

BP, GBW and JK conceived the ideas and designed methodology; BP, GBW, SS and JK collected the data; AG and UFvL designed, trained and validated the CNN; BP analysed the data; BP and JK led the writing of the manuscript, AG wrote the sections describing the CNN. All authors contributed critically to the drafts and gave final approval for publication.

Competing interests

The authors declare no competing interests.

Data Accesibility

GitHub repository accessible under https://github.com/ExPlEcoGreifswald/Root-Detecto

Permissions and Licenses

Permissions for all field work were obtained in advance.

All experiments were conducted in compliance with relevant institutional, national, and international guidelines on specimen collection.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI*, 265–283. https://doi.org/10.5555/3026877.3026899

Allen, M. F., & Kitajima, K. (2013). In situ high-frequency observations of mycorrhizas. *New Phytologist*, *200*(1), 222–228. https://doi.org/10.1111/nph.12363

Austin, P. C., & Hux, J. E. (2002). A brief note on overlapping confidence intervals. *Journal of Vascular Surgery*, *36*(1), 194–195. https://doi.org/10.1067/mva.2002.125015

Blume-Werry, G. (2021). The belowground growing season. *Nature Climate Change*, *25*, 1922. https://doi.org/10.1038/s41558-021-01243-y

Blume-Werry, G., Wilson, S. D., Kreyling, J., & Milbau, A. (2016). The hidden season: Growing season is 50% longer below than above ground along an arctic elevation gradient. *New Phytologist*, *209*(3), 978–986. https://doi.org/10.1111/nph.13655

Clemmensen, K. E., Bahr, A., Ovaskainen, O., Dahlberg, A., Ekblad, A., Wallander, H., . . . Lindahl, B. D. (2013). Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science*, *339*(6127), 1615–1618. https://doi.org/10.1126/science.1231923

Defrenne, C. E., Childs, J., Fernandez, C. W., Taggart, M., Nettles, W. R., Allen, M. F., . . . Iversen, C. M. (2021). High-resolution minirhizotrons advance our understanding of root-fungal dynamics in an experimentally warmed peatland. *New Phytologist*, *3*(5), 640–652. https://doi.org/10.1002/ppp3.10172

Eissenstat, D. M. & Yanai, R. D. (1997). The Ecology of Root Lifespan. Advances in Ecological Research, 27, 1–60. https://doi.org/10.1016/S0065-2504(08)60005-7

Freschet, G. T., Pagès, L., Iversen, C. M., Comas, L. H., Rewald, B., Roumet, C., . . . McCormack, M. L. (2021). A starting guide to root ecology: Strengthening ecological concepts and standardising root classification, sampling, processing and trait measurements. *The New Phytologist*, *232*(3), 973–1122. https://doi.org/10.1111/nph.17572

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Hansson, A. C., Steen, E., & Andren, O. (1992). Root-growth of daily irrigated and fertilized barley investigation with ingrowth cores, soil cores and minirhizotrons. *Swedish Journal of Agricultural Research*, *22*(4), 141–152.

Ioffe, S., & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, 448–456

Jurasinski, G., Ahmad, S., Anadon-Rosell, A., Berendt, J., Beyer, F., Bill, R., . . . Wrage-Mönnig, N. (2020). *From Understanding to Sustainable Use of Peatlands: The WETSCAPES Approach. Soil Syst.* **2020**, *4*, 14. https://doi.org/10.3390/soilsystems4010014

Kimura, K., Kikuchi, S., & Yamasaki, S.-i. (1999). Accurate root length measurement by image analysis. *Plant and Soil*, *216*(1/2), 117–127. https://doi.org/10.1023/A:1004778925316

Kingma, D. P., & Ba, J. (2015). *Adam: A Method for Stochastic Optimization*. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. 2015 Retrieved from http://arxiv.org/pdf/1412.6980v9

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., . . . Dollár, P. (2014). *Microsoft COCO: Common Objects in Context*. Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, H., Wang, H., Li, N., Shao, J., Zhou, X., van Groenigen, K. J., & Thakur, M. P. (2021). Phenological mismatches between above- and belowground plant responses to climate warming. *Nature Climate Change*, *1*, 14. https://doi.org/10.1038/s41558-021-01244-x

Lynch, J. M., & Whipps, J. M. (1990). Substrate flow in the rhizosphere. *Plant and Soil*, *129*(1), 1–10. https://doi.org/10.1007/BF00011685

Majurski, M., Manescu, P., Padi, S., Schaub, N., Hotaling, N., Simon, C., & Bajcsy, P. (2019, June 16–17). Cell image segmentation using generative adversarial networks, transfer learning, and augmentations. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1114–1122). IEEE. https://doi.org/10.1109/CVPRW.2019.00145

Matamala, R., Gonzàlez-Meler, M. A., Jastrow, J. D., Norby, R. J., & Schlesinger, W. H. (2003). Impacts of fine root turnover on forest NPP and soil C sequestration potential. *Science*, *302*(5649), 1385–1387. https://doi.org/10.1126/science.1089543

McCormack, M. L., Crisfield, E., Raczka, B., Schnekenburger, F., Eissenstat, D. M., & Smithwick, E. A. H. (2015). Sensitivity of four ecological models to adjustments in fine root turnover rate. *Ecological Modelling*, *297*, 107–117. https://doi.org/10.1016/j.ecolmodel.2014.11.013

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016, October 25–28). V-Net: Fully Convolutional Neural Networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565–571). IEEE. https://doi.org/10.1109/3DV.2016.79

Mokany, K., Raison, R. J., & Prokushkin, A. S. (2006). Critical analysis of root:shoot ratios in terrestrial biomes. *Global Change Biology*, *12*(1), 84–96. https://doi.org/10.1111/j.1365-2486.2005.001043.x

Narisetti, N., Henke, M., Seiler, C., Junker, A., Ostermann, J., Altmann, T., & Gladilin, E. (2021). Fully-automated root image analysis (faRIA). *Scientific Reports*, *11*(1), 16047. https://doi.org/10.1038/s41598-021-95480-y

Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, *3*, 34. https://doi.org/10.1093/jis/3.1.34

Robinson, D. (2004). Scaling the depths: Below-ground allocation in plants, forests and biomes. *Functional Ecology*, *18*(2), 290–295. https://doi.org/10.1111/j.0269-8463.2004.00849.x

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Lecture Notes in Computer Science. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Schwieger, S., Blume-Werry, G., Peters, B., Smiljanić, M., & Kreyling, J. (2019). Patterns and drivers in spring and autumn phenology differ above- and belowground in four ecosystems under the same macroclimatic conditions. *Plant and Soil*, *205*, DOI: 10.1007/s11104-019-04300-w. https://doi.org/10.1007/s11104-019-04300-w

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 1106. https://doi.org/10.1186/s40537-019-0197-0

Smith, A. G., Petersen, J., Selvan, R., & Rasmussen, C. R. (2020). Segmentation of roots in soil with U-Net. *Plant Methods*, *16*, 13. https://doi.org/10.1186/s13007-020-0563-0

Sokol, N. W., & Bradford, M. A. (2019). Microbial formation of stable soil carbon is more efficient from belowground than aboveground input. *Nature Geoscience*, *12*, 46–53. https://doi.org/10.1038/s41561-018-0258-6

Träger, S., Wilson, S. D., & Kudo, G. (2017). Root heterogeneity along an arctic elevational gradient: The importance of resolution. *Functional Ecology*, *31*(2), 480–487. https://doi.org/10.1111/1365-2435.12721

Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A. S., . . . Jin, J. (2019). SegRoot: A high throughput segmentation method for root image analysis. *Computers and Electronics in Agriculture*, *162*, 845–854. https://doi.org/10.1016/j.compag.2019.05.017

Warren, J. M., Hanson, P. J., Iversen, C. M., Kumar, J., Walker, A. P., & Wullschleger, S. D. (2015). Root structural and functional dynamics in terrestrial biosphere models – evaluation and recommendations. *New Phytologist*, *205*(1), 59–78. https://doi.org/10.1111/nph.13034

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org.

528 Zhang, T.Y., & Suen, C.Y. (1984). A fast parallel algorithm for thinning digital patterns. Commun. ACM, 27(3):236239, Mar. 1984.
530 Zheng, L., Yang, Y., & Tian, Q. (2018). Sift meets CNN: A decade survey of instance retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(5), 1224–1244.
532 https://doi.org/10.1109/TPAMI.2017.2709749

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PetersetalRootDetectormethodsSciRepSupplements.odt