Learning from Synthetic Dataset for Crop Seed Instance Segmentation

Yosuke Toda[1,2*], Fumio Okura[1,3], Jun Ito[4], Satoshi Okada[5], Toshinori Kinoshita[2], Hiroyuki Tsuji[4], and Daisuke Saisho[5]

[1]Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
[2]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Chikusa, Nagoya 464-8602, Japan
[3]Department of Intelligent Media, Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
[4]Kihara Institute for Biological Research, Yokohama City University, Maioka 641-12, Totsuka, Yokohama, 244-0813 Japan
[5]Institute of Plant Science and Resources, Okayama University, Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan

*Correspondence should be addressed to Yosuke Toda; tyosuke@aquaseerser.com

Incorporating deep learning in the image analysis pipeline has opened the possibility of introducing precision phenotyping in the field of agriculture. However, to train the neural network, a sufficient amount of training data must be prepared, which requires a time-consuming manual data annotation process that often becomes the limiting step. Here, we show that an instance segmentation neural network (Mask R-CNN) aimed to phenotype the barley seed morphology of various cultivars, can be sufficiently trained purely by a synthetically generated dataset. Our attempt is based on the concept of *domain randomization*, where a large amount of image is generated by randomly orienting the seed object to a virtual canvas. After training with such a dataset, performance based on recall and the average Precision of the real-world test dataset achieved 96% and 95%, respectively. Applying our pipeline enables extraction of morphological parameters at a large scale, enabling precise characterization of the natural variation of barley from a multivariate perspective. Importantly, we show that our approach is effective not only for barley seeds but also for various crops including rice, lettuce, oat, and wheat, and thus supporting the fact that the performance benefits of this technique is generic. We propose that constructing and utilizing such synthetic data can be a powerful method to alleviate human labor costs needed to prepare the training dataset for deep learning in the agricultural domain.

Introduction

Deep learning is a process that involves neural network parameter optimization to solve a specific task of interest[1]. While traditional machine learning requires a user predefined feature extraction, the neural network itself can learn the most suitable representation from the dataset and can therefore exert its power on high content data. In ImageNet Large Scale Visual Recognition Challenge of 2012[2], a convolutional neural network (CNN)-based architecture, namely AlexNet, outperformed the human image classification accuracy to classify 1000 categories[3]. Since then, deep learning has gathered wide attraction in both the scientific and industrial communities. Initially, deep learning was actively applied to image classification, however in recent years, it has been further developed to process various tasks in computer vision, such as semantic segmentation[4,5], object detection [6,7], and instance segmentation[8].

Such deep-learning-based image analysis has also been influencing the field of agriculture. This involves image-based phenotyping including weed detection[9], crop disease diagnosis[10,11], fruit detection[12], and many other applications as listed in the recent review[13]. Meanwhile, it is not only features from images but also a combination of environmental variables, that functionalized a neural network to predict plant water stress for automated control of greenhouse tomato irrigation[14]. Utilizing the numerous and high context data generated in the relevant field seems to have high affinity with deep learning.

However, one of the drawbacks of using deep learning is the need to prepare a large amount of labeled data. The ImageNet dataset as of 2012 consists of 1.2 million and 150,000 manually classified images in the training dataset and validation/test dataset, respectively[2]. Meanwhile, the COCO 2014 Object Detection Task constitutes of 328,000 images containing 2.5 million labeled object instances of 91 categories[15]. This order of annotated dataset is generally difficult to prepare for an individual or a research group. In the agricultural domain, it has been reported that sorghum head detection network can be trained with a dataset consisting of 52 images with an average of 400 objects per image[16], while a crop stem detection network was trained starting from 822 images[17]. These case studies imply that the amount of data required in a specialized task may be less compared to a relatively generalized task such as ImageNet classification and COCO detection challenges. Nonetheless, the necessary and sufficient amount of annotation data to train a neural network is generally unknown. The annotation process is highly stressful for researchers, as it is like running a marathon without knowing the goal.

In such cases, domain adaptation (e.g. using ImageNet trained weights as initial network parameter for the tasks in different domains; also known as transfer learning or fine-tuning) and image augmentation (e.g. image flipping and rotating) have been the most commonly adopted techniques to compensate for the lack of data. More recently, several reports have highlighted the challenges with incorporating active learning or other approaches that loops the annotation and model training to minimize the labor cost [18–20].

On the lines of domain adaptation and data augmentation, learning from synthetic (e.g. CG-generated) images has been highlighted, which is occasionally referred to as the *sim2real* transfer. One of the important advantages of using synthetic dataset for training is that the ground truth annotations can be automatically obtained without the need for human labor. A successful example can be found in person image analysis method, that uses the image dataset with synthetic human models[21] for various uses such as person

pose estimation[22]. One drawback of the sim2real approach are the gaps between the synthesized images and the real scenes, e.g. non-realistic appearances. To counter this problem, many studies attempt to generate realistic images from synthetic datasets, such as by domain adaptation techniques using generative adversarial networks (GAN)[23]. Although the methods for generating realistic images from synthetic images were well studied in the CG community[24], GAN-based approaches are recently being paid attention for generation of training dataset[25].

While the GAN-based approaches still require a large set of real images, another set of approaches that are bridging the sim2real gap *domain randomization*, which trains the deep networks using large variations of synthetic images with randomly sampled physical parameters. Although domain randomization is somewhat related to data augmentation, synthetic environment enables representation of variations under many conditions, which is generally difficult to attain by straightforward data augmentation techniques for real images. An early attempt at domain randomization was made by generating the images using different camera positions, object location, and lighting conditions; which is similar to the technique applied to control robots[26]. For object recognition tasks, Tremblay et al.[27] proposed a method to generate images with a randomized texture on synthetic 3D models.

Such sim2real approaches have also been used for the preparation of training data for plant image analysis. While Isokane et al [28] used the synthetic plant models for the estimation of branching pattern, Giuffrida et al. used GAN-generated images to train a neural network for *Arabidopsis* leaf counting[29]. Similarly, Arsenovic et al. used StyleGAN[30] to create training images for the plant disease image classification[31]. Meanwhile, Ward et al. generated artificial images of *Arabidopsis* rendered from 3D models and utilized them for neural network training in leaf segmentation[32]. As far as difficulties in the collection and annotation of training datasets is concerned, the use of synthetic images has a huge potential in the plant phenotyping research field.

Among various crop phenotypes, seed morphology has been one of the most important traits. This is because the seed shape directly influences the crop yield[33]. Several studies report identification of genes that enhance rice yield by utilizing Quantitative Trait Locus (QTL) involved in seed width[34,35]. Moreover, several studies utilized elliptic Fourier descriptors which enables to handle the seed shape as variables representing a closed contour, successfully characterizing the characters of various species[36–39]. Focusing on morphological parameters of seeds seems to be powerful metrics for both crop yield improvement and for biological studies. However, including the said reports, many of the previous studies have evaluated the seed shape either by qualitative metrics (e.g. whether the seeds are similar to the parental phenotype), by vernier caliper, or by manual annotation using an image processing software. The phenotyping is generally labor-intensive and cannot completely exclude the possibility of quantification errors that differ by the annotator. To execute a precise and large-scale analysis, automation of the seed phenotyping step was preferred.

In recent years, several studies have been reported to systematically analyze the morphology of plant seeds by image analysis. Ayoub et al. focused on barley seed characterization in terms of area, perimeter, length, width, F-circle, and F-shape based on digital camera captured images[40]. Herridge et al. utilized a particle analysis function of ImageJ (https://imagej.nih.gov/ij/) to quantify and differentiate the seed size of

Arabidopsis mutants from the background population[41]. *SmartGrain* software has been developed to realize the high throughput phenotyping of crop seeds, successfully identifying the QTL that is responsible for seed length of rice[42]. Moreover, commercially available products such as *Germination Scanalyzer* (Lemnatec; Germany) and *PT portable tablet tester* (Greenpheno, China) also aim or have the ability to quantify the morphological shape of seeds. However, the aforementioned approaches require the seeds to be sparsely oriented for efficient segmentation. When seeds are physically touching or overlapping each other, they are often detected as a unified region, leading to an abnormal seed shape output. This requires the user to reorient the seeds in a sparse manner, which is a potential bar to secure sufficient amount of biological replicate in the course of high throughput analysis. In such situations, utilizing deep learning-based instance segmentation was used to overcome such a problem by segmenting the respective seed regions regardless of their orientation. Nonetheless, the annotation process as described previously was thought to be the potential limiting step.

In this paper, we show that utilizing a synthetic dataset that the combination and orientation of seeds are artificially rendered, is sufficient to train an instance segmentation of deep neural network to process real-world images. Moreover, applying our pipeline enables us to extract morphological parameters at a large scale with precise characterization of barley natural variation at a multivariate perspective. The proposed method can alleviate the labor-intensive annotation process to realize the rapid development of deep learning-based image analysis pipeline in the agricultural domain as illustrated in Fig. 1. Our method is largely related to the sim2real approaches with the domain randomization, where we generate a number of training images by randomly locating the synthetic seeds with actual textures by changing its orientation and location.

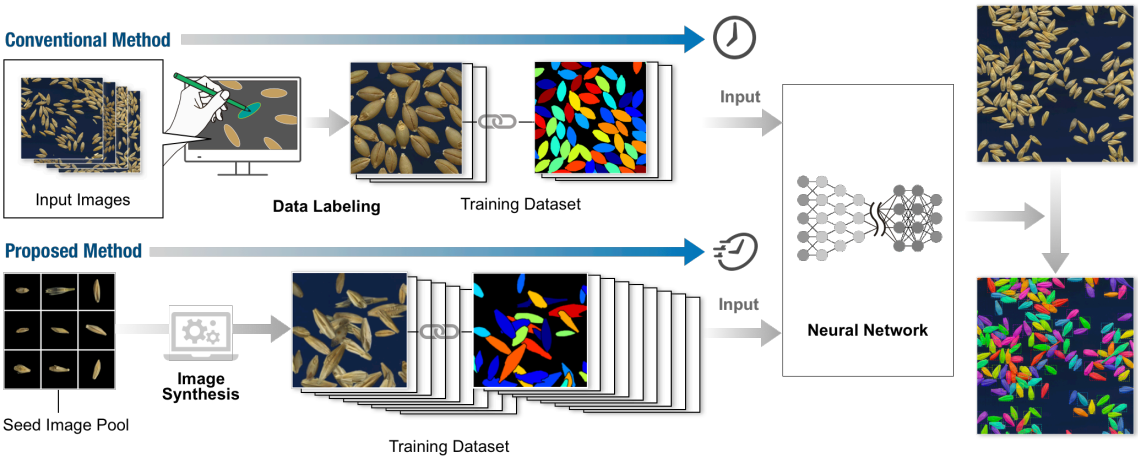

Fig. 1 | Overview of the proposed training process of crop seed instance segmentation.

Contribution: The contribution of this study is two folds. First, this is the first attempt to utilize a synthetic dataset (i.e., a sim2real approach) with domain randomization for the crop seed phenotyping, which can significantly decrease the manual labor for data creation (Fig. 1). Second, we propose a first method that can be used against the densely sampled (i.e., touching or overlapping) seeds using instance segmentation.

Methods

Plant Materials

Barley seeds used in this research are 19 domesticated barley (*Hordeum vulgare*) accessions and one wild barley (*H. spontaneum*) accession: B669, Suez (84); C319, Chichou; C346, Shanghai 1; C656, Tibet White 4; E245, Addis Ababa 40 (12-24-84); E612, Ethiopia 36 (CI 2225); I304, Rewari; I335, Ghazvin 1 (184); I622, H.E.S. 4 (Type 12); I626, Katana 1 (182); J064, Hayakiso 2; J247, Haruna Nijo; J647, Akashinriki; K692, Eumseong Covered 3; K735, Natsudaikon Mugi; N009, Tilman Camp 1 (1398); T567, Goenen (997); U051, Archer; U353, Opal; and H602, wild barley. All the details of the said cultivars can be obtained at the National BioResource Project (NBRP) (https://nbrp.jp). Meanwhile, seeds of rice (*Oryza Sativa*, cv. Nipponbare), oat (*Avena sativa*, cv. Negusaredaiji), Lettuce (*Lactuca sativa*, cv. Great Lakes), and wheat (*Triticum aestivum* cv. CS, Chinese Spring; N61, Norin 61; AL, Arina; and Syn01, a synthetic hexaploid wheat line Ldn/KU-2076 which is generated by a cross between tetraploid wheat *Triticum turgidum* cv. Langdon and *Aegilops tauchii* strain KU-2076) [43] were used in this report.

Image Acquisition

All the barley seeds were threshed using a commercial table-top threshing system (BGA-RH1, OHYA TANZO SEISAKUSHO & Co., Japan). The seed images were captured on an EPSON GT-X900 A4 scanner with the supplied software without image enhancement. Seeds were spread uniformly on the glass, scanned at 7019 x 5100 px at 600 dpi using a blue colored paper background. For the image acquisition of seeds of rice, oat, lettuce, and wheat, an overhead scanner ScanSnap SV600 (Fujitsu, JAPAN) was used with the image size of 3508 x 2479 at 300 or 600 dpi.

Synthetic Image Generation

A total of 20 single seed images per cultivar were isolated and saved as an individual image file. The background regions were removed such that the pixel value other than the area of the seed will be (0,0,0) in RGB color value. As a result, a total of 400 (20 seed images for 20 cultivars) background clean images were prepared to constitute a “seed image pool”. For the background image, four images at the fixed size of 1024 x 1024 were cropped from the actual background used in the seed scanning process and were prepared as a “background image pool”.

The synthetic image generation process is described as follows. First, an image was randomly selected from the background image pool and pasted to the virtual canvas of size 1024 x 1024. Second, another image was randomly selected from the seed image pool. Image rotation angle was randomly set upon selection. After rotation, the x and y coordinate at which the image was to be pasted was randomly determined, however, the coordinate value was restricted to a certain range so that the image does not exceed the canvas size, which its values were dependent on the selected seed image size and its rotation angle. Third, the seed

image was pasted to the canvas according to the determined values described above. When pasting, alpha masks were generated and utilized in alpha blending such that the area outside of the seed will be transparent and does not affect the canvas image. Moreover, utilizing the alpha mask, the seed perimeter was gaussian blurred to decrease the artifacts resulting from the background removal process of the seed image. Notably, if the region where the image was to be pasted in the canvas already had a seed image, the overlapping proportion of the area of the seeds was calculated. If the calculated value exceeded the ratio of 0.25, pasting was canceled, and another coordinate was chosen again. A maximum of 70 pasting trials were performed to generate a single image.

During the synthetic image generation, a mask that has the same image size as the synthetic image was created by first creating a black canvas and coloring the seed region with unique colors based on the coordinate of the placing object. The coloring was performed when the seed were randomly placed in the synthetic image. If a seed to be placed were overlapping an existing seed, the colors in the corresponding region in the mask image were replaced by the foreground color.

The above procedure generates an image size of 1024 x 1024 with seeds randomly oriented inside the canvas region. While in real-world images, seeds that are adjacent to the border of the image are cut off. To replicate such a situation, the borders of synthetic images were cropped to obtain the final image. The generated synthetic dataset constitutes 1200 set of data pairs of synthetic and mask image, in which each image has a size of 768 x 768, that were used for neural network training.

Model Training

We used a Mask R-CNN[8] implementation on the Keras/Tensorflow backend (https://github.com/matterport/Mask_RCNN). Configuration predefined by the repository was used including the network architectures and losses. The residual network ResNet101[44] was used for the feature extraction. From the initial weights of ResNet101 obtained by training using MS COCO dataset, we performed fine-tuning using our synthetic seed image dataset for 40 epochs by stochastic gradient descent optimization with a learning rate of 0.001 and batch size of 2. Within the 1200 images of the synthetic dataset, 989 were used for training, 11 for validation, and 200 for the test dataset. No image augmentation was performed during training. The synthetic training data has a fixed image size of 768 x 768; however, the input image size for the network was not exclusively defined such that variable sizes of the image can be fed upon inference. The network outputs a set of bounding boxes and seed candidate mask regions with a probability value. A threshold value of 0.5 was defined to isolate the final mask regions.

Test dataset for Model Evaluation

We prepared a test dataset consisting of 20 images and each image contained seeds derived from a homogeneous population (Fig. 3a). Each image had a size of 2000 x 2000. AP_{50}, AP_{75}, and AP@[.5:.95] per image (cultivar), as well as the mean AP of all images, was calculated. As the seeds to be detected per image averages to approximately 100 objects per image and images itself were acquired under the same experimental condition, we used one image per cultivar for model evaluation. For reference, we also prepared

200 synthetic images for testing (synthetic test dataset), which were not used for the model training or validation.

Metrics for Model Evaluation

To assess the accuracy of object detection using Mask R-CNN, we evaluated using two metrics, which were also used in the evaluation of the original report[8]. While they are commonly-used measures in object recognition and instance segmentation, such as in MS COCO[15] and Pascal VOC[45] dataset, we briefly recap our evaluation metrics for clarity. During the experiment, the evaluation metrics were calculated using the Mask R-CNN distribution.

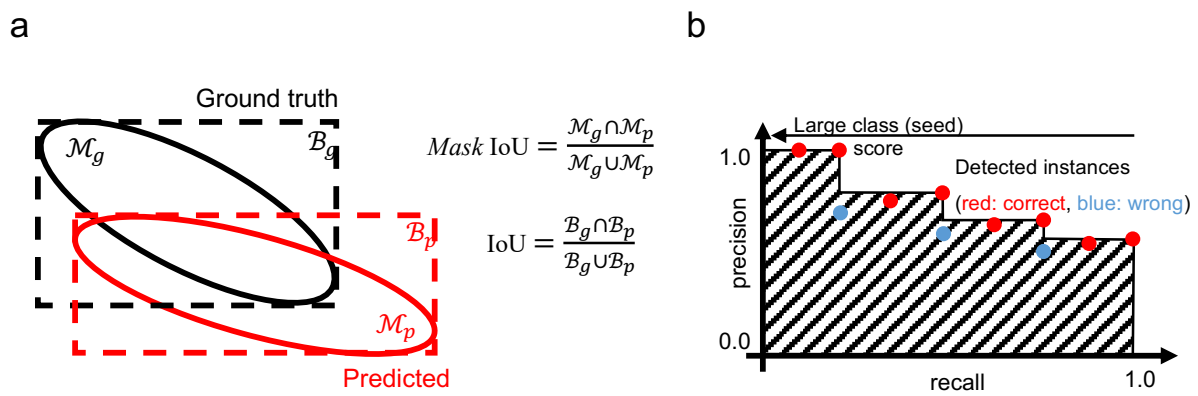

Fig. 2 | Evaluation metrics for object detection accuracy. (a) The intersection-over-union (IoU) definitions for bounding boxes and masks. (b) The average precision (AP) defined as the area under the curves (AUC); shown as the area marked with slanted lines.

Recall. We first measured the recall, which evaluates how well the objects (i.e., seeds) are detected, which can be obtained by the ratio of true positive matches over the total number of ground-truth objects. To calculate the recall values, we determined the correct detection when the detection threshold of the intersection-over-union (IoU) between the ground-truth and predicted bounding boxes is over 0.5 (see Figure 2). In other words, for each ground-truth bounding box, if a detected bounding box overlaps over 50%, it was counted as the true positive. Hereafter, we denote the recall measures as Recall_{50}.

Average precision (AP) using *mask* IoUs. The drawbacks of the recall measure include penalizing the false positive detections and evaluating using the overlaps of bounding boxes that are poor approximation of the object shape. We, therefore, calculated the average precision (AP) using *mask* IoUs, which can be a measure of the detection accuracy (in terms of both recall and precision) as well as providing a rough measure of mask generation accuracy. During the computation of APs, we first compute the IoU between the instance masks (*mask* IoU), as shown in Figure 2a. AP can be obtained based on the number of correct (i.e., true positive) and wrong (i.e., false positive) detection determined using a certain threshold of mask IoUs. Figure 2b

summarizes the computation of the AP. We sort the detected instances using the class score (i.e., the confidence that the detected object is a seed, in our case) in the descending order. For the n-th instance, the precision and recall, based on the mask IoU threshold, are calculated for the subset of instances from 1st to n-th detections. By repeating the process for each of the instances, we obtain a receiver operating characteristics (ROC) curve shown in Figure 2b. The AP is defined as the ratio of the rectangle approximations of the area under the curve (AUC), which is shown as the area marked by slanted lines in the figure. APs thus takes the value from 0.0 to 1.0 (i.e., 100%). We evaluated APs using multiple mask IoU thresholds. AP_{50} and AP_{75} are computed using the mask IoU threshold of 0.5 and 0.75, respectively. AP_{75} becomes a stricter measure than AP_{50}, because AP_{75} requires the correct matches with more accurate instance masks. Similar to MS COCO evaluation, we also measured AP@ [.5:.95], which is the average value of APs with IoU thresholds from 0.5 to 0.95 with an the interval of 0.05.

Quantification of Seed Morphology

The main application of the seed instance segmentation is to quantify phenotypes of seeds for analyzing and comparing morphological traits. In the mask image, morphological variables of seed shape such as area, width, and height were calculated using the *measure.regionprops* module of the scikit-image library, respectively. To analyze the characteristics of seeds across different cultivars, principal component analysis (PCA) was applied to the variables. In the result section, we briefly present the analysis using different types of descriptors, computed by elliptic Fourier descriptors (EFD) and variational autoencoder (VAE) both of which are described below.

Post-processing: Selection of isolated seeds. The instance segmentation network outputs a set of bounding boxes and seed area candidates as mask images, where some seeds overlap with each other. To analyze the seed morphology (or use for further phenotyping applications), it is required to select the seeds that are isolated (i.e., not partly hidden) from neighboring seed instances. To select such seeds, post-processing step was introduced. First, the bounding box coordinates were checked whether it resides inside the 5 px margin of the image. The bounding boxes that protrude the margin were removed. Second, using the solidity (ratio of the region of interest area against its convex hull area) of the respective mask as a metric, the 25% lower quantile threshold was determined and used to remove the outliers. Similarly, further outliers were removed by a 5% lower and 95% higher quantile threshold of length-to-width ratio. The threshold was empirically determined during the analysis.

Elliptic Fourier descriptors (EFD). EFD[46] has been used to quantify the contour shape of seeds[36], which approximate the contour shape as the set of different ellipses. During the computation of EFD, segmented seed images were first converted to binary mask image where the background pixel value was 0 and the seed area is 1. Next, the contour of the seed was detected by the *find_contours* module of the scikit-image library. The detected contours were converted to EFD coefficients using the *elliptic_fourier_descriptors* module of *pyefd* library (https://github.com/hbldh/pyefd) under the condition of harmonics 20 and with normalization

so as to be rotation and size-invariant. The output was flattened, which converted the shape of the array from 4 x 20 to 80. As the first three coefficients are always or nearly equal to 1, 0, 0 due to the normalization process, they were discarded upon further analysis. A total of 77 variables were used as descriptors for principal component analysis (PCA).

Variational autoencoder (VAE). Autoencoder (AE) is a type of neural network with an encoder-decoder architecture that embeds a high-dimensional input data (e.g., images) to a low-dimensional latent vector, to correctly decode the input data from the low-dimensional vector. Variational autoencoder (VAE)[47] is a variant of AE, where the distribution in the latent space is generated to fit a prior distribution (e.g., Gaussian distribution, N(0,1)). In a generative model, the low-dimensional parameters in the latent space is often used as the nonlinear approximation (i.e., dimensional reduction) of the dataset. Similar to other approximation methods like PCA, the parameters in the latent space estimated by VAE can be used for interpolation for the data distribution; while the input data with different characteristics (e.g., different species) is often well separated in the space (see [48] for details) compared to the conventional methods (e.g., PCA), without using the ground-truth labels for the classes during the training. We used a VAE with a CNN-based encoder-decoder network to visualize the latent space. In brief, the network receives an RGB image which has a shape of 256 x 256 x 3. For the encoder, input data were first passed through 4 layers of convolution with a filter number of 32, 64, 128, 256, respectively. Since we fit the latent space to the Gaussian distribution, the log variance and the mean of the latent space are computed after full-connection layers. For the decoder, the output of the encoder was passed through 4 layers of deconvolution with filter number of 256, 128, 64, 32, respectively. Finally, the convolution layer with 3 filters was added to convert the data back to an RGB image with its shape identical to the input image. In our analysis, we utilized the two-dimensional latent space (i.e., the final output of the encoder of VAE) to visualize the compressed features of the input image.

Software Libraries and Hardware

Computational analysis in this study was performed using Python 3.6. Keras (ver.2.2.4) with Tensorflow (ver. 1.14.0) backend for deep learning related processes. OpenCV3 (ver. 3.4.2) and scikit-image (ver. 0.15.0) was used for operations in morphological calculations of the seed candidate regions as well as basic image processing. A single GPU was used for network training and inference. R (ver. 3.5.1) was used for ANOVA and Tukey post hoc HSD test analysis.

Results

Preparation of Barley Seed Synthetic Dataset

Examples of seed images captured by the scanner are shown in Figure 3a. The morphology of barley seeds is highly variable between cultivars, in terms of size, shape, color, and texture. Moreover, the seeds randomly come in contact with or partially overlap each other. Determination of the optimal threshold for binarization may enable isolation of the seed region from the background; however, conventional segmentation methods such as watershed fail to segment the seed area for morphological quantification (see the results shown in Figure S1), indicating that employing a sophisticated segmentation method (in our case, instance segmentation using Mask R-CNN[8]) is indeed required for successful separation of the individual seeds. However, Mask R-CNN requires annotations of bounding boxes—which circumscribes the seed—and mask images that necessarily and sufficiently cover the seed area (Fig. S2). Given that the numbers of seeds per image are abundant (see Figure 3a), the annotation process has been predicted to be labor-intensive.

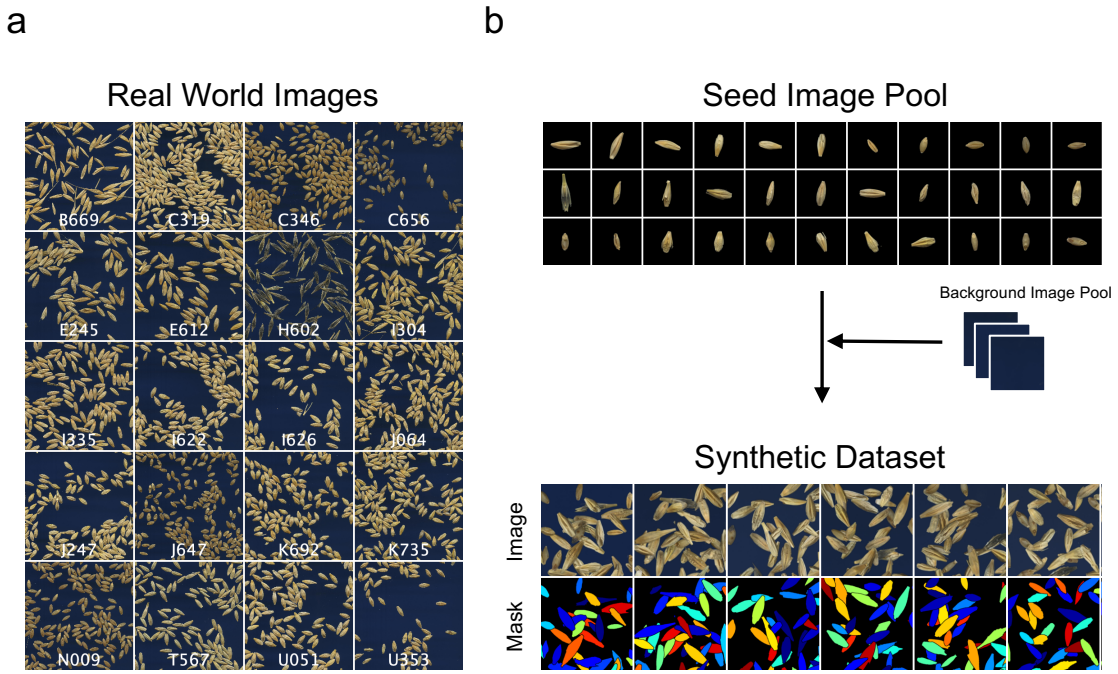

Fig. 3 | Data prepared in this study. (a) Images of barley seeds scanned from 20 cultivars. Cultivar names are described in white text in each image. These images were also used as a real-world test dataset shown in Table 1. (b) Scheme of generating synthetic images. Images are generated by combining real images of scanned seeds with the background images on to the virtual canvas. Simultaneously generated ground truth label (Mask) is shown at the bottom, wherein each seed area is marked with a unique color.

Figure 3b shows the seed image pool and synthesized dataset obtained using the proposed method (see Methods for details). Instead of labeling real-world images for use as a training dataset, Mask R-CNN

was trained using the synthetic dataset (examples shown at the bottom of Figure 3b), which is generated from the seed image pool and background image pool (Figure 3b top) using a domain randomization technique.

Model Evaluation

We show herein the visual results and a quantitative evaluation of object detection and instance segmentation by Mask R-CNN. The trained Mask R-CNN model outputs a set of bounding box coordinates and masks images of seed regions (Raw Output) (Figure 4a top). Examples of visualized raw output obtained from the real-world images show that the network can accurately locate and segment the seeds regardless of their orientation (Figure 4b and Figure S3). Table 1 summarizes the quantitative evaluation using the recall and AP measures (see Method section for details). The efficacy of seed detection was evaluated using the recall values computed for bounding box coordinates at 50% Intersection of Union (IoU) threshold ($Recall_{50}$). The model achieved an average of 95% and 96% on the synthetic and real-world test datasets, respectively. This indicates that the trained model can locate the position of seeds with very low false negative rate. From the Average Precision (AP) values, which were computed based on mask regions at varying mask IoU thresholds, comparable AP_{50} were achieved between the synthetic (96%) and real-world (95%) datasets. For higher IoU threshold (AP@ [.5:.95] and AP_{75}), the values of the synthetic test dataset (73%) exceeded that of the real-world test dataset (59%). These results suggest that the model's ability to segment the seed region is better in the case of the synthetic than the real-world images; however, considering the visual output interpretation (Figure 4B) and the values of AP_{50} (95%), we judged that seed morphology can be sufficiently determined from real-world images. The relatively low AP in high IoU in the real-world test dataset is possibly derived from the subtle variation in the manual annotation of seed mask regions.

		Object Detection Metric	Mask Region Metrics		
		$Recall_{50}$	AP@[.5:.95]	AP_{50}	AP_{75}
Synthetic Test Dataset		0.95	0.73	0.96	0.93
Real World Test Dataset	B669	0.92	0.56	0.92	0.84
	C319	0.95	0.62	0.91	0.86
	C346	0.98	0.64	0.97	0.89
	C656	0.96	0.61	0.95	0.92
	E245	0.95	0.63	0.94	0.84
	E612	0.96	0.66	0.98	0.89
	H602	0.87	0.42	0.78	0.41
	I304	0.99	0.64	0.98	0.88
	I335	0.97	0.67	0.93	0.92
	I622	0.93	0.62	0.93	0.87
	I626	0.96	0.65	0.95	0.89
	J064	0.93	0.65	0.97	0.86
	J247	0.94	0.65	0.97	0.86
	J647	0.98	0.62	0.98	0.92
	K692	0.98	0.69	0.98	0.93
	K735	0.95	0.62	0.92	0.86
	N009	0.99	0.63	0.99	0.91
	T567	0.98	0.63	0.98	0.88
	U051	0.96	0.65	0.96	0.89
	U353	1.00	0.65	0.98	0.89
	Average	0.96	0.59	0.95	0.86

Table 1 | Model Evaluation. Table describing the evaluation result of the trained Mask R-CNN raw output. Recall values at the IoU threshold of 50% ($Recall_{50}$) and Average Precision (AP) at the IoU 50% (AP_{50}), 75% (AP_{75}), and the mean value from IoU 50% to 95% with the step size of 5% (AP@ [.5:.95]) are shown.

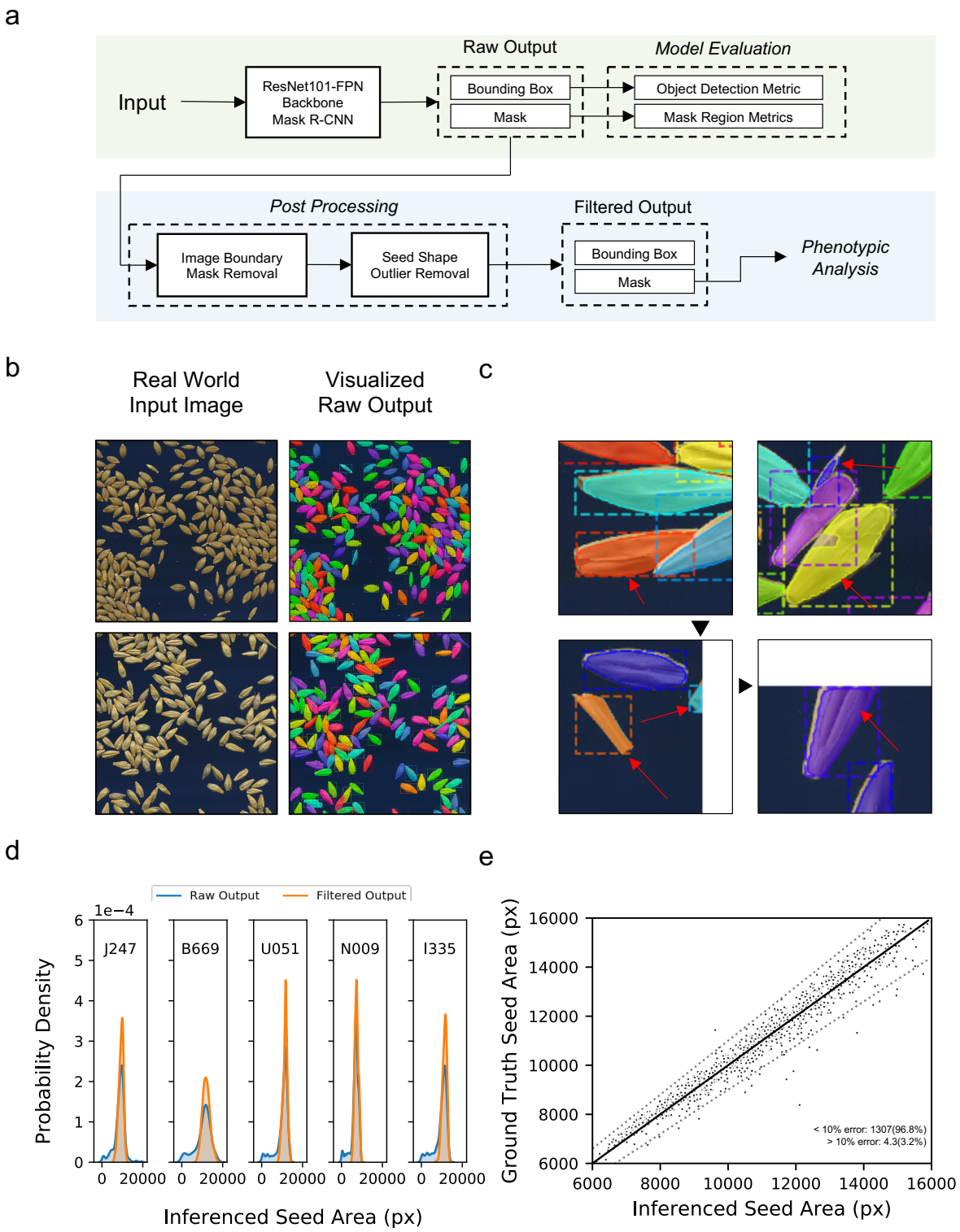

Fig. 4 | Image Analysis pipeline. (a) Summary of the image analysis pipeline. (b) Examples of graphical output of the trained Mask R-CNN on real-world images. Different colors indicate an individual segmented seed region. Note that even though the seeds overlap or touch each other, the network can still distinguish them as independent objects. (c) Examples of detected candidate regions to be filtered in the post-processing step, indicated using red arrows. Black arrowheads indicate the input image boundary. (e) Probability density of the seed areas of the raw output and filtered output. (f) Scatterplot describing the correlation of the seed

area that was measured by the pipeline (Inferenced Seed Area) and by manual annotation (Ground Truth Seed Area). Each dot represents the value of a single seed. Black and gray lines indicate the identity line and the 10% error threshold line, respectively. The proportion of the seeds exhibiting error lower or higher than the 10% mark is also displayed.

Post-Processing

As described in the Methods section, we introduced a post-processing step to the raw output to eliminate detections that are not suitable for further analysis. This process removes seed occlusion due to physical overlap, incomplete segmentation by the neural network, non-seed objects such as dirt or awn debris, or the seeds which were partly hidden due to the location being outside the scanned area (Figure 4c). Figure 4d and 4e show the distribution of the seed area before and after post processing. Even though the seed area itself was not used as a filtering criterion, the area values in the respective cultivars shift from a long-tailed to a normal distribution, which well reflects the characteristics of a homogenous population (Figure 4d). A comparison of the filtered output (Inferenced Seed Area) and hand-measured (Ground Truth Area) values displays a strong correlation, where the Pearson correlation value is 0.97 (Figure 4e). These results suggest that the filtered output values obtained from our pipeline are reliable for further phenotypic analyses.

Morphological Characterization of Barley Natural Variation

Our pipeline learns from synthetic images, which eases the training dataset preparation process. This pipeline enables large-scale analysis across multiple cultivars or species. To highlight the important advantages of the proposed pipeline, we herein demonstrate an array of analyses to morphologically characterize the natural variation of barley seeds, which highlights the crucial biological features that will provide guidance for further investigation. We selected 19 out of 20 cultivars which were used to train the neural network; however, we have acquired a new image that was not used for training or testing in further analysis. One accession, H602, was excluded from the analysis because the rachis could hardly be removed by husk threshing; therefore, the detected area did not reflect the true seed shape. From the pipeline, we obtained 4,464 segmented seed images in total (average of 235 seeds per cultivar).

As simple and commonly used morphological features, the seed area, width, length, and length-to-width ratio per cultivar were extracted from the respective images and are summarized in Figures 3a-d. With a sufficient number of biological replicates, we can not only compare the inter-cultivar difference (e.g. median or average) but also consider the intra-cultivar variance. We applied the Analysis of variance (ANOVA) with Tukey's post-hoc test to calculate the statistical difference between cultivars. Interestingly, many cultivars that visually display similar distribution patterns or medians are grouped into statistically different clusters (e.g., K735 and K692 in Figure 5a). The categorization into numbers of clusters in the respective morphological features suggest that they are regulated by multiple quantitative trait loci. To gain further insight into the morphology of barley cultivars characterized by various descriptors, we performed a multivariate analysis.

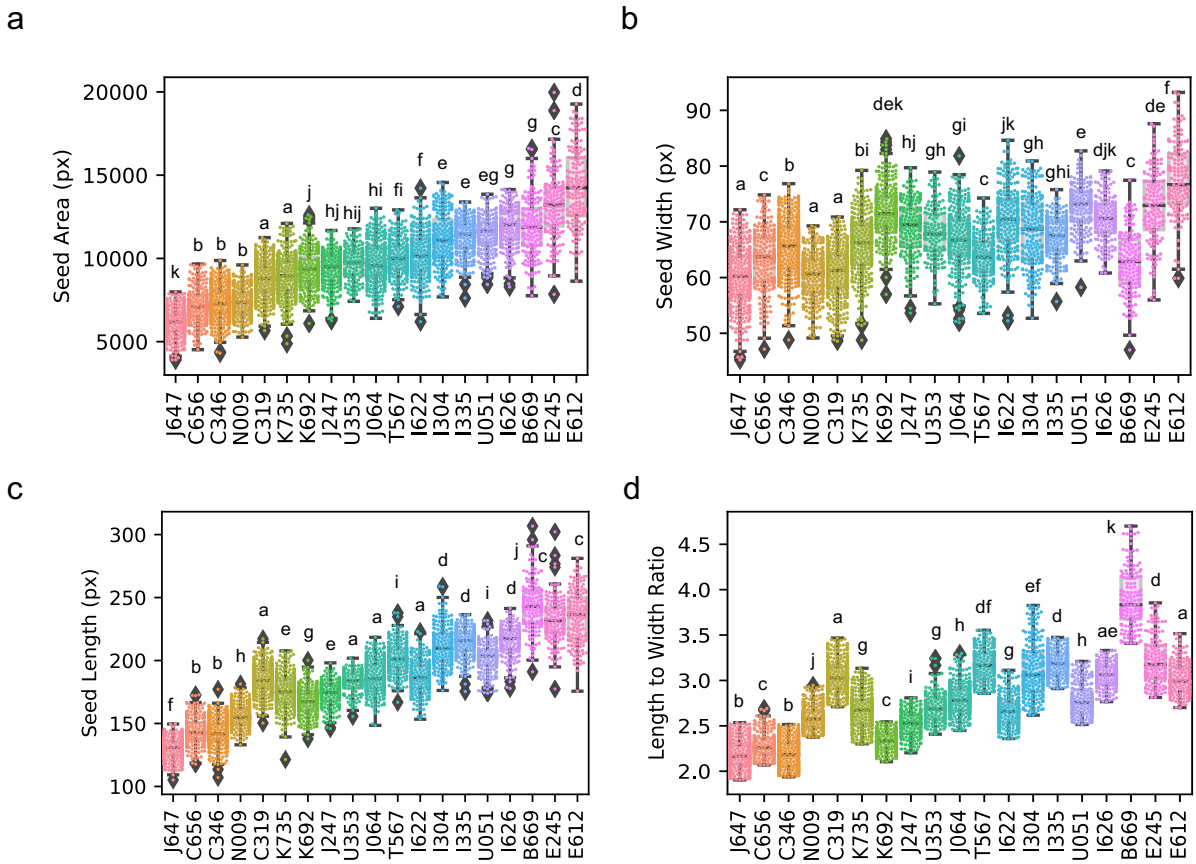

Fig. 5 | Analysis of natural variation of barley seed morphology. Whiskerplot overlaid with a swarmplot (colored dot) grouped by barley cultivars. (a) Seed area, (b) seed width, (c) length, and (d) length-to-width ratio. Diamonds represent outliers. Statistical differences were determined by one-way ANOVA followed by Tukey post-hoc analysis. Different letters indicate significant differences ($p < 0.05$).

First, we show the results of a principal component analysis (PCA) using eight predefined descriptors (area, width, length, length-to-width ratio, eccentricity, solidity, perimeter length, and circularity). The first two principal components (PC) could explain 88.5% of the total variation (Fig. 6a, b). Although there were no discrete boundaries, the data points tended to form a cluster unique to the cultivar in the latent space, indicating that cultivars can be classified to a certain extent according to the said descriptors. (Fig. 6a). Variables such as seed length (L) and perimeter length (PL) mainly constituted the first PC, with seed circularity (CS) oriented towards the opposite direction, while seed width (W) and length-to-width ratio had a major influence in PC2 (Fig. 6b). This is exemplified by the distribution of the slenderest B669 and the circular-shaped J647 at the far-right and far-left orientation in the latent space. Notably, while width (W) mainly constituted PC2, the direction of its eigenvector differs from that of length (L). Along with the moderate value of Pearson's correlation between length and width (0.5, $p < 0.01$) (Fig. S4), it is implied that genes that control both or either of size and length may coexist in the determination of barley seed shape, as reported in rice[49].

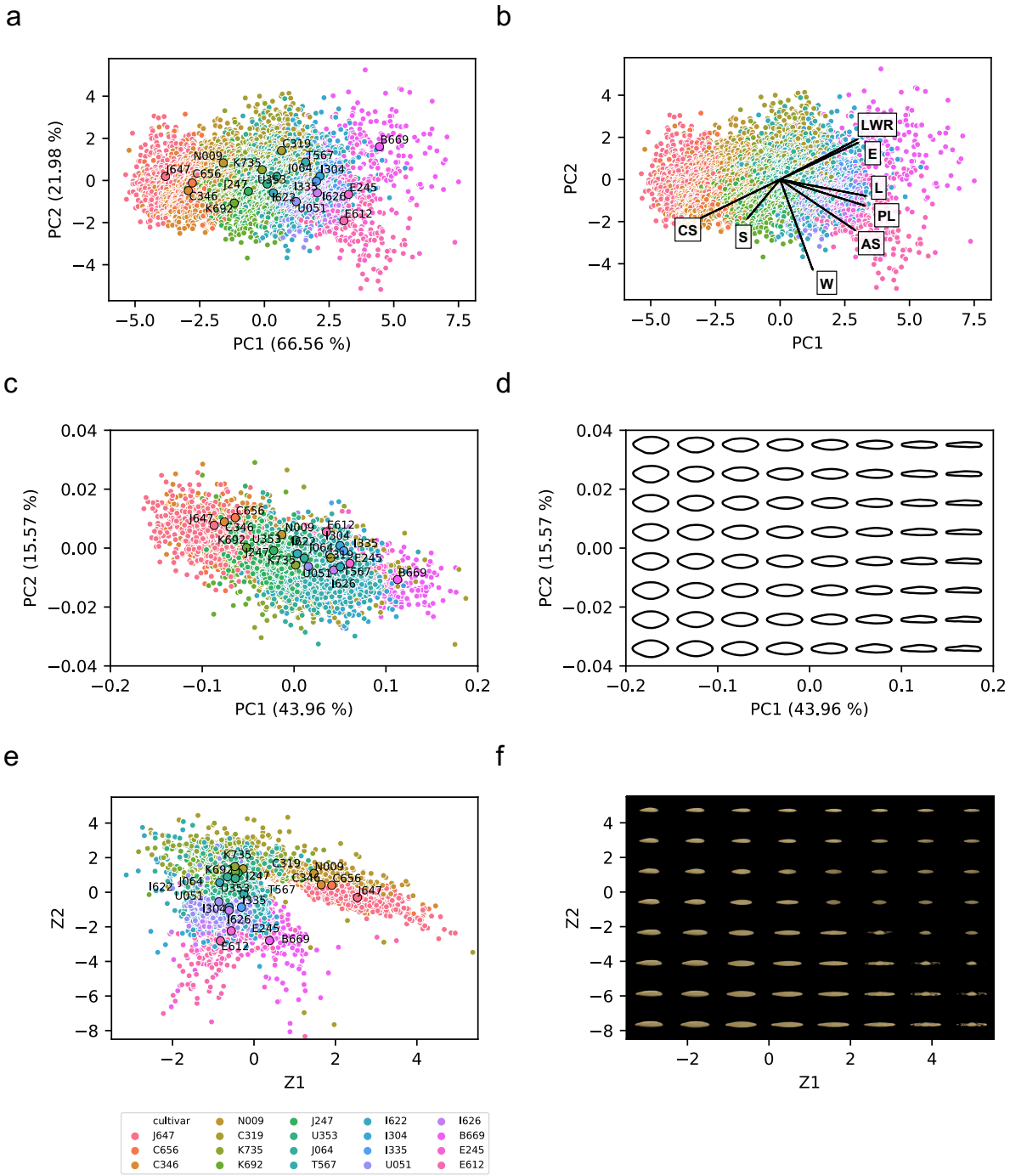

Fig. 6 | Multivariate analysis of barley seed morphology. (a,b) Principal Component analysis (PCA) with morphological parameters of barley seeds. Each point represents the data point of the respective seed. The colors correspond to those defined in the color legend displayed below (e). Mean PC1 and PC2 values of each cultivar are plotted as large circles with text annotations in (a). Eigenvectors of each descriptor are drawn as arrows in (b). LWR, length-to-width ratio; E, eccentricity; L, seed length; PL, seed perimeter length; AS, seed area; W, seed width; S, solidity; CS, seed circularity. (c,d) PCA with elliptic Fourier descriptors (EFD). The colors and point annotated in (c) follow those of (a). Interpolation of the latent space followed by reconstruction of the contours are displayed in (d). (e,f) Latent space visualization of Variational Autoencoders (VAE). The colors and point annotated in (e) follows those of (a). Interpolation of the latent

space followed by image generation using the generator of VAE are displayed in (f).

Next, we extracted the contour shapes of seeds using elliptic Fourier descriptors (EFDs) followed by PCA (Fig. 6b,c), which is also used in other studies for seed morphological analysis[36,37]. Compared to the PCA based on the eight morphological descriptors in Fig. 6a, the distributions of respective seeds were relatively condensed, while the clusters by cultivars were intermixed (Fig. 6c), possibly because the size information is lost upon normalization; therefore, EFD can utilize only the contour shape. Interpolating the latent space in the PC1 axis direction clearly highlights the difference in slenderness of the seed (Fig. 6d and Fig. S4a, left). PC2 did not show an obvious change in shape when compared to PC1 (Fig. 6d); however, it seemed to be involved in the sharpness of the edge shape in the longitudinal direction (Fig. S5a, right). Although further verification is required, rendering the average contours which represent the shapes of the respective cultivars implies the difference in such metrics (Fig. S5b).

Finally, we trained a variational autoencoder (VAE) for latent space visualization[47]. Unlike other methods using the shape descriptors (i.e., eight simple features or EFDs), the VAE inputs the segmented seed images, which can thus obtain a representation that well describes the dataset without feature predefinition (see Methods for details). The learned representation can be visualized into a two-dimensional scatterplot similar to a PCA (Fig. 6e). Compared to the PCA-based methods, VAE seems to cluster the cultivar in the latent space more explicitly. While the predefined morphological descriptors extract a limited amount of information from an image, VAE can handle an entire image itself; hence, the latter theoretically can learn more complex biological features. Overall, Z1 tend to be involved in the seed color (i.e. brightness) and size, while Z2 is in seed length (Fig.6f). Generally, unsupervised learning, utilizing deep neural networks such as VAE, requires a sufficient amount of data to fully exert its power to learn the representation of the dataset. The large-scale analysis across various cultivars provides researchers with a novel option to execute such analyses as demonstrated.

Application in various crop seeds

We further extended our method to verify the efficacy of our approach for other crop seeds. In this report, we newly trained our model to analyze the seed morphology of wheat, rice, oat, and lettuce, with the respectively generated synthetic datasets (Figure 7, top row). Processing the real-world images resulted in a clear segmentation of each species, regardless of seed size, shape, texture, or color, and background (Fig. 7 middle and bottom rows). In conclusion, these results strongly suggest the high generalization ability of our presented method.

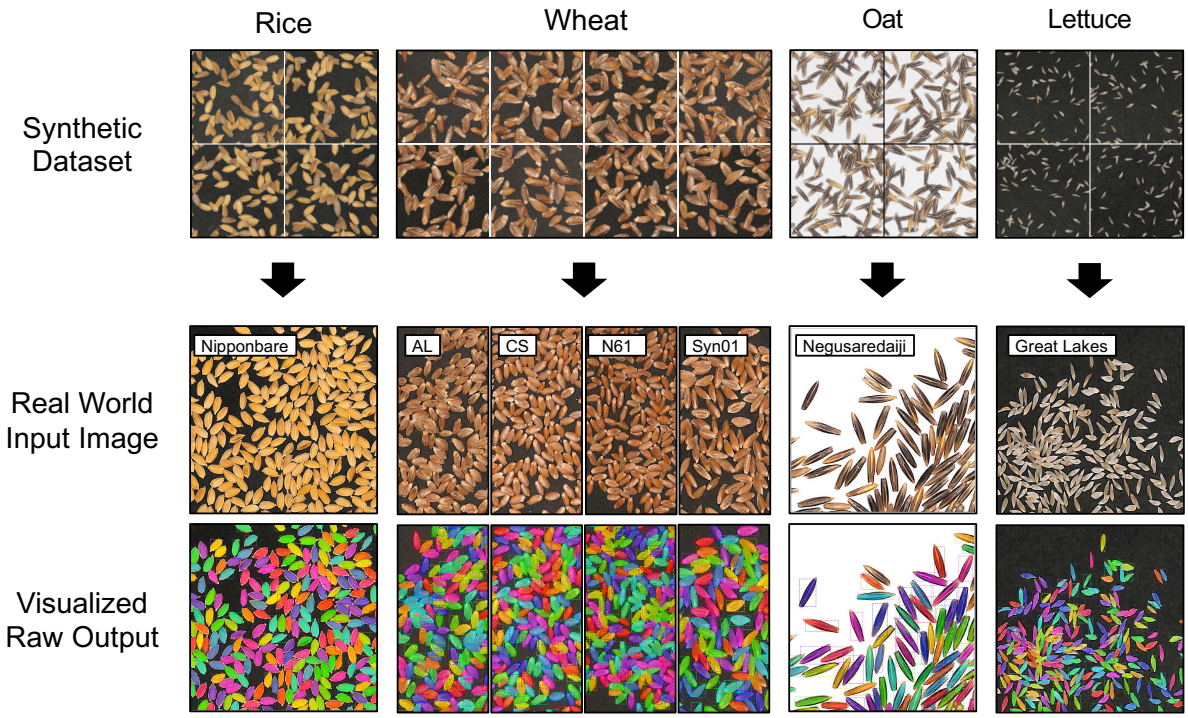

Fig. 7 | Application of our proposed pipeline to seeds of various species. Synthetic data of respective species were generated (top row) and the neural networks were independently trained. The inference result against the real-world input images (middle row) were visualized (bottom row). The name of the cultivar per species is overlaid, respectively.

Discussions

In this research, we showed that utilizing a synthetic dataset can successfully train the instance segmentation neural network to analyze the real-world images of barley seeds. The values obtained from the image analysis pipeline was comparable to that of manual annotation (Figure 4e and Table 1), thus achieving high throughput quantification of seed morphology in various analysis. Moreover, our pipeline requires a limited number of synthesized images to be added to the pool for creating a synthetic dataset. This is labor cost-efficient and practical compared to labeling numerous amounts of images required for deep learning.

To completely understand the use of synthetic data for deep learning, we must have a precise understanding of "*what type of features are critical to represent the real world dataset*". In the case of seed instance segmentation, we presumed that the network must learn the representation that is important for segregating physically touching or overlapping seeds into an individual object. Therefore, in the course of designing synthetic images, we prioritized the dataset to contain numerous patterns of seed orientation, rather than to contain massive patterns of seed textures. Based on the result that the model showed sufficient result against the test dataset (Fig. 4B, S3, and Table 1), it is suggested that our presumption was legitimate to a certain extent. However, because the neural network itself is a *black box*, we cannot discuss more than *ex post facto* reasoning. Recently, there have been challenges to understand the representation of biological

479 context by various interpretation techniques, exemplified in plant disease classification[11,50]. Extending such
480 techniques applicable to an instance segmentation neural network as used in our study will help verify the
481 authenticity of both the synthesized dataset and the trained neural network in future studies.

482 We introduced post-processing to exclude nonintegral mask regions prior to phenotypic analysis (Fig.
483 4a, bottom row and 4c,d). Theoretically, if we can add a category label to the synthetic dataset to determine
484 whether the respective regions are suitable for analysis, the neural network may acquire the classification
485 ability to discriminate such integrity. However, the complexity of synthetic data generation increases, and
486 miss-detected or incomplete mask regions cannot be excluded. We presume heuristic-based post-processing
487 is a simple yet powerful approach. Nonetheless, our outlier removal process is based on the assumption that
488 the seed population is homogeneous. It is important to verify if such filtering is valid against the
489 heterogeneous population. Notably, *SmartGrain* also introduces a post-processing step, involving a repetitive
490 binary dilation and erosion. Those processes were reported to be effective in analyzing the progenies of two
491 cultivars in rice upon QTL analysis[42]. As the post-processing is independent of the neural network in our
492 pipeline, designing and verifying various methods are important for expanding the functionality of the
493 analysis pipeline.

494 The shape and size of seeds (grains) are important agronomic traits that determine the quality and
495 the yield of crops [33]. In recent years, a number of genes have been identified and characterized through genetic
496 approach, accompanied by laborious phenotyping. In previous studies, researchers manually measured the
497 shape and size of seeds, which is time-consuming and erroneous; it restricts the number of seeds that the
498 researcher can analyze. The researchers used to manually select several seeds that seemed to represent the
499 population in a subjective manner and for this reason, small phenotypic differences between genotypes could
500 not be detected. Our pipeline can phenotype a large number of seeds without the need to consider the seed
501 orientation to be sparse in image acquisition and thereby can obtain large amount of data in a short period of
502 time. This allows easy and sensitive detection of both obvious and subtle phenotypic differences between
503 cultivars supported by statistical verification (Fig. 5). This will be a breakthrough in identifying
504 agronomically important genes, especially for molecular genetic research such as genome-wide association
505 study (GWAS), quantitative trait locus (QTL) analysis, and mutant screening. Thus, will open a new path to
506 identify genes that were difficult to isolate by conventional approaches.

507 Moreover, the application of our pipeline is not restricted to barley but can be extended to various
508 crops such as seeds of wheat, rice, oats, and lettuce (Fig. 7). Our results strongly suggest that our approach
509 is applicable to any varieties or species in principle, thus is expected to accelerate research in various fields
510 with similar laborious issues. One example can be an application in characterization of and gene isolation
511 from seeds of wild species. Cultivated lines possess limited genetic diversity due to bottlenecks in the process
512 of domestication and breeding, therefore many researchers face challenges to identify agronomically
513 important genes from wild relatives as a source of genes for improving agronomic traits. As the appearance
514 of the seeds of wild species is generally more diverse than that of cultivated varieties, development of a
515 universal method to measure both traits were difficult. Another example is in understanding the development
516 of seed morphology of wheat. Although the shapes of small florets can be manually quantified from the image

of a scanned spikelet, the automated quantification has not been realized owing to excess non-seed objects (e.g., glume, awn, and rachis) in the image. Applying another domain of randomization for synthesizing a training dataset can be utilized to functionalize a neural network to quantify seed phenotype from such images.

Collectively, we have shown the efficacy of utilizing the synthetic data, based on the concept of *domain randomization* to train the neural network for real-world tasks. Recent technical advances in the computer vision domain have enabled us to generate a realistic image, or even a realistic "virtual reality" environment, thus will provide more possibilities to give solutions to current image analysis involved challenges in the agricultural domain. We envision that a collaboration with plant and computer scientists will open a new point of view for generating a workflow that is valuable for plant phenotyping, leading to a further understanding of the biology of plants through the complete use of machine learning/deep learning methods.

Data Availability

Synthetically generated datasets and real-world test datasets can be obtained from the following Github repository (https://github.com/totti0223/crop_seed_instance_segmentation). Code to reproduce the deployment of the trained Mask R-CNN and multivariate analysis is formatted as IPython notebooks and can also be obtained from the same repository. Other data and information regarding the manuscript are available upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Author Contribution Statement

YT directed and designed the study, wrote the program codes, generated the synthetic test dataset, performed the experiments with assistance from FO, HT, DS, and KT. HT and DS collected and scanned the barley seed images and JI collected wheat images. YT annotated the test dataset. YT, HT, and DS was involved in the conceptualization of this research. YT, FO, and HT wrote the manuscript with assistance from DS, KT, JI, and SO, furthermore with verification of scientific validity from all the coauthors.

Acknowledgments

We thank Ms. Yoko Tomita at Nagoya University for assistance in the labor-intensive annotation to generate a ground truth test dataset. We also thank Dr. Miya Mizutani for a comprehensive discussion and critical reading of the manuscript. The graphical abstract was rendered by Dr. Issey Takahashi who is a member of the Research Promotion Division in ITbM of Nagoya University. Dr. Shunsaku Nishiuchi provided Nipponbare rice seeds used in this study. Dr. Toshiaki Tameshige amplified and provided wheat seeds. Dr.

Kentaro Shimizu amplified and provided wheat Arina seeds and Drs. Shigeo Takumi and Yoshihiro Matsuoka established, amplified and provided synthetic wheat Ldn/KU-2076 (Syn01) seeds. This work was supported by Japan Science and Technology Agency (JST) PRESTO [Grants nos. JPMJPR17O5 (YT) and JPMJPR17O3 (FO)], JST CREST [Grant Number JPMJCR16O4 (HT, DS, SO)], MEXT KAKENHI [Numbers 16H06466 and 16H06464 (HT), 16KT0148 (DS), and 19K05975 (JI)], and JST ALCA [Number JPMJAL1011 (TK)]. All the barley materials are provided by the National BioResource Project (NBRP: Barley).

Figure Legends

Fig. 1 | Overview of the proposed training process of crop seed instance segmentation.

Fig. 2 | Evaluation metrics for object detection accuracy. (a) The intersection-over-union (IoU) definitions for bounding boxes and masks. (b) The average precision (AP) defined as the area under the curves (AUC); shown as the area marked with slanted lines.

Fig. 3 | Data prepared in this study. (a) Images of barley seeds scanned from 20 cultivars. Cultivar names are described in white text in each image. These images were also used as a real-world test dataset in Table 1. (b) Scheme of generating synthetic images. Images are generated by combining actual scanned seed images over the background images on to the virtual canvas. Simultaneously generated ground truth label (Mask) is shown at the bottom in which each seed area is marked with a unique color.

Fig. 4 | Image Analysis pipeline. (a) Summary of the image analysis pipeline. (b) Examples of graphical output of the trained Mask R-CNN on real-world images. Different colors indicate an individual segmented seed region. Note that even if the seeds are overlapping or touching each other, the network can discriminate them as an independent object. (c) Examples of detected candidate regions to be filtered in the post-processing step indicated in red arrows. Black arrowheads indicate the input image boundary. (e) Probability density of the seed areas of the raw output and filtered output. (f) Scatterplot describing the correlation of the seed area that was measured by the pipeline (Inferenced Seed Area) and by manual annotation (Ground Truth Seed Area). Each dot represents the value by a single seed. Black and gray lines indicate the identity line and the 10% error threshold line, respectively. The proportion of the seeds that have lower or higher than the 10% error is also displayed.

Fig. 5 | Analysis of natural variation of barley seed morphology. Whiskerplot overlaid with a swarmplot (colored dot) grouped by barley cultivars. (a) Seed area, (b) seed width, (c) length, and (d) length to width ratio. Diamonds represent outliers. Statistical differences were determined by one-way ANOVA followed by Tukey post hoc analysis. Different letters indicate significant differences ($p<0.05$).

Fig. 6 | Multivariate analysis of barley seed morphology. (a,b) Principal Component analysis (PCA) with morphological parameters of barley seeds. Each point represents the data point of respective seed. The colors correspond to that defined in the color legend displayed below (e). Mean PC1 and PC2 values of each cultivar are plotted as large circle with text annotations in (a). Eigenvectors of each descriptor are drawn as arrows in (b). LWR, length to width ratio; E, eccentricity; L, seed length; PL, seed perimeter length; AS, seed area; W, seed width; S, solidity; CS, seed circularity. (c,d) PCA with elliptic Fourier descriptors (EFD). The colors and point annotated of (c) follows that of (a). Interpolation of the latent space followed by reconstruction of the contours are displayed in (d). (e,f) Latent space visualization of Variational Autoencoders (VAE). The colors and point annotated of (e) follows that of (a). Interpolation of the latent space followed by image generation using the generator of VAE are displayed in (f).

Fig. 7 | Application of our proposed pipeline to seeds of various species. Synthetic data of respective species were generated (top row) and the neural networks were independently trained. The inference result against the real-world input images (middle row) were visualized (bottom row). The name of the cultivar per

species is overlaid, respectively.

Table 1 | Model Evaluation. Table describing the evaluation result of the trained Mask R-CNN raw output. Recall values at the IoU threshold of 50% ($Recall_{50}$) and Average Precision (AP) at the IoU 50% (AP_{50}), 75% (AP_{75}), and the mean value from IoU 50% to 95% with the step size of 5% (AP@ [.5:.95]) are shown.

References

1. Goodfellow I., Bengio, I., Courville, Y. & Aaron. *Deep Learning*. (The MIT Press, 2016).
2. Russakovsky, O. *et al.* ImageNet large scale visual recognition challenge. *Int J Comput Vis* **115,** 211–252 (2015).
3. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun ACM* **60,** 84–90 (2017).
4. Ronneberger, O., Fischer, P. & Brox, T. in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) **9351,** 234–241 (Springer International Publishing, 2015).
5. Shelhamer, E., Long, J. & Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39,** 640–651 (2017).
6. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. in *IEEE Conference on Computer Vision and Pattern Recognition* 580–587 (IEEE, 2014). doi:10.1109/CVPR.2014.81
7. Girshick, R. Fast R-CNN. in *IEEE International Conference on Computer Vision (ICCV)* 1440–1448 (IEEE, 2015). doi:10.1109/ICCV.2015.169
8. He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask R-CNN. in *IEEE International Conference on Computer Vision (ICCV)* 2980–2988 (IEEE, 2017). doi:10.1109/ICCV.2017.322
9. Milioto, A., Lottes, P. & Stachniss, C. REAL-TIME BLOB-WISE SUGAR BEETS VS WEEDS CLASSIFICATION FOR MONITORING FIELDS USING CONVOLUTIONAL NEURAL NETWORKS. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **IV-2/W3,** 41–48 (2017).
10. Mohanty, S. P., Hughes, D. P. & Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **7,** 1419 (2016).
11. Ghosal, S. *et al.* An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. USA* **115,** 4613–4618 (2018).
12. Bresilla, K. *et al.* Single-Shot Convolution Neural Networks for Real-Time Fruit Detection Within the Tree. *Front. Plant Sci.* **10,** 611 (2019).
13. Kamilaris, A. & Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* **147,** 70–90 (2018).
14. Kaneda, Y., Shibata, S. & Mineno, H. Multi-modal sliding window-based support vector regression for predicting plant water stress. *Knowledge-Based Systems* (2017). doi:10.1016/j.knosys.2017.07.028
15. Lin, T.-Y. *et al.* in *European Conference on Computer Vision (ECCV)* (eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) **8693,** 740–755 (Springer International Publishing, 2014).
16. Guo, W. *et al.* Aerial Imagery Analysis - Quantifying Appearance and Number of Sorghum Heads for Applications in Breeding and Agronomy. *Front. Plant Sci.* **9,** 1544 (2018).
17. Jin, X. *et al.* High-Throughput Measurements of Stem Characteristics to Estimate Ear Density and Above-Ground Biomass. *Plant Phenomics* **2019,** 1–10 (2019).

18. Ghosal, S. *et al.* A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* **2019,** 1–14 (2019).
19. Chandra, A. L., Desai, S. V., Balasubramanian, V. N., Ninomiya, S. & Guo, W. Active Learning with Weak Supervision for Cost-Effective Panicle Detection in Cereal Crops. *arXiv* **arXiv:1910.01789,** (2019).
20. Nath, T. *et al.* Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* **14,** 2152–2176 (2019).
21. Varol, G. *et al.* Learning from Synthetic Humans. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 4627–4635 (IEEE, 2017). doi:10.1109/CVPR.2017.492
22. Doersch, C. & Zisserman, A. Sim2real transfer learning for 3D pose estimation: motion to the rescue. in *Annual Conference on Neural Information Processing Systems (NeurIPS)* (2019).
23. Goodfellow, I. J. *et al.* Generative Adversarial Networks. in *Annual Conference on Neural Information Processing Systems (NeurIPS)* (2014).
24. Johnson, M. K. *et al.* Cg2real: improving the realism of computer generated images using a large collection of photographs. *IEEE Trans Vis Comput Graph* **17,** 1273–1285 (2011).
25. Shrivastava, A. *et al.* Learning from Simulated and Unsupervised Images through Adversarial Training. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2242–2251 (IEEE, 2017). doi:10.1109/CVPR.2017.241
26. Peng, X. B., Andrychowicz, M., Zaremba, W. & Abbeel, P. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. in *IEEE International Conference on Robotics and Automation (ICRA)* 3803–3810 (IEEE, 2018). doi:10.1109/ICRA.2018.8460528
27. Tremblay, J. *et al.* Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 1082–10828 (IEEE, 2018). doi:10.1109/CVPRW.2018.00143
28. Isokane, T., Okura, F., Ide, A., Matsushita, Y. & Yagi, Y. Probabilistic Plant Modeling via Multi-view Image-to-Image Translation. in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2906–2915 (IEEE, 2018). doi:10.1109/CVPR.2018.00307
29. Giuffrida, M. V., Scharr, H. & Tsaftaris, S. A. ARIGAN: Synthetic Arabidopsis Plants using Generative Adversarial Network. *BioRxiv* (2017). doi:10.1101/184259
30. Karras, T., Laine, S. & Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2019).
31. Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A. & Stefanovic, D. Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* **11,** 939 (2019).
32. Ward, D., Moghadam, P. & Hudson, N. Deep Leaf Segmentation Using Synthetic Data. in (2018).
33. Shomura, A. *et al.* Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **40,** 1023–1028 (2008).
34. Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. & Lin, H.-X. A QTL for rice grain width and weight

encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39,** 623–630 (2007).

35. Weng, J. *et al.* Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **18,** 1199–1209 (2008).
36. Williams, K., Munkvold, J. & Sorrells, M. Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (Triticum aestivum L.). *Euphytica* **190,** 99–116 (2013).
37. Ohsawa, R., Tsutsumi, T., Uehara, H., Namai, H. & Ninomiya, S. Quantitative evaluation of common buckwheat (Fagopyrum esculentum Moench) kernel shape by elliptic Fourier descriptor. *Euphytica* **101,** 175–183 (1998).
38. Iwata, H., Ebana, K., Uga, Y., Hayashi, T. & Jannink, J.-L. Genome-wide association study of grain shape variation among Oryza sativa L. germplasms based on elliptic Fourier analysis. *Mol. Breeding* **25,** 203–215 (2010).
39. Eguchi, M. & Ninomiya, S. Evaluation of soybean seed shape by elliptic Fourier descriptors. in *World Conference on Agricultural Information AND IT* (2008).
40. Ayoub, M., Symons, J., Edney, J. & Mather, E. QTLs affecting kernel size and shape in a two-rowed by six-rowed barley cross. *Theor. Appl. Genet.* **105,** 237–247 (2002).
41. Herridge, R. P., Day, R. C., Baldwin, S. & Macknight, R. C. Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. *Plant Methods* **7,** 3 (2011).
42. Tanabata, T., Shibaya, T., Hori, K., Ebana, K. & Yano, M. SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol.* **160,** 1871–1880 (2012).
43. Takumi, S., Nishioka, E., Morihiro, H., Kawahara, T. & Matsuoka, Y. Natural variation of morphological traits in wild wheat progenitor Aegilops tauschii Coss. *Breed Sci* **59,** 579–588 (2009).
44. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016). doi:10.1109/CVPR.2016.90
45. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int J Comput Vis* **88,** 303–338 (2010).
46. Kuhl, F. P. & Giardina, C. R. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* **18,** 236–258 (1982).
47. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. in *International Conference on Learning Representations (ICLR)* (2014).
48. Doersch, C. Tutorial on variational autoencoders. *arXiv preprint* **arXiv:1606.05908,** (2016).
49. Li, N., Xu, R., Duan, P. & Li, Y. Control of grain size in rice. *Plant Reprod.* **31,** 237–251 (2018).
50. Toda, Y. & Okura, F. How Convolutional Neural Networks Diagnose Plant Disease. *Plant Phenomics* **Article ID 9237136,** 14 pages (2019).