Germination Detection of Seedlings in Soil: A System, Dataset and Challenge

Hanno Scharr[1] (✉), Benjamin Bruns[1], Andreas Fischbach[1],

Johanna Roussel[1,2], Lukas Scholtes[1], and Jonas vom Stein[1]

[1]Institute of Bio- and Geosciences (IBG), IBG-2: Plant Sciences
Forschungszentrum Jülich, Germany
`h.scharr@fz-juelich.de`
[2] Medical Engineering and Technomathematics
University of Applied Sciences Aachen, Jülich, Germany

Abstract. In phenotyping experiments plants are often germinated in high numbers, and in a manual transplantation step selected and moved to single pots. Selection is based on visually derived germination date, visual size, or health inspection. Such values are often inaccurate, as evaluating thousands of tiny seedlings is tiring. We address these issues by quantifying germination detection with an automated, imaging-based device, and by a visual support system for inspection and transplantation. While this is a great help and reduces the need for visual inspection, accuracy of seedling detection is not yet sufficient to allow skipping the inspection step. We therefore present a new dataset and challenge containing 19.5k images taken by our germination detection system and manually verified labels. We describe in detail the involved automated system and handling setup. As baseline we report the performances of the currently applied color-segmentation based algorithm and of five transfer-learned deep neural networks.

Keywords: Transplantation guidance, high-throughput plant phenotyping, automation, Arabidopsis, two leaf stadium

1   Introduction

We present a new dataset and challenge containing 19.5k images taken by an automated germination detection system and manually verified labels.

A plant's phenotype is constituted by its traits as reaction to diverse environmental conditions. Plant phenotyping has been identified to be key for progress in plant breeding and basic plant science [26]. To increase throughput of plant experiments and overcome the phenotyping bottleneck [10] many automated technologies have been and are developed. Such systems are often image-based and image analysis builds a performance bottleneck [17].

When running larger-scale experiments with hundreds of plants in soil, due to space limitations, it is reasonable to sow seeds in smaller pots until germination
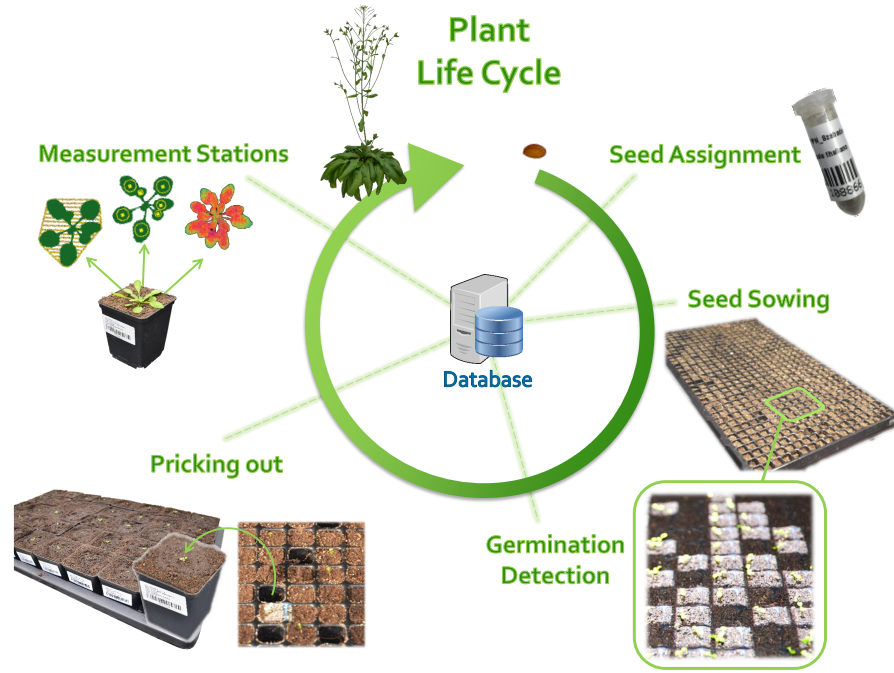
Fig. 1. Life Cycle of a Plant. In a typical plant phenotyping experiment seeds are selected and assigned to the experiment, they are sown, plants germinated, seedlings pricked out and transplanted to bigger pots, and then measured at different stations with a selection of measurement systems. Some plants are then brought to flowering to produce seeds, closing the circle.

and transplant them later into larger pots (see Figure 1). We seed into special multi-well-plates with 576 wells of size $15\,mm \times 15\,mm$. They are too small for the final experiment, but allow for undisturbed germination in soil and, importantly, undisturbed transplantation of a seedling to its final typically $7\,cm \times 7\,cm$ pot. The number of seeds to sow strongly depends on germination rates which can differ from genotype to genotype. Often the three- to fourfold amount of required seedlings are sown to ensure enough plants germinate within a given time-frame needed for synchronisation of treatments in an experiment. For typical medium to high-throughput experiments several thousand seeds need to be sown.

For larger seedlings, like e.g. Tobacco plants, visual germination detection and manual selection of seedlings for subsequent experiments is routinely and reliably done. For tiny seedlings like Arabidopsis, with a typical diameter of 1 mm to 2 mm, it is tiring and frequent errors are unavoidable. However, this inspection is crucial for subsequent experiments, as seedlings are selected according to the germination date, size or health and thus it should to be automated to increase reliability and throughput [14,19,7,20]. Further, without an automated system,

size and health are just available by visual inspection and after transplantation no mapping between seedlings' and plants' identities is available. However, when data-handling is intended to cover the whole plant life cycle, plants' identities need to be tracked through this process.

We designed and built two systems, allowing for image-based automated germination detection, support of visual seedling inspection and guided transplantation of seedlings. Imaging for germination detection, or more exactly, detection of the two leaf stadium, is done by a variant of GrowScreen [27,13]. Further, we designed a special workplace, a *Handling Station*, equipped with a camera and video projector for visual seedling inspection and guided transplantation. The systems are coupled via a database (see Section 2.1), storing e.g. layout of multi-well-plates or trays, seed properties and identification number (seed ID) for each well, germination date and seedling size when detected etc. Germination can be verified and transplantation automatically documented in the database using the Handling Station. Both systems are described in detail below (see Section 2.2 and 2.3).

In the workflow individual seeds / seedlings are identified by their location on the soil-filled tray. The germination detection system automatically measures relevant seedling data and stores it in a database. The system allows rule-based selection of seedlings for pricking out and subsequent experiments based on their measured traits. For visual inspection all selected seedlings on a tray are highlighted by the video projector. After validation, the user is guided through transplantation by highlighting only the next single seedling to process. The system generates an adhesive label indicating among other things the ID and (randomized) target position of the newly potted plant. By this a one-to-one mapping between seedling and newly potted plant is generated in the database. Using this setup for germination detection and pricking out reduced time and labor, as well as increased reliability of seedling-to-plant assignment. In addition, the tedious work of selecting the right plant among hundreds of others is now done by the system, allowing for faster and less tiring working.

While the two systems significantly increased manual throughput and enabled reliable seed to plant tracking, it not yet allows for fully automatic seedling selection. Especially reliability of two-leaf stage detection is still too low. We aim at well above 99.8% reliability (i.e. human rater performance), but as we are not yet there, visual inspection of germination detection results is still needed. Therefore, image data of the germination detection system together with binary labels 'positive', i.e. germinated seedling visible versus 'negative', i.e. no germinated seedling visible, are provided together with this paper [21]. As a baseline for further algorithm development on the challenge of predicting binary labels from images, we provide the performance of a simple color-segmentation based algorithm (see Section 2.6), as well as five transfer learned deep neural networks (DNNs). The dataset is described in detail in Section 2.5 and the experiments for the baseline in Section 3.

2   Materials and Methods

2.1   Database

Automated plant phenotyping commonly is performed in an environment of specialized, rarely interconnected systems with highly diverse datasets and custom analysis tools [22,8,13,18,12]. In lab or greenhouse situations, where potted plants are monitored over a longer time span, in different situations, and with different measurement systems, a wealth of often weakly structured data is collected. Keeping track of the individual plants in an experiment can then be cumbersome. Database systems handling individual plants data, e.g. by using unique IDs for each plant, allow to join such data in a user-friendly and reliable fashion [22,3].

We use the information system PhenOMIS [22,3], a distributed information system for phenotyping experiments, and integrated the systems presented here. PhenOMIS allows to track plant histories by acquiring (and guiding) manual treatments during activity, covering plants' whole life cycle (cmp. Figure 1). The information system allows central access to distributed, heterogeneous phenotyping data and integration of spatial, temporal and sensor data. To this end it loosely couples different components, not integrating all data in one database, but rather enabling and supporting data co-existence. Web services for encapsulation allow to further extend functionality.

Here, seeds are sown into multi-pot trays (Figure 2B) using a robotic system [12] assigning each seed a unique ID. Seed ID and pot position in the tray are stored in the database for identification. Subsequent measurements from germination detection are then assigned to this ID.

2.2   Automated Germination Detection System

Imaging. A variant of the *Growscreen Fluoro* [13] system is used for automated germination detection (Figure 2A), fit into a 19" rack. It consists of an x-y-z-moving stage (blue components in Figure 2A), where we mounted a 5 MP RGB color camera (red in Figure 2A) with a 25 mm optical lens, instead of the fluorescence camera. This offers a spatial resolution of 29.0 pixel/mm. A white LED ring is used for illumination.

Different plant species need different pot sizes for germination. The current tray is identified by a bar code, the tray's layouts read from the database, allowing the system to calculate camera positions for suitable image tiling. The camera is automatically moved to the calculated positions by x-y-moving stages. Optimal working distance and thus focus is ensured by the z-moving stage according to the known pot height. Scanning a tray takes several seconds to a few minutes, depending on tray layout. Some sample images are shown in Figure 2C. They are automatically cropped into images showing one pot each from the known tray layout. RGB images in the dataset (see Section 2.5) are collected from this step. Figure 2D top shows a time series of images for one pot.
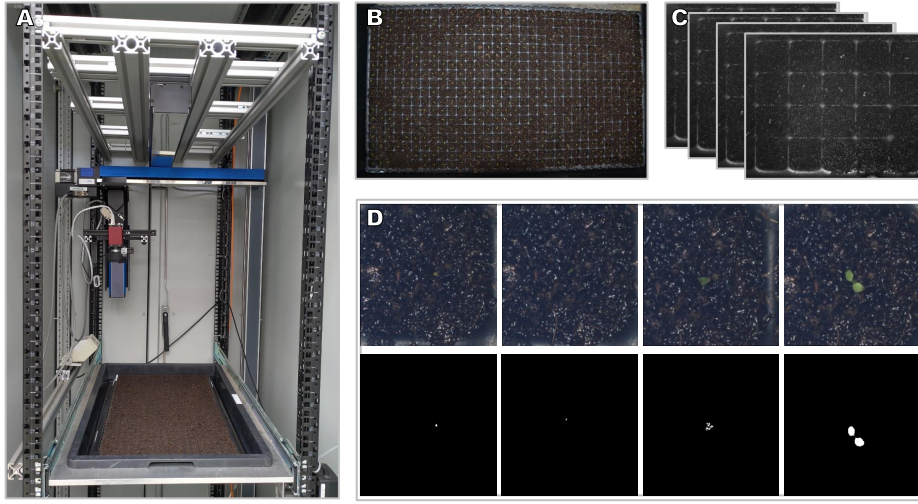
Fig. 2. Germination detection. A: Imaging setup. B: A soil-filled tray used for germination, together with germinated seedlings. C: Acquired images. D: Temporal image sequence of one pot, showing the first time point when green tissue is visible (left) and two leaf stage (right).

Processing. In the system, single pot images are segmented into a binary foreground-background mask using thresholds in HSV-space [27]. Thresholds are manually optimized once per plant species and stored in the database for this species, in order to adapt to seedling color and illumination conditions. Figure 2D bottom shows some segmentation results.

Two points in time need to be derived from the images:

- germination date, defined by the first occurrence of green pixels in the image. To avoid artefacts a size threshold of 10 pixels $(0.0119\,\mathrm{mm}^2)$ is used for Arabidopsis.
- emergence of separate cotyledons, i.e. two leaf stage. This is where the data challenge is aiming at.

In the current system, the two leaf stage is detected using the algorithm described in Section 2.6. In our plant experiments, we require seedlings to reach this stage to be suitable for transplantation and subsequent experiments. In addition, for *Arabidopsis thaliana* the seedlings are required to have a minimum leaf diameter of four pixels $(0.138\,\mathrm{mm})$ and a projected leaf area larger than $100\,\mathrm{pixels}$ $(0.119\,\mathrm{mm}^2)$.

2.3   A Handling Station for Pricking out and Randomization

For visual inspection selected seedlings are highlighted using a video projector. The same method guides the user through the transplantation process by highlighting the next single seedling to pique.
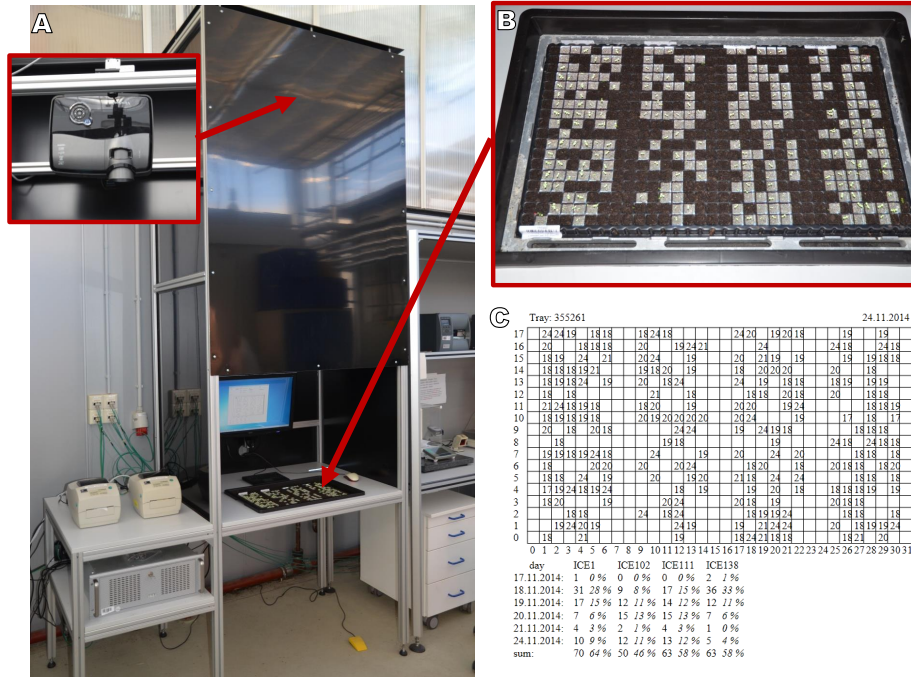
Fig. 3. Visualization of germination results. A: The setup for visual inspection and transplantation guidance. Inset: Video projector for highlighting. B : Germinated seedlings highlighted in their pots. C: print out showing the germination dates and rates for the same tray.

Materials. The handling station consists of a video projector (ViewSonic Pro 8600) and an RGB camera (Point Grey Grashopper GS2-GE-20S4C-C) mounted above a table and aiming towards the table workspace (see Figure 3A). A cutout in the table allows to keep standard trays in a fixed position (Figure 3B) and enables single pot access from below for the pricking out process (Figure 4D).

The camera is used to calibrate the projector to the tray. This is done in two steps. In the first step an image of the table with a tray is acquired. On this image a user marks the four corners of the tray clockwise by mouse-click. From these positions and the known tray dimension a homography $H_{C,T}$ is calculated mapping tray coordinates x_T to camera image coordinates x_C. In the second step five red square markers are projected on the table and imaged by the camera. Four squares are used as calibration markers, the fifth marks the projector origin. For the known positions of the markers in the camera as well as projected image homography $H_{P,C}$ from camera coordinates x_C to projector coordinates x_P is derived. Combining the two homographies the pixel position x_P in the projector's plane can be calculated for a given location x_T on the tray by $x_P = H_{P,C} \cdot H_{C,T} \cdot x_T$. With the tray layout information available
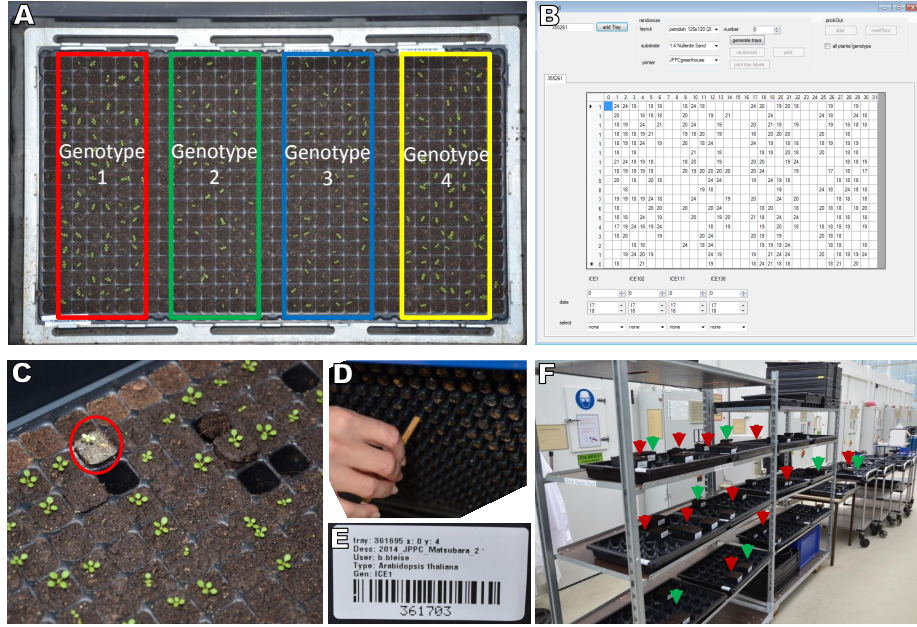
Fig. 4. Transplantation and Randomization. A: Tray with four genotypes. B: Graphical user interface for germination. C: Next seedling to pick. D: Pricking out from below the table without moving the tray. E: Label with destination. F: 22 target trays for 880 plants with randomized destinations, two genotypes symbolically indicated by arrows.

from the database, size and location of a pot or well in the projector image can be calculated. For higher accuracy more markers could be used, however the described procedure enables sufficiently accurate highlighting of single pots.

Visual Inspection. For Visual inspection a tray is set into the cutout and identified by its ID, triggering highlighting of all seedlings with two or more leaves (see Figure 3B). Germination detection can be toggled by mouse click on the respective pot. The toggling event is stored in the database. In case a non-detected seedling is now marked as being in two-leaves stage the time point when the mouse click was performed is used as detection date. Overviews and statistics of the germination dates and rates can be printed and used for further experiment design (Figure 3C).

2.4   Transplantation Guidance

Seedlings are pricked out and transplanted into bigger pots for subsequent experiments. Which seedlings to pick is decided by selection rules, e.g. medium sized or the largest seedlings, germinated at a certain day, with the highest growth rate, or random selection. This improves repeatability of experiments compared

to visual selection of seedlings. Thus, in a first step for transplantation, a graphical user interface allows selection of trays to calculate statistics like seedling size distributions (see Figure 4B) and to define selection rules.

We use a randomized plant order on trays to mitigate border and micro-climate effects. To enable randomization, the desired number of seedlings per genotype need to be provided together with the layout of destination trays. The system automatically randomizes seedlings' target positions over all destination trays and generates labels for each destination tray, which are printed directly.

Transplantation is done source tray after source tray and plant after plant. For each plant (i) the system highlights the seedling (Figure 4C) using the projector, (ii) the user pushes the pot content from below (Figure 4D), and (iii) transplants the seedling into a bigger pot, prefilled with soil. Simultaneously the system generates a new plant object in the database, links it to the seed ID, and prints out the label for the pot in which the seedling is implanted. To simplify randomization, the printed plant label also contains the destination tray ID and position on that tray (Figure 4E). The user sticks the label to the pot, waters the seedling and brings the pot to its destination position (Figure 4F). The destination position printed on the label speeded up the randomization process considerably and reduced mistakes. A foot switch (Figure 4A, yellow object on the floor) is used to step to the next seedling, avoiding dirt on mouse or keyboard, keeping the users hands free for transplantation.

Besides considerable speedup, improved reliability, and less tiring work due to user guidance, the system automatically joins results of seedling- and plant measurements in the database. This is enabled by unique plant and seed identities and appropriate position tracking of pots during the pricking out process.

2.5   Dataset

By using this system, we collected 19,486 images and initially labelled them with 'plant', when one or more healthy seedlings were visible having at least two fully developed and unfolded leaves, i.e. their cotyledons, or 'no plant' else. However, a finer-grained classification may be of interest, when fully automated systems need to assess germination of seedlings. We therefore relabelled the images by visual inspection in four 'no plant' subclasses and four 'plant' subclasses. The table in Figure 6 shows the number of images available per class and subclass and Figure 5 shows example images of the eight subclasses.

The negative 'no plant' subclasses are:

- class 0 'no plants': no plant material visible at all
- class 1 'plants with less than two leaves visible': there is a plant visible, but it clearly does not have two leaves fully unfolded yet; or larger parts of the plant are covered by soil.
- class 2 'multiple plants with less than two leaves visible': same as class 1, but multiple plants are visible.
- class 3 'plants with almost two leaves visible': This class contains the hard to decide cases, where seedlings are very close to have two fully developed leaves, but are just not yet there.
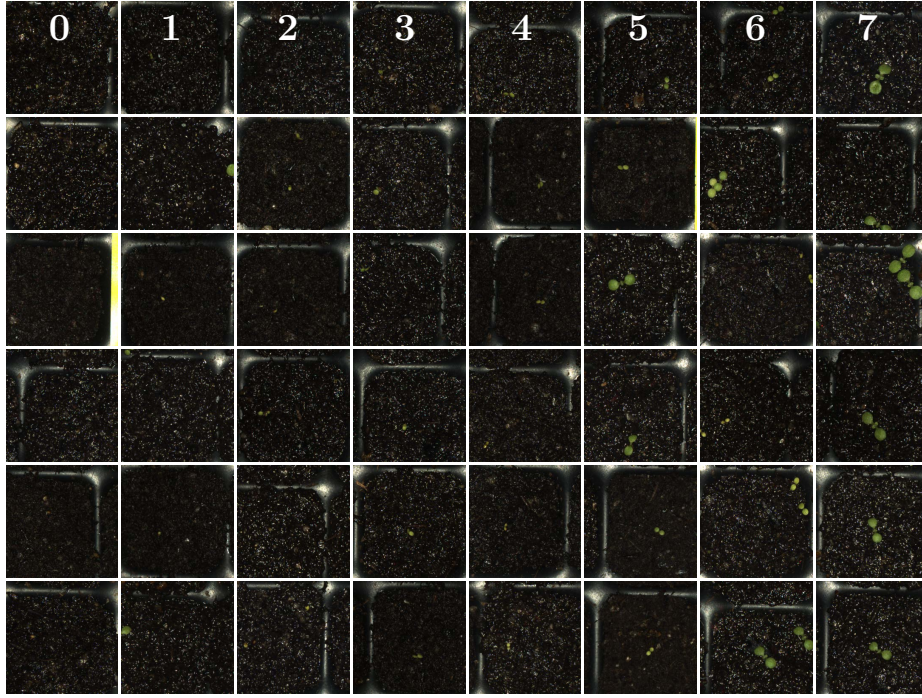
Fig. 5. Example images from each subclass, ordered from left to right. Each column shows examples from the same subclass.

The positive 'plant' subclasses are:

– class 4 'plants with just two leaves': This class contains the hard to decide cases, where seedlings have just unfolded their cotyledons. They may be tiny or by are fully developed.
– class 5 'plants with two leaves': clear cases of plants in two-leaves stage.
– class 6 'multiple plants with two leaves': like class 5, but multiple plants visible.
– class 7 'plants with four leaves': at least one of the visible seedlings has more than two leaves.

The cases in subclasses 3 and 4 are the borderline cases of the two main classes. They are hard to distinguish even for human raters, as the opening of cotyledons is a gradual process. However, the more accurate the transition between still closed and already opened cotyledons can be pinpointed, the higher the temporal resolution possible in an automated system. Currently, imaging takes place once a day and visual inspection only just before transplantation.

Labelling was done by five human raters. Per image assignments to the different subclasses are represented as normalized probabilities. Their values can
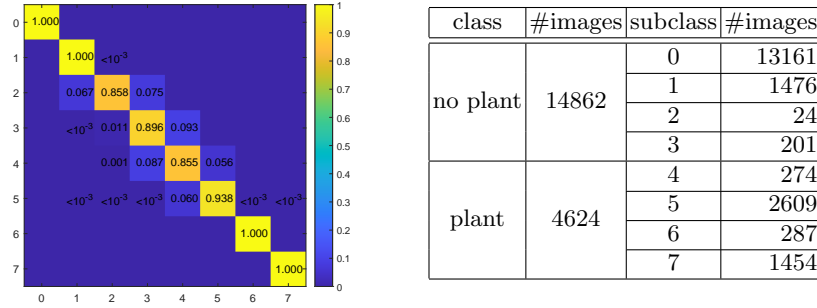
class	#images	subclass	#images
no plant	14862	0	13161
		1	1476
		2	24
		3	201
plant	4624	4	274
		5	2609
		6	287
		7	1454

Fig. 6. Confusion matrix for subclasses as labelled by human raters and number of images per subclass.

be found in the 'ImageDescription' tag of each image, stored in tiff format. A confusion matrix is shown in Figure 6. For binary 'plant' (positive) or 'no plant' (negative) assignment we summed up probabilities of the four respective sub-classes. An image then was assigned to the 'plant' or 'no plant' subclass with the highest probability. In doubt, the class with lower index was selected. Number of images per class are also given in Figure 6.

True positive and true negative rates are 99.44% and 99.87%, respectively, and the overall accuracy for the whole dataset being above 99.77%. The target performance of an automated solution should therefore be similarly high.

In Figure 5 we observe that cropping of captured images to images of single pots according to the tray layout does not always yield well centered pots. A more precise mechanical setup or image-based pot detection and cropping may help. Please note, that non-perfect pot alignment does not significantly influence the presented detection task and challenge.

2.6   Algorithm for Detection of Two-Leaves Stadium

Input RGB images are segmented into a binary foreground-background mask using thresholds in HSV-space [27] (cmp. Section 2.2). On the binary mask a Euclidean distance map is computed (using *openCV* [2]), containing distances between the object pixel and the nearest background pixel. Local maxima in the distance map are potential leaf center points x_i. A maximum value gives the radius r_i of a leaf candidate i, more specifically, the radius r_i of the biggest inscribed circle. Candidates i with a radius r_i smaller than a predefined, species-specific threshold (minimal leaf radius) are deleted. The threshold is needed to delete maxima detected on the petiole. Images with two or more detected leaves are labelled as 'germinated'. Leaf diameters $d_i = 2r_i$, center points x_i, detected projected leaf area (i.e number of foreground pixels) etc. are then assigned to the seed ID and written to the database.

3   Experiments

Experiments involve transfer learning [28] of five well established deep neural networks, namely VGG16 [24], VGG19 [24], DenseNet121 [11], and InceptionV3 [25], where we distinguish two different optimizer configurations for InceptionV3. The networks have been pre-trained for image classification on ImageNet [6], i.e. for classification of 1000 different classes of single dominant foreground objects. Seedling detection can be considered as classification of just two (or eight) other classes. The selected networks are known to perform well on ImageNet and their pretrained versions are easily available e.g. in Keras [4].

For all tested architectures, we keep the pretrained feature extraction layers fixed. The original classifier layers are removed and the simple classifier – dense layer 256, dropout, dense layer 1 – introduced in [23] is used instead. As optimizer we either use RMSProp (setup 'A', with VGG's and InceptionV3) or Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (setup 'B', with DenseNet121 and InceptionV3). Best batch size, step size and number of epochs are derived by grid search. Training is done with class weighted cross entropy loss, as usual.

For the challenge, we provide a fixed split of the data into training, validation and test set in an 80/10/10 split. In order to ensure that each of the sets represent the same distribution, we perform a 5-fold cross-validation within the grid search. To this end, we split the data into 5 sets $S_i, i \in \{1, ..., 5\}$ of approximately same size. For the ith validation run, we leave set S_i out as validation set and keep the other 4 sets for training of our baseline transfer learned neural nets. In addition, we split the 20% validation set into two subsets being validation and test set candidates for the final split. The grid search results are shown in supplemental Figures S1–S3. In Figure 7 we show boxplots of the accuracies[1] and losses of our five different network configurations, trained using the best values from the grid search. Shown are results for training as well as 20% validation set and its split into 10% test and 10% validation subsets. We observe that for all tested networks all validation values lie close together, respectively, indicating that the different data splits are equivalent. We provide the split with least deviation from the average performance. With the resulting 80/10/10 split, we transfer learned the neural nets, again, where we reinitialized the trainable parameters. Results are shown in Figure 8 as 'confusion' wrt. the binary labels, i.e. per subclass accuracy. The nets achieve accuracies between 96.1% (VGG16) and 98.3% (DenseNet121). We observe, that wrong classification mostly happens in the hard to decide subclasses 3 and 4, but even for the clear cases in subclasses 0 and 6, 7, accuracies are mostly not 100%.

For comparison we also show the results of our classic detection algorithm from Section 2.6 and results of human raters. We see that results of the classic algorithm are excellent in terms of recall (99.8%) at the cost of a lower precision (71.7%). High recall is preferred by human raters manually cleaning the data using the handling stations. Overall accuracy is 90.6%. The deep networks

---

[1] For definitions of accuracy, precision and recall, please see Section B in the supplemental material.
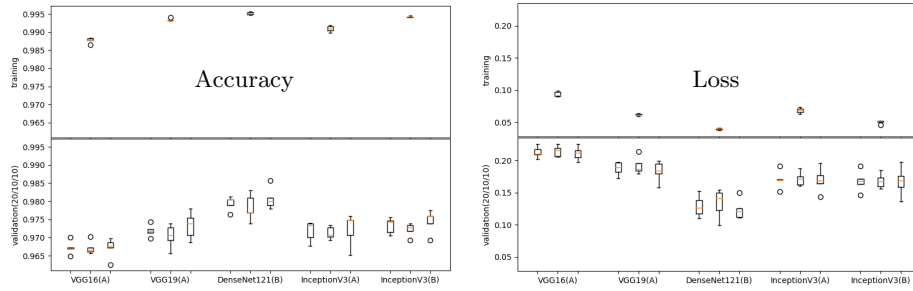
Fig. 7. Results of cross-validation experiment. Plots show results per net. Top part shows training results, bottom the results for the 20% validation set and its splits into two 10% sets. Left: Accuracy. Right: Loss

perform better in terms of accuracy, and may gain from fine-tuning and more sophisticated classifiers, which we will test in future work. However, available solutions so far are not yet close to human raters with well above 99.7% accuracy. Even the best achieved training accuracy of 99.5% (DenseNet121) is not yet there. For fully automated germination detection there is still a considerable gap to fill, which we hope to achieve by challenge results.

4   Conclusions

Fully automated above ground germination detection, i.e. detection of full opening of cotyledons, seems to be well in reach. However, detection reliability is not yet high enough to allow for fully automated operation. We presented a suitable imaging-based system, where we hope and believe that image processing can still be improved. Hence we make our image data and labels freely available and pose the germination detection problem as a new challenge. We hope that this data complements other available computer vision challenges in plant phenotyping [16,15,9,5,1], allowing a further improvement of urgently needed methods [17].

In conjunction with the germination detection system, we presented a handling system supporting manual greenhouse work in plant phenotyping experiments. It not only facilitates visual inspection and transplantation of seedlings, making such work less tiring, less error-prone, but at the same time increases throughput even at a reduced human workload. We observed that transplantation can be done at 15s per plant over hours without slowing down when teams of two people cooperate. The system was designed along well established greenhouse procedures which increased usability and user acceptance. Especially the avoidance of data to be keyed in during transplantation was very well received.

The system allows for automated, measurement-based seedling selection, making this process quantifiable and less subjective. Heterogeneous data describing the plants status is automatically kept consistent with the database during the transplantation process. This enables reliable tracking of plants from seed to
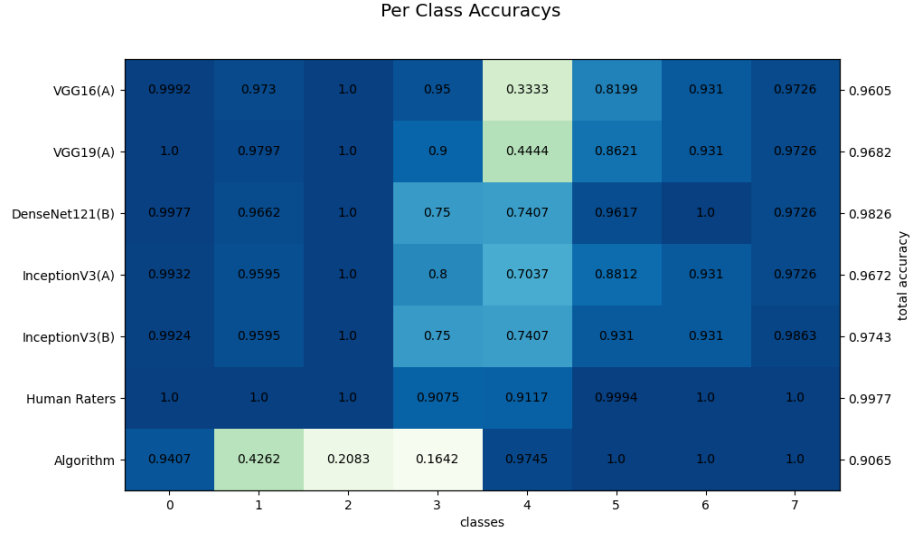
Per Class Accuracys

Fig. 8. Confusion in the subclasses for the transfer learned nets using the 80/10/10 split. Human raters and the algorithm (see Section 2.6) are evaluated on the whole dataset.

plant, an experimental option we want to explore in future plant phenotyping experiments.

Authors' contributions

HS initiated the research leading to this manuscript and drafted the manuscript. HS, BB, AF, and JR have made substantial contributions to conception and design of the overall system described here and contributed to algorithmic solutions and interfaces. BB designed, implemented, and tested the database system. AF designed, implemented, and tested the germination detection system. JR and AF designed, implemented, and tested the pricking out station jointly. JvS and LS performed the deep learning studies. All authors contributed to writing the manuscript and proofreading. All authors read and approved the final manuscript.

Acknowledgements

References

1. J. Bell and H. Dee. Aberystwyth Leaf Evaluation Dataset. https://doi.org/10.5281/zenodo.168158, 2016.
2. G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
3. B. Bruns, H. Scharr, and F. Schmidt. Entwicklung einer Multi-Plattform-Benutzerschicht zur tätigkeitsbegleitenden Verwaltung von Phänotypisierungsexperimenten und Pflanzenbestandsdaten. In *Komplexität versus Bedienbarkeit Mensch-Maschine-Schnittstellen, Referate der 35. GIL-Jahrestagung, 23. - 24. Februar 2015, Geisenheim, Germany*, pages 1–4, 2015.
4. F. Chollet et al. Keras. https://keras.io, 2015.
5. J.A. Cruz, Xi Yin, Xiaoming Liu, S.M. Imran, D.D. Morris, D.M. Kramer, and Jin Chen. Multi-modality imagery database for plant phenotyping. *Machine Vision and Applications*, 27:735–749, 2016.
6. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
7. F. Fiorani and U. Schurr. Future scenarios for plant phenotyping. *Annual review of plant biology*, 64:267–291, 2013.
8. C. Granier, L. Aguirrezabal, K. Chenu, SJ. Cookson, M. Dauzat, P. Hamard, JJ. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau, and F. Tardieu. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist*, 169(3):623–635, 2006.
9. Wei Guo, E. David, S. Madec, I. Stavness, H. Aasen, A. Hund, P. Sadeghi-Tehran, S. Chapmann, and Shouyang Liu. Global WHEAT dataset. http://www.global-wheat.com/, 2020.
10. D Houle, D R Govindaraju, and S Omholt. Phenomics: the next challenge. *Nat Rev Genet*, 11(12):855–866, December 2010.
11. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.
12. S. Jahnke, J. Roussel, T. Hombach, J. Kochs, A. Fischbach, G. Huber, and H. Scharr. phenoseeder-a robot system for automated handling and phenotyping of individual seeds. *Plant physiology*, 172(3):1358–1370, 2016.
13. M. Jansen, F. Gilmer, B. Biskup, K.A. Nagel, U. Rascher, A. Fischbach, S. Briem, G. Dreissen, S. Tittmann, S. Braun, I. De Jaeger, M. Metzlaff, U. Schurr, H. Scharr, and A. Walter. Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in Arabidopsis thaliana and other rosette plants. *Functional Pant Biology, Special Issue: Plant Phenomics*, 36(10/11):902–914, 2009.
14. N. MacLeod, M. Benfield, and P. Culverhouse. Time to automate identification. *Nature*, 467(7312):154–155, 09 2010.
15. M. Minervini, A. Fischbach, H. Scharr, and S.A. Tsaftaris. Plant phenotyping datasets. http://www.plant-phenotyping.org/datasets, 2015.
16. M. Minervini, A. Fischbach, H. Scharr, and S.A. Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern recognition letters*, 81:80–89, 2016.

17. M. Minervini, H. Scharr, and S.A. Tsaftaris. Image analysis: The new bottleneck in plant phenotyping. *IEEE Signal Process. Mag.*, 32(4):126–131, July 2015.
18. K.A. Nagel, A. Putz, F. Gilmer, K. Heinz, A. Fischbach, J. Pfeifer, M. Faget, S. Blossfeld, M. Ernst, C. Dimaki, B. Kastenholz, A.-K. Kleinert, A. Galinski, H. Scharr, F. Fiorani, and U. Schurr. GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. *Functional Plant Biology*, 39:891–904, 2012.
19. J.A. Poland and R.J. Nelson. In the eye of the beholder: The effect of rater variability and different rating scales on qtl mapping. *Phytopathology*, 101(2):290–298, 2010.
20. D. Rousseau, H. Dee, and T. Pridmore. Imaging Methods for Phenotyping of Plant Traits. In Jitendra Kumar, Aditya Pratap, and Shiv Kumar, editors, *Phenomics in Crop Plants: Trends, Options and Limitations*, pages 61–74, New Delhi, 2015. Springer India.
21. H. Scharr, B. Bruns, A. Fischbach, J. Roussel, L. Scholtes, and J. vom Stein. Juelich dataset for germination detection of soil-grown plants. http://dx.doi.org/10.25622/FZJ/2020/1, 2020.
22. F. Schmidt, B. Bruns, T. Bode, H. Scharr, and A. B. Cremers. A distributed information system for managing phenotyping mass data. In *Massendatenmanagement in der Agrar- und Ernährungswirtschaft, Erhebung - Verarbeitung - Nutzung, Referate der 33. GIL-Jahrestagung, 20. - 21. Februar 2013, Potsdam, Germany*, pages 303–306, 2013.
23. Xiaolei Shen, Jiachi Zhang, Chenjun Yan, and Hong Zhou. An automatic diagnosis method of facial acne vulgaris based on convolutional neural network. *Scientific Reports*, 8(1):5839, 2018.
24. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.
25. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.
26. F. Tardieu and U. Schurr. White paper on plant phenotyping. In *EPSO workshop on plant phenotyping*, Jülich, November 2009. http://www.plantphenomics.com/phenotyping2009.
27. A. Walter, H. Scharr, F. Gilmer, R. Zierer, K.A. Nagel, M. Ernst, A. Wiese, O. Virnich, M.M. Christ, B. Uhlig, S. Jünger, and U. Schurr. Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytologist*, 174(2):447–455, 2007.
28. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3320–3328. Curran Associates, Inc., 2014.