Received: 25 January 2024 | Accepted: 4 September 2024

DOI: 10.1002/ppj2.70002

The Plant Phenome Journal OPEN ACCESS

ORIGINAL ARTICLE

Sorghum segmentation and leaf counting using in silico trained deep neural model

Ian Ostermann[1] | **Bedrich Benes**[1] | **Mathieu Gaillard**[1] | **Bosheng Li**[1] | **Jensina Davis**[2,3] | **Ryleigh Grove**[2,3] | **Nikee Shrestha**[2,3] | **Michael C. Tross**[2,3] | **James C. Schnable**[2,3]

[1]Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

[2]Center for Plant Science Innovation and Department of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, Nebraska, USA

[3]Complex Biosystems Graduate Program, University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Correspondence
Bedrich Benes, Department of Computer Science, Purdue University, West Lafayette, IN, USA.
Email: bbenes@purdue.edu

Present address Ian Ostermann, Department of Computer Science, Purdue University, West Lafayette, IN, USA.

Assigned to Associate Editor Dr. Weizhen Liu.

Funding information
National Science Foundation, Grant/Award Numbers: 2412928, 2412930, 2417510; Division of Biological Infrastructure; Division of Integrative Organismal Systems; National Institute of Food and Agriculture, Grant/Award Numbers: 2021-67021-35329, 2024-67013-42449, 2024-67021-42879

Abstract

This paper introduces a novel deep neural model for segmenting and tracking the number of leaves in sorghum plants in phenotyping facilities. Our algorithm inputs a sequence of images of a sorghum plant and outputs the segmented images and the number of leaves. The key novelty of our approach is in training the deep neural model. Manual annotations are tedious, and we have developed a procedural three-dimensional (3D) sorghum model that provides detailed geometry and texture to generate photorealistic 3D models. The overall shape of the sorghum leaf geometry is determined by its skeleton, and it is detailed by a procedural model that varies its curvature, width, length, and overall shape. The color is determined by using a Monte Carlo path tracer. We mimic the illumination of the phenotyping facility and use reflectance and transmittance on sorghum surfaces to determine the color of the leaves. The 3D procedural model allows us to generate photorealistic and segmented images that we use to train a deep neural model. Our segmentation provides a mean intersection over union score of 0.51, resulting in leaf counting accuracy within the 90% confidence interval for the human counts.

1 | INTRODUCTION

Understanding the physiological and genetic basis of both how plants grow and develop and how they adapt to environmental stresses and perturbations across time has driven the development of highly automated plant phenotyping facilities around the globe (Chen et al., 2014; Fahlgren et al., 2015; Ge et al., 2016; Yang et al., 2014; Zhang et al., 2017). These fully or semi-automated facilities host dozens to

Abbreviations: 2D, two dimensional; 3D, three dimensional; AI, artificial intelligence; RGB, red green blue; BRDF, bi-directional reflectance distribution function; BTDF, bi-directional transmittance distribution function; CPU, central processing unit; DL, deep learning; GPU, graphics processing unit; IoU, intersection over union; RMSE, root mean squared error; SAM, Segment Anything Model; SOTA, state-of-the-art.

thousands of plants in controlled environments and incorporate the capability to image these plants on regular intervals, often from multiple angles and multiple imaging modalities. The inclusion of high-resolution red green blue (RGB) (visible light) cameras is nearly universal. Still, different facilities may incorporate thermal infrared, chlorophyll fluorescence, hyperspectral reflectance cameras, and other data collection instruments such as gravimetric measurement of water use or depth information from the time of flight cameras (Li et al., 2021). A single medium-scale automated plant phenotyping facility can generate multiple terabytes of image data each week for processing, including multiple photos taken from different perspectives of each of hundreds of plants.

The ideal output of the visual data processing would be a timestamped, highly detailed three-dimensional (3D) plant geometry with color information available for each point on the 3D plant geometry's surface. Such data would allow automatic measurements of plant surface area, volume, branching angles, length of leaves, etc. The temporal data would allow for capturing the plant's growth and allow tests of the correlation of different plant phenotypes with different experimental treatments. Despite the tremendous progress in the field of full 3D reconstruction in phenotyping facilities (McCormick et al., 2016; Thapa et al., 2018; Wu et al., 2022), 3D models reconstructed from small numbers of images per plant are often not detailed enough to be used for deep learning (DL) models. Higher accuracy has been achieved but requires capturing 80–160 images per plant from different viewing angles (Li et al., 2022). From the viewpoint of signal theory, plant two-dimensional (2D) RGB images, the most common output of phenotyping facilities, are considered noisy signals with high variability. Various state-of-the-art algorithms can reconstruct certain parts (usually low-frequency data), such as the stem of broad leaves. However, such reconstruction often works only from particular views. This is because the plant leaves are often very thin, which makes them difficult to capture in side views. While the lighting conditions are controlled, the plant's complex geometry causes complex shading that adds to the uncertainty during the reconstruction, especially for color calibration.

While completely reconstructing a 3D model of a plant is complex and data intensive, recent algorithms in artificial intelligence (AI) and specifically in deep learning (DL) allowed for answering targeted and well-articulated questions, such as the branching angles of sorghum (Tross et al., 2021), leaf count (Gaillard et al., 2023; Giuffrida et al., 2018; Miao et al., 2021; Ubbens et al., 2020), detection and quantification of panicles (Ghosal et al., 2019; Lin & Guo, 2020), stomata classification and quantification (Zhang et al., 2021), genotypic prediction based on phenotype (Zhang et al., 2022), quantification of plant disease (Johnson et al., 2021; Nagasubramanian et al., 2019; Stewart et al., 2019), biomass distribution (Gage et al., 2019), identifying grape varieties (Pereira et al., 2019), estimating flowering time in wheat (Wang et al., 2019), quantifying wheat spikes and estimating rice density for yield estimation (Hasan et al., 2018; Liu et al., 2020), identifying root and shoot features (Pound et al., 2017), identifying and quantifying plant stress (Ghosal et al., 2018; Singh et al., 2018), quantifying seeds in soybean pods (Uzal et al., 2018), and plant growth modeling (Shibata et al., 2020). These approaches are promising because they answer specific research questions and do not require the entire plant 3D reconstruction.

Core Ideas

- An automatic method to determine the number of leaves of sorghum for phenotyping facilities.
- An automatic method to segment leaves of sorghum for phenotyping facilities.
- Procedural model to generate three-dimensional sorghum geometry and appearance to bridge the sim-to-real gap in deep learning.

Leaf count is a crucial component of many agronomic traits, making it a valuable indicator for various characteristics. It serves as a dependable proxy for different traits, facilitating early selection in plant breeding programs. For instance, studies have demonstrated that variations in temperature influence leaf growth rates, leaf numbers, and net photosynthetic rates in different races of maize, underscoring the importance of leaf count as a criterion for resistance against temperature stress and overall plant health (Duncan & Hesketh, 1968). Moreover, leaf number is correlated with flowering time, as distinct varieties of the same species tend to produce varying numbers of leaves before entering the reproductive stage. This indicates that leaf count has a substantial genetic relationship with flowering time in maize (Miao et al., 2021). Notably, genes regulating leaf number and flowering time have been found to be in pleiotropy, further supporting the interconnectedness of these traits. Leaf number also plays a pivotal role in the selection process for biomass yield in various plants (Ciancolini et al., 2013; Plénet et al., 2000). Additionally, it is widely used to determine the growth stages of various plants such as sorghum (Ottman et al., 2001; Boyes et al., 2001), and many others. Leaf count has previously been estimated at a given point in time from 2D images for maize (Miao et al., 2021) and rosette-shaped plants (Giuffrida et al., 2018; Miao et al., 2021) using methods such as direct regression over the input image or explicitly identifying and counting features such as leaf tips. These existing methods for estimating the leaf count from 2D images—even when using trained human annotators—tend to undercount the total leaves present. This is partly because no single viewing angle

will have unobscured views of all extant leaves. In addition, the change in position, size, and color of individual leaves throughout development, combined with the senescence of leaves later in development, makes tracking the identity of leaves across multiple time points challenging and creates a systematic bias toward underestimating the total leaf number and leaf appearance rate.

Complete 3D plant reconstruction is not the goal addressed by this paper. Still, it focuses on a particular problem of employing multiple time point data from high-throughput phenotyping to segment and quantify the number of leaves (leaf number) of individual plants over time. Both total leaf count and leaf emergence rates are critical variables within crop growth models used to estimate how crops will perform in different growing environments or to respond to different management decisions by a farmer (Hammer et al., 2010; Jiang et al., 2019). Leaf number and leaf emergence rate can differ between different varieties of the same plant species (Allen et al., 1973; Kim et al., 2017) and between genetically identical plants grown in varying environments (Brooking et al., 1995; Tollenaar & Hunter, 1983). Genotype-specific crop growth models incorporating differences in phenological parameters such as leaf number and leaf emergence rate can provide significant improvements in predictive accuracy over generic models that employ average values for a whole crop species (Kumar et al., 2009; Technow et al., 2015).

The constraining factor in adopting more accurate plant models is the cost and difficulty of obtaining accurate, genotype-specific measurements for large numbers of plants in multiple environments. Manual leaf number and emergence rate data collection is time intensive and error prone. Individual leaves are initially very small and may be obscured by other recently emerged leaves. Over time, leaves grow in size and alter their angle, position, and coloration as they mature. Later in development, early emerging leaves may senesce and die. As a result, measurements at multiple points throughout development and accurate tracking of leaf identity from one-time point to the next allow us to accurately determine the total number of leaves produced by a given plant throughout its life cycle.

Our approach's key contribution is generating novel synthetic data to train a panoptic segmentation network and to segment and count the number of leaves in images of real plants. Manually counting and segmenting leaves is tedious work, and our approach improves upon three critical drawbacks of traditional 2D image-based counting.

The first improvement is that our method uses segmentation to implicitly count entire leaves, as opposed to the state-of-the-art (Miao et al., 2021) that focuses on only identifying leaf tips. This approach significantly reduces the probability that the leaf is missed in the input image, consequently increasing the precision.

The second improvement our method offers over the limitations of previous methods is that our method implicitly learns to count leaves with complex occlusions because the synthetic training data correctly models plant geometry with a range of modeled occlusive conditions. Our method improves upon the tendency of previous automated methods to undercount the leaf number, which we believe to be due to our method segmenting entire leaves even when they become occluded further down the leaf in the image space. Confirming the relationship to occlusion is not the purpose of this work and is a topic for future work.

Lastly, our method does not require manual segmentation as it relies purely on synthetically generated datasets for training and synthetically segmented real data pairs in the fine-tuning step. This allows for the generation of large datasets of training data, and we provide one for future experiments (Li, 2022).

2 | MATERIALS AND METHODS

2.1 | Overview

An overview of this process is shown in Figure 1. Our work's first key innovation is how the training data was generated. We use a computer graphics-based synthetic algorithm to bridge the sim-to-real gap in deep learning (Zhao et al., 2020) by providing highly detailed 3D geometric models of sorghum with spectral reflectance on the surface. Our method also models the phenotyping facility, its camera, and its illumination settings. We combine the plant and the 3D scene by rendering the images with an algorithm that considers light interaction with the leaves on a photonic level. This results in an arbitrary number of realistic images paired with their corresponding segmented masks using solid colors for the leaves. Each plant is displayed from multiple views, thus simulating the imaging from various angles. The procedural model allows us to generate the tens of thousands of image pairs needed to train the deep neural model. We use the state-of-the-art DL-based image segmentation algorithm Axial-Deeplab (Wang et al., 2020) implemented in the Deeplab2 library (Weber et al., 2021).

Images of real plants and corresponding temporal data from the University of Nebraska–Lincoln Greenhouse Innovation Center phenotyping facility are used in the inference step. Each plant has been imaged from multiple angles, including the top view, and has also been imaged over time. We independently run these real images through the trained deep neural model to obtain segmentation masks. Then, they have their leaves counted using a data processing step that estimates the total leaf count at each measurement and uses temporal coherence to establish the new leaf emergence. The output of our algorithm is the number of leaves per plant image.

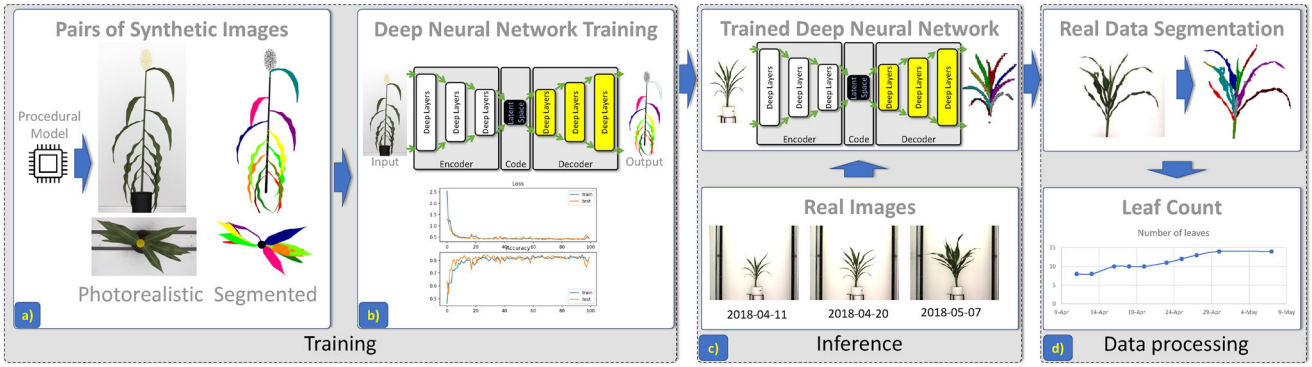

FIGURE 1 Method overview. The procedural model (a) generates three-dimensional synthetic sorghum models that are rendered to provide pairs of images: a photorealistic (left) and its segmented counterpart (right). Tens of thousands of pairs of images are used to train the deep neural model (b). The trained model segments real sorghum images from time sequences, providing the segmented images and the leaf number per plant, per day, and per view (c). The data is processed by providing the number of leaves over time (d).

2.2 | Procedural sorghum model

One of the main contributions of this work is providing a flexible computer-generated synthetic 3D sorghum model that can generate an arbitrary number of 3D plant geometries with labeled parts. Computer graphics has focused on 3D plant modeling for over forty years (see, e.g., Prusinkiewicz & Lindenmayer, 1990, or a review by Pirk et al., 2016). These approaches mainly focus on providing a detailed appearance that includes plant geometry and surface properties (Baranoski & Rokne, 1997, 2001). This work builds on the related work by providing an algorithmic model that is easy to control and provides 3D shapes at varying stages of development. We provide both shape and surface properties (texture) for the procedurally generated sorghum. To calibrate the 3D model, we used 3D partial reconstruction of plants into voxels (Gaillard et al., 2020b) and 3D skeletons (Gaillard et al., 2020a) that use multiple views of sorghum from the phenotyping facility to reconstruct the 3D model. The 3D data from real plants guides the procedural model.

2.2.1 | Leaf model

The procedural model generates the entire plant as a set of concentric leaves. The 3D synthetic sorghum leaf model and its parts are shown in Figure 2a. The leaf skeleton is the key procedural driver, defining the overall leaf appearance. The skeleton is modeled as a cubic spline curve and partitioned into a sheath and blade for each leaf. The sheath is sampled concurrently for all leaves to join them smoothly. Each subpart of the skeleton is further uniformly subdivided into the desired level of detail. To reveal the general shape of the leaf, we need a minimum of 16 segments. A leaf is modeled as a half-opened generalized cylinder that closes in the proximity of the branching point. It starts as a circle at the sheath, and the circle breaks into an open curve at the auricle and then gradually expands to a curved line to form the blade of the leaf. At the end of each leaf, the defining sweeping curve degenerates into a single point, forming the tip of the leaf. We used at least 128 segments per leaf to ensure a smooth surface for photorealistic rendering. Each part of the model is fully controlled by procedural parameters, allowing for generating leaf variations (see Figure 2b,c). The default values of the most important parameters (leaf width, waviness, and length) are based on measured data with mean and variance from Gaillard et al. (2020a, 2020b), and the rest of the parameters are set manually by experts to make leaves visually similar to real leaves. The possible imprecision of the generated leaves does not affect the training because we generate tens of thousands of leaves statistically similar to the real-world measured data.

2.2.2 | Plant model

The leaves are concentrically arranged to make the whole plant, like the example in Figure 3. To arrange each leaf, we first generate the 3D skeleton of the entire plant, branching with the variance and waviness observed from the measured data. The plant skeleton has common parts (stem) and branching points corresponding to the auricle. The spacing and the branching angles are obtained from measured data from Gaillard et al. (2020a, 2020b). Our model then generates individual leaves around the skeleton. The plant's root includes all leaves wrapping around the skeleton, and each leaf further sweeps the skeleton until the branching point.

2.2.3 | Leaf surface rendering

Generating a representative leaf color and texture in the final image is one of the most critical parts needed for

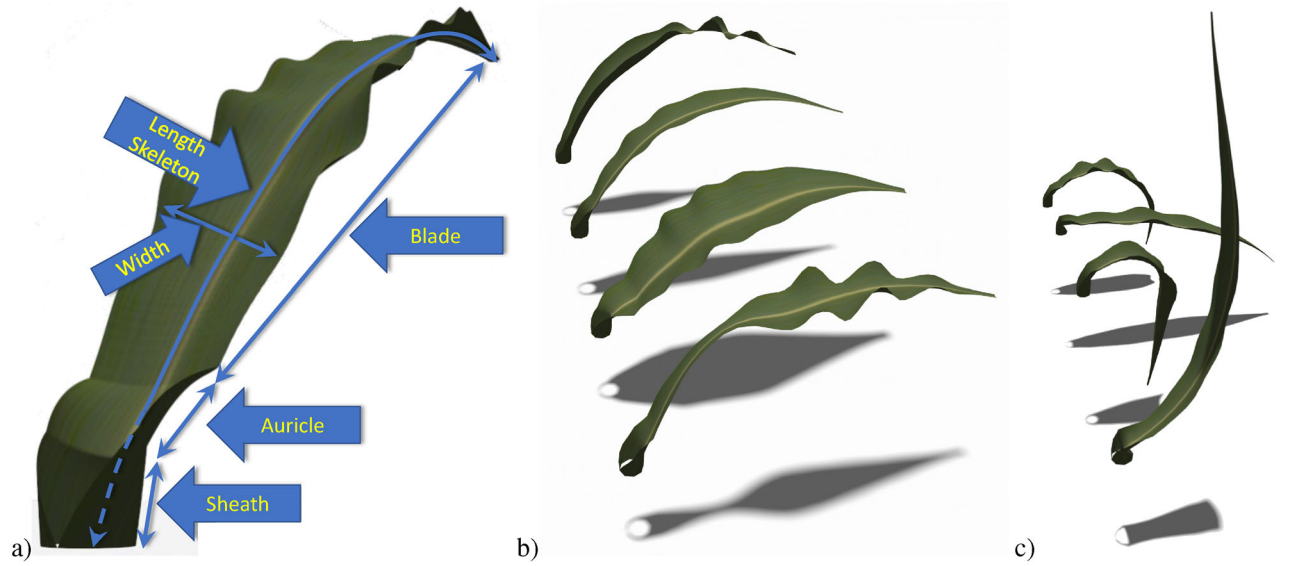

FIGURE 2 The three-dimensional (3D) sorghum procedural leaf model and some of its main control parameters (a). Variation of 3D leaves generated by the model by altering the parameters (b,c). With a fixed skeleton, we can alter the width, the waviness, and the curling angle along each leaf (b). With a fixed leaf shape, we can alter the branching angle, the bending amount, the overall width, and the length of the leaf (c).

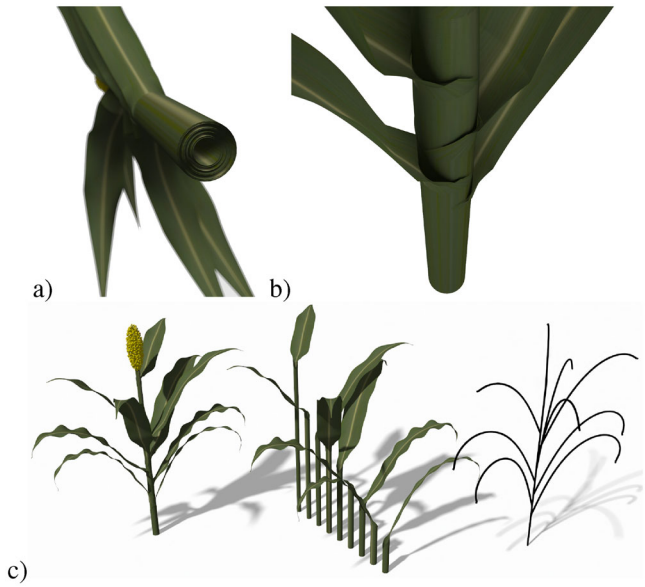

FIGURE 3 Detailed views of a procedurally generated three-dimensional sorghum plant model from the bottom (a) detail of the stem (b), and a plant with its leaves and the skeleton (c).

correctly leaping over the sim-to-real gap in deep learning (Zhao et al., 2020) training model generation. The leaf color results from a complex interaction of the leaf light reflectance and transmittance properties. They are described in terms of the bi-directional transmittance distribution function (BTDF) and bi-directional reflectance distribution function (BRDF). These functions take the direction under which a photon of a particular wavelength reaches the surface as input and output the probability of the same photon leaving the surface in another direction. We use the procedural reflectance and transmittance model from Baranoski and Rokne (1997).

Another critical portion of correctly calculating the final leaf color is mimicking the light conditions of the phenotyping facility. We created a 3D model of the facility with preset lights with approximately uniform spectral characteristics and illumination.

A final rendering pass determines the final color of the leaf in our model. We use our in-house developed bi-directional path tracer that simulates the light interaction on the photonic level. An example in Figure 4 shows a side-by-side comparison of our synthetic rendered model and the real images.

2.2.4 | Reference sorghum sampling

One advantage of using our procedural model for generating synthetic data is the expressiveness and ability to generate geometry for a wide range of phenotypes. Images were generated using several different ranges for the input parameters to represent different sorghum genotypes and to capture a range of characteristics during training. The input parameters included the stem height for the entire plant and per-leaf height, length, and angle. Branching angles were measured from the midrib underneath the auricle. If the leaf midrib is underneath the auricle, the branching angle was found using the top of the leaf as a reference. Leaf height was measured from the highest point on the dewlap of the leaf. Stem height was measured from the rooting zone to the leaf height of the youngest unfurled leaf.

FIGURE 4 A side-by-side comparison of rendered three-dimensional sorghum procedural plant (b,d) and real models from the phenotyping facility (a,c). Note that we do not attempt to reconstruct the plants. We generate their geometry.

The values for each leaf are sampled from a distribution of values generated from manual measurements of the following genotypes: As4601 Pawaga-2, Tx3197-1, Tx3197-2, BTx623-2, RTx430-1, BTx623-1, SC301-1, and SC301-2. These plants were measured from approximately the v2 stage through grain fill at the University of Nebraska's Greenhouse Innovation Center (Ge et al., 2016) from April 26 to August 2, 2022.

2.3 | Datasets

Our algorithm utilizes four image datasets:

1. Synthetic image–mask pairs generated from 3D plant models for training the deep neural network (Li, 2022).
2. Real images paired with synthetically generated masks using the methods of Gaillard et al. (2020b, 2020a) for fine tuning the deep neural network.
3. Real images for evaluation from Miao et al. (2021). Some of these images had leaf count values annotated by experts.
4. Real images that were hand segmented on a per-pixel basis to evaluate the segmentation model's performance separately from leaf counting.

2.3.1 | Synthetic training dataset generation

Training deep neural models traditionally requires tens of thousands of pairs of images for robust results. Phenotyping facilities can provide many images of plants from varying angles at different stages of development. However, the plant part labeling is not commonly accessible. The usual task is to label the images manually (Gaggion et al., 2021; Lu et al., 2017; Minervini et al., 2016; Miao et al., 2021) or with the help of software (Miao et al., 2021). This tedious task is prone to error and does not scale well. Moreover, each image needs to be labeled multiple times by different people and cross-validated to diminish possible human error.

Generating a dataset in our pipeline mimics the imaging process within the real phenotyping facility. First, we generate the skeleton of sorghum at the given age, randomize its parameters, and then generate the complete 3D sorghum plant. An example of the skeleton and synthetic plant can be found in Figure 3c. The synthetic sorghum is then placed into the virtual phenotyping facility under simulated lighting conditions. The intrinsic and extrinsic camera parameters are carefully tuned to avoid image distortion and displacement. We collect RGB images for each sorghum with raytraced light and the colored mask as the label. We generate different colors for each leaf and the stem. The sorghum is rotated within the scene by a step of 36°, and for each sorghum, we capture five images and five masks for side views and one image and one mask for the top view. Figure 5 shows a 3D synthetic sorghum plant from multiple views and its masks, which are all the output of this imaging process.

We have generated 10k different 3D sorghum plants alongside a total of 60k image–mask pairs. Each plant was rendered with a resolution of 512×512 pixels from five side views at angles: $\{0, 36, 72, 108, 144\}$ degrees and a top view. The age of procedurally generated plants ranges from 1 to 8 weeks, with between 4 and 13 leaves. The number of leaves for each synthetic plant in the dataset is saved as ground truth to validate our method of leaf counting. Of these 60k pairs generated, 1800 were withheld for validation and were not used to train the model.

2.3.2 | Real training dataset generation

We have generated 600 image–mask pairs for use in the fine-tuning step of the training pipeline. While the input images were real images from the phenotyping facility, the masks were automatically generated using the segmentation procedure from Gaillard et al. (2020b, 2020a), where the real

FIGURE 5 A photorealistic rendering and mask of a sorghum plant rotating by 36° and its top view.

images were registered and used to calculate a 3D voxelization of the plant which is then turned into a skeleton via thinning and finally segmented. These segmented 3D models were then re-projected onto the real images to give an approximate mask where each leaf is represented as a different color.

2.3.3 | Real validation data

We used a dataset of real plants imaged at the University of Nebraska–Lincoln Greenhouse Innovation Center phenotyping facility in 2017 for validation data. All images were acquired using the RGB imaging setup (see Gaillard et al. (2020b) for details). Each plant is imaged from five sides and one top view using a camera with a 2454×2056 pixels resolution.

This sorghum image dataset is originally from the work of Miao et al. (2021) and consists of 343 unique sorghum plants, representing 295 inbred lines, imaged from July to August 2017 over 37 days. The total number of images collected for this dataset is 27,770, and side view images were taken at angles: $\{0, 36, 72, 108, 144\}$ degrees. Images were cropped to remove the imaging scaffold; thus, their resolutions may vary.

In the previous work of Miao et al. (2021), experts were asked to annotate the number of visible leaves in 5554 images. Only the view at 108° of the plant was annotated for this dataset because it should have the least occlusion and, therefore, the most visible leaf tips. Also, some images were annotated twice to estimate the agreement between experts. These annotations are used in our statistical analysis to quantify the human counting error and evaluate our method on leaf counting. The important point is that the experts did not always agree with the leaf count, as the leaves are not always easily discernible.

This dataset was pruned to exclude those plants that were injured, tilted, out of frame, or had other abnormalities that made them significantly different than the training data. The valid subset totaled 1750 images.

2.3.4 | Real segmented data

Trained experts manually segmented 10 sorghum plants to calculate image segmentation performance. Each image was segmented and stored by three participants. The participants manually colored each leaf a unique color while the stem/branch was black and the background was white.

2.4 | Implementation

2.4.1 | Neural network architecture

The chosen network for our work is Axial-Deeplab (Wang et al., 2020), a modification of the Panoptic-DeepLab network from Cheng et al. (2020), which relies on an encoder–decoder architecture. The decoder portion of Panoptic-DeepLab

FIGURE 6 Segmentation comparison on real plants: an input image is hand segmented by a human, the label predicted by our model ("AI"), and the binary difference between the human and the prediction. The rows represent a segmentation with the highest (0.70) (a), an average (0.54) (b), and the lowest (0.34) intersection over union (c) when averaged across the three human labelers.

remains unchanged from the original. It consists of two data paths: one to predict the semantic label and another to predict the instance center labels and center regression. The instance centers are predicted as a set of blurred points on the image, and then for each pixel, an offset is predicted such that if that pixel were to be moved by that offset, it would lie in the center of the instance it belongs to. The three outputs are then fused to create a panoptic prediction consisting of each pixel labeled with an `objectID` and an `instanceID` by shifting the pixels by their offset, finding which center they are closest to, and then accounting for which semantic class that center lies on. Axial-DeepLab improves upon this architecture by providing a different backbone for the encoder portion of the network by replacing the residual blocks of the ResNet backbone with axial self-attention modules. This network was chosen as it was the most performant network on our task at the time of the manuscript submission. However, computer vision moves quickly, and our method is agnostic to the architecture as long as the model can be trained on image–mask pairs (as almost all current models are) and can output a pixel-level segmentation.

2.4.2 | Pipeline

We implemented our system in Python using TensorFlow and Axial-Deeplab from Deeplab2 (Weber et al., 2021). It has been executed on a server equipped with two NVIDIA TITAN Xp 12 GB graphics processing units (GPUs) and two Intel Xeon E5-2680 v3 central processing units (CPUs). The batch size used during training was eight images of size 512×512, which took approximately 2.5 weeks to converge. The model trained on synthetic image–mask pairs was then fine tuned using real image–mask pairs where an image from the phenotyping facility was paired with an automatically generated mask produced using the methods from Gaillard et al. (2020b, 2020a). This fine tuning further trained the model to match the color and textures of real sorghum plants from the greenhouse and overcome differences between our synthetic textures and real sorghum images. The fine-tuning data was completely synthetically generated without humans needing to hand segment any masks. The machine used for fine tuning was a desktop computer equipped with an AMD Ryzen 9 5900X, 32 GB of RAM, and an NVIDIA GeForce RTX 3070 with 8 GB of video memory. The trained model was then used to perform the panoptic segmentation of the sorghum images, and the output represents a prediction of which pixels correspond to which leaves.

Our implementation modifies the Deeplab system so that it outputs a raw `numpy` array of $length \times width \times 3$, where the first channel is the semantic ID of each plant: (0-background, 1-leaf, 2-stem, and 3-panicle) and the second channel is the instance ID. This modification was a performance

optimization because the larger raw arrays could be loaded into our system faster than comparatively smaller regular images when image decompression was accounted for.

To count just the leaves in the image, ignoring the panicle, background, and stem, we then count unique instances of the pixels with semantic ID = 1. The leaf number L_t for the plant at timestamp t is the number of distinct semantic IDs associated with leaves. This is analogous to counting how many colors are in the output segmentation and ignoring the background, panicle, and stem color, but it was more performant.

2.5 | Evaluation techniques

2.5.1 | Leaf segmentation

The performance of the network's segmentation is validated by comparing images labeled by human annotators, and Figure 6 shows side-by-side examples of varying quality segmentation. A total of 10 images were chosen from the real image dataset, where 5 were randomly selected from those our method had a correct leaf count, and another 5 were selected from those on which our method had the worst performance. Each of the 10 images selected was annotated three times by three different human annotators. The annotators were provided a background segmentation mask as a starting point to aid their annotations. The images were provided at the 512×512 resolution seen by the network and a full-resolution version of the image to help the annotators decide on difficult areas. The hand-labeled images were then manually color matched to the colors produced by the model. For each color, the intersection over union (IoU) was calculated by dividing the area of pixels labeled correctly in both the ground truth and the prediction by the area of pixels labeled that color by either the ground truth or the prediction. The IoU values for each color were then averaged to create an IoU score for the entire image, which was then averaged across the three human annotators.

Due to discrepancies in definitions between fields, the biological definition of the distinction between leaf and stem differs from the one used to generate our procedural model. We do not pay close attention to the segmentation of pixels in the stem. To express this concern numerically, a pixel is part of the intersection if it satisfies either of two cases: (1) it is the same color in the annotation and the prediction, or (2) it is labeled stem in the annotation but the leaf of interest in the prediction.

2.5.2 | Statistical analyses

Our algorithm is validated on the manually counted data from Miao et al. (2021). This dataset contains plants taken at their 108° view. It excludes those injured, tilted, out of frame, or had other abnormalities that made them significantly different from the training data. The valid subset totaled 1750 images. It is important to note that human annotators do not precisely agree with their counts. The root mean squared error (RMSE) between two annotators on the same image is 0.71 leaves as reported by Miao et al. (2021).

A fixed integer number of human-counted leaves cannot be used because it varies among participants. Instead, we created a linear model that approximates the input data. The ground-truth values for each image were calculated from the human counts of all sorghum plants present in the Sorghum 2017 dataset with more than three leaves using a mixed linear model. The model is used to compare our method's predictions with a confidence interval for the true leaf number calculated from the human annotations to account for the underlying uncertainty for human counts. The equation fit by the model was:

$$y = X\beta + Zu + \epsilon, \tag{1}$$

where y is the fit number of leaves for a particular image; $X\beta$ represents the design matrix and coefficients for the fixed effects in the model provided by the user ID of the annotator; the Zu is the design matrix, and random effects coefficients provided by the counts for each image; and the ϵ is the unique intercept for each image. In our analysis, the fixed effects term, $X\beta$, was found to be negligible in the final model, which was to be expected if particular human annotators perform similarly between each other when counting leaves.

Equation (1) gives as output a new leaf count for each image calculated from the distribution of human counts and the standard error of human counts. We used the standard error and the new predicted leaf counts to generate confidence intervals for each image and report the percentage of counts across a curated subset where our model performs within those confidence intervals.

When counting the number of leaves on the real images of plants (see Figure 7c,d), we compare both the exact human annotations and the ground-truth values estimated by a mixed linear model fit from all sorghum plants counted.

2.5.3 | Tracking leaf count over populations and time

Our method has been used to track the leaf number across a population being imaged several times during a growing period. To validate the leaf tracking, we plot a comparison of the average number of leaves counted on each day imaging took place (see Figure 8). We plot the error bars between the human-counted average number of leaves and our method's

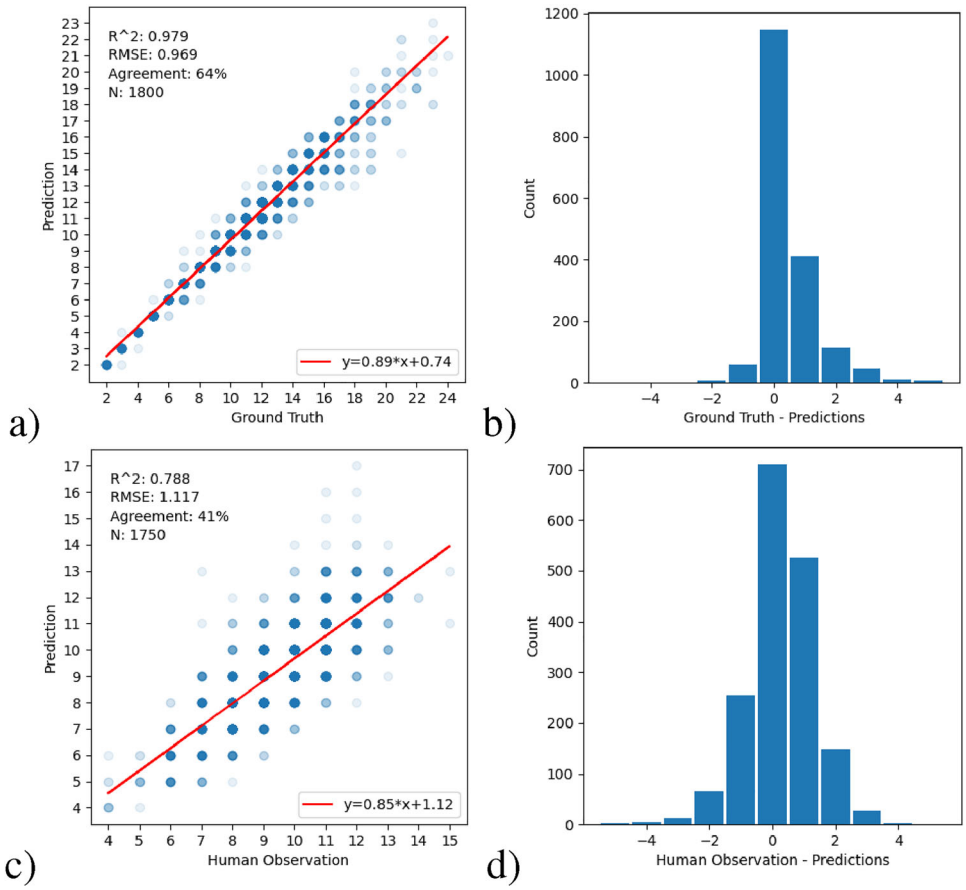

FIGURE 7 Scatter plots (a,c) and histograms (b,d) showing the prediction performance of our method. Comparison to 1800 images of unseen synthetic data (a,b) and 1750 real and manually annotated sorghum plants from Miao et al. (2021).

automatic count per plant. The error bounds are given by the 95% confidence interval calculated from the number of plants imaged that day and the standard deviation produced by the statistical model in Equation (1).

3 | RESULTS

3.1 | Leaf segmentation

The IoU performance is reported separately for the plants counted correctly and the plants counted incorrectly. The correctly counted plants had mean IoU scores ranging from 0.35 to 0.71, with a mean value of 0.56 across the five plants. Our method counted most incorrectly in plants with mean IoU scores from 0.34 to 0.58, with an average of 0.47 across those five plants. We note that the segmentation behavior of different human annotators can vary in decisions made at boundaries, resulting in one plant having a mean IoU from one annotator of 0.13 while another annotator label calculated to 0.58 for the same image.

3.2 | Leaf counting

Two different plots are used to represent the results: (1) a scatter plot of the predicted leaf number w.r.t. the human-annotated leaf number, and (2) a histogram of the distribution of prediction errors, that is, the difference between the human annotation and the prediction. In addition to the scatter plot, we compute a set of metrics: a linear regression (red line) with the R^2, the RMSE, and the agreement, which is the percentage of exact prediction over the entire dataset. This evaluation is executed on both the synthetic and the real dataset, and the results are shown in Figure 7.

3.2.1 | Synthetic dataset validation

Figure. 7a,b shows the trained and fine-tuned network's performance evaluated on a set of 1800 synthetic sorghum images not seen by the model. The ground-truth number of leaves for these images is calculated from the synthetic segmentation mask by counting the number of unique leaf colors. Our

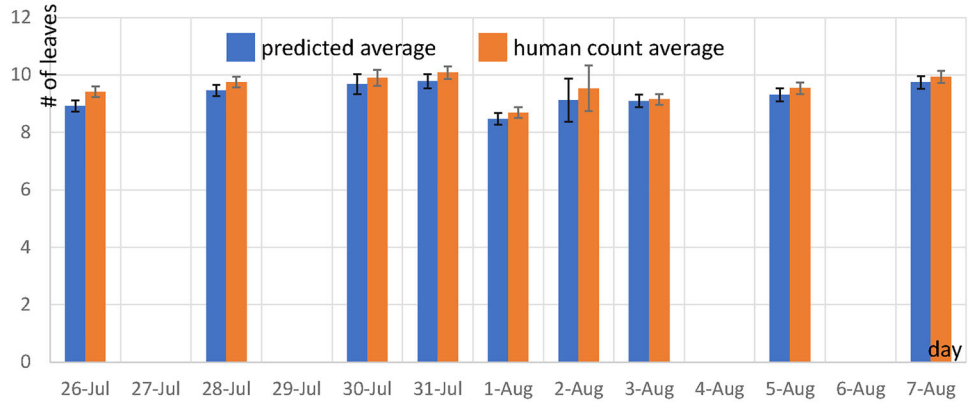

FIGURE 8 Comparison of the average leaf count on a particular day between our method's predictions and ground-truth human counts. The *x*-axis shows the day of 2017, and the *y*-axis is the count and the standard deviation. The error bars correspond to the 95% confidence interval around each mean value.

method produces counts within one leaf of the correct number in 93% of images and has an exact agreement of 64%. The RMSE for the synthetic leaf counts is smaller than one leaf at 0.969 leaves over the 1800 plants. The model is expected to perform well on the synthetic data because these validation images are generated using the same procedural model as the training images.

3.2.2 | Real dataset validation

We agree exactly with a human annotator in 41% of images. Still, since the RMSE between two human annotators on the same image is 0.71 leaves, we also compare to the ground-truth estimations produced by the linear model. Our method predicts leaves within the 90% confidence interval leaf counts in 97.12% of images. For an 80% confidence interval, we predict within it for 87.69% of images. Compared to our predictions, the RMSE for the fitted mixed linear model was 1.11 leaves, and the model had a standard error of 1.47 leaves.

Annotated images captured from the angle of 108˚ show the most details from the plant and the least occlusions of other views in the dataset. We have trained our deep neural model on images from multiple directions. We used the number of leaves detected from the angle 108˚ as the point of comparison to the count from different angles, and Figure 9 shows the results. The maximum error is for the top view, and for the side views, the error decreases as the angle gets closer to the view in which most of the detail is captured. The number of leaves present in the predictions at the 108˚ images ranged from 4 to 20 leaves. The curated dataset of 1750 images was expanded with the images corresponding to the other five views for the 108˚ views chosen, which increases the dataset size for Figure 9 to 10,500 total images.

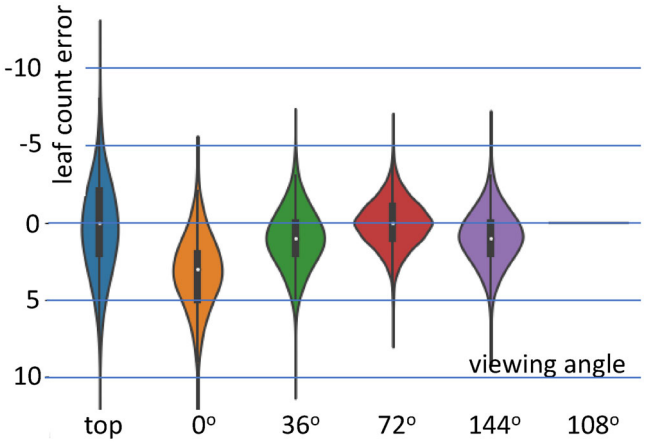

FIGURE 9 Error of the leaf count from different viewing directions.

3.3 | Tracking leaf number across populations and time

We compare the leaf count over time to the human observations to show the application of leaf counting to leaf number calculation across populations. We additionally account for the variability in the number of plants imaged that day to calculate the estimator for the standard error in the mean and display the 95% confidence interval. Figure 8 shows that although the exact number of leaves predicted does not match the predictions and human observations, the 95% confidence intervals around the mean number of leaves overlap between predictions and the human annotations on all but 2 days that were imaged. These plots show our approach is viable for counting changes in the relative number of leaves in a group of plants over time.

Figure 10 shows the detection of the number of leaves and the corresponding plant in a period of leaf emergence. This

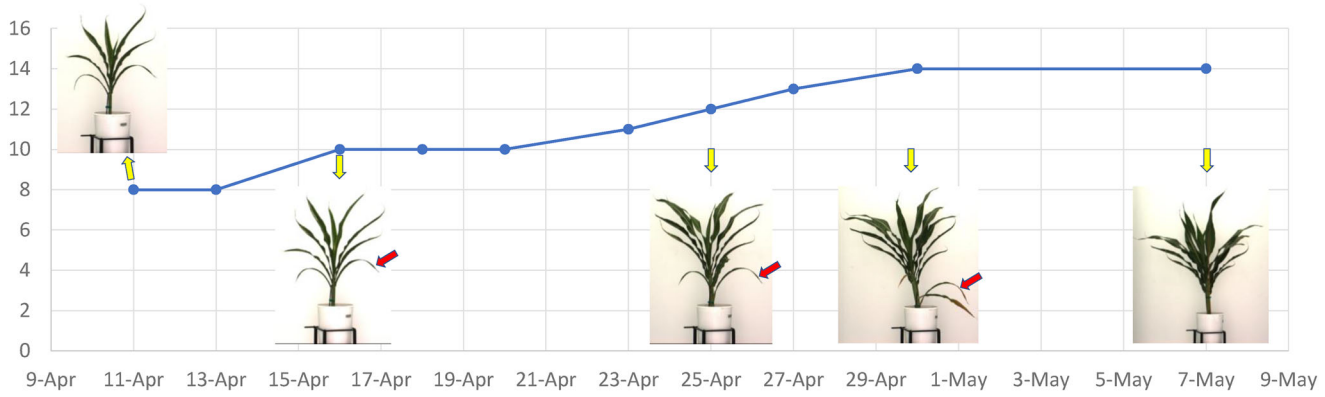

FIGURE 10 Leaf count over a rapid leaf emergence period in 2018. While the leaf count between May 1st and May 7th is the same, some leaves have died off (red arrow), and new leaves emerged.

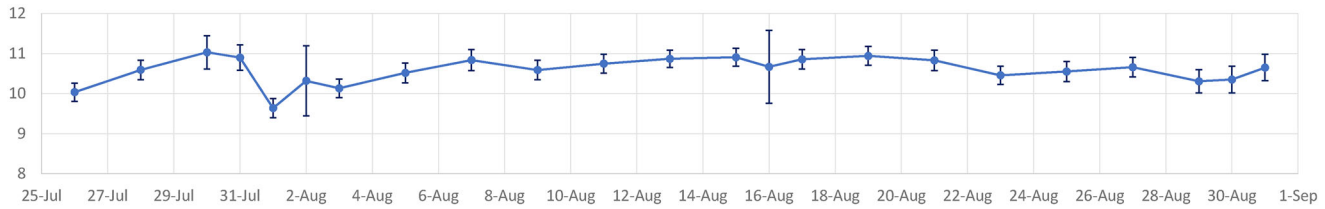

FIGURE 11 Leaf count over 1 month period of 343 plants.

sequence was captured in April–May 2018 and shows the leaf count and the corresponding plant. The red arrow indicates a dying leaf. Interestingly, the number of leaves is constant between May 1 and 7, but the leaves vary in the plant.

Figure 11 shows the detection of the number of leaves in 2 months of data from the phenotyping facility between July 25 and August 31, 2017. A total of 23 data points were calculated for 343 plants. Some images were discarded because the plant was tipped over, cropped, etc. We evaluated, on average, 241 plant images per day, with a minimum of 22 on August 2 and a second minimum of 24 on August 16. Figure 11 shows wider error bars on these days because of the smaller sample size. One can notice that of the 343 total plants and an average of 241 plants imaged per imaging day, the choice of phenotypes and the age of the individuals imaged that day can affect the mean leaf count on that day.

4 | DISCUSSION

4.1 | Leaf segmentation

By checking the segmentation performance on images where our method had both the best and worst leaf counts, we can check if the segmentation performance measured by IoU roughly corresponds with leaf counting performance. The minimum IoU values in the well-counted and poorly counted five plants are similar and poor. This is likely due to taking the average across all the leaves in the image and, second, all the annotators, which exacerbates differences in how the annotators segment the boundary of the leaves. Specifically, there was disagreement among the human annotations with defining the boundaries between the background and the edge of leaves, which get slightly blurred in the image downsampling process from an arbitrary but larger, cropped resolution to the 512×512 seen by the model. The large amount of effort needed to give pixel-level annotations for sorghum leaves in images accurately further motivated our creation of an automatic method. The exact leaf segmentation is not needed to count leaves. Indeed, the IoU performance indicates that it is not a perfect segmentation. The error comes from the uncertainty in highly ambiguous and occlusive parts, such as the middle region of mature plants, where leaf boundaries are impossible to discern from single images exactly. The leaf tips are well segmented, and the segmentation propagates down the leaves to ambiguous areas, as shown in Figure 6b in an example of an average case segmentation. However, our method correctly segments even disconnected components which is an essential property needed to count leaves in highly occluded environments.

4.2 | Ablation study and comparison to the state-of-the-art

4.2.1 | Comparison to state-of-the-art (SOTA)

Computer vision and image segmentation are quickly evolving. Our method serves as a framework where a model trained on synthetic data generates the segmented leaf output. Any state-of-the-art (SOTA) model architecture, such as the Segment Anything Model (SAM) (Kirillov et al., 2023), could be used once trained on specific segmentation masks for leaves. The need for this additional training on our synthetic dataset is illustrated in Figure 12. SAM is a general segmentation with outstanding results on general tasks. However, it is not trained on leaf-level masks because such datasets require intractable amounts of hand labeling. Because of this data gap in existing segmentation models, the masks produced are limited to background separation and are insufficient for leaf counting.

4.2.2 | Ablation study

We performed an ablation study to test the necessity of fine tuning on real data in our pipeline. To achieve this, we segmented and counted portions of our pipeline with a model trained on only synthetic data. Figure 13 shows the improvement in counting performance that the fine tuning gives. The non-fine-tuned model tends to overcount relative to humans, while the fine-tuned model centers more on human counts. We believe this difference is partly due to the real fine-tuning data helping the network learn the exact lighting and texture characteristics for the sorghum in the greenhouse and, as a result, performing better with ambiguous areas that tend to over-segment and thus overcount.

Due to the large number of images needed to train segmentation networks, training on real data would require more images than is feasible to gather. This limitation motivated our work to use a 3D synthetic model to generate thousands of labeled images to train such a network.

4.3 | Leaf counting

Creating an automatic method that has high agreement with a single human annotator is difficult due to the uncertainty of how to define when exactly a new leaf emerges and when senescent leaves fall off. This discrepancy becomes more problematic when multiple experts count an image to contain different numbers of leaves and can be measured with the RMSE between annotators being 0.71 and the agreement being 72% as reported by Miao et al. (2021). The most performant model from that work agreed with human counts of 0.43 compared to our 0.41. However, their model is trained on human-annotated data instead of synthetically generated data that we use. Additionally, the disagreement between human annotators makes comparisons to exact leaf count via agreement and RMSE more difficult, so we instead compare our predictions to the confidence intervals for the leaf number from our mixed linear model. The high proportion of plants that fall within reasonable confidence intervals to human counts on the same image validates that our method counts sorghum leaves with similar accuracy to humans despite any artificial error inherently added by using synthetic training data and stochastic processes.

4.4 | Tracking leaf count over populations and time

The imprecision inherent to leaf count can be mitigated by averaging across populations and time. Figure 8 demonstrates how our method can be used to automatically estimate leaf count across a population even if not all plants are counted exactly correctly on each sampling day. This result validates our technique as a tool to replace expensive manual measurements of a population despite the agreement that a single human counter on single images is insufficient for exact counting.

4.5 | Procedural sorghum model

Our method relies upon the generation of synthetic data from our procedural model. The images produced by our synthetic sorghum image pipeline could be added as data augmentation to an existing dataset of real images. This would enable existing techniques to become more performant by increasing their dataset sizes with images very close to their imaging environment and tweaking synthetic sorghum plants to match the phenotypes being imaged. The procedural model needs to correctly represent only the traits required for the task in question. In our case, we focused on the shape and reflectance of leaves. Researchers could generate synthetic training data by taking measurements of only a representative sample of individuals' geometric traits for each genotype of interest and then generate synthetic sorghum images that fall within the distribution of characteristics for the population of interest.

5 | CONCLUSIONS

This work introduces a novel sorghum segmentation and leaf counting model based on a training dataset generated by a procedural sorghum generator. We show that the in silico model has a good potential to replace the tedious human counting

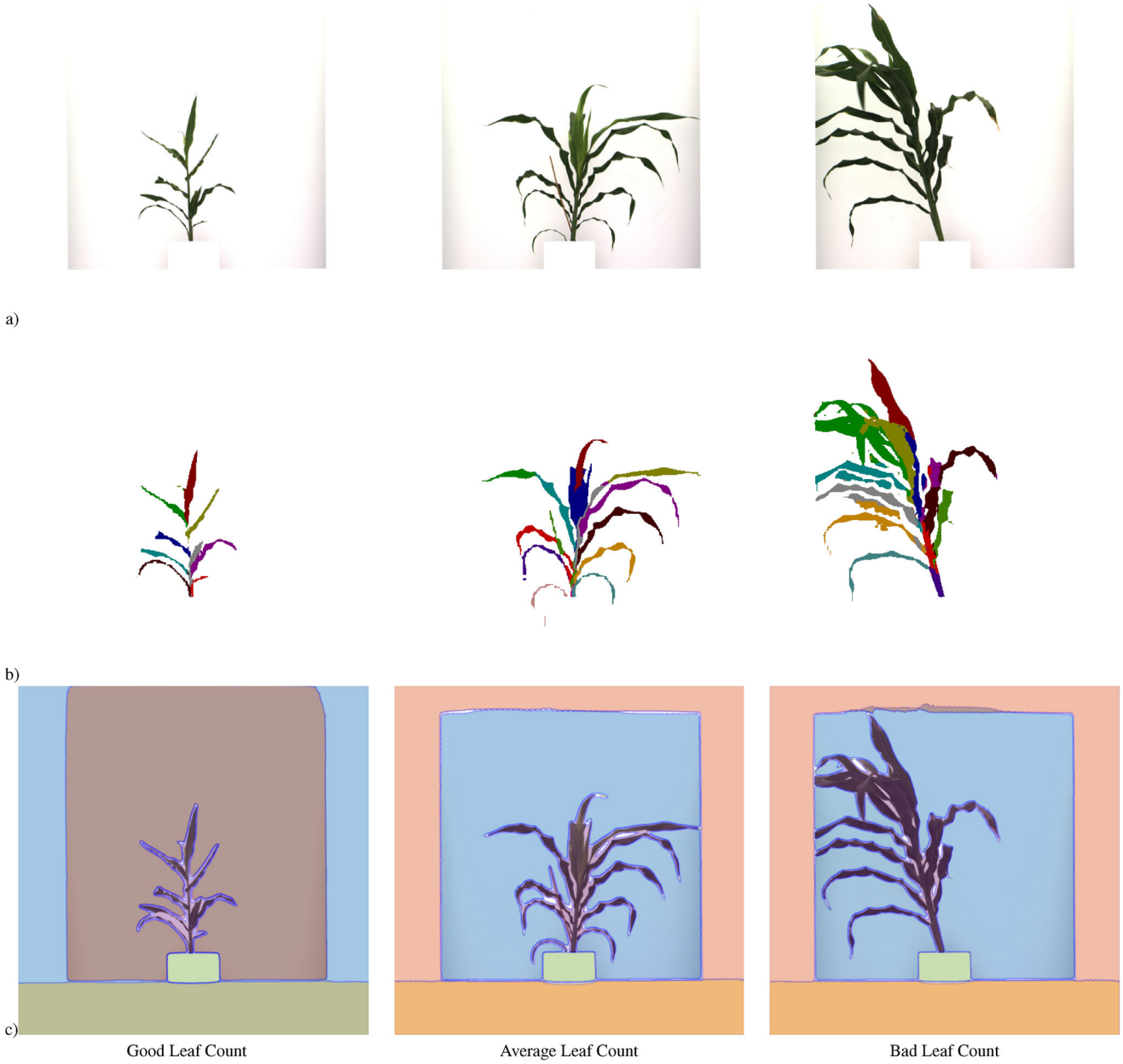

FIGURE 12 Segmentation comparison between our method and Segment Anything Model (SAM) (Kirillov et al., 2023). The rows represent the input image (a), the segmentation from our method (b), the segmentation from SAM which is the current SOTA for general image segmentation (c).

by generating training data that allow deep neural models to perform comparably to humans.

Our method has several limitations. We do not determine the exact location of the emergent leaf, but only the timestamp of its emergence. Theoretically, we could cross-examine the segmented images from multiple views and determine the differences. While this seems like an obvious choice, sorghum's complexity, its variance over time, and the changes in viewing conditions make this task rather complex. The second limitation is the unknown set of parameters that need to be tuned for the procedural model to work well. This approach is trial-and-error and not entirely predictable. The third limitation is a dependence on real-world ground-truth data. The data provided by the multiple annotators had some discrepancies (the leaf count was off by even four leaves for some images) that made it difficult to validate the deep neural model. A fourth limitation is that our approach is trained only on images replicating a particular environment. While this makes our method performant for the particular phenotyping facility that we attempted to imitate, the model could perform poorly with changes in the imaging setup, such as lighting conditions, focal length, and chamber geometry. Our method would also

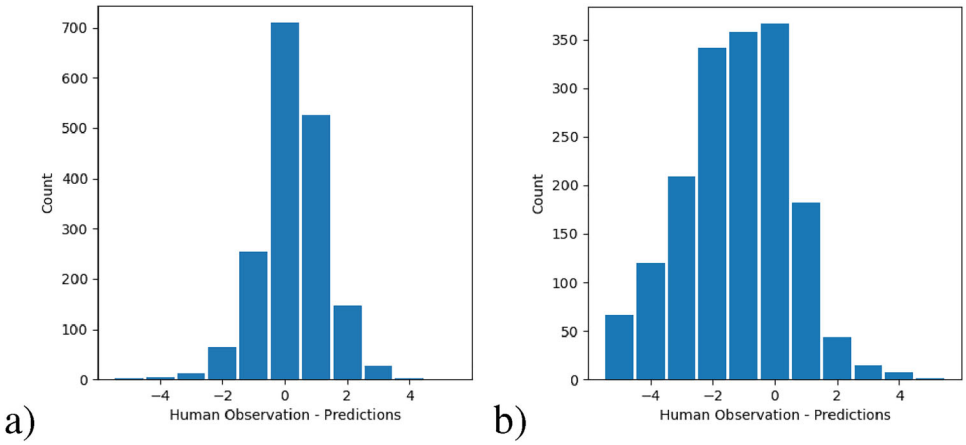

FIGURE 13 Histograms (a) and (b) show the distribution of the leaf counting error relative to human leaf counts for the fine-tuned model (a) and the not-fine-tuned model (b). Note that Figure 7d has been repeated here as figure (a) for ease of visual comparison.

fail when used in a field setting. However, the flexibility of our procedural sorghum model would allow for new training data to be generated to replicate other imaging scenarios with some additional modeling effort to match the target environment closely. This flexibility also allows our method to be extended to capture a wider range of phenotypes through the manual collection of geometric traits and their subsequent input to our procedural model before generating a wider variety of training data, including those phenotypes. Our method's precision is comparable to multiple human counting as it performs within the expected error tolerance that is found when two people count leaves in the same image as reported by Miao et al. (2021). Being sufficiently close to the distribution of multiple human counting makes our method suitable for tracking the number of leaves across a population of plants over time. Future work could focus on improving the precision to perform better than human subjects.

AUTHOR CONTRIBUTIONS

Ian Ostermann: Programming; writing. **Bedrich Benes**: Writing—review and editing; supervision; funding acquisition; project administration. **Mathieu Gaillard**: Formal analysis; data curation; writing. **Bosheng Li**: Programming; writing. **Jensina Davis**: Data acquisition and validation; leaf segmentation. **Ryleigh Grove**: Data acquisition and validation; leaf segmentation. **Nikee Shrestha**: Data acquisition and validation; leaf segmentation. **Michael C. Tross**: Investigation; writing—review and editing. **James C. Schnable**: Writing–review and editing; supervision; funding acquisition; project administration.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. National Science Foundation under awards No. 2412928, 2412930, and 2417510. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation. This work is based upon efforts supported by the USDA-NIFA grants 2024-67021-42879, 2024-67013-42449, and 2021-67021-35329. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US government or NRCS. The US government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. We thank Chenyong Miao for motivating this work. We also thank NVidia for the HW support.

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY STATEMENT

The dataset of synthetic sorghum images is available at https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Benes_SegmentedProceduralSorghumPlants_2022 (Li, 2022). The code for our leaf counting pipeline is available upon request.

ORCID

Ian Ostermann https://orcid.org/0009-0000-9171-8322
Nikee Shrestha https://orcid.org/0000-0003-4195-7549
Michael C. Tross https://orcid.org/0000-0002-4410-9679
James C. Schnable https://orcid.org/0000-0001-6739-5527

REFERENCES

Allen, J. R., McKee, G. W., & McGahen, J. H. (1973). Leaf number and maturity in hybrid corn. *Agronomy Journal*, *65*(2), 233–235.
Baranoski, G. V., & Rokne, J. G. (1997). An algorithmic reflectance and transmittance model for plant tissue. *Computer Graphics Forum*, *16*, C141–C150.

Baranoski, G. V., & Rokne, J. G. (2001). Efficiently simulating scattering of light by leaves. *The Visual Computer*, *17*(8), 491–505.

Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, A. J., Hoffman, N. E., Davis, K. R., & Gorlach, J. (2001). Growth stage–based phenotypic analysis of arabidopsis: A model for high throughput functional genomics in plants. *The Plant Cell*, *13*(7), 1499–1510.

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.

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., & Klukas, C. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *The Plant Cell*, *26*(12), 4636–4655.

Cheng, B., Collins, M. D., Zhu, Y., Liu, T., Huang, T. S., Adam, H., & Chen, L.-C. (2020). Panoptic-DeepLab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12475–12485). IEEE.

Ciancolini, A., Alignan, M., Pagnotta, M. A., Vilarem, G., & Crinò, P. (2013). Selection of italian cardoon genotypes as industrial crop for biomass and polyphenol production. *Industrial Crops and Products*, *51*, 145–151.

Duncan, W., & Hesketh, J. (1968). Net photosynthetic rates, relative leaf growth rates, and leaf numbers of 22 races of maize grown at eight temperatures. *Crop Science*, *8*(6), 670–674.

Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., Hill, S. T., McEntee, C. J., Warnasooriya, S. N., Kumar, I., Ficor, T., Turnipseed, S., Gilbert, K. B., Brutnell, T. P., Carrington, J. C., Mockler, T. C., & Baxter, I. (2015). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in setaria. *Molecular Plant*, *8*(10), 1520–1535.

Gage, J. L., Richards, E., Lepak, N., Kaczmar, N., Soman, C., Chowdhary, G., Gore, M. A., & Buckler, E. S. (2019). In-field whole-plant maize architecture characterized by subcanopy rovers and latent space phenotyping. *The Plant Phenome Journal*, *2*(1), 1–11.

Gaggion, N., Ariel, F., Daric, V., Lambert, É., Legendre, S., Roulé, T., Camoirano, A., Milone, D. H., Crespi, M., Blein, T., & Ferrante, E. (2021). Chronoroot: High-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. *GigaScience*, *10*(7), giab052.

Gaillard, M.Benes, B.Tross, M. C., & Schnable, J. C. (2023). Multi-view triangulation without correspondences. *Computers and Electronics in Agriculture*, *206*, 107688. https://doi.org/10.1016/j.compag.2023.107688

Gaillard, M., Miao, C., Schnable, J., & Benes, B. (2020a). Sorghum segmentation by skeleton extraction. In A. Bartoli & A. Fusiello (Eds.), *Computer Vision – ECCV 2020 Workshops* (pp. 296–311). Springer International Publishing.

Gaillard, M.Miao, C.Schnable, J. C., & Benes, B. (2020b). 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.

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, *115*(18), 4613–4618.

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., & Guo, W. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics*, *2019*, 1–14. https://doi.org/10.34133/2019/1525874

Giuffrida, M. V., Doerner, P., & Tsaftaris, S. A. (2018). Pheno-deep counter: A unified and versatile deep learning architecture for leaf counting. *The Plant Journal*, *96*(4), 880–890.

Hammer, G. L., van Oosterom, E., McLean, G., Chapman, S. C., Broad, I., Harland, 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.

Hasan, M. M., Chopin, J. P., Laga, H., & Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods*, *14*(1), 1–13.

Jiang, R., He, W., Zhou, W., Hou, Y., Yang, J., & He, P. (2019). Exploring management strategies to improve maize yield and nitrogen use efficiency in northeast china using the DNDC and DSSAT models. *Computers and Electronics in Agriculture*, *166*, 104988.

Johnson, J., Sharma, G., Srinivasan, S., Masakapalli, S. K., Sharma, S., Sharma, J., & Dua, V. K. (2021). Enhanced field-based detection of potato blight in complex backgrounds using deep learning. *Plant Phenomics*, *2021*, 9835724.

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.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. arXiv:2304.02643.

Kumar, S. R., Hammer, G. L., Broad, I., Harland, P., & McLean, G. (2009). Modelling environmental effects on phenology and canopy development of diverse sorghum genotypes. *Field Crops Research*, *111*(1-2), 157–165.

Li, B. (2022). *Segmented procedural sorghum plants*. CyVerse Data Commons. https://doi.org/10.25739/4fh7-y927

Li, D., Quan, C., Song, Z., Li, X., Yu, G., Li, C., & Muhammad, A. (2021). High-throughput plant phenotyping platform (ht3p) as a novel tool for estimating agronomic traits from the lab to the field. *Frontiers in Bioengineering and Biotechnology*, *8*, 623705.

Li, Y., Wen, W., Miao, T., Wu, S., Yu, Z., Wang, X., Guo, X., & Zhao, C. (2022). Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Computers and Electronics in Agriculture*, *193*, 106702.

Lin, Z., & Guo, W. (2020). Sorghum panicle detection and counting using unmanned aerial system images and deep learning. *Frontiers in Plant Science*, *11*, 534853.

Liu, L., Lu, H., Li, Y., & Cao, Z. (2020). High-throughput rice density estimation from transplantation to tillering stages using deep networks. *Plant Phenomics*, *2020*, 1–14. https://doi.org/10.34133/2020/1375957

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). Tasselnet: Counting maize tassels in the wild via local counts regression network. *Plant Methods*, *13*(1), 1–17.

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.

Miao, C., Guo, A., Thompson, A. M., Yang, J., Ge, Y., & Schnable, J. C. (2021). Automation of leaf counting in maize and sorghum using deep learning. *The Plant Phenome Journal*, *4*(1), e20022.

Miao, T.Wen, W.Li, Y.Wu, S.Zhu, C., & Guo, X. (2021). Label3DMaize: toolkit for 3D point cloud data annotation of maize shoots. *GigaScience*, *10*(5), giab031. https://doi.org/10.1093/gigascience/giab031

Minervini, M., Fischbach, A., Scharr, H., & Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, *81*, 80–89.

Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., & Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3d deep learning on hyperspectral images. *Plant Methods*, *15*(1), 1–10.

Ottman, M., Kimball, B., Pinter, P., Wall, G., Vanderlip, R., Leavitt, S., LaMorte, R., Matthias, A., & Brooks, T. (2001). Elevated CO_2 increases sorghum biomass under drought conditions. *New Phytologist*, *150*(2), 261–273.

Pereira, C. S., Morais, R., & Reis, M. J. (2019). Deep learning techniques for grape plant species identification in natural images. *Sensors*, *19*(22), 4850.

Pirk, S., Benes, B., Ijiri, T., Li, Y., Deussen, O., Chen, B., & Měch, R. (2016). Modeling plant life in computer graphics. In *ACM SIGGRAPH 2016 Courses*, SIGGRAPH '16 (pp. 18:1–18:180). ACM. http://doi.acm.org/10.1145/2897826.2927332

Plénet, D., Mollier, A., & Pellerin, S. (2000). Growth analysis of maize field crops under phosphorus deficiency: ii. Radiation-use efficiency, biomass accumulation and yield components. *Plant and Soil*, *224*(2), 259–272.

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

Prusinkiewicz, P., & Lindenmayer, A. (1990). *The algorithmic beauty of plants*. Springer–Verlag.

Shibata, S.Mizuno, R., & Mineno, H. (2020). Semisupervised deep state-space model for plant growth modeling. *Plant Phenomics*, *2020*, 4261965. https://doi.org/10.34133/2020/4261965

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends in Plant Science*, *23*(10), 883–898.

Stewart, E. L., Wiesner-Hanks, T., Kaczmar, N., DeChant, C., Wu, H., Lipson, H., Nelson, R. J., & Gore, M. A. (2019). Quantitative phenotyping of northern leaf blight in UAV images using deep learning. *Remote Sensing*, *11*(19), 2209.

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.

Thapa, S., Zhu, F., Walia, H., Yu, H., & Ge, Y. (2018). A novel LiDAR-based instrument for high-throughput, 3D measurement of morphological traits in maize and sorghum. *Sensors*, *18*(4), 1187.

Tollenaar, M., & Hunter, R. (1983). A photoperiod and temperature sensitive period for leaf number of maize. *Crop Science*, *23*(3), 457–460.

Tross, M. C., Gaillard, M., Zweiner, M., Miao, C., Grove, R. J., Li, B., Benes, B., & Schnable, J. C. (2021). 3D reconstruction identifies loci linked to variation in angle of individual sorghum leaves. *PeerJ*, *9*, e12628.

Ubbens, J.Cieslak, M.Prusinkiewicz, P.Parkin, I.Ebersbach, J., & Stavness, I. (2020). Latent space phenotyping: Automatic image-based phenotyping for treatment studies. *Plant Phenomics*, *2020*, 5801869. https://doi.org/10.34133/2020/5801869

Uzal, L. C., Grinblat, G. L., Namías, R., Larese, M. G., Bianchi, J. S., Morandi, E. N., & Granitto, P. M. (2018). Seed-per-pod estimation for plant breeding using deep learning. *Computers and Electronics in Agriculture*, *150*, 196–204.

Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A., & Chen, L.-C. (2020). Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation. In *European Conference on Computer Vision* (pp. 108–126). Springer International Publishing.

Wang, X., Xuan, H., Evers, B., Shrestha, S., Pless, R., & Poland, J. (2019). High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat. *GigaScience*, *8*(11), giz120.

Weber, M., Wang, H., Qiao, S., Xie, J., Collins, M. D., Zhu, Y., Yuan, L., Kim, D., Yu, Q., Cremers, D., Leal-Taixe, L., Yuille, A. L., Schroff, F., Adam, H., & Chen, L.-C. (2021). DeepLab2: A TensorFlow library for deep labeling. arXiv:2106.09748.

Wu, S., Wen, W., Gou, W., Lu, X., Zhang, W., Zheng, C., Xiang, Z., Chen, L., & Guo, X. (2022). A miniaturized phenotyping platform for individual plants using multi-view stereo 3D reconstruction. *Frontiers in Plant Science*, *13*, 897746.

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., Wang, G., Luo, Q., Zhang, Q., Liu, Q., & Xiong, L. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications*, *5*(1), 1–9.

Zhang, W., Calla, B., & Thiruppathi, D. (2021). Deep learning-based high-throughput phenotyping accelerates gene discovery for stomatal traits. *Plant Physiology*, *187*(3), 1273–1275.

Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., Wang, K., Xiao, Y., Chen, G., Liu, Q., Xiong, L., Yang, W., & Yan, J. (2017). High-throughput phenotyping and qtl mapping reveals the genetic architecture of maize plant growth. *Plant Physiology*, *173*(3), 1554–1564.

Zhang, Z.Pope, M.Shakoor, N.Pless, R.Mockler, T. C., & Stylianou, A. (2022). Comparing deep learning approaches for understanding genotype × phenotype interactions in biomass sorghum. *Frontiers in Artificial Intelligence*, *5*, 872858. https://doi.org/10.3389/frai.2022.872858

Zhao, W., Queralta, J. P., & Westerlund, T. (2020). Sim-to-real transfer in deep reinforcement learning for robotics: A survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 737–744). IEEE.

How to cite this article: Ostermann, I., Benes, B., Gaillard, M., Li, B., Davis, J., Grove, R., Shrestha, N., Tross, M. C., & Schnable, J. C. (2024). Sorghum segmentation and leaf counting using in silico trained deep neural model. *The Plant Phenome Journal*, *7*, e70002. https://doi.org/10.1002/ppj2.70002