Received: 28 March 2023 | Accepted: 17 July 2023

DOI: 10.1002/aps3.11548

Applications in Plant Sciences

APPLICATION ARTICLE

From leaves to labels: Building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2

William N. Weaver | **Stephen A. Smith**

Department of Ecology and Evolutionary Biology, University of Michigan, 1105 N. University Ave., Ann Arbor 48109, Michigan, USA

Correspondence

William N. Weaver, Department of Ecology and Evolutionary Biology, University of Michigan, 1105 N. University Ave., Ann Arbor, Michigan 48109, USA.
Email: willwe@umich.edu

This article is part of the special issue "Advances in Plant Imaging across Scales."

Abstract

Premise: Quantitative plant traits play a crucial role in biological research. However, traditional methods for measuring plant morphology are time consuming and have limited scalability. We present LeafMachine2, a suite of modular machine learning and computer vision tools that can automatically extract a base set of leaf traits from digital plant data sets.

Methods: LeafMachine2 was trained on 494,766 manually prepared annotations from 5648 herbarium images obtained from 288 institutions and representing 2663 species; it employs a set of plant component detection and segmentation algorithms to isolate individual leaves, petioles, fruits, flowers, wood samples, buds, and roots. Our landmarking network automatically identifies and measures nine pseudo-landmarks that occur on most broadleaf taxa. Text labels and barcodes are automatically identified by an archival component detector and are prepared for optical character recognition methods or natural language processing algorithms.

Results: LeafMachine2 can extract trait data from at least 245 angiosperm families and calculate pixel-to-metric conversion factors for 26 commonly used ruler types.

Discussion: LeafMachine2 is a highly efficient tool for generating large quantities of plant trait data, even from occluded or overlapping leaves, field images, and non-archival data sets. Our project, along with similar initiatives, has made significant progress in removing the bottleneck in plant trait data acquisition from herbarium specimens and shifted the focus toward the crucial task of data revision and quality control.

KEYWORDS

digital extended specimen, digital specimen voucher, herbarium, machine learning, morphometrics, neural networks, phenology, traits

Two decades ago, molecular sequencing experienced the beginning of what would be several revolutions in the generation of molecular data that ushered in a paradigm shift in biology. Unfortunately, quantitative and morphological data have not experienced an equivalent development. Although the global network of herbaria and natural history collections have been diligently digitizing collections over this time, rapid means of extracting morphological information from those images have yet to be developed. Specimens that often sat dormant in cabinets became easily accessible through data portals like the Global Biodiversity Information Facility (GBIF; https://www.gbif.org) and iDigBio (https://www.idigbio.org). However, using specimen images from this veritable

forest of nearly 400 million preserved plant specimens to address a specific research question poses many challenges, including the extraction of quantitative traits and phenology data (Heberling, 2022). While it is possible to use existing manual or semi-automated methods to extract quantitative traits from specimen images, the process tends to be labor intensive and does not scale beyond a few dozen or a few hundred images. The adoption of robust computer vision and machine learning workflows promises to augment and expedite researchers' ability to parse, measure, and review digital specimens at any project scale.

We previously published LeafMachine as a first step toward large-scale analysis of herbarium specimens (Weaver et al., 2020). LeafMachine used pixel-wise semantic segmentation (binning each pixel into a predetermined class) to split the image into five classes for further processing. This process was effective, but it was also error-prone and could not handle complex specimens. We were able to locate objects of interest like leaves, stems, text, fruit, and flowers, but leaves that overlapped or were obstructed by mounting tape could not be measured. When a leaf candidate was identified, we used a support vector machine to analyze its shape traits, which greatly limited the number of supported taxa given the relatively low degree of diversity captured by our small training data set. The original training data set had only 425 images, thus generalizability was also poor. LeafMachine struggled to process images containing lobed leaves, poor lighting, and cluttered backgrounds. We also lacked autonomous methods to determine specimen-specific pixel-to-metric conversion factors (CF; please see Table 1 for a full list of definitions), so the process was not fully autonomous. At the time, our lack of sufficient high-quality training data and underestimation of the degree of heterogeneity in preserved plant data sets prevented us from achieving our goal of extracting trait and phenology data from all publicly available images.

Since the publication of the original LeafMachine software, the barrier to entry into machine-assisted biological research has been significantly lowered as a result of the prolific use of machine learning in nearly all aspects of modern life (Martens, 2018; Safadi and Watson, 2023). Thanks to the monumental efforts of open-source collaborations and the equally monumental funding funneled into the field by corporate backers, there are now highly accurate plug-and-play machine learning architectures for object detection, instance segmentation, panoptic segmentation, scene detection, facial recognition, and pose estimation (He et al., 2017; Wu et al., 2019; Kirillov et al., 2020; Jocher et al., 2022). In combination with transfer learning, these prebuilt network architectures allow researchers to focus their efforts on generating high-quality training data sets (Yang et al., 2020).

In this paper, we introduce LeafMachine2, a modular suite of computer vision and machine learning algorithms that enables efficient identification, location, and measurement of vegetative, reproductive, and archival components from digital plant data sets (Figure 1). For LeafMachine2, we took full advantage of this new paradigm by heavily utilizing Meta AI's (New York, New York, USA) PyTorch implementation of Detectron2 (Mask R-CNN) and the Ultralytics (Los Angeles, California, USA) implementation of YOLOv5, one of many YOLO variants (He et al., 2017; Wu et al., 2019; Jocher et al., 2022). These frameworks are extremely flexible, well-supported, and surprisingly approachable. As a result, many recent projects have also coalesced around these two frameworks with great success, including efforts to segment leaves (Younis et al., 2020; Triki et al., 2020, 2021; Guo et al., 2021; Hussein et al., 2021b; Gu et al., 2022; Ott and Lautenschlager, 2022), segment plant tissue (Love et al., 2021; Goëau et al., 2022; Milleville et al., 2023), isolate plant organs (Davis et al., 2020; Pearson et al., 2020; Triki et al., 2020; Ott and Lautenschlager, 2022), extract specimen label data (Milleville et al., 2023), isolate diseased or damaged leaf tissue (Kaur et al., 2022; Mu et al., 2022; Kavitha Lakshmi and Savarimuthu, 2023), measure bird skeletons (Weeks et al., 2023), isolate preserved snakes (Curlis et al., 2022), segment fossils (Panigrahi et al., 2022), or remotely monitor phenology (Mann et al., 2022). However, rather than relying on a single machine learning architecture to extract trait and archival data from specimens, we developed a modular framework of seven different machine learning algorithms that work in tandem to comprehensively process each image (Table 2, Figure 1). We designed LeafMachine2 to emulate the way a human might extract data from a plant specimen—by breaking down a complex problem into multiple discrete steps.

LeafMachine2 is one of several recently developed tools that are aimed at extracting quantitative trait data from herbarium images (for a summary of the current state of machine learning in a herbarium setting see Hussein et al., 2022). Most methods have been semi-autonomous, requiring some form of human intervention to measure traits. For example, TraitEx requires users to draw a border around leaves of interest to aid with segmentation, like some ImageJ workflows, and can take approximately 10 minutes to measure each leaf (Maloof et al., 2013; Kommineni et al., 2021). These workflows rely on computer vision algorithms, typically superpixel or graph cut segmentation, to extract leaf masks (Zhang et al., 2018; Alajas et al., 2021). Manual intervention is required because these methods are "static" and cannot self-adjust to handle variable input (changes in location, size, or color of leaves). As a result, static computer vision methods limit utility in non-ideal scenarios and cannot segment an individual leaf from among a group of overlapping leaves or produce usable results if the leaf is bisected with mounting tape without additional post-processing (Kommineni et al., 2021).

To overcome these limitations, many groups turned to machine learning algorithms, typically some kind of

TABLE 1 Definitions of specialized and abbreviated terms.

Term	Definition
Archival component detector (ACD)	A YOLOv5x6 object detection network trained to place bounding boxes around text, labels, barcodes, rulers, color correction cards, attached items, envelopes, maps, photos, and paper weights. LeafMachine2 uses these predicted bounding boxes to crop components from the full specimen image for component-specific analyses.
COCO format	A standardized JSON format for object instance segmentation annotations widely used in the computer vision community, introduced by the COCO (Common Objects in Context) data set.
Conversion factor (CF)	The image-specific ratio of pixels per metric unit. For LeafMachine2, all conversion factors refer to the number of pixels in an image that correspond to 1 cm.
Convolutional neural network (CNN)	A type of deep learning model primarily used for image recognition and segmentation that uses convolutional layers to filter inputs for useful information. Convolutional layers in a neural network are like a set of digital filters that scan across an image to detect and learn patterns, like how human eyes perceive different shapes and textures.
Detectron2	A popular open-source software system developed by Fundamental AI Research (FAIR) that implements state-of-the-art object detection algorithms, including Mask R-CNN.
Graphics processing unit (GPU)	A specialized type of processor designed for handling the computations required for 3D graphics rendering and machine learning tasks.
Instance segmentation, leaf segmentation	A machine learning task that involves identifying and delineating each distinct object of interest appearing in an image down to the pixel level. Each class can contain numerous instances.
JavaScript Object Notation (JSON)	A lightweight data format (a dictionary) that is easy for humans to read and write and easy for machines to parse and generate. The JSON file for a simple spreadsheet may look like: {“Header 1”: “Value 1”, “Header 2”: “Value 2”}
Large language model (LLM)	A class of machine learning models that are trained on a large corpus of text data, which can generate human-like text based on the input they receive. ChatGPT is arguably the best-known implementation, but there are many variants. This field of machine learning is developing rapidly.
Machine-assisted labeling (MAL)	The process of using automated or semi-automated systems to apply labels to data to expedite the creation of training data sets.
Mask R-CNN	A convolutional neural network–based model designed for object instance segmentation, which both detects objects in an image and generates a segmentation mask for each instance.
Mean average precision (mAP)	A common metric for measuring the accuracy of object detectors like Mask R-CNN and YOLO. mAP averages the precision scores at different recall levels, providing a single summary measure of a model's performance across all threshold levels. mAP is an overall score of how well a system can identify and correctly label objects in an image.
Plant component detector (PCD)	A YOLOv5x6 object detection network trained to place bounding boxes around ideal leaves, partial leaves, leaflets, individual fruits and seeds, groups of fruits and seeds, individual flowers, groups of flowers, buds, roots, wood samples, and all plant material. LeafMachine2 uses these predicted bounding boxes to crop components from the full specimen image for component-specific analyses. Ideal leaves are sent to the PLD and leaf segmentation networks.
PointRend	A module that can be added to existing segmentation models to enhance the edge detection of CNNs, generating more precise and detailed segmentation masks. LeafMachine2 uses PointRend to refine the edges of masks produced by the Detectron2 Mask R-CNN instance segmentation network to retain fine details in a leaf outline, like leaf teeth.
Pseudo-landmark	Landmarks refer to biologically homologous points that consistently represent specific aspects of an organism's morphology (Chitwood and Sinha, 2016; Klein and Svoboda, 2017). We designed LeafMachine2 to look for visually similar prominent points, like tracing points along the midvein, for a wide range of taxa. Therefore, we use the term pseudo-landmarks because the homology of our detected traits is unknown and varies by taxa, although we sometimes use the term “landmark” interchangeably in the text.
Pseudo-landmarks detector (PLD)	A YOLOv5x6 object detection network trained to place fixed dimension bounding boxes at points that correspond to pseudo-landmark locations including apex and base angles, midvein length, lamina length and width, lobe locations, and petiole length.
Ruler conversion	The process of determining the number of pixels between unit markers on a scale bar or ruler, producing a conversion factor that can be applied to pixel-based measurements to yield a metric result.

(Continues)

TABLE 1 (Continued)

Term	Definition
Semantic segmentation	A machine learning task that involves classifying each pixel in an image into a specific category or class.
YOLOv5, YOLOv5x6	The "You Only Look Once" (YOLO) real-time object detection network. The YOLOv5x6 variant offers a larger model size for increased performance for a larger computational cost.

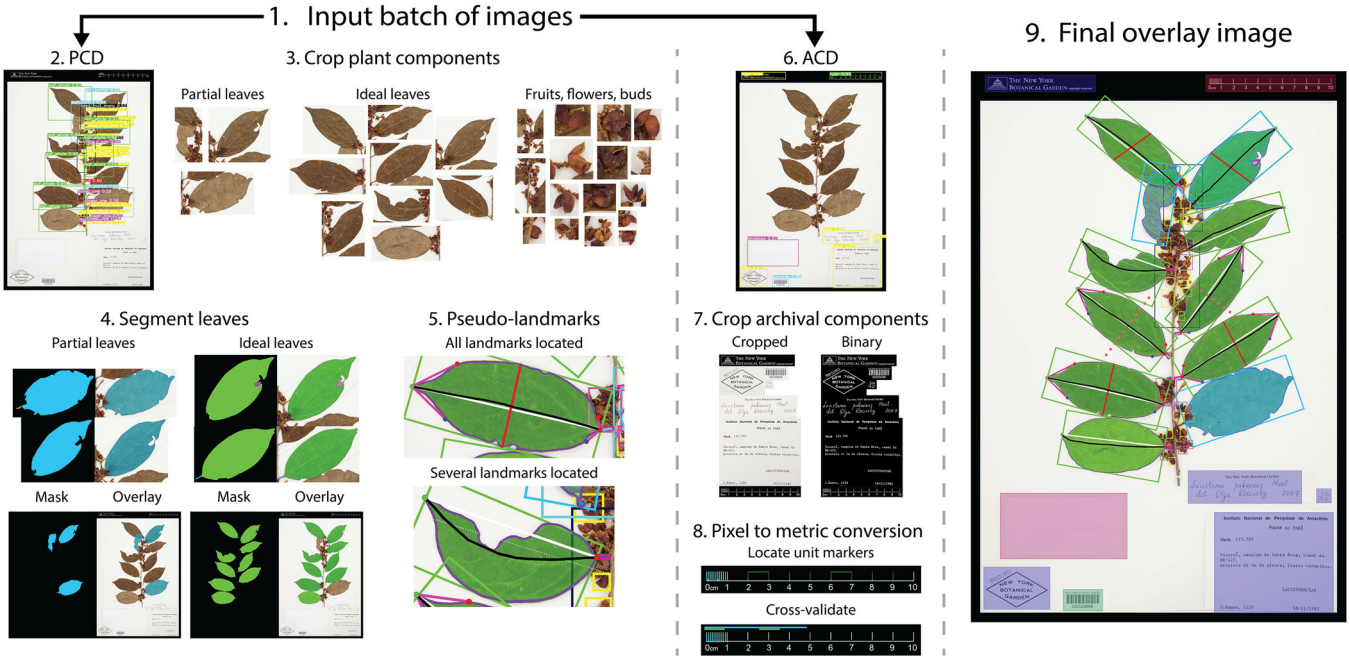

FIGURE 1 LeafMachine2 workflow. A batch of images is processed by the plant component detector (PCD) (2) and archival component detector (ACD) (6) networks. (2) Bounding boxes identifying predicted plant components. Each bounding box identifies a unique component, directing it to the appropriate processing pipeline. (3) The PCD produces cropped images of each plant component. (4) Individual cropped leaves undergo instance segmentation by the Detectron2 network, producing leaf outline masks for ideal leaves (green) and optionally partial leaves (blue). The first set of images shows individual leaves, while the second set shows the compilation of the individual leaves back onto the full specimen image. (5) Cropped ideal leaves are processed by the pseudo-landmarks detector (PLD) and individual landmarks are measured. Please see Figure 2 for a description of each landmark annotation. (6) Bounding boxes identifying predicted archival components. (7) Cropped archival components from the ACD are processed and cleaned into binary images for downstream applications, like optical character recognition (OCR) or interpretation by large language models (LLMs). (8) The cropped ruler image is processed by our scanline or template matching algorithms to identify unit markers. Located tick marks are shown as colored dots. Green and cyan lines indicate the converted 1- and 5-cm distances for quality control purposes. For more information about pixel-to-metric conversion, please see Appendices S2 and S3. (9) The final overlay image shows all machine-derived masks, measurements, and identified components. All the visuals in this figure are sourced directly from the output files produced by LeafMachine2.

convolutional neural network (CNN), which can categorize individual pixels as members of discrete classes (Ott et al., 2020; Weaver et al., 2020; Younis et al., 2020; Triki et al., 2020, 2021; Goëau et al., 2020, 2022; Guo et al., 2021; Love et al., 2021; Hussein et al., 2021b; Gu et al., 2022; Ott and Lautenschlager, 2022; Milleville et al., 2023). For the task of isolating and measuring individual leaves, semantic segmentation algorithms still lack the power to resolve complex situations (e.g., overlapping leaves) because they produce one mask that contains all leaf pixels and require postprocessing to obtain usable results (Weaver et al., 2020; Hussein et al., 2021b, 2022). Instance segmentation algorithms improve on this as they can directly isolate a single leaf from nearby leaves (Guo et al., 2021; Triki et al., 2021). While full-image instance segmentation is promising, it requires substantial effort to create a suitable training data set because every leaf in a training image must be manually segmented. Some groups have implemented human-in-the-loop workflows to manage this task, iteratively winnowing away partial leaves until only training masks for ideal leaves remain (Mora-Fallas et al., 2019).

LeafMachine2 offers an alternative approach by separating the task of identifying ideal and partial leaves from the task of segmenting leaves. We use two YOLOv5 networks to first isolate (place bounding boxes around) plant and archival components and then use an array of component-specific processing tools, including Mask R-CNN instance segmentation, to generate measurements

TABLE 2 Machine learning components. Data sets in parentheses indicate the parent data set, "L" data sets are cropped from full specimen images. For a more detailed description of each machine learning component, see Appendix S1.

Component	Training data sets	ML architecture	ML type	No. of specimens	No. of annotations
Plant component detector	D-5GENUS, D-TARGET	YOLOv5x6	Object detection	3001	321,406
Archival component detector	All D sets	YOLOv5x6	Object detection	5573	101,374
Ruler classifier	R-CLASS	ResNet18	Object classifier	5573	12,242
Ruler segmentation	R-DOC	DocEnTr small 8×8 patch	Semantic segmentation	778	2852
Ruler binary classifier	R-BINARY	ResNet18	Object classifier	778	8622
Leaf segmentation	L-SEG (D-5GENUS, D-TARGET)	Detectron2 + PointRend	Instance segmentation	1183	15,606
Landmarks identifier	L-LAND (D-TARGET)	YOLOv5x6 Object detection	Object detection	1381	32,664
Label segmentation	R-DOC	DocEnTr small 8×8 patch	Semantic segmentation	778	2852
			TOTAL	**5573**	**494,766**

(He et al., 2017; Wu et al., 2019; Jocher et al., 2022). Our modular approach brings several primary benefits that elevate the utility of LeafMachine2, sidestepping several hurdles faced by previous attempts at automated plant trait extraction. First, each network is specialized for a specific task, so fine-tuning performance for edge cases or expanding support to more taxa or specimen preparation styles is more manageable because new training data can remain focused on a discrete task. Second, leaf segmentation and landmark detection algorithms only run on individual ideal leaf candidates, significantly lowering computational requirements while producing exceptional leaf masks. Third, manually generating ground truth segmentation training data is streamlined because humans labeling the images only need to focus on one leaf at a time. This approach also promotes diversity in leaf sampling, as we can subset leaves from each specimen for manual annotation, thus expanding the number and variety of taxa and specimens we can label given the constraints associated with time and funds dedicated to manual annotation.

LeafMachine2 is a broadly useful tool that can enhance a range of research areas, from botany to ecology to agriculture. With LeafMachine2, we support at least six different use cases. (1) LeafMachine2 can measure quantitative leaf traits by identifying and isolating (cropping) ideal and partial leaf candidates. Cropped leaf images are sent to our leaf segmentation algorithm to generate leaf masks, which are used to measure shape properties, including elliptic Fourier descriptors, for every leaf (Neto et al., 2006). Leaf candidates are also processed with our landmark detection algorithm to identify and validate pseudo-landmarks including apex and base angles, midvein length, lamina length and width, lobe locations, and petiole length. (2) LeafMachine2 can detect the presence or absence of plant organs by isolating individual fruits, groups of fruits, individual flowers, groups of flowers, roots, wood samples, and buds, scoring the presence or absence of these traits for most flowering taxa. (3) We currently support the identification of 37 ruler types and can determine specimen-specific CFs for 26 ruler types, with more support in future iterations. (4) LeafMachine2 uses machine learning algorithms to isolate and binarize (clean) text contained within the specimen image, increasing the efficiency and effectiveness of optical character recognition (OCR) and large language model interpretation techniques. (5) By processing a batch of images, LeafMachine2 can screen for the presence of several archival components including attached items and envelopes that may contain tissue or seeds, maps printed alongside specimen labels, photographs attached to specimen sheets, or even paperweights. Fruits and seeds can be identified even if contained by plastic bags. (6) We have also found that LeafMachine2 can be used to generate training data for new machine learning networks because it can be configured to save and record vast amounts of intermediate metadata. We find metadata extremely useful for training other machine learning networks to perform novel tasks or for diagnosing unexpected results. For example, all leaf masks can be exported in the Common Objects in Context (COCO; Lin et al., 2014) format for training instance segmentation algorithms

(Figure 1, section 4). LeafMachine2 is a multifaceted tool with the ability to transform botanical research by streamlining data extraction, organ detection, image processing, and even aiding in the development of new machine learning algorithms.

METHODS

In the following sections, we outline the LeafMachine2 workflow and our training data sets, followed by a description of each of the seven machine learning components. Finally, we test LeafMachine2's performance across angiosperms and end with a discussion of outstanding challenges.

LeafMachine2 overview

LeafMachine2 v.2.1 was developed with Python 3.8.4 and requires PyTorch 1.11 (Paszke et al., 2019) and CUDA 11.3 (Nvidia, Santa Clara, California, USA). We have tested and validated performance on both Windows 10 (Microsoft, Redmond, California, USA) and Ubuntu 20.04 (https://ubuntu.com/) workstations with at least one GPU. We have not tested and do not recommend running LeafMachine2 without a discrete Nvidia GPU. Full installation instructions and source code can be found at https://github.com/Gene-Weaver/LeafMachine2 (see Data Availability Statement). Currently, LeafMachine2 is run from the command line, but users can adjust approximately 100 different configurable parameters with a configuration file. LeafMachine2 is designed to process locally stored images and supports downloading images using Darwin Core Archive (DWC) occurrence and image files. Users can query an online data portal like GBIF, download the corresponding DWC files, and point LeafMachine2 to these files. LeafMachine2 will begin downloading images (parallelized for increased speed) and immediately begin processing the images. With default settings (Appendix S1; see Supporting Information with this article), LeafMachine2 can process 150–200 images per hour, on average, using a consumer-grade GPU with 8 GB of VRAM and at least 8 CPU processing cores, depending on the number of leaves in each image and the input image resolution. The plant component detector (PCD) and archival component detector (ACD) process all images in the project, and the project is then split into batches based on how much system RAM is available. Batches can be parallelized across up to 8 CPU cores for improved performance. For quality control purposes, a summary image showing all identified components and measured traits is produced for every specimen and saved as a page in a PDF (Figure 1, section 9). For each run, LeafMachine2 saves a copy of the configuration file and logs for reference or debugging along with a multitude of configurable output files. Data are exported as a CSV file and can be merged with the parent DWC files.

Specimen training data sets

Training data set development was a major focus of this project. To ensure generalizability, we prioritized taxonomic diversity, institutional diversity, and diverse specimen quality (Table 3). We queried GBIF and downloaded the DWC records for the 7,204,118 preserved Magnoliopsida specimens that had both images and geospatial coordinates (Appendix 1). We sampled these records to create four specimen image data sets (Table 3). To bolster institutional diversity, we also obtained DWC files for 195 herbaria (some were duplicated in GBIF) and included up to 10 randomly sampled images from each herbarium in the data set D-HERB. We chose 51 species, based on their apparent morphological diversity and frequent representation in GBIF, to represent our D-TARGET data set, which includes herbaceous and woody taxa. We randomly sampled 50 images per species to account for intraspecific diversity. For Lecythidaceae, we split the 50 images between two species. The D-5GENUS data set contains up to five randomly chosen species per genus of North American woody perennials (165 genera), one image per species, which adds morphological breadth to our training data set. Overall, our specimen data sets include 5648 specimens representing 2663 species from 288 institutions.

Data set annotation

Manually annotating training data is an arduous and time-consuming endeavor. Our team of seven labelers has logged more than 2000 hours to generate the 494,766 annotations used to train LeafMachine2. To our knowledge, this is the most comprehensive manually annotated training data set for herbarium specimen analysis to date (Hussein et al., 2022). We labeled images using an academic license for the Labelbox platform (https://labelbox.com), which enabled our labeling team to annotate images remotely, programmatically manage large data sets with the Labelbox API, employ machine-assisted labeling (MAL), and review labels. For efficiency, we employed MAL whenever possible. This involved manually labeling enough images to train a rudimentary version of a given machine learning network and then processing another batch of images using the machine learning network to generate annotations that were uploaded into Labelbox for revision. For plant components, this roughly tripled productivity; 15 minutes per specimen was reduced to five minutes. For archival components, however, we saw a seven-fold increase in productivity because the archival labels required substantially less editing. We implemented a comparable method for both segmentation and landmark labels. We utilized the built-in

TABLE 3 Training data sets.

Data set name[a,b]	No. of images	No. of species	No. of herbaria	Train/ validation/test	Resolution[c]		
					Minimum	Average	Maximum
D-HERB	1755	1287	277	80/10/10	573 × 800 (0.5)	3159 × 4637 (14.6)	5000 × 7500 (37.5)
D-3FAM	831	831	65	80/10/10	2927 × 5000 (14.6)	3924 × 5669 (22.2)	5000 × 7500 (37.5)
D-5GENUS	562	562	47	80/10/10	2960 × 5000 (14.8)	4024 × 5803 (23.4)	5000 × 7500 (37.5)
D-TARGET	2500	51	75	80/10/10	2610 × 3781 (9.9)	3786 × 5644 (21.4)	5232 × 7500 (39.2)
Total unique	**5648**	**2663**	**288**				
R-CLASS	12,242	12,242	277	80/20/0	19 × 127 (0.002)	170 × 1337 (0.23)	924 × 7360 (6.8)
R-BINARY	8622	8622	277	80/20/0	19 × 127 (0.002)	170 × 1337 (0.23)	924 × 7360 (6.8)
R-DOC	2852	133,801	277	80/10/10	19 × 127 (0.002)	170 × 1337 (0.23)	924 × 7360 (6.8)
L-SEG	5105	298	71	80/10/10	13 × 27 (0.001)	499 × 768 (0.38)	3980 × 4848 (19.3)
L-LAND[d]	5761 (2132)	202 (15)	42	80/20/0	16 × 31 (0.001)	525 × 808 (0.42)	3553 × 4749 (16.9)

[a]"D" data sets are full herbarium specimen images. "R" data sets are rulers cropped from full herbarium specimen images. "L" data sets are ideal leaves cropped from full herbarium specimen images.

[b]D-HERB = data sets containing institutional diversity; D-3FAM = up to three random species from 341 families, one image each; D-5GENUS = one random image per species, up to five random species per genus, for 165 genera of North American woody perennials; D-TARGET = select group of 51 species of herbaceous and woody angiosperm species, 50 images each; R-CLASS = cropped ruler images from data set D-HERB; R-DOC = binary image subset of data set R-CLASS, up to 50 images per ruler class; R-BINARY = same images as data set R-DOC but prepared for two-class prediction (i.e., pass or fail); L-SEG = data set of segmented leaf, petiole, and leaf hole masks; L-LAND = subset of cropped ideal leaves from data set D-TARGET for landmark detection.

[c]Image resolutions are reported in pixel dimensions with parentheses around the approximate megapixels.

[d]Numbers in parentheses for data set L-LAND report counts excluding the 188 Icacinaceae species that are part of the landmarking data set.

segmentation tools of Labelbox to create segmentation masks, which we further refined to produce high-quality training masks. With this labeling workflow, it is relatively simple to add support for new classes and export the annotations to retrain a custom version of our component detectors.

Labeling procedure

Our staff labelers were trained to identify nine archival components and 11 plant organs (Table 4). Archival components were mostly uncontroversial, due to the relative homogeneity of class instances when compared to plant components. We bounded rulers tightly, minimizing white space and focusing on unit markers. For the label class, we annotated all text found within the image, except for text within a barcode bounding box and ruler unit labels. If we encountered long text blocks, we split the text section between multiple bounding boxes based on the principle of minimizing white space. We differentiated between envelopes and attached items based on their visual appearance, even though both classes serve the same utility (i.e., containing loose tissue). Extraneous objects were either labeled as weights or ignored. The diverse range of plant organs, combined with the broad taxonomic scope of our training data set, posed significant challenges. Several plant organ instances defied straightforward classification, presenting us with difficult decisions. For example, some specimens are collected primarily for their floral or fruit characteristics and often display mixed stages of reproductive development. This necessitated subjective decisions regarding the classification of intermediate developmental stages, which is likely the root cause for much of the uncertainty in the final PCD.

Component detection

Archival component detector

Specimen vouchers typically contain additional archival components in addition to plant material including barcodes, labels, text, and rulers. Data contained within barcodes and labels may already be present in the specimen's DWC record, but the global backlog of millions of non-databased specimens leaves room for computational assistance (Davis, 2022; Hardisty et al., 2022; Heberling, 2022). One approach is to isolate archival components from the full image. Working with smaller images improves the performance of downstream processes and improves the efficiency of label transcription by humans, or soon by large language models (LLM). Isolating rulers also enables image-specific CF determination. We use the ACD for this task, which is a YOLOv5x6 object detector that is modified to support nine classes and our bounding box dimensions (Jocher et al., 2022). To maintain broad generalizability, our training data set was drawn from 288 herbaria and included

TABLE 4 Annotation counts. Total number of ground truth annotations per class sorted by machine learning component.

Type	Annotation	Count
Plant components	Ideal leaf	41,748
	Partial leaf	90,607
	Leaflet	70,665
	Seed/fruit one	24,573
	Seed/fruit many	1356
	Flower one	56,601
	Flower many	6388
	Bud	22,233
	Specimen	6299
	Roots	895
	Wood	41
	Total	**321,406**
Archival components	Label	59,880
	Ruler	14,045
	Barcode	13,399
	Color card	8628
	Envelope	3430
	Attached item	943
	Photo	70
	Weights	680
	Map	299
	Total	**101,374**
Landmarks	Lamina tip	3247
	Lamina base	3325
	Petiole tip	2924
	Lobe tip	7507
	Width	3323
	Midvein trace	3322
	Petiole trace	2771
	Apex angle	3089
	Base angle	3156
	Total	**32,664**
Segment	Lamina	9709
	Petiole	1710
	Hole	4187
	Total	**15,606**
Ruler	R-CLASS	12,242
	R-BINARY	8622
	R-DOC	2852
	Total	**23,716**
	Total annotations	**494,766**

some non-standard specimen images like book pages, cleared leaves, and field images (Table 3). We trained the ACD in three stages, following MAL procedures, for a total training duration of 600 epochs (~300 h) on an Ubuntu system with 128 GB of system memory and two Nvidia Quadro P6000s with a total of 48 GB of VRAM. All LeafMachine2 networks were trained on this machine. We achieved a final mean average precision (mAP) of 94.7% and recall of 90.3%. For more detailed training information, see Appendix S1.

Plant component detector

To locate plant components, we employ another network using the same architecture as the ACD but trained to isolate 11 common plant organs (Table 4). The PCD was trained to bin reproductive structures into four categories: "fruit" (e.g., acorn, hickory nut), "fruits" (e.g., a fruiting cluster of grapes), "flower" (a single flower), and "flowers" (an inflorescence). Where possible, we annotated individual flowers or fruits within a larger fruiting cluster or inflorescence. In LeafMachine2 v.2.1, we treat leaflets as simple leaves. Our PCD was trained to identify and isolate compound leaves (pinnate, bipinnate, and palmate) and individual leaflets, but we found that the PCD would frequently classify simple leaves as leaflets. In future iterations, it may be possible to include several PCD networks, each trained to classify one leaf type.

The PCD was trained in two stages, first with half of the images from data set D-TARGET, then with data sets D-TARGET plus D-5GENUS. After training the PCD for 450 epochs, we achieved a mAP of 45.2% and a recall of 40%. While these metrics are substantially lower than the ACD, the data set is significantly more heterogeneous and both mAP and recall will improve as we add more specimens to our training data sets. Tracking performance metrics is helpful for selecting the best-trained network, but we are most interested in generalizability and how consistently the PCD performs as part of the LeafMachine2 framework for taxa not represented in the training data sets. The ACD and PCD networks are responsible for feeding cropped images to downstream processes. If the PCD fails to identify a leaf, then that leaf will not be segmented or processed for pseudo-landmarks.

Processing rulers and labels

Obtaining an accurate CF is crucial for utilizing trait measurements obtained from digital images. The three most common methods for obtaining CFs are (1) using a tool like ImageJ or TraitEx to manually place two points in an image to capture a known distance (Maloof et al., 2013; Kommineni et al., 2021), (2) including a high-contrast object of known dimensions into images and then extracting its pixel dimensions in post-processing (Easlon and Bloom, 2014), or (3) correlating a known metric distance with image resolution given rigid imaging procedures (Weeks et al., 2023). These methods are either labor intensive or require a uniform imaging environment, which is a serious impediment when processing herbarium specimens at scale. Few dynamic methods that rely on machine learning and computer vision techniques have been developed to automate CF determination. One study used an object detection algorithm to locate the number "2" and the number "3" on rulers to compute CFs, but was limited to only two ruler types; the authors advocated for an approach that located tick marks directly (Karnani et al., 2022). With LeafMachine2, we required a more generalizable procedure for obtaining image-specific CFs. While developing this project, we observed significant discrepancies in ruler quality between herbaria. Some herbaria place high-quality, high-contrast, machine-readable rulers in their images, while others used "rulers of convenience," or even toys (Appendix S2: "Interesting rulers and failed conversions"). Moreover, we also observed faded, bent, damaged, occluded, and poorly imaged rulers that make it difficult, if not impossible, to autonomously obtain CFs. We encourage herbarium curators to scrutinize our CF methods and results and hope that future digitization efforts make use of high-contrast machine-readable rulers (Appendix S2).

To meet this need, we developed a system of three machine-learning networks to preprocess rulers, increasing the performance and precision of calculated CFs. The previously described ACD first places a bounding box around a ruler's unit markers, minimizing unwanted text or noise; rulers are then sent to an image classifier to predict the ruler type so that unit markers can be interpreted appropriately. To obtain a CF, we first convert the ruler into a binary image where unit markers are white and everything else is black. We binarize each ruler in three different ways: threshold sweep, segmentation (DocEnTr; Souibgui et al., 2022), and skeletonization. Finally, we use another machine learning network, an image classifier, to determine whether the binarization was successful. The stack of three binary images is processed by our scanline or template matching algorithms to identify and cross-validate distances between unit markers. We compute the mean CF from all identified unit markers (typically hundreds of points) and overlay 1-cm, 5-cm, and 1-inch reference lines in a summary image for quality control. For a more detailed description of this procedure, please refer to Appendix S3.

Leaf segmentation

We designed LeafMachine2 such that ideal leaves and, optionally, partial leaves (Figure 1, blue masked leaves) cropped from the full image by the PCD undergo instance segmentation individually. This works similarly to the way that one can blur the background in a Zoom call, but in this case, our goal is to separate the foreground (leaf) from everything else. To obtain training images, we processed data sets D-5GENUS and D-TARGET with the PCD and randomly sampled 5105 ideal leaves from among more than 15,000 leaves identified by the PCD. Our labelers used the Labelbox auto-segment tool as a starting point and then manually refined the ground truth masks, including segmenting each leaf's petiole and any internal holes. Critically, if an extraneous object intruded inside the leaf outline, we included the obstruction in the mask with the goal of teaching the network how to properly segment tricky leaves. We used the Labelbox API to export the annotations and converted them to the COCO format in preparation for training (https://cocodataset.org). After training a modified Detectron2 implementation of Mask R-CNN, enhanced with PointRend for greater mask precision, for 100,000 iterations, the network achieved a Mask R-CNN training accuracy of 99.1% and a PointRend training accuracy of 95.7% (He et al., 2017; Wu et al., 2019; Kirillov et al., 2020). For more training information, see Appendix S1.

Pseudo-landmark detection

Segmentation masks are useful for measuring traits that are derived from a leaf's outline but lack the information necessary to measure many other distinguishing structural traits like venation, angles, or interior distances; they are also susceptible to errors when measuring lamina length and width (Ellis et al., 2009). In Figure 2, Leaves E, G, and R demonstrate that some trilobed and deltoid leaves confound minimal bounding box algorithms, resulting in incorrect laminar dimensions for length and width. Similar problems occur with curved leaves. In Figure 2, Leaf T shows that tracing the midvein (black line) provides a more refined measurement of lamina length. To bridge this performance gap, we developed a pseudo-landmarks detector (PLD) using the same YOLOv5x6 machine learning architecture as described for both the ACD and PCD. As a starting point, we identified a set of pseudo-landmarks shared by most broadleaf plants that include the lamina tip and base, lamina width, apex and base angles, midvein length, petiole length, and lobe locations. Our labeling team annotated this landmark set on 2132 cropped leaf images representing 15 species selected from data set D-TARGET, along with 3629 images representing 188 Icacinaceae species, totaling 5761 images and 202 species. We converted these points to fixed dimension bounding boxes, ranging in size from 9–27 pixels depending on image resolution, and trained the PLD for 200 epochs using the same settings as with our PCD.

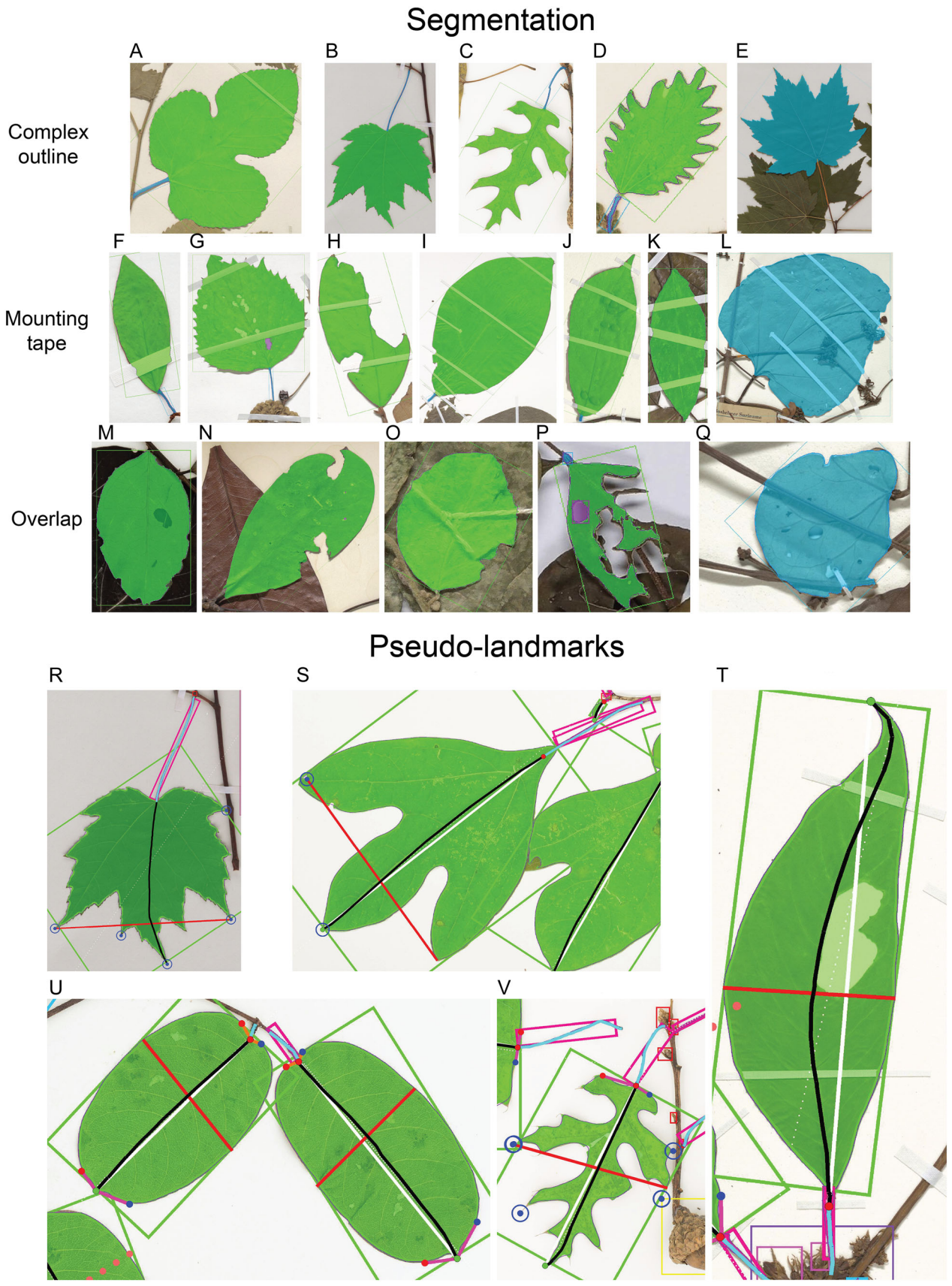

FIGURE 2 (See caption on next page).

RESULTS

Testing ruler conversion performance

Using default LeafMachine2 settings, we processed all 10,619 cropped ruler images in the R-CLASS data set that were used to train the ruler classifier and visually assessed the ruler summary images (Figure 1, section 8) to determine the proportion of correct CFs by ruler class. Correct CFs were determined for at least 80% of ruler occurrences for seven ruler classes and were correct for at least 50% of ruler occurrences for 16 ruler classes (Figure 3A, colored boxes). Overall, the best-performing rulers were high-contrast metric rulers.

Ruler conversion validation

For a more detailed test, we processed data set D-3FAM, which contained 20 ruler types (represented by colored dots in Figure 3B), using default LeafMachine2 settings to test the performance of our CF determination algorithms with a simulated real-world data set. These rulers were part of the ruler classifier training data set but were "unseen" to our template matching and scanline algorithms. To serve as a baseline, our image labeling team manually placed points on every unit marker for each unit type present in each of the 1274 rulers in data set D-3FAM. This is more rigorous than typical manual methods where only two points are placed at the beginning and end of a known distance (Weaver et al., 2020; Kommineni et al., 2021). For each manually annotated ruler, we calculated the standard deviation of the points used to determine the 1-cm CF, and the resulting average residual standard deviation (avgRSD) of 0.8% served as the baseline for acceptable performance. Figure 3B shows the results of a *t*-test between the manually labeled and autonomously generated CFs. We removed known unsupported rulers prior to the *t*-test. If an autonomous CF had a pooled standard deviation greater than 10%, then the ruler class was deemed to be unsupported, leaving 708 of the original 1275 rulers. The *t*-statistic was 2.784 and the *P* value was 0.006, clearly demonstrating that there is room for improvement, as our autonomous methods generated significantly different CFs compared to the manual method. However, there are a few notable takeaways. First, 58.8% of the autonomous CFs fell within the bounds of the avgRSD, with most belonging to Rulers 2 and 7. Ruler 2 is used by the New York Botanical Garden (NYBG), and Ruler 7 was developed for the Global Plants Initiative and is widely used by many herbaria. Second, of the best-performing rulers (Figure 3B, indices 0–500), most of the outlier values correspond to poor-quality images and display either low resolution, bad lighting, or damaged rulers. Poor image quality translates to inconsistency. Of the best-performing autonomously determined CFs (Figure 3B, indices 0–400), the pooled standard deviation for each ruler is lower than the corresponding manually determined CF. With algorithmic refinement, this consistency can likely be extended to more ruler types. Third, identifying unit markers directly (e.g., using a modified version of our landmark detection algorithm) would likely improve consistency, particularly for poor-quality images, as this would bypass the need to create binary ruler images, which were a common failure point.

Qualitative performance of leaf segmentation

In Figure 2, leaves A–Q, we demonstrate LeafMachine2's segmentation ability in difficult circumstances; all leaves in Figure 2 are from data set D-3FAM and were not part of the training data set. These leaves were selected as exemplars, but more than 8000 ideal leaf segmentations extracted from the D-3FAM test data sets can be viewed at https://zenodo.org/record/7764379 (see Data Availability Statement). Exceptional masks are produced for a wide variety of leaf shapes, even for lobed and toothed taxa. LeafMachine2 successfully ignores mounting tape and returns complete leaf masks, bypassing the need for shape matching or connected component analyses as is required by other methods (Hussein et al., 2021a). Accurate segmentation of individual leaves is possible from a group of leaves, even when obstructions are present (Figure 2, leaves E, K–Q). Green leaf masks indicate an ideal leaf candidate, as predicted by the PCD, while blue masks indicate partial leaves. As seen in Figure 2, partial leaves can also produce usable leaf masks, which is another benefit of our modular approach of decoupling leaf identification from leaf segmentation. Users can take advantage of this depending on the project requirements. If the data set is large and it is preferable to minimize the data curation workload, then users can restrict LeafMachine2's PCD by using a high confidence threshold (90%) and opting to only segment

FIGURE 2 Segmentation and pseudo-landmark examples. All leaves are from the D-3FAM data set and were not part of the segmentation of landmarking data sets. Ideal leaves, as predicted by the PCD, are green masks while partial leaves are blue masks. (Leaves A–Q) A sample of leaves demonstrating segmentation performance when leaves have complex outlines; these are obstructed by mounting tape, overlapping leaves, or a combination of obstructions, notably leaves L, P, and Q. (Leaves R–V) A sample of leaves showing pseudo-landmark performance. For landmark overlay images, the red line is lamina width, the cyan line traces the petiole, the solid black line traces the midvein, the dotted white line is the line of best fit for the points that comprise the midvein, the solid white line is the base to tip length, blue bullseye points are lobe tips, pink angles are less than 180 degrees, orange angles are reflex angles greater than 180 degrees, the green dot is the lamina tip, and the solitary red dot is the lamina base. Green bounding boxes are the minimal rotated bounding box. Petioles are either pink or orange masks, and holes are purple. Leaf V shows bounding boxes around fruit and buds.

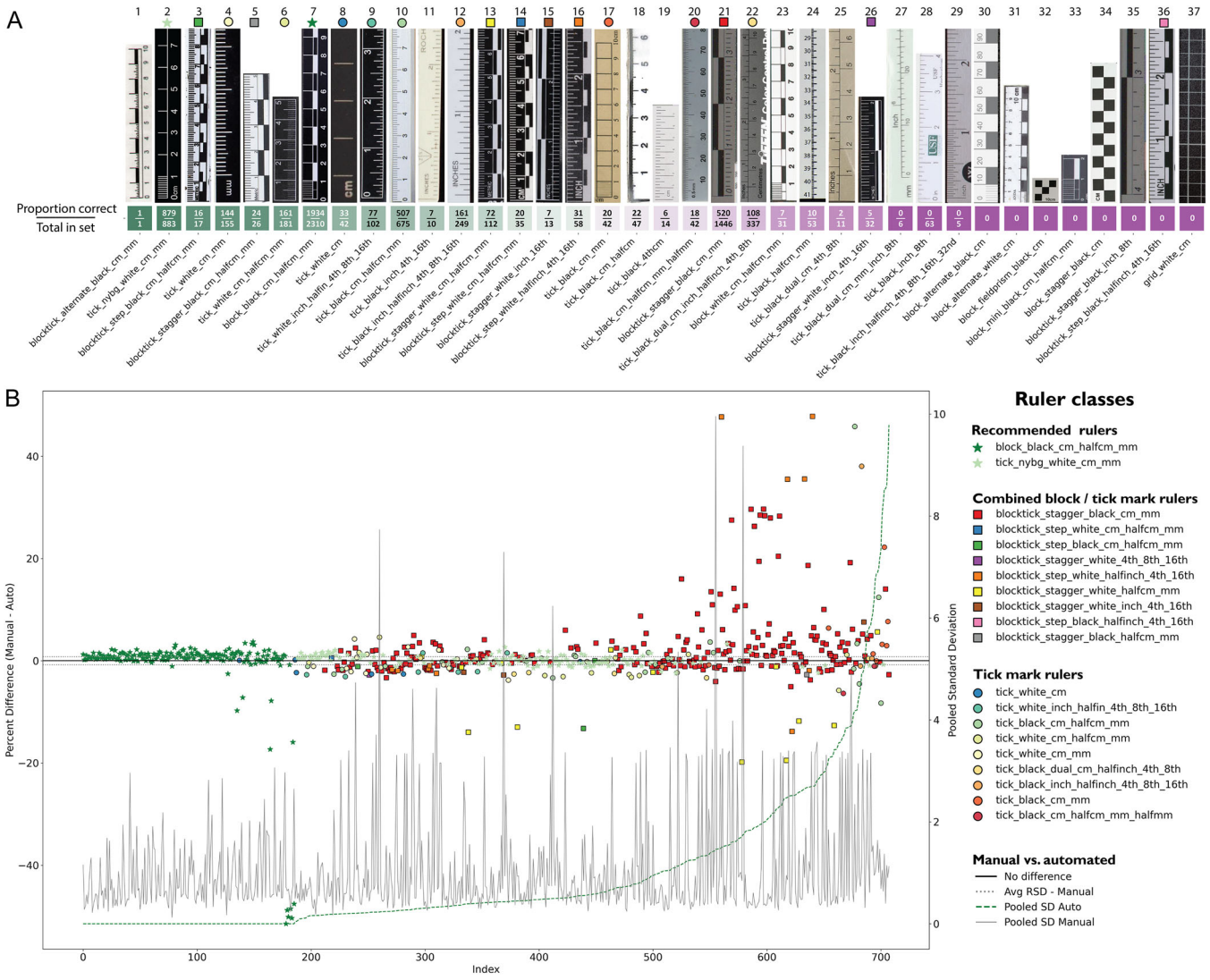

FIGURE 3 Ruler conversion performance. (A) The 37 ruler types that our ruler classifier was trained to recognize, arranged from best performing to worst (left to right). Rulers 30–37 are block-based rulers that can be identified but not converted; however, they are well-suited for our template-matching procedures and will be supported in future iterations. The colored boxes below each ruler correspond to the conversion factor (CF) determination success rate within the data set R-CLASS. The numerator is the proportion visually assessed to be a correct conversion based on the quality control output (see Appendix S2, images 1–38), and the denominator is the total number of rulers of that class present in the data set R-CLASS. Rulers with a zero can be identified by the ruler classifier but were not present in R-CLASS. Colored shape identifiers are placed above each ruler image for the ruler classes that are present in both data sets R-CLASS and D-3FAM. (B) Results of a *t*-test between manually obtained CFs and autonomously generated CFs for 708 rulers in the test data set D-3FAM. The *y*-value of each point is the percent difference from the manually converted CF (left *y*-axis). Points are sorted by autonomous CF pooled standard deviation, with lower values to the left and higher values to the right (right *y*-axis). Inconsistently converted rulers have higher index values, while consistent rulers have lower index values. Accurate autonomous conversions fall between the average residual standard deviation (avgRSD) dotted lines. The two recommended ruler types (rulers 2 and 7) are denoted by green star-shaped markers.

ideal leaves. This will yield comparatively fewer leaves, but these will be of high quality and high confidence; therefore, this option is best suited for large-scale projects. If the priority is maximizing leaf extractions, then both ideal and partial leaves can be segmented and measured at a lower PCD confidence threshold (10%); this option is also useful for data sets that contain damaged or incomplete leaves that would otherwise be overlooked. Figure 4 illustrates changing the PCD confidence threshold; higher confidence values return few leaves, but the quality of mask segmentation remains unchanged in the leaves that are identified at all confidence levels, which is a benefit over traditional Mask R-CNN implementation where the confidence of a leaf identification is linked to the quality of the generated mask. This is true not only for herbarium specimens, but also for field images processed with FieldPrism (Weaver and Smith, 2023), the Leafsnap data set (Kumar et al., 2012), and even iNaturalist-style photographs (Figure 4). Given these promising results, we will continue to explore use-cases beyond standard herbarium vouchers.

For each segmented leaf, we automatically calculate standard shape metrics including area, perimeter, convex hull,

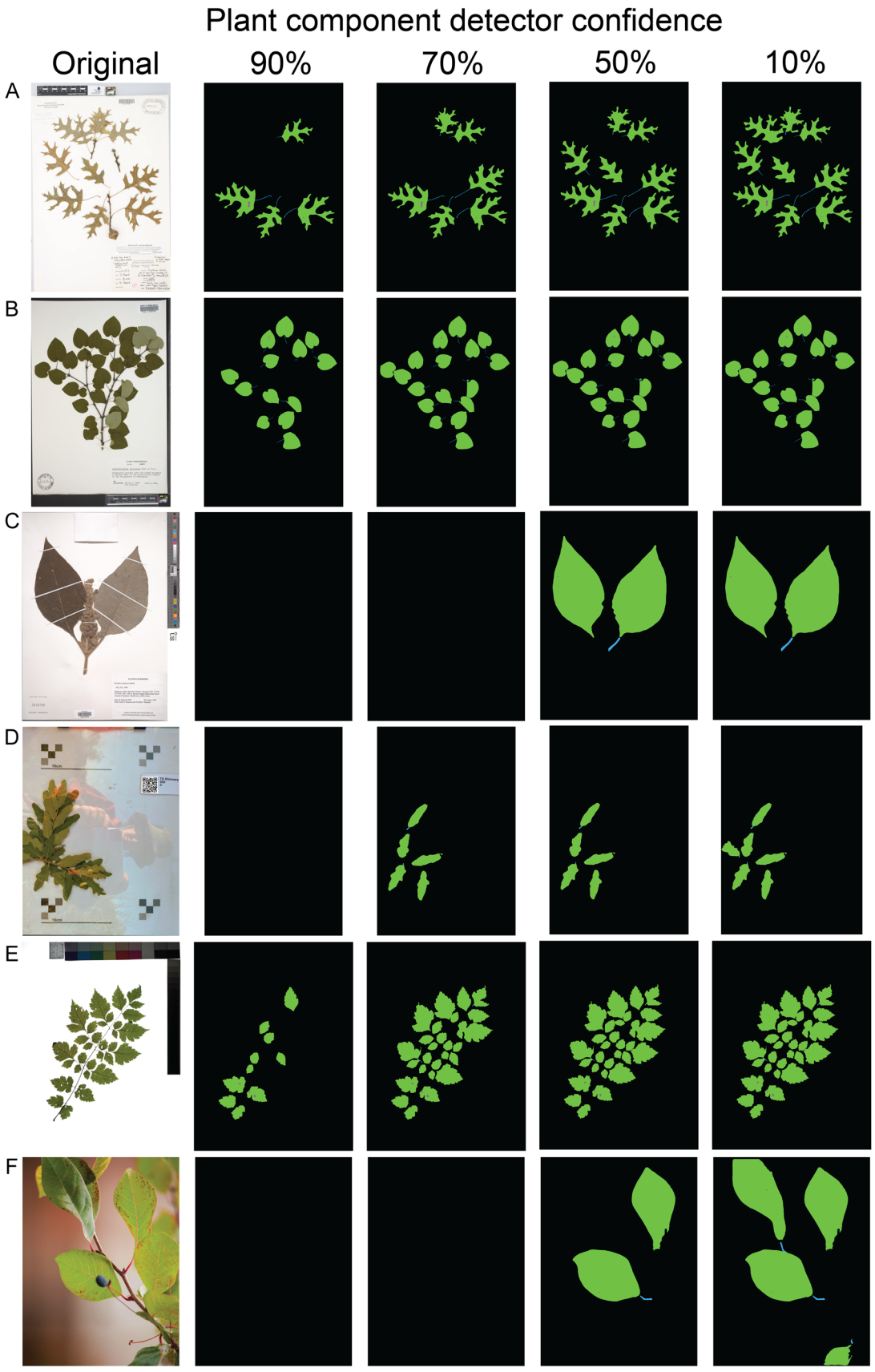

FIGURE 4 (See caption on next page).

length and width (by fitting a rotated bounding box based on the diameter of a minimum bounding circle), centroid, convexity, concavity, circularity, and aspect ratio. Users can optionally set LeafMachine2 to calculate elliptic Fourier descriptors for each leaf outline, which are plotted as a dark purple line in all summary images (Figure 1, sections 5 and 9). LeafMachine2 can also calculate measurements for segmented petioles and can attempt to locate holes in the lamina, although the latter has shown inconsistent results. Meineke et al. (2020) demonstrate a more nuanced approach for identifying leaf holes and damage that could be incorporated into LeafMachine2 in future iterations.

Qualitative performance of pseudo-landmark detection

After training for 200 epochs, our PLD achieved a final mAP of 20.9% and a recall of 29.6%. These values are underwhelming but were expected. Our labelers may have only placed 20 points along a midvein, and the PLD is meant to replicate those exact 20 points. However, the pixel information that describes the midvein between each ground truth point appears nearly indistinguishable from the pixel information of actual ground truth points. If the PLD predicts the location of a midvein point 10 pixels to the left of the ground truth point, it still lies on the midvein. Therefore, while the PLD may suggest reduced certainty, we found the results to be quite usable (Figure 5A, middle column). Unfortunately, with some landmarks (e.g., lamina width), we observed such low confidence due to this behavior that the PLD was overly conservative and often refrained from making a prediction (Figure 1, section 5, bottom image).

Testing LeafMachine2 across angiosperms

To stress test LeafMachine2, we assembled a test data set (D-3FAM) consisting of one image for up to three species per angiosperm family. Both the species and images were randomly chosen. We did not reject any images as we wanted to see how well LeafMachine2 performed on a set of uncurated images. D-3FAM contained 831 images from 65 herbaria representing 831 species from 341 purported angiosperm families; 51 of the D-3FAM families were also present in the training data set, although none of the images were shared between data sets. We did not clean or curate taxonomy. With this test data set, we wanted to gain deeper insight into LeafMachine2's performance on unseen taxa (not in the training data set), assess its real-world utility, and uncover common pitfalls as we continue to develop algorithms and expand training data sets. We used default settings to process D-3FAM; notably, only ideal leaves with a 50% PCD confidence threshold were segmented and landmarked (Figure 4). Below, we describe our qualitative assessment of LeafMachine2's ability to segment leaves, identify pseudo-landmarks, and identify the presence or absence of organs. Summary images for this run can be viewed at the previously mentioned Zenodo repository (see Data Availability Statement).

Scoring phenology

Reproductive structures among flowering plants are highly heterogeneous, and our training data set does not adequately capture their immense diversity. Even so, we found that LeafMachine2 performed beyond our expectations and managed to accurately score the presence or absence (although not necessarily every instance) of all plant organs present in the image for 245 of the 341 families in D-3FAM (Figure 5A, green boxes). Lowering the PCD confidence to 10% detects still more occurrences of non-laminar organs at the cost of class accuracy. The PCD can also successfully isolate small structures like vegetative and reproductive buds, which has been challenging for other methods (Triki et al., 2020).

Identifying and segmenting leaves

We further scrutinized a random sample of 100 images from D-3FAM by counting instances of failure and success. Within this sample, we counted 953 true-positive leaf masks, of which 118 (12.4%) were incomplete or spilled beyond the actual leaf edge. It is important to note that true-positive leaf masks first depend on the PCD to isolate the leaf and then the segmentation algorithm to accurately extract the leaf outline. We also observed 54 instances (5.6%) where LeafMachine2 segmented an object that was not a leaf, typically a flower or miscellaneous vegetative material. This is in line with our training metrics, which showed a 6% error rate for confusing leaves with non-leaf objects (Appendix S1). LeafMachine2's PCD failed to identify 43 leaves that met our definition of an ideal leaf and failed to fully capture the entire outline of six leaves (0.6%), leading to an incomplete segmentation. Of the 953

FIGURE 4 Leaf detection with archival and non-archival data sets, with varying PCD confidence. The left column is the original image. Ordered by decreasing levels of PCD confidence from left to right are full image masks of ideal leaves (or leaflets). (A) Herbarium voucher of *Quercus coccinea* (Fagaceae). (B) Herbarium voucher of *Pilostyles blanchetii* (Apodanthaceae). (C) Herbarium voucher of *Brookea tomentosa* (Plantaginaceae). (D) FieldPrism-processed field image of *Quercus havardii* (Fagaceae) (courtesy of the Morton Arboretum). (E) Leafscan image of *Koelreuteria paniculata* (Sapindaceae). (F) iNaturalist-style photograph of *Nyssa sylvatica* (Nyssaceae) (photo credit William Weaver).

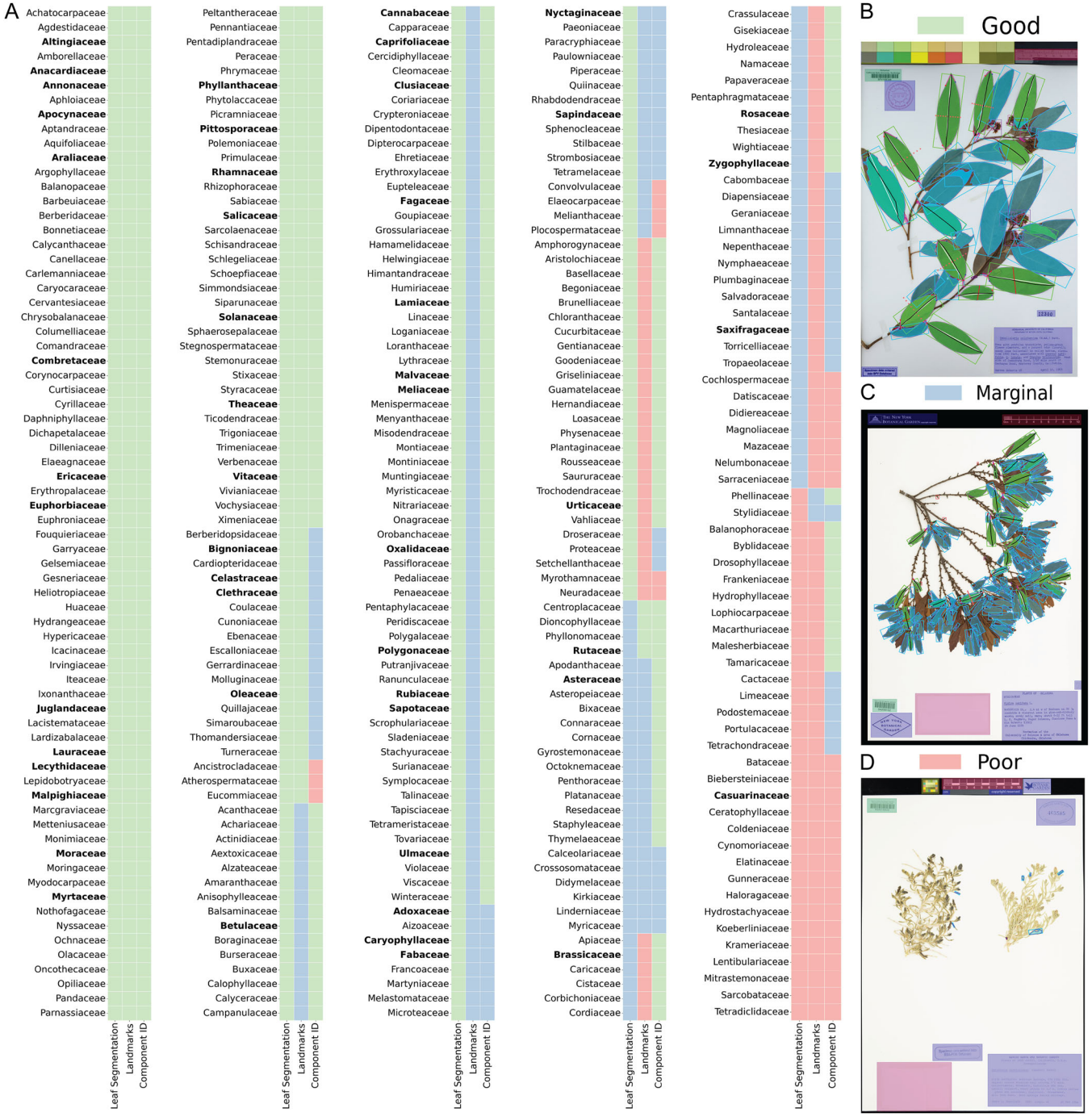

FIGURE 5 Qualitative performance of LeafMachine2, by family and task, across 341 plant families, as identified by the home herbaria. We visually inspected LeafMachine2's quality control summary images for the 831 species/images in the D-3FAM test data set produced with default settings and a PCD confidence of 50%. (A) We followed a power ranking scheme to assign qualitative ratings to families with more than one image, conservatively rounding down in the case of split ratings between the images. For leaf segmentation, a "good" rating indicates that most leaf masks are high-quality, a "marginal" rating indicates that usable masks are present but require manual filtering, and a "poor" rating indicates that no usable masks are present. For landmarks, a "good" rating indicates that at least one usable and accurate landmark skeleton was present, a "marginal" rating indicates that only partial landmark skeletons were present, and a "poor" rating means that no landmarks could be identified. For component identification, a "good" rating means that LeafMachine2 scored the presence of all non-laminar organs, but not necessarily all instances of each organ. A "marginal" rating indicates that some non-laminar organs were not identified, while "poor" means that LeafMachine2 misidentified or failed to identify most non-laminar organs. Bolded families were included in the LeafMachine2 training data set. (B) An image of *Umbellularia californica* (Lauraceae) as an example of "good" ratings in all categories. (C) An image of *Morella cerifera* (Myricaceae) as an example of "marginal" ratings in all categories. (D) An image of *Sarcobatus vermiculatus* (Sarcobataceae) as an example of "poor" ratings in all categories.

true-positive leaf masks, LeafMachine2 located at least four pseudo-landmarks for 200 leaves (21%), corresponding to the "good" category in Figure 5. We attribute this low success rate to the minimal taxonomic diversity in the L-LAND data set and to the drawbacks of our PLD, as previously discussed. With future iterations, the inclusion of more ground truth points or increased bounding box dimensions for high-resolution images could result in a higher success rate. Modified pose estimation or facial recognition algorithms are also promising for this task.

DISCUSSION

LeafMachine2 offers unique capabilities that allow for the extraction of quantitative trait measurements from a broad range of taxa. Our PCD focuses on identifying and isolating leaves with complete outlines (i.e., ideal leaves), including those bisected by tape or stems but excluding those partially concealed by other leaves or objects. Each ideal leaf undergoes processing by our leaf segmentation algorithm, generating an outline mask. This allows for the measurement of multiple traits (e.g., area, convex hull, perimeter, length, width, centroid, convexity, concavity, circularity, aspect ratio, lobedness, toothedness), as well as the calculation of Fourier descriptors. Moreover, our PLD also processes each ideal leaf, enabling the measurement of pseudo-landmarks that are challenging or impossible to determine from an outline mask, including tracing petiole and midvein lengths, counting lobes, and measuring apex and base angles. LeafMachine2's workflow modularity allows for specific task optimization while offering the tools necessary to measure a fundamental set of traits for a wide range of angiosperm species (refer to Figure 5). Below, we outline future adaptations and suggested improvements to further increase the reliability and scope of LeafMachine2's capabilities.

Extending LeafMachine2 methods

Our PLD methods are quite flexible and could be readily adapted to more specialized and focused applications. For example, we experimented with detecting other features including sinus angles for oak leaves and measuring prickle and spine dimensions for *Acacia* specimens. For these tasks, we manually labeled a relatively small number of images, about 200 for each set, to serve as new training data. In both cases, we leveraged transfer learning by replacing the final classification layer with the new classes while retaining the original weights of the PLD algorithm to aid with generalizability. This principle could be extended to other scenarios to yield more consistent results, such as identifying genus-specific landmarks by drawing training images from a single genus. Future iterations of LeafMachine2 will include PLD versions trained on more taxa and specialized versions for the detection of taxa-specific traits.

Expanding the training data set

While the versatility of LeafMachine2 is demonstrated in Figure 5, we also see that LeafMachine2 would be improved by a focused expansion of training data for poor-performing taxa (families with mostly red or blue boxes). Future sampling strategies should target more herbaceous and non-woody taxa to bolster the quality of measurements for taxa that display morphologies that are underrepresented in the current training data set. When inspecting the 8000 leaves generated for our validation test, we observed a bias toward small leaves. While we sampled taxa uniformly, the number of leaves present on specimen sheets varies. Taxa with small leaves are disproportionately represented in our PCD training data set and are therefore more likely to be correctly identified by the PCD when processing new images (Appendix S1, Figure H); future sampling strategies should include more specimen images for large-leaf taxa to compensate. At times, leaves that could potentially be segmented successfully are not forwarded to the segmentation algorithm (Appendix S1, Figures J and K). Additional information relating to sampling and training biases can be found in Appendix S1.

Recommendations for future digitization efforts

While developing automated algorithms for LeafMachine2, we observed several specimen preparation and imaging practices that negatively affect the quality of quantitative trait measurements. First, the rulers used by herbaria for archival digitization are not broadly standardized, which presents challenges for projects that aim to extract quantitative trait data. While parsing our results, two rulers stand out for their reliable machine readability (i.e., the consistency with which pixel distance is converted into metric distance): the ruler used by NYBG (Figure 3A, ruler 2) and the JSTOR Plants ruler (Figure 3A, ruler 7). These rulers are high contrast, simple, and provide unit markers for unit cross-validation. Other ruler types may have these features but can be overly complex or detailed (Figure 3A, rulers 21 and 22; Appendix S2). Curators should consider the machine readability of rulers and how automated systems might interpret them. For future digitization efforts, we recommend that curators adopt either the NYBG or JSTOR ruler types and permanently affix rulers to the copy stand so that the ruler is rectilinear with the camera, contains no reflections, and is *not* on top of the specimen sheet. The inclusion of multiple rulers or rulers with both metric and imperial units should be avoided. Please see Appendix S2 for additional ruler recommendations and a discussion of challenges associated with different ruler types.

Second, specimen sheets for taxa with compound leaves were often too cluttered to be usable for anything other than leaflet measurements. We found that few specimen sheets

present solitary compound leaves, but instead present interlaced and overlapping compound leaves, making the identification of the whole leaf too challenging for existing algorithms. This task is better suited for a more discrete analysis where the data set is curated or created to include only solitary leaves, which would then require only minimal adjustment of the existing LeafMachine2 algorithms. For future collection efforts of taxa with compound leaves, we recommend that accessions include examples of solitary compound leaves or are arranged in a way that presents at least one solitary leaf, where possible.

Looking ahead

We demonstrate that LeafMachine2, and similar projects, are beginning to reduce the phenological trait acquisition bottleneck in biological research but are also introducing a new challenge—the curation of machine-derived trait measurements. Machine learning tools and their generated data are increasingly commonplace in many research fields. Within the scope of natural history collections, we see two pressing issues that must be addressed by the larger scientific community: (1) how to maintain, review, and revise the torrent of machine observations, measurements, and annotations that will soon exist and (2) how research groups should effectively compare or compile data given that unique specimens will be processed repeatedly using different methods.

Regarding the first issue, LeafMachine2 alone can generate thousands of data points per specimen. There is currently no infrastructure capable of supporting these data as part of a digital extended specimen, much less one that can integrate the measurements with those produced by other projects. We need new, scalable, and flexible data management standards and infrastructure. One possible approach for validating machine-derived data is to take a similar approach as iNaturalist (https://www.inaturalist.org/) with taxa identification: crowd-sourced voting. While this would bolster confidence, it too would struggle to keep pace with trait extraction. We have already tested LeafMachine2 on tens of thousands of specimens. On average, LeafMachine2 locates 10 leaves per image and calculates 20 measurements per leaf; thus, a data set of 10,000 images (a small fraction of the images available through GBIF and other institutions) will yield two million data points. The only data curation solution at this scale is more computational filtration, validation, and comparison, likely powered by machine learning. To ensure that human resources are allocated most efficiently, it is imperative that we develop robust validation procedures before we start to process specimens en masse.

To the second issue of how we can effectively compare or compile data given that specimens will be processed repeatedly by different methods, we chose to use only GBIF images because of duplicated specimens between data portals. This phenomenon means that the total number of unique specimen images present across multiple portals is lower than the number of images claimed by portal queries (Kommineni et al., 2021). As an example, iDigBio also houses herbarium vouchers, but numerous institutions deposit images into both portals, resulting in the same image being assigned two distinct identifiers. We support calls for a global federated specimen identification system as a part of a larger movement toward a flexible and comprehensive digital extended specimen concept to enable the effective mobilization of machine-derived data at scale (Lendemer et al., 2020; Hardisty et al., 2022).

LeafMachine2 will continue to evolve as we add support for more traits and taxa, contributing data to answer endless biological research questions. Even so, we look ahead to when every digitized herbarium specimen has a comprehensive set of measured traits and ask, Will it be enough? We suspect that this fantastic corpus of botanical descriptions will be revolutionary but will also amplify known contemporary sampling and taxonomic biases (Loiselle et al., 2008; Willis et al., 2017; Daru et al., 2018; Kozlov et al., 2021; Meineke and Daru, 2021; Davis, 2022; Heberling, 2022) and shift focus toward traits that cannot be captured by two-dimensional images of preserved tissue (Borges et al., 2020). Therefore, let this resurgence in attention paid to herbaria also serve as a catalyst for preserving not only physical specimens, but also digital-only collections. These collections may include snapshot vouchers (i.e., non-destructive, photogrammetrically validated field images of living plant tissue; Weaver and Smith, 2023) or three-dimensional scans of living tissue (James et al., 2023). Herbaria are indeed a bastion of global biodiversity memory, but modern questions also require integrative data sets, and maintaining the status quo will not be enough.

AUTHOR CONTRIBUTIONS

W.N.W. planned and designed the project, wrote the software, labeled and reviewed specimen annotations, and wrote the initial manuscript. S.A.S. contributed to project planning, design, and testing. Both authors contributed to revising and editing the text. Both authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank Jorge Jaime-Rivera, Megan Kastelen, Sophie Farr, Sky Teal, Maria Donlin, Arielle Sutherland, Valerie Xu-Friedman, and Claire Henley for assisting with labeling training data sets and the members of the Smith Lab (University of Michigan) for providing insightful feedback and encouragement. The authors also thank Julienne Ng and Rob Laport (College of Idaho) for their contributions and initial conception of LeafMachine. The authors thank Charles Cannon and the Morton Arboretum for providing field images and feedback. Funding for W.N.W. and S.A.S. was provided by the National Science Foundation (DEB 2217116), Michigan Institute for Data

Science, and Michigan Institute for Computational Discovery and Engineering.

DATA AVAILABILITY STATEMENT

The LeafMachine2 source code, examples, machine learning networks, and user manual are available at https://github.com/Gene-Weaver/LeafMachine2 and https://www.LeafMachine.org. Sample images from the test data sets are available at https://zenodo.org/record/7764379.

ORCID

William N. Weaver http://orcid.org/0000-0003-0633-5066
Stephen A. Smith http://orcid.org/0000-0003-2035-9531

REFERENCES

Alajas, O. J., R. Concepcion, E. Dadios, E. Sybingco, C. H. Mendigoria, and H. Aquino. 2021. Prediction of grape leaf black rot damaged surface percentage using hybrid linear discriminant analysis and decision tree. 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India. https://doi.org/10.1109/CONIT51480.2021.9498518

Borges, L. M., V. C. Reis, and R. Izbicki. 2020. Schrödinger's phenotypes: Herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. *Methods in Ecology and Evolution* 11(10): 1296–1308. https://doi.org/10.1111/2041-210X.13450

Chitwood, D. H., and N. R. Sinha. 2016. Evolutionary and environmental forces sculpting leaf development. *Current Biology* 26(7): R297–R306. https://doi.org/10.1016/j.cub.2016.02.033

Curlis, J. D., T. Renney, A. R. Davis Rabosky, and T. Y. Moore. 2022. Batch-Mask: Automated image segmentation for organisms with limbless or non-standard body forms. *Integrative and Comparative Biology* 62(4): 1111–1120. https://doi.org/10.1093/icb/icac036

Daru, B. H., D. S. Park, R. B. Primack, C. G. Willis, D. S. Barrington, T. J. S. Whitfeld, T. G. Seidler, et al. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytologist* 217(2): 939–955. https://doi.org/10.1111/nph.14855

Davis, C. C. 2022. The herbarium of the future. *Trends in Ecology & Evolution* 38: P412–P423. https://doi.org/10.1016/j.tree.2022.11.01

Davis, C. C., J. Champ, D. S. Park, I. Breckheimer, G. M. Lyra, J. Xie, A. Joly, et al. 2020. A new method for counting reproductive structures in digitized herbarium specimens using Mask R-CNN. *Frontiers in Plant Science* 11: 1129. https://doi.org/10.3389/fpls.2020.01129

Easlon, H. M., and A. J. Bloom. 2014. Easy Leaf Area: Automated digital image analysis for rapid and accurate measurement of leaf area. *Applications in Plant Sciences* 2(7): e1400033. https://doi.org/10.3732/apps.1400033

Ellis, B., D. Daly, L. Hickey, K. Johnson, J. Mitchell, P. Wilf, and S. Wing. 2009. Manual of Leaf Architecture. Cornell University Press, Ithaca, New York, USA.

Goëau, H., A. Mora-Fallas, J. Champ, N. L. R. Love, S. J. Mazer, E. Mata-Montero, A. Joly, and P. Bonnet. 2020. A new fine-grained method for automated visual analysis of herbarium specimens: A case study for phenological data extraction. *Applications in Plant Sciences* 8(6): e11368. https://doi.org/10.1002/aps3.11368

Goëau, H., T. Lorieul, P. Heuret, A. Joly, and P. Bonnet. 2022. Can artificial intelligence help in the study of vegetative growth patterns from herbarium collections? An evaluation of the tropical flora of the French Guiana Forest. *Plants* 11(4): 530. https://doi.org/10.3390/plants11040530

Gu, X., Y. Zhu, S. Ren, and X. Shu. 2022. BCMask: A finer leaf instance segmentation with bilayer convolution mask. *Multimedia Systems* 29: 1145–1159. https://doi.org/10.1007/s00530-022-01044-z

Guo, R., L. Qu, D. Niu, Z. Li, and J. Yue. 2021. LeafMask: Towards greater accuracy on leaf segmentation. ArXiv:2108.03568 [Cs] [Preprint]. Available at https://doi.org/10.48550/arXiv.2108.03568 [posted 8 August 2021; accessed 8 September 2023].

Hardisty, A., E. Ellwood, G. Nelson, B. Zimkus, J. Buschbom, W. Addink, R. Rabeler, et al. 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. *BioScience* 72: 978–987. https://doi.org/10.1093/biosci/biac060

He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969.

Heberling, J. M. 2022. Herbaria as big data sources of plant traits. *International Journal of Plant Sciences* 183(2): 87–118. https://doi.org/10.1086/717623

Hussein, B. R., O. A. Malik, W.-H. Ong, and J. W. F. Slik. 2021a. Automated extraction of phenotypic leaf traits of individual intact herbarium leaves from herbarium specimen images using deep learning based semantic segmentation. *Sensors* 21(13): 4549. https://doi.org/10.3390/s21134549

Hussein, B. R., O. A. Malik, W.-H. Ong, and J. W. F. Slik. 2021b. Reconstruction of damaged herbarium leaves using deep learning techniques for improving classification accuracy. *Ecological Informatics* 61: 101243. https://doi.org/10.1016/j.ecoinf.2021.101243

Hussein, B. R., O. A. Malik, W.-H. Ong, and J. W. F. Slik. 2022. Applications of computer vision and machine learning techniques for digitized herbarium specimens: A systematic literature review. *Ecological Informatics* 69: 101641. https://doi.org/10.1016/j.ecoinf.2022.101641

James, N., A. Adkinson, and A. Mast. 2023. Rapid imaging in the field followed by photogrammetry digitally captures the otherwise lost dimensions of plant specimens. *Applications in Plant Sciences* 11(5): e11547.

Jocher, G., A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, et al. 2022. YOLOv5 Realtime Instance Segmentation [posted 22 November 2022]. Available at Zenodo repository: https://doi.org/10.5281/zenodo.7347926 [accessed 23 August 2023].

Karnani, K., J. Pepper, Y. Bakiş, X. Wang, H. Bart Jr., D. E. Breen, and J. Greenberg. 2022. Computational metadata generation methods for biological specimen image collections. *International Journal on Digital Libraries* 23(4): https://doi.org/10.1007/s00799-022-00342-1

Kaur, P., S. Harnal, V. Gautam, M. P. Singh, and S. P. Singh. 2022. An approach for characterization of infected area in tomato leaf disease based on deep learning and object detection technique. *Engineering Applications of Artificial Intelligence* 115: 105210. https://doi.org/10.1016/j.engappai.2022.105210

Kavitha Lakshmi, R., and N. Savarimuthu. 2023. DPD-DS for plant disease detection based on instance segmentation. *Journal of Ambient Intelligence and Humanized Computing* 14: 3145–3155. https://doi.org/10.1007/s12652-021-03440-1

Kirillov, A., Y. Wu, K. He, and R. Girshick. 2020. PointRend: Image Segmentation as Rendering. arXiv:1912.08193 [Preprint]. Available at https://doi.org/10.48550/arXiv.1912.08193 [posted 17 December 2019; accessed 8 September 2023].

Klein, L. L., and H. T. Svoboda. 2017. Comprehensive methods for leaf geometric morphometric analyses. *Bio-Protocol* 7(9): e2269. https://doi.org/10.21769/BioProtoc.2269

Kommineni, V. K., S. Tautenhahn, P. Baddam, J. Gaikwad, B. Wieczorek, A. Triki, and J. Kattge. 2021. Comprehensive leaf size traits dataset for seven plant species from digitised herbarium specimen images covering more than two centuries. *Biodiversity Data Journal* 9: e69806. https://doi.org/10.3897/BDJ.9.e69806

Kozlov, M. V., I. V. Sokolova, V. Zverev, and E. L. Zvereva. 2021. Changes in plant collection practices from the 16th to 21st centuries: Implications for

the use of herbarium specimens in global change research. *Annals of Botany* 127(7): 865–873. https://doi.org/10.1093/aob/mcab016

Kumar, N., P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares. 2012. Leafsnap: A computer vision system for automatic plant species identification. *In* A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid [eds.], Computer Vision – ECCV 2012, Vol. 7573, 502–516. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-642-33709-3_36

Lendemer, J., B. Thiers, A. K. Monfils, J. Zaspel, E. R. Ellwood, A. Bentley, K. LeVan, et al. 2020. The extended specimen network: A strategy to enhance us biodiversity collections, promote research and education. *BioScience* 70(1): 23–30. https://doi.org/10.1093/biosci/biz140

Lin, T.-Y., M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, et al. 2014. Microsoft COCO: Common Objects in Context [Preprint]. Available at https://doi.org/10.48550/arXiv.1405.0312 [posted 1 May 2014; accessed 14 September 2023].

Loiselle, B. A., P. M. Jørgensen, T. Consiglio, I. Jiménez, J. G. Blake, L. G. Lohmann, and O. M. Montiel. 2008. Predicting species distributions from herbarium collections: Does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 35(1): 105–116. https://doi.org/10.1111/j.1365-2699.2007.01779.x

Love, N. L. R., P. Bonnet, H. Goëau, A. Joly, and S. J. Mazer. 2021. Machine learning undercounts reproductive organs on herbarium specimens but accurately derives their quantitative phenological status: A case study of *Streptanthus tortuosus*. *Plants* 10(11): 2471. https://doi.org/10.3390/plants10112471

Maloof, J. N., K. Nozue, M. R. Mumbach, and C. M. Palmer. 2013. LeafJ: An ImageJ plugin for semi-automated leaf shape measurement. *Journal of Visualized Experiments: JoVE* 71: e51128. https://doi.org/10.3791/50028

Mann, H. M. R., A. Iosifidis, J. U. Jepsen, J. M. Welker, M. J. J. E. Loonen, and T. T. Høye. 2022. Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning. *Remote Sensing in Ecology and Conservation* 8(6): 765–777. https://doi.org/10.1002/rse2.275

Martens, B. 2018. The importance of data access regimes for artificial intelligence and machine learning. JRC Digital Economy Working Paper 2018-09. https://doi.org/10.2139/ssrn.3357652

Meineke, E. K., and B. H. Daru. 2021. Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution* 36(12): 1071–1082. https://doi.org/10.1016/j.tree.2021.08.003

Meineke, E. K., C. Tomasi, S. Yuan, and K. M. Pryer. 2020. Applying machine learning to investigate long-term insect–plant interactions preserved on digitized herbarium specimens. *Applications in Plant Sciences* 8(6): e11369. https://doi.org/10.1002/aps3.11369

Milleville, K., K. K. T. Chandrasekar, and S. Verstockt. 2023. Automatic extraction of specimens from multi specimen herbaria. *Journal on Computing and Cultural Heritage* 16: 3575862. https://doi.org/10.1145/3575862

Mora-Fallas, A., H. Goëau, S. Mazer, N. Love, E. Mata-Montero, P. Bonnet, and A. Joly. 2019. Accelerating the automated detection, counting and measurements of reproductive organs in herbarium collections in the era of deep learning. *Biodiversity Information Science and Standards* 3: e37341. https://doi.org/10.3897/biss.3.37341

Mu, W., Z. Jia, Y. Liu, W. Xu, and Y. Liu. 2022. Image segmentation model of pear leaf diseases based on Mask R-CNN. 2022 International Conference on Image Processing and Media Computing (ICIPMC), Xi'an, China, 41–45. https://doi.org/10.1109/ICIPMC55686.2022.00016

Neto, J. C., G. E. Meyer, D. D. Jones, and A. K. Samal. 2006. Plant species identification using Elliptic Fourier leaf shape analysis. *Computers and Electronics in Agriculture* 50(2): 121–134. https://doi.org/10.1016/j.compag.2005.09.004

Ott, T., and U. Lautenschlager. 2022. GinJinn2: Object detection and segmentation for ecology and evolution. *Methods in Ecology and Evolution* 13(3): 603–610. https://doi.org/10.1111/2041-210X.13787

Ott, T., C. Palm, R. Vogt, and C. Oberprieler. 2020. GinJinn: An object-detection pipeline for automated feature extraction from herbarium specimens. *Applications in Plant Sciences* 8(6): e11351. https://doi.org/10.1002/aps3.11351

Panigrahi, I., R. Manzuk, A. Maloof, and R. Fong. 2022. Improving fine-grain segmentation via interpretable modifications: A case study in fossil segmentation. arXiv:2210.03879 [Preprint]. Available at https://doi.org/10.48550/arXiv.2210.03879 [posted 8 October 2022; accessed 8 September 2023].

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, et al., 2019. PyTorch: An imperative style, high-performace deep learning library [Preprint]. Available at https://doi.org/10.48550/arXiv.1912.01703 [posted 3 December 2019; accessed 14 September 2023].

Pearson, K. D., G. Nelson, M. F. J. Aronson, P. Bonnet, L. Brenskelle, C. C. Davis, E. G. Denny, et al. 2020. Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70(7): 610–620. https://doi.org/10.1093/biosci/biaa044

Safadi, H., and R. T. Watson. 2023. Knowledge monopolies and the innovation divide: A governance perspective. *Information and Organization* 33(2): 100466. https://doi.org/10.1016/j.infoandorg.2023.100466

Souibgui, M. A., S. Biswas, S. K. Jemni, Y. Kessentini, A. Fornés, J. Lladós, and U. Pal. 2022. DocEnTr: An End-to-End Document Image Enhancement Transformer. arXiv:2201.10252 [Preprint]. Available at https://doi.org/10.48550/arXiv.2201.10252 [posted 25 January 2022; accessed 8 September 2023].

Triki, A., B. Bouaziz, W. Mahdi, and J. Gaikwad. 2020. Objects Detection from Digitized Herbarium Specimen based on Improved YOLO V3. *In* Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications VISIGRAPP, Vol. 4, p. 529. https://doi.org/10.5220/0009170005230529

Triki, A., B. Bouaziz, J. Gaikwad, and W. Mahdi. 2021. Deep leaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images. *Pattern Recognition Letters* 150: 76–83. https://doi.org/10.1016/j.patrec.2021.07.003

Weaver, W. N., and S. A. Smith. 2023. FieldPrism: A system for creating snapshot vouchers from field images using photogrammetric markers and QR codes. *Applications in Plant Sciences* 11(5): e11545.

Weaver, W. N., J. Ng, and R. G. Laport. 2020. LeafMachine: Using machine learning to automate leaf trait extraction from digitized herbarium specimens. *Applications in Plant Sciences* 8(6): e11367. https://doi.org/10.1002/aps3.11367

Weeks, B. C., Z. Zhou, B. K. O'Brien, R. Darling, M. Dean, T. Dias, G. Hassena, et al. 2023. A deep neural network for high-throughput measurement of functional traits on museum skeletal specimens. *Methods in Ecology and Evolution* 14(2): 347–359. https://doi.org/10.1111/2041-210X.13864

Willis, C. G., E. R. Ellwood, R. B. Primack, C. C. Davis, K. D. Pearson, A. S. Gallinat, J. M. Yost, et al. 2017. Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology & Evolution* 32(7): 531–546. https://doi.org/10.1016/j.tree.2017.03.015

Wu, Y., A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. 2019. Detectron2. Available at https://github.com/facebookresearch/detectron2 [accessed 8 September 2023].

Yang, Q., Y. Zhang, W. Dai, and S. J. Pan. 2020. Transfer learning. Cambridge University Press, Cambridge, United Kingdom.

Younis, S., M. Schmidt, C. Weiland, S. Dressler, B. Seeger, and T. Hickler. 2020. Detection and annotation of plant organs from digitised herbarium scans using deep learning. *Biodiversity Data Journal* 8: e57090. https://doi.org/10.3897/BDJ.8.e57090

Zhang, S., H. Wang, W. Huang, and Z. You. 2018. Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. *Optik* 157: 866–872. https://doi.org/10.1016/j.ijleo.2017.11.190

SUPPORTING INFORMATION

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

Appendix S1. Training information for LeafMachine2 networks.

Appendix S2. Ruler conversion examples.

Appendix S3. Ruler conversion procedures.

How to cite this article: Weaver, W. N., and S. A. Smith. 2023. From leaves to labels: Building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2. *Applications in Plant Sciences* 11(5): e11548. https://doi.org/10.1002/aps3.11548

Appendix 1. A list of the data sets that provided training images for our machine learning algorithms. We sampled from among these data sets to create our annotated data sets of herbarium specimen vouchers. The data set is available from GBIF at https://doi.org/10.15468/dl.bh9dem.

Herbarium data set	Data set DOI
Allan Herbarium	10.15468/x5ucvh
Appalachian State University	10.15468/ivsxey
Arizona State University	10.15468/a2o8vy
Arkansas Natural Heritage Commission Herbarium	10.15468/v94jsu
Artportalen (Swedish Species Observation System)	10.15468/kllkyl
Asociación Jardín Botánico La Laguna	10.15468/gfwydn
Auckland Museum Botany Collection	10.15468/mnjkvv
B.M. Kozo-Polyansky VSU	10.15468/xyqng3
Bell Museum	10.15468/bihrxd
Berea College	10.15468/hcwetj
Black Hills State University Herbarium	10.15468/ptcrqx
Botanical Collections of the Abo Akademi	10.15468/mpsjrk
Botanical Museum Berlin-Dahlem	10.15468/ed17cn
Botanische Staatssammlung München	10.15468/ni5yho
Botanische Staatssammlung München	10.15468/sookye
Botanische Staatssammlung München	10.15468/zinzhd
Botanische Staatssammlung München	10.15468/zdcclb
Botanische Staatssammlung München	10.15468/lqetda
Botanische Staatssammlung München	10.15468/dixlft
Botanischer Garten und Botanisches Museum Berlin-Dahlem Herbarium	10.15468/tgwryf
Brauckmann at the Botanische Staatssammlung München	10.15468/onfqgb
BRI AVH	10.15468/jsffsa
Brown University	10.15468/kpsj8r
Brown University Herbarium	10.15468/njgg1a
Bush Heritage - Carnarvon Station Reserve	10.15468/q0dhpr
Cal Poly State University	10.15468/mypdjd
Cal State LA Herbarium	10.15468/36qz6p
California Botanic Garden Herbarium	10.15468/0yosx9
California State University Fullerton	10.15468/1uvzxh
California State University San Bernardino	10.15468/t885ps
California State University, Long Beach	10.15468/3y25yl
California State University, Northridge	10.15468/nrcdx7
Canadian Museum of Nature Herbarium	10.15468/kowta4
Cape Breton University Collection	10.15468/7dtqgn
Capture of Primary Biodiversity Data for West African Plants	10.15468/9czcig
Carnegie Museum of Natural History Herbarium	10.15468/d51v1f
CAS Botany	10.15468/7gudyo
CBNA	10.15468/oc5zh7
CEN herbarium	10.15468/wasmx9
Central Michigan University Herbarium	10.15468/iykbez
Central Siberian Botanical Garden	10.15468/qdfdqq
Central Siberian Botanical Garden Herbarium	10.15468/5wcerp
Centro de Pesquisas do Cacau	10.15468/vg8rjh
Charles University Prague	10.15468/8xrt7r
CHAS Botany Collection (Arctos)	10.15468/ji4vbl
Chico State Herbarium	10.15468/ckxw7v
Clamence Lortet herbarium	10.15468/e64fbk
Clemson University Herbarium	10.15468/srjd22
Colección de plantas vasculares del herbario de la Universitat de Valencia	10.15468/xmki52
Colección Herbario Federico Medem Bogota	10.15472/ighftu
Colorado Plateau Museum of Arthropod Biodiversity	10.15468/du1hci
CRI Herbarium	10.15468/vvctbg
CRSN herbarium from Kahuzi-Biega National Park	10.15468/bhvwem
CRSN Herbarium	10.15468/ra9vp0
CRSN herbarium	10.15468/exh7vo
CSBG	10.15468/c1y9q2
CSBG SB RAS	10.15468/67ouin

Herbarium data set	Data set DOI
CSBG SB RAS	10.15468/be6owh
CSBG SB RAS	10.15468/zw7jnn
CSBG SB RAS Digital Herbarium	10.15468/6f3ybc
CSBG SB RAS Digital Herbarium	10.15468/7anvyu
CSBG SB RAS Herbarium Collections	10.15468/sunx5n
Dataflos	10.15468/dcc6j8
Desert Botanical Garden Herbarium	10.15468/abe1lg
Dr. Sultan Ahmad Herbarium	10.15468/xaju4z
E. C. Smith Herbarium	10.15468/zc4csq
Earth Sciences Collection (Arctos)	10.15468/4n2ev3
Eastern Kentucky University	10.15468/fy8dsi
Estonian Museum of Natural History	10.15468/bquqpv
Estonian University of Life Sciences	10.15468/m3x9uu
Fairchild Tropical Botanic Garden	10.15468/hdpruf
Field Museum of Natural History	10.15468/pyjtoc
Field Museum of Natural History	10.15468/4nodxs
Field Museum of Natural History	10.15468/nxnqzf
Flora of Sumatra: ANDA Herbarium	10.15468/ue7xyn
Flora of the Korean Peninsula	10.15468/0vcvsq
Flora of the Korean Peninsula	10.15468/fyxnsd
Flora Sumatra: (ANDA)-Part 2	10.15468/55evew
Forest Herbarium Ibadan	10.15468/uhnd5n
Forest Herbarium Ibadan Nigeria	10.15468/rhbyxz
Fresno State Herbarium	10.15468/puyrj8
Genus Medicago in CSBG Herbarium	10.15468/jvrxeh
George Mason University	10.15468/t8ar55
Georgian Academy of Sciences	10.15468/6tbhmd
Harvard University Herbaria	10.15468/o3pvnh
Harvard University Herbarium	10.15468/29fhdy
Herbario Joao de Carvalho e Vasconcellos	10.15468/olfpjv
Herbario Museo de La Salle Bogota	10.15472/ppzpea
Herbarium Berolinense, Berlin	10.15468/dlwwhz
Herbarium Fennoscandicum	10.15468/ekpyfw
Herbarium GAT	10.15468/hiiw6b
Herbarium Generale	10.15468/dg4cb4
Herbarium Generale	10.15468/83cb4a
Herbarium Hamburgense	10.15468/31iaih
Herbarium Horti Botanici Pisani	10.15468/soyil7
Herbarium of Andalas University	10.15468/sncpxn
Herbarium of Numto Nature Park	10.15468/g4gcrg
Herbarium of the University of Granada	10.15470/k97bjm
Herbarium of Yugra State University	10.15468/z8mpt5
Herbarium Senckenbergianum	10.15468/ucmdjy
Herbarium Willing at Herbarium Berolinense, Berlin	10.15468/abcz8i
Herbier du Québec	10.5886/jd11sg3p
Herbier Louis-Marie	10.5886/3p8ltbg7
Herbiers Universitaires de Clermont-Ferrand	10.15468/9axq0b
Humboldt State University	10.15468/qguk7r
HVASF herbarium	10.15468/kz6y6z
IAN herbarium	10.15468/cv2dmt
IICT Herbario	10.15468/iinlqm
Institut Botanic de Barcelona	10.15468/pff0t6
Institut Scientifique Mohamed V University	10.15468/48pwft
Institute of Biological Problems of the North, Far East Branch RAS	10.15468/ms9q2t
Instituto do Meio Ambiente do Estado de Alagoas	10.15468/mu8w57
Intermountain Herbarium	10.15468/t43wjj
Jardins botaniques and Conservatoire Botanique of Nancy	10.15468/g1zohr
JOI Herbarium	10.15468/pf6pv2
Kathryn Kalmbach Herbarium	10.15468/axrelr
Kenai National Wildlife Refuge (Arctos)	10.15468/ycpd7y
Komarov Botanical Institute	10.15468/udzn9d
Komi Republic	10.15468/336sdv
KULPOL Herbarium	10.15468/h9qfje
KUZ Herbarium	10.15468/4ru3f6
Lajitietokeskus FinBIF	10.15468/4g56tp
Lord Fairfax Community College Herbarium	10.15468/c2gj2t
MAG Herbarium	10.15468/ahqbdc
Marie-Victorin Herbarium	10.5886/rzav8bu2
Masaryk University	10.15468/soarvd
McGill University Herbarium	10.5886/srzbj7
Meise Botanic Garden Herbarium	10.15468/wrthhx
MEL AVH	10.15468/rhzrxw
Melu AVH	10.15468/2yyu7i
MHA Herbarium	10.15468/827lk2
Ministerio del Medio Ambiente de Chile	10.15468/ezyu58
Missouri Botanical Garden	10.15468/mmbcpb

(Continues)

Herbarium data set	Data set DOI
Moscow University Herbarium	10.15468/cpnhcc
MUFAL herbarium	10.15468/viuv6v
Muséum National d'Histoire Naturelle, Paris	10.15468/kw8pex
Muséum National d'Histoire Naturelle	10.15468/nc6rxy
Museu Botanico Municipal Curitiba	10.15468/v52pmc
Museu de Biologia Mello Leitao	10.15468/dmkg7b
Museu Paraense Emalio Goeldi	10.15468/rdq4nx
Museum d'Histoire Naturelle of Aix-en-Provence	10.15468/fqykeb
National Academy of Sciences of Republic of Armenia	10.15468/xn64eb
National Museum of Natural History Luxembourg	10.15468/s2iu7d
Natural History Museum	10.5519/0002965
Natural History Museum Rotterdam	10.15468/kwqaay
Natural History Museum, Vienna	10.15468/5sl7sh
Naturalis Biodiversity Center	10.15468/ib5ypt
Naturhistorisches Museum Mainz	10.15468/l0wmu8
Naturhistoriska Riksmuseet	10.15468/jbcsfu
NCSM Herbarium Collection	10.36102/dwc.12
NEON Biorepository	10.15468/ggrfcb
NEON Biorepository	10.15468/bmmg36
NEON Biorepository	10.15468/bmmdp5
Newhaven Sanctuary Observations	10.15468/mwgsdh
NHMD Vascular Plants Collection	10.15468/4zygkn
Nitraria komarovii	10.15468/jp2qco
NMNH	10.15468/hnhrg3
North Carolina State University	10.15468/9ufthy
Northern Arizona University	10.15468/b7tfpa
Norwegian Species Observation Service	10.15468/zjbzel
Nova Scotia Museum of Natural History	10.15468/tl3cde
NSW AVH data	10.15468/jf3yae
NSW South Coast	10.15468/px2xfi
Plant Resources Center	10.15468/g85t8z
Qarshi Botanical Garden	10.15468/pjxa84
Quaid-i-Azam University Herbarium	10.15468/bp6jy3
Queensland Museum	10.15468/lotsye
R. L. McGregor Herbarium	10.15468/htptzr
Real Jardín Botanico	10.15468/mug7kr
Rhoen and Vogelsberg	10.15468/hbhfi3
Rio de Janeiro Botanical Garden Herbarium	10.15468/7ep9i2
Rio de Janeiro Botanical Garden Herbarium Collection	10.15468/bbsqoa
Royal Botanic Garden Edinburgh Herbarium	10.15468/ypoair
Royal Botanic Gardens, Kew	10.15468/ly60bx
Royal Botanic Gardens, Kew	10.15468/rvrsru
Royal Ontario Museum Green Plant Herbarium	10.5886/g7j6gct1
Rutgers University	10.15468/1n787c
Rutgers University	10.15468/hhnd4h
Sagehen Herbarium	10.15468/fl8uov
SAMES herbarium	10.15468/l0hdtn
San Diego Natural History Museum	10.15468/lneqwn
San Diego State University Herbarium	10.15468/8sx2ag
San Francisco State University	10.15468/6zdzvc
San Jose State University	10.15468/t3a60p
SANT Herbarium	10.15468/dgbpla
Santa Barbara Botanic Garden	10.15468/adb2bb
Sociata des Sciences Naturelles et Mathamatiques de Cherbourg	10.15468/lmznjw
South Australian Museum Australia	10.15468/wz4rrh
South-Siberian Botanical Garden	10.15468/y6xmme
Species recordings from the Danish National portal Arter.dk	10.15468/q3yy4u
Staten Island Museum	10.15468/ctqpb5
Steiermarkisches Landesmuseum Joanneum	10.15468/dmdck6
SVER Herbarium	10.15468/xwzszg
SVER Herbarium	10.15468/5npjcc
Tallinn Botanic Garden	10.15468/hfs8d4
Terre d'huiles	10.15468/wr1vhd
Texas Tech University	10.15468/uyakmh
The Exsiccatal Series	10.15468/qxmief
The James C. Parks Herbarium at Millersville University	10.15468/qdatdf
The New York Botanical Garden	10.15468/5y84ye
The New York Botanical Garden Herbarium	10.15468/6e8nje
The Vascular Plant Collection at the Botanische Staatssammlung Manchen	10.15468/vgr4kl
TKM Herbarium	10.15468/sfxvrv
Towson University	10.15468/podgza
TRH, NTNU University Museum	10.15468/zrlqok
Tropicos Specimen Data	10.15468/hja69f
TUL Herbarium	10.15468/ca08cm

Herbarium data set	Data set DOI
TULGU Herbarium	10.15468/5nret6
Turku University	10.15468/nsyt4y
UAM Herbarium (Arctos)	10.15468/iawody
UC Davis Herbarium	10.15468/on4axg
UiT Tromsa Museum	10.15468/14epds
Universidad del Valle de Guatemala	10.15468/u339qt
Universidade de Sao Paulo	10.15468/nt6dng
Universidade Estadual de Feira de Santana	10.15468/gsy3jn
Universidade Estadual do Norte Fluminense	10.15468/qsaagd
Universidade Estadual do Oeste do Parana	10.15468/eqp1dr
Universidade Federal da Bahia	10.15468/tbtrr3
Universidade Federal de Goias	10.15468/fw6hdt
Universidade Federal de Parana	10.15468/fpf5j6
Universidade Federal de Rondania	10.15468/5cjyj6
Universidade Federal de Sergipe	10.15468/9xujh5
Universidade Federal de Uberlandia	10.15468/cshs8n
Universidade Federal do Ceara	10.15468/s8xuen
Universidade Federal do Espírito Santo	10.15468/kasze8
Universidade Federal do Oeste do Para	10.15468/ztzkde
Universidade Federal do Rio Grande do Norte	10.15468/gtxawd
Universidade Federal do Rio Grande Do Sul	10.15468/suhqjx
Universidade Federal Rural do Rio de Janeiro	10.15468/0svt7m
Universidade Regional de Blumenau	10.15468/vse5f3
Universidade Tecnológica Federal do Parana	10.15468/4b74v2
Universita de Montpellier	10.15468/gyvkrn
Universita Lyon	10.15468/7m584w
University of Balochistan Herbarium	10.15468/qrau0v
University of British Columbia Herbarium	10.5886/rtt57cc9
University of California Santa Barbara Herbarium	10.15468/qpxmw0
University of California Santa Cruz	10.15468/uavt0t
University of California, Los Angeles Herbarium	10.15468/33k42a
University of California, Riverside	10.15468/ai1kou
University of Cincinnati	10.15468/bhgpmq
University of Cincinnati	10.15468/xkca3p
University of Colorado Museum of Natural History	10.15468/wyofjv
University of Florida Herbarium	10.15468/v5wjn7
University of Gothenburg	10.15468/asgd85
University of Graz Institute of Plant Sciences	10.15468/axtkuz
University of Hargeisa Herbarium	10.15468/qvbvdp
University of Jena, Herbarium Haussknecht	10.15468/8arhjc
University of Kentucky	10.15468/fi4vfu
University of Lethbridge Herbarium	10.5886/wrt547hq
University of Manitoba Herbarium	10.5886/2fva5p4r
University of Michigan Herbarium	10.15468/nl8bvi
University of New Mexico Herbarium	10.15468/dlvoyt
University of North Carolina at Chapel Hill	10.15468/63vxjd
University of Sargodha Herbarium	10.15468/n4k5s9
University of South Carolina	10.15468/fmj4at
University of South Florida Herbarium	10.15468/mdnmzb
University of Tartu Natural History Museum	10.15468/5hqb2z
University of Tartu Natural History Museum and Botanical Garden	10.15468/d59dmk
University of Tennessee	10.15468/64w2b1
University of Tennessee Fungal Herbarium	10.15468/da30il
University of Tennessee Vascular Herbarium	10.15468/ok8qvz
University of Vermont	10.15468/crnsua
University of Vermont	10.15468/zsgiog
University of Vienna, Institute for Botany	10.15468/tnj8wm
UTEP Plants (Arctos)	10.15468/yhb6ky
Vascular Plant Herbarium, Oslo	10.15468/wtlymk
Vascular Plant Herbarium, UiB	10.15468/ofn0lf
Western Carolina University Herbarium	10.15468/sk26v2
Yale Peabody Museum	10.15468/hrztgn
Yale Peabody Museum	10.15468/0lkr3w