

ORIGINAL RESEARCH

Automation of leaf counting in maize and sorghum using deep learning

Chenyong Miao^{1,2,3,4}  | Alice Guo¹ | Addie M. Thompson⁵ | Jinliang Yang^{1,2}  | Yufeng Ge⁶ | James C. Schnable^{1,2,3} 

¹ Center for Plant Science Innovation, Univ. of Nebraska-Lincoln, Lincoln, NE, 68588, USA

² Department of Agronomy and Horticulture, Univ. of Nebraska-Lincoln, Lincoln, NE, 68503, USA

³ Quantitative Life Sciences Initiative, Univ. of Nebraska-Lincoln, Lincoln, NE, 68503, USA

⁴ Current affiliation: Bayer Crop Science, Chesterfield, MO, 63017, USA

⁵ Dep. of Plant, Soil and Microbial Sciences, Michigan State Univ., East Lansing, MI, 48824, USA

⁶ Dep. of Biological Systems Engineering, Univ. of Nebraska-Lincoln, Lincoln, NE 68503, USA

Correspondence

James Schnable, Center for Plant Science Innovation, Univ. of Nebraska-Lincoln, Lincoln, NE 68588, USA.
Email: schnable@unl.edu

Assigned to Associate Editor Seth Murray.

Funding information

National Science Foundation Award, Grant/Award Number: OIA-1557417; Biological and Environmental Research Program, Grant/Award Number: DE-SC0020355

Abstract

Leaf number and leaf emergence rate are phenotypes of interest to plant breeders, plant geneticists, and crop modelers. Counting the extant leaves of an individual plant is straightforward even for an untrained individual, but manually tracking changes in leaf numbers for hundreds of individuals across multiple time points is logistically challenging. This study generated a dataset including over 150,000 maize and sorghum images for leaf counting projects. A subset of 17,783 images also includes annotations of the positions of individual leaf tips. With these annotated images, we evaluate two deep learning-based approaches for automated leaf counting: the first based on counting-by-regression from whole image analysis and a second based on counting-by-detection. Both approaches can achieve root of mean square error (RMSE) smaller than one leaf, only moderately inferior to the RMSE between human annotators of between 0.57 and 0.73 leaves. The counting-by-regression approach based on convolutional neural networks (CNNs) exhibited lower accuracy and increased bias for plants with extreme leaf numbers which are underrepresented in this dataset. The counting-by-detection approach based on Faster R-CNNs (region based convolutional neural networks) object detection models achieve near human performance for plants where all leaf tips are visible. The annotated image data and model performance metrics generated as part of this study provide large scale resources for the comparison and improvement of algorithms for leaf counting from image data in grain crops.

Abbreviations: CNNs, convolutional neural networks; DAP, days after planting; MSE, mean square error; QTL, quantitative trait locus; R-CNN, region based convolutional neural networks; RMSE, root mean square error.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *The Plant Phenome Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy and Crop Science Society of America

1 | INTRODUCTION

Plant development, unlike that of many vertebrates, incorporates the potential for substantial plasticity in organ numbers. Different varieties of the same species may produce different numbers of leaves prior to flowering (Allen et al., 1973; Kim

et al., 2017). Two genetically identical clones may produce different numbers of leaves prior to flowering when grown in different environments (Brooking et al., 1995; Tollenaar & Hunter, 1983). Variation in both the total number of leaves produced, the rate at which leaves are produced and the phyllochron, or the rate in thermal time at which leaves appear are associated with a wide range of agricultural plant properties. For example, in maize, the number of leaves is correlated with plant height, flowering time, and moisture at harvest (Allen et al., 1973). In switchgrass, a larger number of leaves, indicative of a long duration of vegetative growth prior to reproductive transition, is highly correlated to the biomass yield (Van Esbroeck et al., 1997). In potato, the number of green leaves has been used as indicator to determine drought resistant and susceptible varieties (Deblonde & Ledent, 2001). In perennial ryegrass, the number of leaves was used as a criterion for determining defoliation time (Fulkerson & Slack, 1994). Many crop growth models also incorporate both leaf emergence rate and mature leaf number as crucial parameters which must be determined when these models are fit to new crops or new varieties of existing crops (Hammer et al., 2010; He et al., 2012; Lizaso et al., 2003; Truong et al., 2017). As a result, counting the number of leaves on a given plant is a task frequently undertaken by plant biologists, plant breeders, and agronomists.

However, manually counting leaves can be time consuming and logistically challenging. This is particularly true for large studies where hundreds of genotypes are being evaluated, each with many replicates (e.g., QTL mapping, genome-wide association study, breeding trial evaluations, etc.). Leaves produced early in development can senesce and be undetectable by maturity (Figure 1a and b). As a result, multiple measurements throughout development are necessary not only to track leaf appearance rates but also to have an accurate final count of the total leaves produced. These challenges have limited the number of studies identifying genetic variants influencing the number of leaves or leaf appearance rates (Cao et al., 2016; Hasan et al., 2016; Li et al., 2016; Méndez-Vigo et al., 2010). The same challenges also hinder the development and deployment of genotype-specific crop growth models (He et al., 2012; Technow et al., 2015; Yang et al., 2021). These logistical hurdles have motivated the exploration of automated or semi-automated solutions to plant phenotyping (Araus & Cairns, 2014; Fernandez et al., 2017; Furbank & Tester, 2011; Zhou et al., 2019). Advances in automated plant imaging have produced large plant image datasets (Feldman et al., 2017; Liang et al., 2017; Zhang et al., 2017). However, developing and deploying the most effective algorithms to allow the automated counting or scoring of the number of leaves from raw images of plants is an ongoing challenge within the plant phenotyping community.

A range of deep learning-based approaches have been proposed in recurring plant phenotyping workshops, conferences,

Core Ideas

- Automated leaf counting in maize was previously limited by annotated training data.
- Two deep learning approaches both achieve near human accuracy in counting leaves.
- There is an ongoing need to extend image datasets with phenotypically extreme plants.

and research articles (Aich & Stavness, 2017; Dobrescu et al., 2019; Giuffrida et al., 2018; Madec et al., 2019). Many of these approaches were developed and validated using image and ground truth data from rosette plants (Giuffrida et al., 2016; Kuznichov et al., 2019). Rosette plants are defined by a shared leaf architecture where leaves are arranged in a compact circular pattern along the stem with substantially reduced stem elongation between leaves (Figure 1c). The focus on rosette plants is likely a result of the popularity of *A. thaliana*, a plant genetic model which belongs to the Brassicaceae, a family of plants including many species which grow in a rosette fashion. For plants exhibiting a rosette leaf architecture, it is frequently possible to collect images of the majority of non-occluded leaves from a single top-down view in a short period of time using relatively simple imaging systems (Bell & Dee, 2019; Giuffrida et al., 2016; Minervini et al., 2016; Scharr et al., 2014). Giuffrida and co-workers proposed and validated a unified deep network which could predict leaf numbers across multiple species which share a rosette-style leaf architecture (Giuffrida et al., 2018). For the majority of species of monocots, leaves are sessile with the blade of the leaf directly joining the stem. Maize, wheat, and rice, three monocot crops in the grass family – Poaceae – have elongated internodes between sessile leaves which emerge following alternating phyllotaxy. These crop species also provide roughly half of all calories consumed by humans around the globe (Tester & Langridge, 2010). Leaf counting has been shown to be practical in sorghum when using depth cameras to reconstruct 3D models of individual plants (McCormick et al., 2016; Xiang et al., 2019). However, only limited work has been conducted in leaf counting from 2D images of grain crops such as maize and sorghum (Pound et al., 2017; Zhou et al., 2020), potentially because of the lack of large, annotated, public training datasets for them. Unlike rosette plants, counting leaves from top-down images is not practical in these species (Figure 1d). Instead, the leaves of grasses are generally counted when viewed from the side, which reduces the incidence of occlusion.

In this study, we assembled a maize image dataset using a non-invasive high throughput plant phenotyping platform, consisting of a large greenhouse where plants are grown on

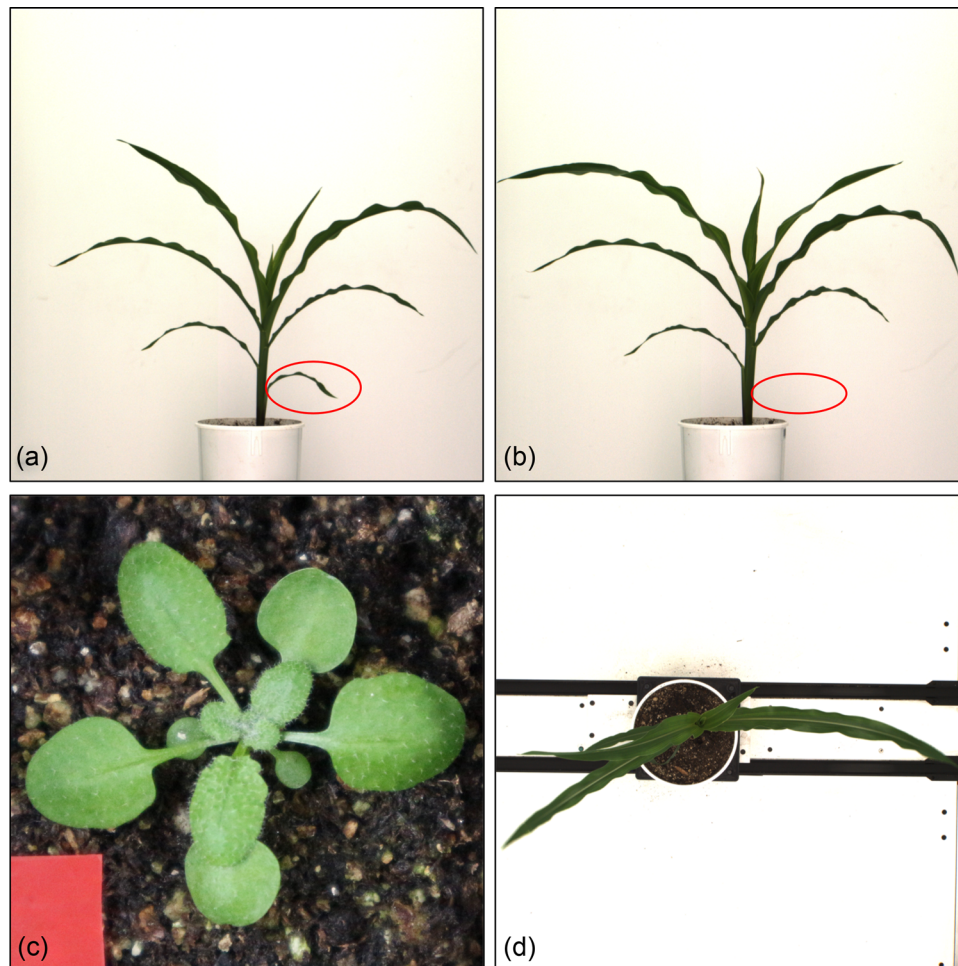


FIGURE 1 Challenges in tracking the number of leaves in plants. Leaves produced early in development can senesce and die over the lifetime of the plant, becoming undetectable at maturity. (a) A maize plant at 32 d after planting (DAP) with eight visible leaves. (b) The same maize plant at 34 DAP. The top three leaves have all lengthened significantly, while the bottom leaf (indicated in red ovals) has senesced and is no longer visible, resulting in only seven visible leaves for this older plant. (c) Illustration of the rosette leaf architecture exhibited by an *Arabidopsis* plant in a top-down photo with all leaves visible. (d) Top-down photo of the same maize plant shown in panels a and b. Leaf occlusion in Panel d is more severe than in panel c

conveyor belts and an imaging system where plants are photographed from multiple angles using a range of cameras with different imaging modalities (Gaillard et al., 2020; Ge et al., 2016). A set of 122,290 RGB images of maize plants, consisting of 10 images per plant per day from 923 unique plants representing 342 unique maize genotypes, were collected as part of two experiments conducted in 2018 and 2019. A parallel sorghum dataset of 27,770 images, consisting of five images per plant per day, was collected from 343 unique plants representing 295 unique sorghum genotypes. The position and classification of individual leaf tips in a subset of these images were annotated by humans using the Zooniverse citizen science/crowd sourcing platform. Leaf counting approaches employing convolutional neural networks (CNNs) trained for regression and Faster R-CNNs (region based convolutional neural networks) object detection models trained for leaf tip detection were evaluated. The release of raw data

including leaf tip position and condition annotations, as well as the initial performance estimates of different models provide the community a benchmark to test against when seeking future improvements in automated leaf counting for grasses and grain crops.

2 | MATERIALS AND METHODS

2.1 | Image acquisition and annotation

Maize image data was taken from two separate experiments conducted at the University of Nebraska-Lincoln's Greenhouse Innovation Center (Lat., 40.83; Long., -96.69) in 2018 and 2019. In the first experiment, 256 lines from the Buckler-Goodman inbred association panel (Flint-Garcia et al., 2005) were planted on 23 July 2018. In the second experiment, 255

lines from the Buckler-Goodman inbred association panel and 33 lines from the Germplasm Enhancement of Maize project (Pollak, 2003) were planted on 25 Mar. 2019. Maize kernels were sown in 2.4 gallon pots with Fafard germination mix supplemented with 1 cup (236 mL) of Osmocote plus and 1 tablespoon (15 mL) of Micromax Micronutrients per 2.8 cubic feet (80 L) of soil. The target photoperiod was 14:10 with supplementary light provided by light-emitting diode (LED) growth lamps from 07:00 to 21:00 each day. The target temperature of the growth facility was between 24–26°C. After growing in the greenhouse for 28 d in the first experiment and 38 d in the second experiment, all the plants were moved to a high throughput phenotyping facility equipped with a plant conveying system and different kinds of imaging chambers (Gaillard et al., 2020; Ge et al., 2016). The conveyor belt transferred each pot to a set of imaging chambers once every 2 or 3 d for imaging, and to a watering station each day. At the watering station plants were weighed and watered back to a target weight to ensure all plants were growing in a good condition. In the imaging chamber, each plant was rotated 36 degrees nine times, and an image was captured for each stop. Therefore, a total of 10 photos with 0°, 36°, 72°, 108°, 144°, 180°, 216°, 252°, 288°, and 324° viewing angles were captured for each plant on the same day (Figure S1a). Plants were photographed from 28 to 73 days after planting (DAP) in the 2018 experiment, and 38 to 70 DAP in the 2019 experiment. These imaging time periods covered both late vegetative growth and flowering stages for the majority of genotypes in the population. A total of 122,290 maize images were acquired from the two experiments and each image had a resolution of 2,454 × 2,056 pixels. All the images described above are now freely accessible, see Data Availability Statement.

Sorghum image data was taken from a previously published image dataset (Miao et al., 2020). This dataset consisted of 27,770 images collected from 343 unique sorghum plants representing 295 inbred lines from the sorghum association panel (Casa et al., 2008). Images were photographed from 26 July 26 to 31 Aug. 2017 over a period of 37 d spanning vegetative and reproductive development for the majority of genotypes in the population. In contrast to 10 viewing angles in the maize dataset, on each imaging date sorghum plants were photographed from only five different viewing angles including 0°, 36°, 72°, 108°, and 144°.

As images with different viewing angles for the same plant captured on the same day share the same leaf number, only the image with the 108° viewing angle was uploaded to the crowd sourcing platform Zooniverse for annotation. All the leaves in each image were grouped to three classes: (a) leaves with visible healthy leaf tip; (b) leaves with visible but damaged or cut leaf tip; (c) leaves with leaf tips that are not visible due to occlusions. For the first two classes where leaves had visible leaf tips, annotators were asked to click corresponding leaf tip positions. For leaves whose leaf tips were

not visible, annotators were asked to click anywhere in the leaf (Supplemental Figure S2). The total number of leaves in each image was calculated as the sum of the number of annotations for each of these three classes. We estimate it took approximately 132 total person hours to complete all annotations for 12,229 maize images and 5,554 sorghum images. Annotation for this project was conducted primarily by paid undergraduate researchers. All the annotation results including the coordinates of tip positions and the leaf numbers have been deposited in FigShare (See *Data and Code Availability Statement*).

2.2 | Image processing

All images were preprocessed by being cropped to a maximum bounding rectangle containing all plant pixels as shown in Figure S1b. Removing uninformative parts in the image such as imaging chamber frames and the bottom pot increased plant portion in the image, which made it easier for deep learning models to learn plant related features. A pixel was classified as a plant pixel if the pixel value in the green channel was larger than 130. This cutoff was set after manually checking the performances of various threshold values on images across different genotypes and time points. Among the 10 images collected from a single maize plant on the same day, the two images where the distance between the furthest left and furthest right plant pixel was greatest were included in the ‘best views’ category and the other eight images were treated as ‘other views’ for downstream analyses (Figure S1b).

The preprocessed images were split to training, validation, and testing images at the genotype level, rather than at the plant or image level. This practice was adopted to avoid issues of data leakage which can occur when very similar images appear in both training and testing processes (Samala et al., 2020). All images from a set of 50 maize inbred lines generated as part of this study were set aside as testing data for evaluating the final accuracy of models. Image data for the remaining 292 lines were split for five-fold cross-validation with five sets of training and validation data. The first two sets contain all images from 233 inbred lines in training dataset and all images from 59 maize inbred lines in the validation dataset. The remaining three sets contain all images from 234 and 58 inbred lines in training and validation dataset respectively.

2.3 | Training and evaluation of CNNs and faster R-CNNs

Two sets of maize CNNs were trained in this study: CNNs trained on all 10 viewing images, and CNNs trained only on the best viewing images (Figure S1). In each case, five

models were trained through five-fold cross-validation. All the CNNs were trained using transfer learning by fine tuning a pre-trained Resnet18 model (He et al., 2016) to fit customized maize images for leaf counting tasks. The pre-trained Resnet18 model adopted in this study was trained on the 1,000-class ImageNet dataset (Russakovsky et al., 2015) with 1,000 output features in the last fully connected layer. In order to adapt the model to the task of leaf counting, the number of output features was set to 1 representing one estimated leaf number for each input image. All the preprocessed images were resized to 224×224 pixels as the input image size during training evaluating CNNs. In addition, two types of image transformation approaches were applied on the input training images. The first transformation approach changes input images by horizontally flipping images randomly with a given probability of 0.5. The second transformation approach randomly changes the brightness, contrast and saturation of the input images. The mean square error (MSE) was used as the loss function considering leaf counting as a regression problem. The stochastic gradient descent optimizer was used with the 'learning rate' set to 0.001 and 'momentum' set to 0.9. The maximum epoch was set to 500 and the early stopping was applied during the training with the 'patience' argument set to 50. This setting means that training will stop early if the overall MSE fails to improve for 50 continuous epochs. In this case the model with the smallest loss value will be saved and used for downstream evaluations.

A set of sorghum CNN models were also trained through transfer learning based on the CNNs trained on the best view maize image dataset. However, a different transfer learning strategy was adopted. Instead of fine tuning all the pre-trained weights, only the weights in the last fully connected layer were retrained using the best viewing sorghum images. All the sorghum images went through the same preprocessing steps as the maize images, including the removal of uninformative portions of each image. Sorghum images were also split on the genotype level to avoid data leakage, with all best view images from 250 sorghum lines used for training and all best viewing images for the remaining 44 lines used for testing. The hyperparameters for retraining sorghum models were the same as in training maize models.

A pre-trained Faster R-CNN object detection model (Ren et al., 2015) with Resnet50 architecture as the backbone was adopted to detect both healthy and damaged leaf tips in maize RGB images. The backbone was pre-trained on the ImageNet dataset and used as fixed feature extractor in Faster R-CNN. The Faster R-CNN model was pre-trained on the COCO 2017 instance dataset, which includes 91 "stuff categories" including person, vehicles, fruits, and animals. To fit our leaf tip dataset, the output number of classes was set to three representing healthy leaf tip, damaged leaf tip, and background classes. Five object detection models were trained through five-fold cross-validation with the same splitting strategy used

for CNNs, but only the directly annotated images were considered because the leaf tip positions were only annotated in these images. Each leaf tip coordinate indicated by (x, y) in a directly annotated image were converted to a bounding box coordinate (x-15, y-15, x+15, y+15) with a 30×30 square centered around the leaf tip. The images used for both training and testing are the original plant photos without cropping as shown in Figure S1a. The stochastic gradient descent optimizer was used during training with the 'learning rate' set to 0.001, 'momentum' set to 0.9 and 'weight decay' set to 0.0005. The maximum epoch was set to 200 and early stopping was applied when both the average precision and average recall were converged. The object detection model was evaluated in the directly annotated images from the 50 genotypes set aside as held out test data. The number of leaves estimated by the model was the sum of all the detected leaf tips with a classification score higher than 0.5.

Three metrics were adopted in this study to evaluate model performances: the square of the correlation coefficient (r^2), the root mean square error (RMSE), and the agreement rate defined by the proportion of perfect predictions. A perfect prediction was called when the difference between ground truth and prediction was less than 0.5 so rounding would produce the same predicted integer value for leaves as reported ground truth. The mean and standard deviation of each metric were calculated from the five-fold cross-validation for the eventual model comparisons. The code for training and evaluating both CNN models and Faster R-CNN models implemented using the PyTorch framework has been deposited on GitHub (See *Data and Code Availability Statement*).

3 | RESULTS & DISCUSSION

3.1 | Assessment of the leaf annotation dataset

Ten images for a maize plant were photographed from 10 different viewing angles each day (Supplemental Figure S1a). Ground truth leaf numbers should not vary among images of the same plant taken on the same day, however apparent leaf numbers may vary as a result of differences in occlusion among images taken from different perspectives. The viewing angles of 108° and 288° were identified as the most straightforward to annotate, with leaves tending to have comparatively fewer occlusions or crossovers among leaves than other viewing angles. The images taken from the viewing angle of 108° were selected for the manual annotation. Annotators were asked to divide leaves into three categories: (a) leaves where the leaf tip was not visible (e.g. out of frame or occluded by another leaf); (b) leaves where the leaf tip was visible but mechanically damaged or cut; and (c) leaves where the leaf tip was visible and healthy (Supplemental Figure S2). The total

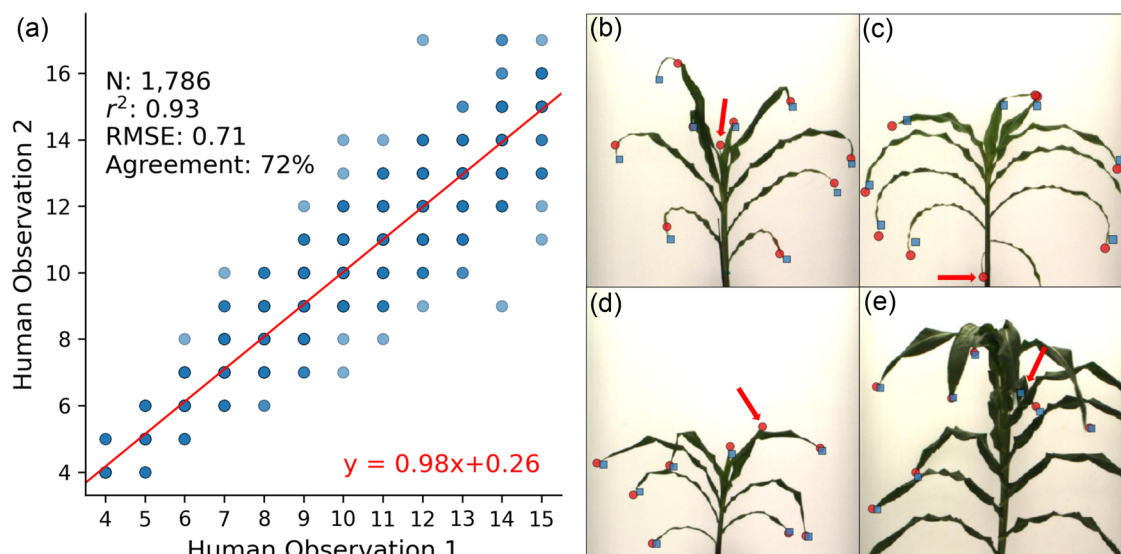


FIGURE 2 Human annotation errors on leaf counting: (a) Concordance in leaf number annotation between independent annotations of the same image by different observers. Dark blue points represent observations from larger numbers of distinct images while light blue points represent observations from smaller numbers of distinct images. The total tested images, square of the correlation coefficient (r^2), root of mean square error (RMSE), and the agreement rate are shown in the upper left corner. The best linear regression line is shown in red and the equation for that line is given in the bottom right. (b) An example of an image where two observers disagreed on the identification of a new leaf emerging from the whorl of a maize plant (red arrow). Selected positions of leaf tips for one observer are indicated with red circles and the other with blue squares. (c) An example of an image where two observers disagreed about the annotation of a senescing and potentially damaged leaf (red arrow). (d) An example of an image where one observer identified a partially occluded or overlapping leaf which was missed by a second observer (red arrow). (e) An example of a plant where phyllotaxy has shifted for upper leaves, making it difficult to annotate from any one side viewing angle. Images for panels b–e were cropped to aid viewing at a reduced figure size. Annotations which were made outside of the cropped portion of the image included in the figure are not shown

number of leaves for each annotated image was calculated by summing up all three leaf categories. The calculated number of leaves for the annotated image was also assigned to the other nine images with different viewing angles but captured for the same plant on the same day. Details of the annotation process and the definition of each of the three categories are provided in the methods section. Annotated leaf numbers for maize plants in this dataset ranged from 3 to 20 with reduced representation at both extremes (Supplemental Figure S3A). All images and annotations are being released, but in the analyses below, we focus only on images of plants with between 4 and 15 leaves with each category containing at least 3,000 images of plants (Supplemental Figure S3A).

A subset of 1,768 images were shown to two or more human observers. In 72% of cases where an image was seen by two observers, both observers agreed on the total number of leaves (RMSE = 0.71) (Figure 2a). Inconsistencies between observers were more common in images with larger total numbers of leaves. Manual follow-ups using the positional information recorded as part of the annotation process found that observer disagreements were primarily explained by three issues: (a) differences when a new leaf, emerging from the top of the whorl, was included in the leaf count for the plant as shown in Figure 2b; (b) differences whether senescing or severely damaged leaves at the base of the stalk were

still included in the leaf count for the plant (Figure 2c); and (c) partially occluded leaves which were missed by one observer but not another as shown in Figures 2d and e. Among assignments to the three different classes of leaves – healthy leaf tip, damaged leaf tip, and non-visible leaf tip – the one error we observed frequently was that at least some human annotators tended to annotate leaves which extended out of frame as healthy leaf tips rather than the non-visible leaf tips.

Data leakage can occur when an annotated image dataset includes individual images sharing many common features and common labels and these images are randomly partitioned into training and testing datasets. Data leakage can result in CNNs with overly optimistic prediction accuracy, i.e., the estimated results from CNNs are impossible to achieve in real world scenarios (Samala et al., 2020). At least two potential sources of data leakage exist in this leaf number/leaf position dataset. Firstly, this dataset includes images of the same plant collected only several days apart. These images will tend to look much more similar to each other than random pairs of images, and tend to contain plants with close or identical numbers of leaves. Secondly, variation in leaf numbers is under genetic control (Cao et al., 2016; Li et al., 2016). The dataset employed here includes images of multiple plants from the same genotype. Images of different, genetically identical individual plants will tend to be more similar to each other than

images from random unrelated individual plants, and will be more likely to share close or identical numbers of leaves at the same stage in development. To avoid problems of data leakage, we elected to conduct data splitting at the genotype level, rather than splitting at the level of individual images or individual plants when preparing data for training and testing (see Methods).

3.2 | Counting leaves by regression

Multiple deep learning based approaches have been applied to the task of counting the number of leaves in plant images (Dobrescu et al., 2019; Pound et al., 2017). Among them, the CNNs that treat the leaf counting task as a regression problem are most widely deployed for the direct estimation of leaf numbers. In this study, two different sets of training data were used when training CNNs to predict the correct number of leaves based on a single image of a maize plant. The first set of CNNs were trained using maize images with all 10 viewing angles in the training dataset. We acknowledge that even for human annotators, accurately counting leaves in images taken from some perspectives (e.g. viewing angles parallel to the plane of phyllotaxy) will be more challenging than others (e.g. viewing angles perpendicular to the plane of phyllotaxy) (Supplemental Figure S1b). Therefore, a second set of CNNs were trained only using those images with the best views in the training data. The best views were the two images where the distances between the left most and right most plant pixel were greatest among all 10 viewing angles, whereas the rest of eight images were treated as other views (Supplemental Figure S1b). While viewing angles 108° and 288° produced the most images classified as “best,” the best images were selected on a plant by plant basis and for some individual plants different angles were selected as best (Supplemental Table S1). The trained CNNs were then evaluated on the images, including all 10 viewing angles from the 50 genotypes set aside as held out test data. However, model performance was different when evaluated based on the testing images with different viewing angles.

For models trained using data from all 10 viewing angles, performance was significantly higher when evaluated using only the best viewing images than when evaluated using data from the other viewing images in the test data (Table 1, Supplemental Table S1). Based on the greater prediction accuracy of CNNs when evaluating using only best viewing angle images, the decision was made to train a second set of CNNs employing only best viewing angle images as training data. The performance of these new models was comparable to the set of models trained using data from all 10 viewing angles when evaluating them on the other viewing images from the held out test set. However, the new CNN models achieved a significantly greater correlation between predic-

TABLE 1 Performance of regression convolutional neural networks (CNNs) on leaf counting in maize

Model	Testing data	r^2	Agreement rate	RMSE
CNN1 ^a	Other views ^c	0.79 ± 0.04	0.33 ± 0.02	1.28 ± 0.11
CNN1	Best views ^d	0.84 ± 0.03	0.37 ± 0.03	1.11 ± 0.09
CNN2 ^b	Other views	0.78 ± 0.02	0.35 ± 0.02	1.31 ± 0.12
CNN2	Best views	0.87 ± 0.01	0.45 ± 0.01	0.96 ± 0.04

Note. RMSE, root of mean square error.

^aCNN1: CNNs trained using all 10 viewing angle images collected from each plant.

^bCNN2: CNNs trained using only the two best viewing angle images.

^cBest views: The images taken from the two viewing angles where the distance between the furthest left and furthest right plant pixel was greatest.

^dOther views: The remaining eight images taken from other viewing angles.

tion and ground truth ($r^2 = .87$; $p = .011$; t -test), higher agreement rate (45%; $p = .031$; t -test), and lower RMSE (0.96 leaves; $p = .001$; t -test) when tested on the best viewing images from the held out test set (Table 1). The training time required for models trained using best viewing images was also much shorter than the training time required for models trained using all 10 viewing images as the number of best viewing images was only 1/5th of all viewing images.

Outcomes for the single highest performing model, one of the five trained using only the best viewing images, were r^2 of .88, an agreement rate of 46%, and an RMSE of 0.92 (Figure 3a). For most numbers of ground truth leaves, the predictions of the CNN were not obviously biased towards predicting too many or too few leaves. However, the model did tend to overestimate the number of leaves for plant images with only four leaves, while tending to underestimate the number of leaves for plant images with 14 or 15 leaves (Figure 3a). These three ground truth values (4 leaves, 14 leaves, or 15 leaves) were the least common among leaf number groups included in our training dataset (Supplemental Figure S3A). When excluding these three leaf number classes, the RMSE was decreased to 0.88 from 0.92 leaves, although this was still higher than the RMSE observed among human annotators (0.72 leaves) among the same leaf number classes (Supplemental Figures S4, 2A).

3.3 | Application of maize models to sorghum

Maize and sorghum share similar plant architectures prior to flowering. Untrained observers can struggle to tell the difference between vegetative stage maize and sorghum plants, and even trained observers will struggle at the seedling stage. We

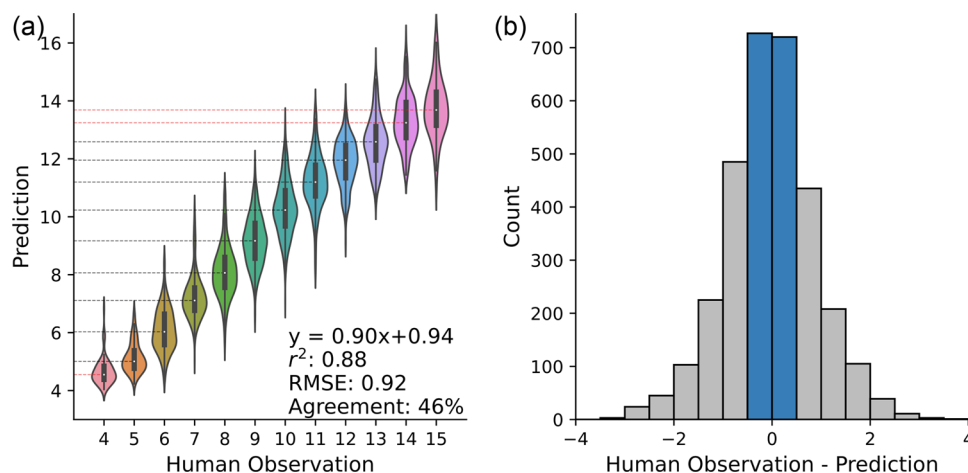


FIGURE 3 Distribution of predictions and errors for counting-by-regression approach in maize: (a) Relationship between predicted leaf number and human annotated number of leaves among the best viewing images for the single best performing convolutional neural networks trained using the best viewing images. The equation of the best linear regression, square of the correlation coefficient (r^2), root of mean square error (RMSE), and the agreement rate are indicated in the bottom right corner. Dashed lines indicate the median predicted leaf number values for images with each different number of leaves identified by human annotators. Cases where the median value for images of plants with a given number of leaves diverges from the annotated number of leaves by >0.5 leaves are indicated in red, all other cases indicated in black. (b) Distribution of error (difference between human observation and predicted number of leaves) for the same single best performing model shown in panel a. The absolute difference of less than 0.5 indicated by the blue bars was used for the calculation of agreement rate

TABLE 2 Performance of CNNs on leaf counting in sorghum images

How maize model was used	r^2	Agreement rate	RMSE
Directly	0.54 ± 0.04	0.35 ± 0.03	1.14 ± 0.05
Transfer learning	0.64 ± 0.04	0.39 ± 0.02	1.06 ± 0.06

Note. RMSE, root of mean square error.

generated annotations for a published set of sorghum images which were collected using the same automated imaging facility employed to generate the maize images described above (Miao et al., 2020). In the sorghum data, only five images were collected from five side viewing angles instead of 10 in the maize data. The sorghum dataset annotated in this study exhibited a narrower distribution of leaf numbers than the annotated maize images (Supplemental Figure S3B). At least 500 images were available of sorghum plants with between 6 and 13 leaves, and images of plants with these numbers of leaves were employed below. Models trained using best viewing maize images and directly applied to sorghum exhibited a decline in performance relative to the performance of those same models in maize across all three metrics (Table 2). The r^2 between annotated and predicted leaf numbers decreased to 0.54 from 0.87, the agreement rate decreased to 35% from 45%, and the RMSE increased to 1.14 from 0.96. The decline in r^2 is likely explained in part by the reduced variability in leaf numbers among plants in the sorghum image dataset rel-

ative to the maize dataset (Supplemental Figure S3B). However, this difference cannot explain the decline in the agreement rate or the increase in the RMSE. A second approach was evaluated where a subset of the annotated sorghum data was employed to retrain the last fully connected layer of the same adopted maize models. This transfer learning based approach resulted in small but significant improvements in all three evaluations of model performance. The mean RMSE of the new models through transfer learning was only 1/10th of a leaf less accurate in sorghum than the original models were in maize (Table 2).

3.4 | Counting leaves by detection

Human beings possess and employ at least two mechanisms for counting objects: subitizing, the near instantaneous recognition of object numbers for set of one through four, and a slower counting based mechanism for larger quantities (Trick & Pylyshyn, 1994). The regression-based approach described above is, in some ways, more analogous to subitizing as it seeks to determine the complete number of leaves present in the image at once. The counting-by-regression approach is sensitive to the number of images representing each potential number of leaves a plant can have, with biased performance for rarer and more extreme leaf numbers (Figure 3a, Supplemental Figure S4). An alternative approach to leaf counting is to implement counting-by-detection models (Buzzy et al., 2020; Xu et al., 2018) built on top of object detection frameworks (Redmon et al., 2016; Ren et al., 2015). Compared to a

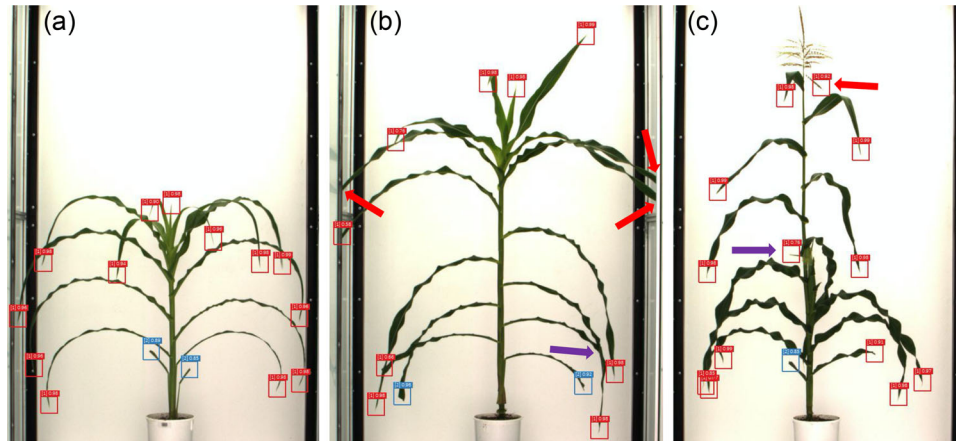


FIGURE 4 Representative leaf tip detection results. (a) Example of a plant with perfect detection, all annotated leaf tips are correctly identified. Undamaged leaf tips detected by the model indicated in red, damaged leaf tips detected by the model indicated in blue. The values above each bounding box indicate the classification confidence score assigned by the faster region based convolutional neural networks model. (b) An example of a plant where multiple leaves were missed by the object detection model either because the leaf tip itself is out of frame (red arrows) or because the leaf tip is occluded behind another leaf of the same plant from this viewing angle (purple arrow). (c) Example of a plant where two erroneous leaf tips were detected by the object detection model. One erroneous detection is centered on a tassel branch (red arrow), although the majority of tassel branches are not misclassified as leaves. The second erroneous detection is centered on a husk leaf associated with a developing ear on this particular maize plant (purple arrow)

single number as the label for each training image in building CNNs, Object detection models are trained using annotations of bounding boxes around objects of interest.

In order to create bounding boxes centering on leaf tips as training data for leaf tip detection, the annotated coordinates of each leaf tip were expanded by 15 pixels in each direction resulting in a 30×30 pixel bounding box centered on each annotated leaf tip. The resulting bounding box annotations were employed to train a set of Faster R-CNN object detection models through five-fold cross-validation using the same splitting strategy employed above in maize. Each Faster R-CNN model employed the Resnet50 architecture as a backbone and was pre-trained on the ImageNet dataset (He et al., 2016; Ren et al., 2015; Russakovsky et al., 2015). All object detection models were trained on the bounding box annotations with intact and damaged leaf tips as two distinct objects. During the training process, model performance was evaluated using standardized metrics: Average precision and recall for the purpose of stopping training process early to avoid over fitting. Final trained models were evaluated using the same metrics used in evaluating CNNs by comparing the leaf numbers annotated by humans and predicted numbers of leaves by summing all detected leaf tips by object detections models. The testing images are in the subset of the directly annotated images for the 50 held out genotypes.

Unlike the CNN based counting-by-regression approach where the number of leaves in an image could be generalized to other images taken of the same plant at the same time, training object detection required information on the position of each leaf tip. As a result, training data consisting solely

TABLE 3 Performance of Faster R-CNNs on leaf counting in maize

Testing data	r^2	Agreement rate	RMSE
Directly annotated	0.78 ± 0.01	0.43 ± 0.01	1.33 ± 0.04
Directly annotated - all tips visible	0.88 ± 0.01	0.56 ± 0.03	1.00 ± 0.06

Note. R-CNNs, region based convolutional neural networks; RMSE, root of mean square error.

of directly annotated images collected from genotypes in the training set. Evaluation of performance was conducted using two subsets of the held out test data. The first subset consisted of all directly annotated images collected from plants belonging to genotypes in the held out test dataset. The second included only the subset of the images in set one where human annotators reported that all leaf tips were in frame and not occluded. When evaluated across all directly annotated images in the held out test set, object detection models achieved an average r^2 of .78, an agreement rate of 43% and an RMSE of 1.33 (Table 3). Manual examination of images where the predicted number of leaves was less than the annotated number of leaves revealed that the majority of missing detections resulted from leaves which were either out of frame or occluded by other leaves (Figure 4b). The latter problem – occlusion of leaves by other leaves – is likely unavoidable in 2D representations of 3D plants. However, leaf tips which

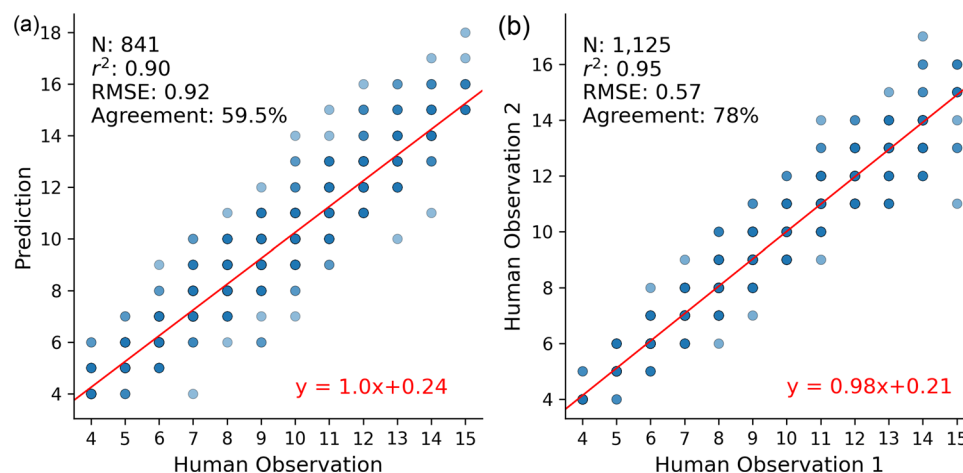


FIGURE 5 Comparison of the performance of the best object detection model on images without leaf tip occlusions to concordance between human annotators. (a) Relationship between predicted leaf number and human annotated leaf number among directly annotated images without leaf tip occlusions for the single best performing object detection model. (b) Relationship between leaf number independently annotated by two different human annotators for the images without leaf tip occlusions. The number of tested images, square of the correlation coefficient (r^2), root of mean square error (RMSE), and the agreement rate are indicated in the upper left corner. The best linear regression line and the corresponding equation were indicated in red color

are missed because the plant is wider than the imaging chamber could be addressed through changes to image acquisition (e.g. make a bigger imaging chamber for bigger plants like mature maize). Manual examination of images where the predicted number of leaves was greater than the annotated number of leaves indicated that the two most common sources of these errors were tassel branches being misclassified as leaves, and the tips of ear husk leaves being detected as leaves. Husk leaves are homologous to normal leaves in maize but traditionally not included when counting leaves in that species (Figure 4c).

When trained object detection models were evaluated solely on the subset of images where all leaf tips were reported to be visible by human annotators, the performance of these models increased substantially with a mean r^2 , agreement rate, and RMSE of 0.88, 56%, and 1.0 respectively (Table 3). The best Faster R-CNN model as part of the five-fold cross-validation process had an r^2 of .90, and agreement rate of 59.5% and an RMSE of 0.92, which is competitive compared to human performances on the same type of testing images in maize (Figure 5).

4 | CONCLUSIONS AND POTENTIAL FOR FURTHER PERFORMANCE IMPROVEMENT

For the counting-by-regression approach, a CNN trained to output a floating point prediction of the number of leaves in a given image performed better on images where the plane of phyllotaxy was perpendicular to the viewing angle of the

camera (Supplemental Figure S1; Table 1). Similar challenges were found among human annotations when the plane of phyllotaxy was parallel to the viewing angle of the camera (Figure 2e). We found a crude metric, the distance from the left most to the right most plant pixel in an image was an effective approach to identify these images that are most amenable to leaf counting. CNN models trained with all images or only these two best viewing images per plant both exhibit better performance on these best viewing images, matching the performance they exhibited when assessed using only the subset of images which were directly annotated (Table 1, Supplemental Table S1). There are two straightforward approaches which could be taken to further improve model performance: New and more optimized model architectures, and further improvements in the size and/or accuracy of labelled training data. The first depends on more advanced models proposed by machine learning experts. Records for leaf counting accuracy of rosette plants continue to be broken as new and better performing models and approaches are proposed and tested (Aich & Stavness, 2017; Dobrescu et al., 2020; Ubbens et al., 2018). The second depends on the data availability and training of human annotators. We identified a number of common inconsistencies between different annotators (Figure 2b and c) which could be incorporated into training for subsequent rounds of annotation of images from maize, sorghum and other grain crop species.

However, the counting-by-regression approach to predicting the number of leaves from CNNs faces another challenge which is somewhat harder to address. For many quantitative genetic and breeding applications the extremes of any phenotypic distribution will be of greatest interest, yet a key

drawback of the CNN based counting-by-regression approach is that performance declined for plants with the smallest or largest numbers of leaves in our dataset (Figure 3a, Supplemental Figure S4). From a computational perspective, this issue can be easily addressed by collecting more images of plants with these numbers of leaves and including them in the dataset. Biologically this approach presents additional challenges. Images of plants with fewer leaves can be collected by starting imaging earlier in development. However, after accounting for the senescence of juvenile leaves, only a small subset of maize genotypes in our study will carry 14+ leaves at any given time point. Additional replication of these genotypes could be included but this would create a bias in the training data with high leaf number plants belonging to a small number of genotypes which exhibit additional genetically controlled differences in appearance. Alternatively, data could be generated using photoperiod sensitive tropical maize lines grown under long day conditions which prolong vegetative development, resulting in the production of greater numbers of leaves (Stevenson & Goodman, 1972). Yet this approach could again mean that genetically and phenotypically distinct lines could represent different portions of the potential range of the total number of leaves observable in maize. Although temperate maize is less photoperiod sensitive than tropical maize, growing existing genotypes under differing photoperiod regimes is another potential alternative approach to generating a greater diversity of plant images for phenotypic extremes of leaf number. In silico approaches to augmenting training image datasets are another potential option which would be explored. Synthetic training data generated using procedural models has been evaluated for leaf counting in both arabidopsis (Ubbens et al., 2018) and maize (Miao et al., 2019). Approaches to generating synthetic training data using generative models such as Variational Autoencoders (VAEs) and the Generative Adversarial Networks (GANs) also show promise (Giuffrida et al., 2017; Goodfellow et al., 2014; Kingma et al., 2013; Shete et al., 2020). Another approach to mitigate this issue would be to experiment with other loss functions. In the future studies, sample weighting could be introduced into the loss function, i.e., assigning a higher weight in the loss function for the extremely small and large leaf number classes (Ling et al., 2011).

In contrast to the CNN-based counting-by-regression approach, the performance of the counting-by-detection approach using leaf tip detection models was consistent across plants with different numbers of leaves. The performance of this approach was modestly poorer than the CNN predictions across all directly annotated images, but achieved higher agreement rates with human annotations and comparable r^2 and RMSE for images without hidden leaf tips (Table 1; Table 3). For images without any hidden leaf tips, the divergence between the annotation of different humans looking at the same images as quantified by RMSE was only

1/3 of a leaf better than the divergence between the best object detection-based leaf counting model and human annotation (Figure 5). False negative leaf detections primarily resulted from non-visible leaf tips, while false positive detections primarily resulted from husk leaves or tassel branches. Including ear shoot/ear and tassel as additional features in subsequent annotations of this or other image datasets would allow these organs to be included as additional objects when training object detection models. This would presumably reduce false positive detections. Leaf occlusions as a result of maize plants which were too large to fit entirely within the field of view of the imaging chamber are a specific constraint of the facility employed in this study (Ge et al., 2016), and are best addressed at the data acquisition level rather than data analysis. However, leaf overlap and self occlusion will be an inherent problem whenever a 3D plant is represented as a 2D image. A similar object detection model was recently employed to estimate leaf number by detecting whole leaves in maize images (Zhou et al., 2020). However, this approach also suffers from issues with occlusion. Performance was reported using mean $\pm$ standard deviation of absolute difference between the predicted and ground truth leaf numbers instead of RMSE. For comparison purposes we calculated the same metric for our best performing leaf tip detection model. The best leaf tip detection model from this study achieved a value of (1.31 ± 1.60) compared to value of (1.60 ± 1.63) reported by Zhou and coworkers. However, this comparison should be treated with caution given the different images used for model evaluations and differences in training data size. By making all of our raw images and annotations publicly available (see *Data Availability Statement*) we hope to reduce the barriers to head to head comparison of different approaches to leaf counting performance in maize, and thus stimulate future improvements in performance. The frequency of self occlusions is likely to vary among genotypes as both leaf curvature and phyllotaxy are under partial genetic control in maize (Ford et al., 2008; Giulini et al., 2004). Leaf occlusion is therefore likely to introduce heritable genetic error (Liang et al., 2017) if counting-by-detection is employed to phenotype maize plants in quantitative genetic or plant breeding projects. Transitions to methods which incorporate 3D information (Gaillard et al., 2020; McCormick et al., 2016; Xiang et al., 2019) and/or the development of approaches to integrate information on leaf tips detected in multiple images of the same plant taken from different angles will likely be ultimately needed. A previous study proposed an approach to detecting leaf tips in wheat using a sliding window-based localization strategy which also provided good performance in leaf tip detection (Pound et al., 2017). It would be intriguing to compare the performance of this method to the regression-based and Faster R-CNN bounding box-based approaches evaluated here. The sliding window-based leaf localization approach is, however,

substantially more computationally intensive, requiring approximately 2 min per image, relative to <1 s per image with the Faster R-CNN model. In addition to the deep learning-based approaches, conventional image processing-based approaches such plant skeletonization or contour mapping can also be potentially applied for leaf counting (Bashyam et al., 2016; Gehan et al., 2017). However, those approaches are largely intolerant of leaf overlap or cross over of leaves within an image which can limit the robustness and applications for leaf counting.

The availability of leaf counting models which can match or exceed human performance would provide significant benefits to plant breeders, plant geneticists, and crop modelers. The number of leaves is employed as a morphological indicator in predicting flowering time and yield (Allen et al., 1973; Van Esbroeck et al., 1997). Many plant geneticists seek to identify the genomic regions and genes controlling natural variation in leaf number among members of the same species. For example, a study was able to map QTL controlling variation in the number of leaves in maize from manually scoring 866 inbred lines from a maize-teosinte BC₂S₃ population with five plants of each inbred scored in each of 2 yr, requiring counting the leaves of at least 8,660 individual maize plants (Li et al., 2016). Automated leaf scoring would enable similar studies to incorporate more genotypes, include additional replication, or conduct a study of equivalent scale using fewer resources. Actual automated scoring of leaf numbers, combined with existing high throughput phenotyping facilities which image plants throughout development, would make it feasible to collect time series data for determining total leaf number, tracking rates of leaf appearance and monitoring how these traits vary across different genotypes, or the same genotype grown under different environmental conditions. Studies of environmental stresses and genotype by environment interactions in crop species can translate to agricultural applications most directly when the studies themselves are conducted in the field. The approach described in this study for analyzing data from individual plants photographed against an uncluttered background in a chamber with constant illumination. A substantial gap remains between these models and models which could conduct accurate leaf counting under field conditions with varying illumination. Here, rosette plants again possess a significant structural advantage as field data is both easier to acquire when imaging from the top down relative to the side. Top down imaging tends to place the plant against a non-uniform but contrasting background of soil (Cao et al., 2016). Automated approaches for collecting side view images in the field are improving (Fernandez et al., 2017; Young et al., 2019); side views tend to result in plants being photographed against a non-contrasting background of other plants in the same field.

ACKNOWLEDGMENTS

This work was supported by a University of Nebraska Agricultural Research Division seed grant to JCS, a National Science Foundation Award (OIA-1557417) to JCS, JY, and YG, and by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research Program under award number DE-SC0020355 to JCS and AMT. This project was completed utilizing the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. We thank Leighton Wheeler, Nate Pester, Isaac Stevens, Elijah Frost, Thomas Hoban, Sierra Conway, Elliot Kadrofske, Emma Chrzanowski, Carolina Freitas, Lou Townsend, and Logan Duryee for the work of image annotations in summer 2020. We thank Rebecca Roston and Jeremy Brown for gifting the Arabidopsis photo used in Figure 1.

AUTHOR CONTRIBUTIONS

Chenyong Miao: Data curation; Formal analysis; Methodology; Validation; Visualization; Writing-original draft. Alice Guo: Data curation. Addie Thompson: Data curation; Methodology; Writing-review & editing. Yufeng Ge: Data curation; Funding acquisition; Writing-review & editing. Jinliang Yang: Data curation; Funding acquisition; Writing-review & editing. James Schnable: Conceptualization; Data curation; Funding acquisition; Project administration; Supervision; Writing-review & editing.

DATA AND CODE AVAILABILITY

All the maize and sorghum images as well as the corresponding leaf number and leaf tip position annotations used in this study have been deposited in Figshare: 1) maize leaf tip pixel annotations and images (<https://doi.org/10.6084/m9.figshare.13056524.v1>); 2) maize leaf number images and annotations (<https://doi.org/10.6084/m9.figshare.13056512.v1>); 3) sorghum leaf number images and annotations (<https://doi.org/10.6084/m9.figshare.13056548.v1>); 4) sorghum leaf tip images and annotations. (<https://doi.org/10.6084/m9.figshare.13056527.v1>). The scripts and source code related to training and evaluating CNN and Faster R-CNN models have been deposited on GitHub (<https://github.com/freemao/MaizeLeafCounting>).

CONFLICTS OF INTEREST

The authors declare no competing interests.

ORCID

Chenyong Miao  <https://orcid.org/0000-0002-0904-3707>
Jinliang Yang  <https://orcid.org/0000-0002-0999-3518>

James C. Schnable  <https://orcid.org/0000-0001-6739-5527>

REFERENCES

- Aich, S., & Stavness, I. (2017). Leaf counting with deep convolutional and deconvolutional networks. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2080–2089.
- Allen, J. R., McKee, G. W., & McGahen, J. H. (1973). Leaf number and maturity in hybrid corn 1. *Agronomy Journal*, 65(2), 233–235. <https://doi.org/10.2134/agronj1973.00021962006500020014x>
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in plant science*, 19(1), 52–61. <https://doi.org/10.1016/j.tplants.2013.09.008>
- Bashyam, S., (2016). *Determination of plant architecture and component phenotyping based time-lapse image analysis* [Master's thesis, University of Nebraska-Lincoln]. <https://digitalcommons.unl.edu/computerscidiss/120/>
- Bell, J., & Dee, H. M. (2019). *Leaf segmentation through the classification of edges*. arXiv. <https://arxiv.org/pdf/1904.03124.pdf>.
- Brooking, I., Jamieson, P., & Porter, J. (1995). The influence of daylength on final leaf number in spring wheat. *Field crops research*, 41(3), 155–165. [https://doi.org/10.1016/0378-4290\(95\)00014-H](https://doi.org/10.1016/0378-4290(95)00014-H)
- Buzzy, M., Thesma, V., Davoodi, M., & Mohammadpour Velni, J. (2020). Real-time plant leaf counting using deep object detection networks. *Sensors*, 20(23), 6896. <https://doi.org/10.3390/s20236896>
- Cao, X.-W., Cui, H.-M., Li, J., Xiong, A.-S., Hou, X.-L., & Li, Y. (2016). Heritability and gene effects for tiller number and leaf number in non-heading chinese cabbage using joint segregation analysis. *Scientia Horticulturae*, 203, 199–206. <https://doi.org/10.1016/j.scienta.2016.03.018>
- Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., Franks, C. D., & Kresovich, S. (2008). Community resources and strategies for association mapping in sorghum. *Crop science*, 48(1), 30–40. <https://doi.org/10.2135/cropsci2007.02.0080>
- Deblonde, P., & Ledent, J.-F. (2001). Effects of moderate drought conditions on green leaf number, stem height, leaf length and tuber yield of potato cultivars. *European Journal of Agronomy*, 14(1), 31–41. [https://doi.org/10.1016/S1161-0301\(00\)00081-2](https://doi.org/10.1016/S1161-0301(00)00081-2)
- Dobrescu, A., Giuffrida, M. V., & Tsafaris, S. A. (2020). Doing more with less: A multitask deep learning approach in plant phenotyping. *Frontiers in Plant Science*, 11.
- Dobrescu, A., Valerio Giuffrida, M., & Tsafaris, S. A. (2019). Understanding deep neural networks for regression in leaf counting. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 4321–4329.
- Feldman, M. J., Paul, R. E., Banan, D., Barrett, J. F., Sebastian, J., Yee, M. - C., Jiang, H., Lipka, A. E., Brutnell, T. P., Dinnyen, J. R., et al. (2017). Time dependent genetic analysis links field and controlled environment phenotypes in the model c4 grass setaria. *PLoS Genetics*, 13(6), e1006841. <https://doi.org/10.1371/journal.pgen.1006841>
- Fernandez, M. G. S., Bao, Y., Tang, L., & Schnable, P. S. (2017). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant physiology*, 174(4), 2008–2022. <https://doi.org/10.1104/pp.17.00707>
- Flint-Garcia, S. A., Thuillet, A. - C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *The Plant Journal*, 44(6), 1054–1064. <https://doi.org/10.1111/j.1365-313X.2005.02591.x>
- Ford, E. D., Cocke, A., Horton, L., Fellner, M., & Van Volkenburgh, E. (2008). Estimation, variation and importance of leaf curvature in zea mays hybrids. *Agricultural and Forest Meteorology*, 148(10), 1598–1610. <https://doi.org/10.1016/j.agrformet.2008.05.015>
- Fulkerson, W., & Slack, K. (1994). Leaf number as a criterion for determining defoliation time for *Lolium perenne*, 1. effect of water-soluble carbohydrates and senescence. *Grass and Forage Science*, 49(4), 373–377. <https://doi.org/10.1111/j.1365-2494.1994.tb02013.x>
- Furbank, R. T., & Tester, M. (2011). Phenomics—technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 16(12), 635–644. <https://doi.org/10.1016/j.tplants.2011.09.005>
- Gaillard, M., Miao, C., Schnable, J. C., & Benes, B. (2020). Voxel carving-based 3d reconstruction of sorghum identifies genetic determinants of light interception efficiency. *Plant Direct*, 4(10), e00255. <https://doi.org/10.1002/pld3.255>
- Ge, Y., Bai, G., Stoerger, V., & Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput rgb and hyperspectral imaging. *Computers and Electronics in Agriculture*, 127, 625–632. <https://doi.org/10.1016/j.compag.2016.07.028>
- Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., & Hoyer, J. S., (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ*, 5, p.e4088. <https://doi.org/10.7717/peerj.4088>
- Giuffrida, M. V., Doerner, P., & Tsafaris, S. A. (2018). Pheno-deep counter: A unified and versatile deep learning architecture for leaf counting. *The Plant Journal*, 96(4), 880–890. <https://doi.org/10.1111/tbj.14064>
- Giuffrida, M. V., Scharr, H., & Tsafaris, S. A. (2017). Arigan: Synthetic arabidopsis plants using generative adversarial network. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2064–2071.
- Giuffrida, M. V., Minervini, M., & Tsafaris, S. A. (2016). Learning to count leaves in rosette plants. In *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*, pages 1.1–1.13. BMVA press.
- Giulini, A., Wang, J., & Jackson, D. (2004). Control of phyllotaxy by the cytokinin-inducible response regulator homologue abphyll1. *Nature*, 430(7003), 1031–1034. <https://doi.org/10.1038/nature02778>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y., (2014). *Generative adversarial networks*. arXiv. <https://arxiv.org/abs/1406.2661>.
- Hammer, G. L., van Oosterom, E., McLean, G., Chapman, S. C., Broad, I., Harlan, P., & Muchow, R. C. (2010). Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. *Journal of Experimental Botany*, 61(8), 2185–2202. <https://doi.org/10.1093/jxb/erq095>
- Hasan, Y., Briggs, W., Matschegewski, C., Ordon, F., Stützel, H., Zetzsche, H., Groen, S., & Uptmoor, R. (2016). Quantitative trait loci controlling leaf appearance and curd initiation of cauliflower in relation to temperature. *Theoretical and Applied Genetics*, 129(7), 1273–1288. <https://doi.org/10.1007/s00122-016-2702-6>
- He, J., Le Gouis, J., Stratonovitch, P., Allard, V., Gaju, O., Heumez, E., Orford, S., Griffiths, S., Snape, J. W., Foulkes, M. (2012). Simulation of environmental and genotypic variations of final leaf number and

- anthesis date for wheat. *European Journal of Agronomy*, 42, 22–33. <https://doi.org/10.1016/j.eja.2011.11.002>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Kim, W., Kim, H. - E., Lee, A. - R., Jun, A. R., Jung, M. G., Ahn, J. H., & Lee, J. - H. (2017). Base-pair opening dynamics of primary mir156a using nmr elucidates structural determinants important for its processing level and leaf number phenotype in arabidopsis. *Nucleic Acids Research*, 45(2), 875–885. <https://doi.org/10.1093/nar/gkw747>
- Kingma, D. P., & Welling, M. (2013). *Auto-encoding variational bayes*. arXiv. <https://arxiv.org/abs/1312.6114>.
- Kuznichov, D., Zvirin, A., Honen, Y., & Kimmel, R. (2019). Data augmentation for leaf segmentation and counting tasks in rosette plants. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2580–2589.
- Li, D., Wang, X., Zhang, X., Chen, Q., Xu, G., Xu, D., Wang, C., Liang, Y., Wu, L., Huang, C. (2016). The genetic architecture of leaf number and its genetic relationship to flowering time in maize. *New Phytologist*, 210(1), 256–268. <https://doi.org/10.1111/nph.13765>
- Liang, Z., Pandey, P., Stoerger, V., Xu, Y., Qiu, Y., Ge, Y., & Schnable, J. C. (2017). Conventional and hyperspectral time-series imaging of maize lines widely used in field trials. *GigaScience*, 7(2), gix117.
- Ling, C. X., & Sheng, V. S., (2011). Class imbalance problem. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning*. Springer. https://doi.org/10.1007/978-0-387-30164-8_110
- Lizaso, J. I., Batchelor, W. D., & Westgate, M. E. (2003). A leaf area model to simulate cultivar-specific expansion and senescence of maize leaves. *Field Crops Research*, 80(1), 1–17. [https://doi.org/10.1016/S0378-4290\(02\)00151-X](https://doi.org/10.1016/S0378-4290(02)00151-X)
- Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., Heritier, E., & Baret, F. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agricultural and Forest Meteorology*, 264, 225–234. <https://doi.org/10.1016/j.agrformet.2018.10.013>
- McCormick, R. F., Truong, S. K., & Mullet, J. E. (2016). 3D sorghum reconstructions from depth images identify qtl regulating shoot architecture. *Plant Physiology*, 172(2), 823–834.
- Méndez-Vigo, B., de Andrés, T., Ramiro, M., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2010). Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in *Arabidopsis thaliana*. *Journal of Experimental Botany*, 61(6), 1611–1623. <https://doi.org/10.1093/jxb/erq032>
- Miao, C., Hoban, T. P., Pages, A., Xu, Z., Rodene, E., Ubbens, J., Stavness, I., Yang, J., & Schnable, J. C. (2019). Simulated plant images improve maize leaf counting accuracy. *BioRxiv*, 706994.
- Miao, C., Xu, Y., Liu, S., Schnable, P. S., & Schnable, J. C. (2020). Increased power and accuracy of causal locus identification in time series genome-wide association in sorghum. *Plant Physiology*, 183(4), 1898–1909. <https://doi.org/10.1104/pp.20.00277>
- Minervini, M., Fischbach, A., Scharr, H., & Tsafaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, 81, 80–89. <https://doi.org/10.1016/j.patrec.2015.10.013>
- Pollak, L. M. (2003). The history and success of the public-private project on germplasm enhancement of maize (GEM). *Advances in Agronomy*, 78, 45–87. [https://doi.org/10.1016/S0065-2113\(02\)78002-4](https://doi.org/10.1016/S0065-2113(02)78002-4)
- Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., Bulat, A., Tzimiropoulos, G., Wells, D. M., Murchie, E. H., et al. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, 6(10), gix083. <https://doi.org/10.1093/gigascience/gix083>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, 91–99.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). ImageNet large scale visual recognition challenge. arXiv. <https://arxiv.org/abs/1409.0575>.
- Samala, R. K., Chan, H. - P., Hadjiiski, L., & Koneru, S. (2020). Hazards of data leakage in machine learning: A study on classification of breast cancer using deep neural networks. *Medical Imaging 2020: Computer-Aided Diagnosis*, 11314.
- Scharr, H., Minervini, M., Fischbach, A., & Tsafaris, S. A. (2014). Annotated image datasets of rosette plants. *European Conference on Computer Vision. Zürich, Suisse*, 6–12.
- Shete, S., Srinivasan, S., & Gonsalves, T. A. (2020). TasselGAN: An application of the generative adversarial model for creating field-based maize tassel data. *Plant Phenomics*, 2020, 8309605. <https://doi.org/10.34133/2020/8309605>
- Stevenson, J., & Goodman, M. (1972). Ecology of exotic races of maize. I. Leaf number and tillering of 16 races under four temperatures and two photoperiods 1. *Crop Science*, 12(6), 864–868. <https://doi.org/10.2135/cropsci1972.0011183X001200060045x>
- Technow, F., Messina, C. D., Totir, L. R., & Cooper, M. (2015). Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS One*, 10(6), e0130855. <https://doi.org/10.1371/journal.pone.0130855>
- Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science*, 327(5967), 818–822. <https://doi.org/10.1126/science.1183700>
- Tollenaar, M., & Hunter, R. (1983). A photoperiod and temperature sensitive period for leaf number of maize 1. *Crop Science*, 23(3), 457–460. <https://doi.org/10.2135/cropsci1983.0011183X002300030004x>
- Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? a limited-capacity preattentive stage in vision. *Psychological Review*, 101(1), 80. <https://doi.org/10.1037/0033-295X.101.1.80>
- Truong, S. K., McCormick, R. F., & Mullet, J. E. (2017). Bioenergy sorghum crop model predicts VPD-limited transpiration traits enhance biomass yield in water-limited environments. *Frontiers in Plant Science*, 8, 335. <https://doi.org/10.3389/fpls.2017.00335>
- Ubbens, J., Cieslak, M., Prusinkiewicz, P., & Stavness, I. (2018). The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods*, 14(1), 6. <https://doi.org/10.1186/s13007-018-0273-z>
- Van Esbroeck, G., Hussey, M., & Sanderson, M. (1997). Leaf appearance rate and final leaf number of switchgrass cultivars. *Crop Science*, 37(3), 864–870. <https://doi.org/10.2135/cropsci1997.0011183X003700030028x>
- Xiang, L., Bao, Y., Tang, L., Ortiz, D., & Salas-Fernandez, M. G. (2019). Automated morphological traits extraction for sorghum plants via 3d

- point cloud data analysis. *Computers and Electronics in Agriculture*, 162, 951–961. <https://doi.org/10.1016/j.compag.2019.05.043>
- Xu, L., Li, Y., Sun, Y., Song, L., & Jin, S. (2018). Leaf instance segmentation and counting based on deep object detection and segmentation networks. *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 180–185.
- Yang, K. W., Chapman, S., Carpenter, N., Hammer, G., McLean, G., Zheng, B., Chen, Y., Delp, E., Masjedi, A., Crawford, M., Ebert, D., Habib, A., Thompson, A., Weil, C., & Tuinstra, M. R. (2021). Integrating crop growth models with remote sensing for predicting biomass yield of sorghum. *in silico Plants*, 1(3), diab001 <https://doi.org/10.1093/insilicoplants/diab001>
- Young, S. N., Kayacan, E., & Peschel, J. M. (2019). Design and field evaluation of a ground robot for high-throughput phenotyping of energy sorghum. *Precision Agriculture*, 20(4), 697–722. <https://doi.org/10.1007/s11119-018-9601-6>
- Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., Wang, K., Xiao, Y., Chen, G., Liu, Q., et al. (2017). High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant physiology*, 173(3), 1554–1564. <https://doi.org/10.1104/pp.16.01516>
- Zhou, S., Chai, X., Yang, Z., Wang, H., Yang, C., & Sun, T. (2020). Maize-PAS: Automated maize phenotyping analysis software using deep learning. *Research Square, Preprint*.
- Zhou, Y., Srinivasan, S., Mirnezami, S. V., Kusmec, A., Fu, Q., Attigala, L., Fernandez, M. G. S., Ganapathysubramanian, B., & Schnable, P. S. (2019). Semiautomated feature extraction from rgb images for sorghum panicle architecture GWAS. *Plant Physiology*, 179(1), 24–37. <https://doi.org/10.1104/pp.18.00974>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Miao, C., Guo, A., Thompson, A. M., Yang, J., Ge, Y., & Schnable, J. C. Automation of leaf counting in maize and sorghum using deep learning. *Plant Phenome J.* 2021;4:e20022. <https://doi.org/10.1002/ppj2.20022>