Applications
in Plant Sciences

APPLICATION ARTICLE

Enhancing plant morphological trait identification in herbarium collections through deep learning–based segmentation

Hanane Ariouat[1] | Youcef Sklab[1] | Edi Prifti[1,2] | Jean-Daniel Zucker[1,2] | Eric Chenin[1]

[1]Institut de Recherche pour le Développement (IRD), Sorbonne Université, UMMISCO, F-93143, Bondy, France

[2]Sorbonne Université, INSERM, Nutrition et Obesities: Systemic approaches, NutriOmique, AP-HP, Hôpital Pitié-Salpêtrière, France

Correspondence
Hanane Ariouat, Institut de Recherche pour le Développement (IRD), Sorbonne Université, UMMISCO, F-93143, Bondy, France.
Email: hanane.ariouat@ird.fr

Abstract

Premise: Deep learning has become increasingly important in the analysis of digitized herbarium collections, which comprise millions of scans that provide valuable resources for studying plant evolution and biodiversity. However, leveraging deep learning algorithms to analyze these scans presents significant challenges, partly due to the heterogeneous nature of the non-plant material that forms the background of the scans. We hypothesize that removing such backgrounds can improve the performance of these algorithms.

Methods: We propose a novel method based on deep learning to segment and generate plant masks from herbarium scans and subsequently remove the non-plant backgrounds. The semi-automatic preprocessing stages involve the identification and removal of non-plant elements, substantially reducing the manual effort required to prepare the training dataset.

Results: The results highlight the importance of effective image segmentation, which achieved an F1 score of up to 96.6%. Moreover, when used in classification models for plant morphological trait identification, the images resulting from segmentation improved classification accuracy by up to 3% and F1 score by up to 7% compared to non-segmented images.

Discussion: Our approach isolates plant elements in herbarium scans by removing background elements to improve classification tasks. We demonstrate that image segmentation significantly enhances the performance of plant morphological trait identification models.

KEYWORDS

deep learning, herbarium scans, semantic segmentation, trait classification

Biodiversity is essential for maintaining ecological balance and supporting life on Earth. Understanding and preserving this biodiversity is critical in the face of extensive challenges, particularly the profound impacts of climate change. There is an urgent need for the scientific community to employ multidisciplinary and innovative strategies, which require comprehensive taxonomic and ecological studies that leverage recent advances in biodiversity informatics, as well as extensive databases on species traits and occurrences.

Integrating these approaches will improve our understanding of biodiversity threats and enable the development of more effective conservation tools (Lorieul et al., 2019).

Natural history collections are invaluable in this endeavor, holding centuries of data on biodiversity evolution and environmental changes. The growing interest in these collections is well founded, as they provide essential insights into understanding and mitigating biodiversity threats (Yue et al., 2017; Lorieul et al., 2019; Younis et al., 2022;

Triki et al., 2022a; Sklab et al., 2024a). Recent advances in digitizing and aggregating specimen-related data have significantly improved the accessibility to these collections. Through extensive digitization efforts and by making this data available online to both the scientific community and the public via platforms like ReColNat (https://explore.recolnat.org/) and the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/), the scientific community is now better equipped to conduct in-depth research into plant behavior and its various manifestations.

Between 2005 and 2007, the French Museum of Natural History (Muséum National d'Histoire Naturelle [MNHN]; https://www.mnhn.fr) in Paris was a pioneer in the field of industrial plant specimen digitization. The museum, which houses one of the world's largest herbaria with more than eight million specimens, has led a collaborative effort to digitize the major herbarium collections in France, resulting in a dataset of around 10 million high-resolution images called ReColNat. Describing this extensive collection for enhanced usability requires the automated capture of comprehensive metadata. To navigate and search efficiently within this massive dataset, users need specific criteria to select the specimens of interest. We focus on criteria based on descriptive characteristics (traits) observable in the specimen images, which relate to plant morphology or the specimen's conservation state. The integration of computer-based processing techniques, particularly deep learning, presents a promising approach to analyzing herbarium specimens (Dhaka et al., 2021). However, herbarium specimens pose several challenges due to their highly variable backgrounds containing diverse elements such as labels, scale bars, color palettes, envelopes, and various notes; these elements constitute visual noise and may negatively impact classification models through shortcut learning (Leksut et al., 2020; Moayeri et al., 2022). Furthermore, the support paper on which they are mounted can darken over time, until the color is similar to the plant specimens themselves (Figure 1). Despite these challenges, deep learning–based methods have shown significant promise in plant image analysis for predictive tasks involving herbarium specimens (Dhaka et al., 2021; Triki et al., 2021, 2022a, 2022b; Thompson et al., 2023; Ariouat et al., 2024; Sklab et al., 2024a). For instance, these methods have facilitated significant strides in plant disease classification, offering new possibilities for enhancing plant health and resilience (Borhani et al., 2022). Another critical application of deep learning is trait recognition, where identifying key characteristics, such as leaf features (Corney et al., 2012; Younis et al., 2018; Triki et al., 2022a) or reproductive structures (flowers, fruits, seeds) (Lorieul et al., 2019), plays a pivotal role in understanding plant biodiversity and attributes (Corney et al., 2012; Younis et al., 2018; Lorieul et al., 2019; Triki et al., 2021).

Multimodal approaches that integrate textual and image data (Sahraoui et al., 2023) represent an innovative strategy to improve plant analysis, bridging the gap between textual

FIGURE 1   Examples illustrating the diversity in paper color, quality, and the non-plant elements present on herbarium sheets.

descriptions and visual representations. For example, Ariouat et al. (2024) and Sklab et al. (in press) developed an improved YOLOv7 (Wang et al., 2023) with an attention mechanism to detect plant organs and non-plant elements, achieving an accuracy of 99%. This integration of deep learning applications highlights its importance in botanical research and conservation efforts.

The phenotype of a plant, as it can be visually perceived through its characteristics, is shaped by the interplay between genetic expression and environmental influences (Fan et al., 2022). The automated identification of such characteristics is challenging in digitized herbarium specimens, and various studies (e.g., Younis et al., 2018) highlight the need to address this issue. Current methods, such as cropping, fail to effectively remove these elements and can sometimes cut off important parts of the specimens.

In the field of segmentation, deep learning techniques have demonstrated their ability in performing leaf segmentation (Triki et al., 2021; Weaver and Smith, 2023; Wilde et al., 2023) and plant segmentation (Hussein et al., 2020; White et al., 2020; Fan et al., 2022; Triki et al., 2022b; Ariouat et al., 2023), facilitating the separation of different elements within a specimen image. For instance, in Fan et al. (2022), two tasks were proposed: plant segmentation to isolate the plant from its background, and leaf counting. Using UNet++ (Zhou et al., 2018) for segmentation and a modified ResNet50 (He et al., 2015) model for leaf counting, the authors analyzed tobacco and *Arabidopsis* (L.) Heynh. plants, which are characterized by their rosette-like growth. These plants display minimal visual artifacts, primarily soil, moss, and pots.

Triki et al. (2022b) proposed a segmentation approach using coarse segmentation in combination with a leaf reconstruction technique. Similar to the study by Fan et al. (2022), their method focuses primarily on leaves and does not explicitly address the preservation of all plant organs. In their coarse segmentation, Triki et al. (2022a) provided annotations for several elements present in specimen sheets. Bounding boxes were used to identify barcodes, text labels, and color palettes, and polygons were used to delineate the plants themselves. The annotations were then converted into color patches and each class was assigned a specific color. An encoder/decoder architecture based on a modified VGG16 backbone (Simonyan and Zisserman, 2014) was then used for segmentation, and the result of the leaf reconstruction segmentation process was used to retrieve images of leaves from herbarium specimens by means of engineering methods. It is important to note that the authors' assumption was that all leaf images from herbarium specimens had transparent backgrounds. White et al. (2020) introduced a procedure for segmenting herbarium images, specifically of ferns, which are distinguished by their specific leaf shapes and the absence of flowers or fruits. The authors used the Otsu method (Otsu, 1979) and Photoshop to create masks for training a segmentation model based on U-Net (Ronneberger et al., 2015) with ResNet34 (He et al., 2015); their approach achieved a Sørensen–Dice coefficient (Dice, 1945;

Sørensen, 1948) of 96%, demonstrating its effectiveness in segmenting fern specimen images. Hussein et al. (2020) proposed two models for removing visual noise in digitized herbarium specimens. In the first model, the authors used DeepLabv3+ (Chen et al., 2018) with pretrained weights, and in the second, they developed and trained from scratch a full-resolution residual network (FRNN-A) (Pohlen et al., 2017). Their results show that FRNN-A slightly outperformed DeepLabv3+, achieving an average intersection over union (IoU) of 99.2%, compared to 98.1% for DeepLabv3+. The authors generated their masks using the image labeler in MATLAB (MathWorks, Natick, Massachusetts, USA), followed by a median filtering technique. However, these masks encompassed only the leaves and stems from various species.

Although recent literature highlights the adaptability of deep learning models and their ability to manage the complexity of herbarium images while abstracting extraneous visual information, there remains a risk of models overfitting to visual noise (Leksut et al., 2020; Moayeri et al., 2022). This overfitting can lead to poor generalization performance when such elements are absent. The primary focus of this paper is on herbarium scan segmentation to separate the plant elements from the background and other non-plant elements, a critical task that involves isolating the plant region within an image by delineating its shape and constituent organs. To the best of our knowledge, no prior work has fully leveraged advanced computer vision techniques for the comprehensive segmentation of whole scanned herbarium specimens, capturing every aspect of the plant from stems to seeds. This innovation distinguishes our work from previous efforts, which primarily focused on segmenting selected plant organs. The novel contributions of our study are four-fold as follows:

1. **Color interval segmentation pipeline**: We developed a color interval segmentation pipeline (CISP) that removes non-plant elements using an object detection algorithm, a hue-saturation-value (HSV) color segmentation algorithm, and a set of morphological transformations to refine the final masks.
2. **Data collection and preparation**: We compiled a dataset of 2277 image–mask pairs, where each image is paired with its corresponding mask. The masks, generated using CISP, were manually refined. We then applied a series of data-augmentation techniques to enhance the dataset.
3. **Plant mask generation approach**: Our approach introduces two segmentation models that delineate entire plants and their individual organs from herbarium scans. We implemented the U-Net model (Ronneberger et al., 2015) through transfer learning, using ResNet101 (He et al., 2015) as the underlying architecture. For the U-Net model, we processed two types of masks: one featuring a white background (WB) and another with a black background (BB).
4. **Identification of plant morphological traits**: Using the segmentation-generated masks, we created images containing only the plant specimens and then conducted

experimental studies using two classification models (ResNet101 [He et al., 2015] and Vision Transformer [ViT; Dosovitskiy et al., 2021]) to identify five distinct morphological traits. The results demonstrated a significant improvement in model performance, with ViT achieving the best results, increasing accuracy by up to 3% and the F1 score by up to 7%.

METHODS

Dataset collection and preprocessing

The dataset used for image segmentation originates from the MNHN in Paris (Sklab et al., 2024b) and encompasses 11 different families and genera (*Amborella* Baill. [91 images], *Castanea* Mill. [161 images], *Desmodium* Desv. [164 images], *Ulmus* L. [352 images], *Rubus* L. [184 images], *Litsea* Lam. [199 images], *Eugenia* L. [219 images], *Laurus* L. [250 images], Convolvulaceae [177 images], *Magnolia* L. [162 images], and Monimiaceae [318 images]) reflecting significant

plant diversity within dicotyledons. We compiled a dataset consisting of 2277 images, each paired with its corresponding mask (Figure 2). These masks result from the methodology described in the following section, based on the previously detailed process by Ariouat et al. (2023). The dataset is divided into a training set of 1821 images (80%) and a test set of 456 images (20%). All images are in red-green-blue (RGB) format and have been resized to 1024×1024 pixels. To enhance the model performance, each resized image is further partitioned into 16 patches of 256×256 pixels. Our experiments show that working with these smaller patches yields better performance compared to using entire images. The masks, initially in RGB format, are converted to binary values for training purposes, i.e., 1 indicates the plant foreground while 0 indicates the background. Our dataset, including both the original images and their corresponding masks, is available on Figshare (Sklab et al., 2024b; see the Data Availability Statement). To further enhance the dataset, various data augmentation techniques have been applied for both the black and the white backgrounds, such as rotations, shifts, flips, zooming, and shearing (Figure 3).
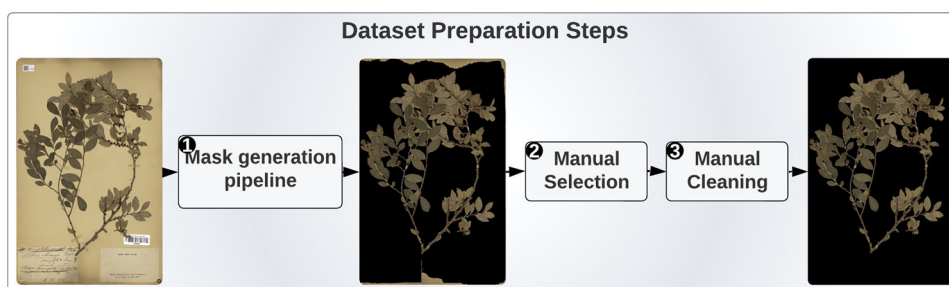

FIGURE 2 The data preparation pipeline. This pipeline defines a semi-automated approach for preparing an initial dataset of segmented images to train deep learning segmentation models. This pipeline builds on the mask generation pipeline (step 1 in Figure 4), which enabled us to build a first dataset of 5000 images, from which we selected 2227 images (step 2) to train the deep learning model. Of these, we had to manually enhance 1500 images (step 3).
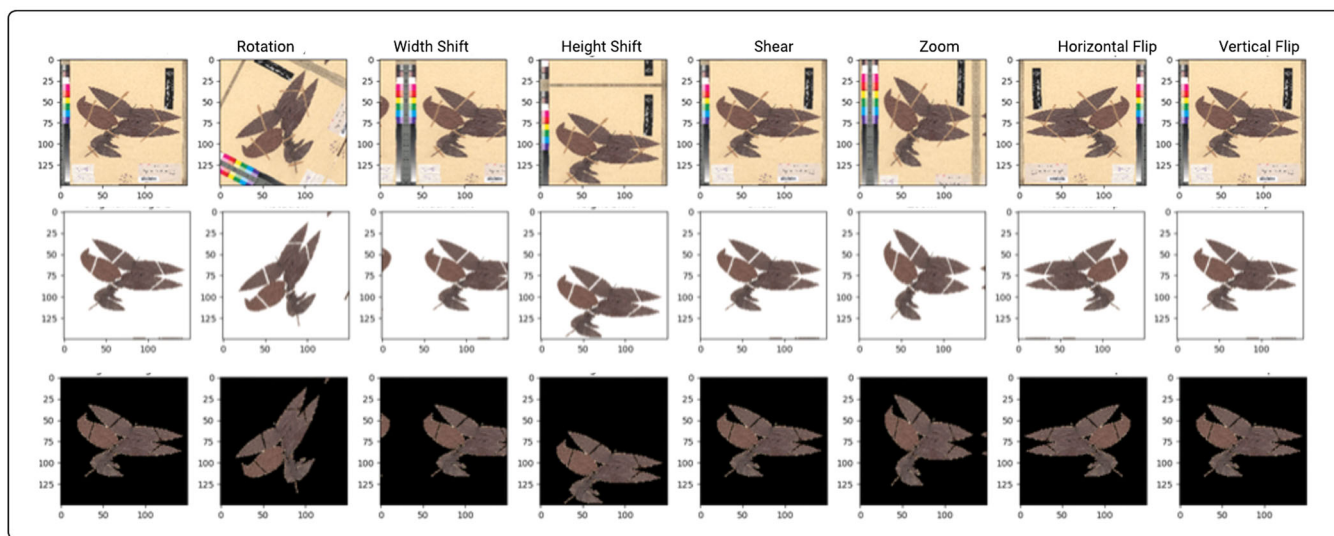

FIGURE 3 Examples from the training dataset showing the augmentation techniques applied to the original images (top row), segmented images with a white background (middle row), and segmented images with a black background (bottom row).

Segmentation approach

Our segmentation approach is implemented in two distinct phases. The first phase, color interval segmentation, is used to create an initial dataset of 9800 images. In the second phase, 2277 images were carefully selected from the initial dataset and manually corrected, and then used for deep learning–based segmentation.

Color interval segmentation

The complete pipeline of the color interval segmentation (Figure 4) comprises four stages: (i) removal of non-plant elements, (ii) HSV color segmentation, (iii) morphological transformations, and (iv) plant mask generation. It generates plant tissue masks leveraging the HSV color space's ability to isolate plant regions, as well as morphological transformations to enhance the mask's accuracy. The process illustrates the step-by-step transformation from RGB to HSV and the application of the defined intervals for green color extraction (Figure 4).

- *Removal of non-plant elements:* In this stage, we focus on detecting non-plant elements within specimen images, a crucial step for isolating plants from potential interfering elements. To achieve this, we prepared a manually annotated dataset of 950 images, specifically targeting images containing non-plant elements such as barcodes, boxes, stamps, color palettes, and rulers. We used this dataset to
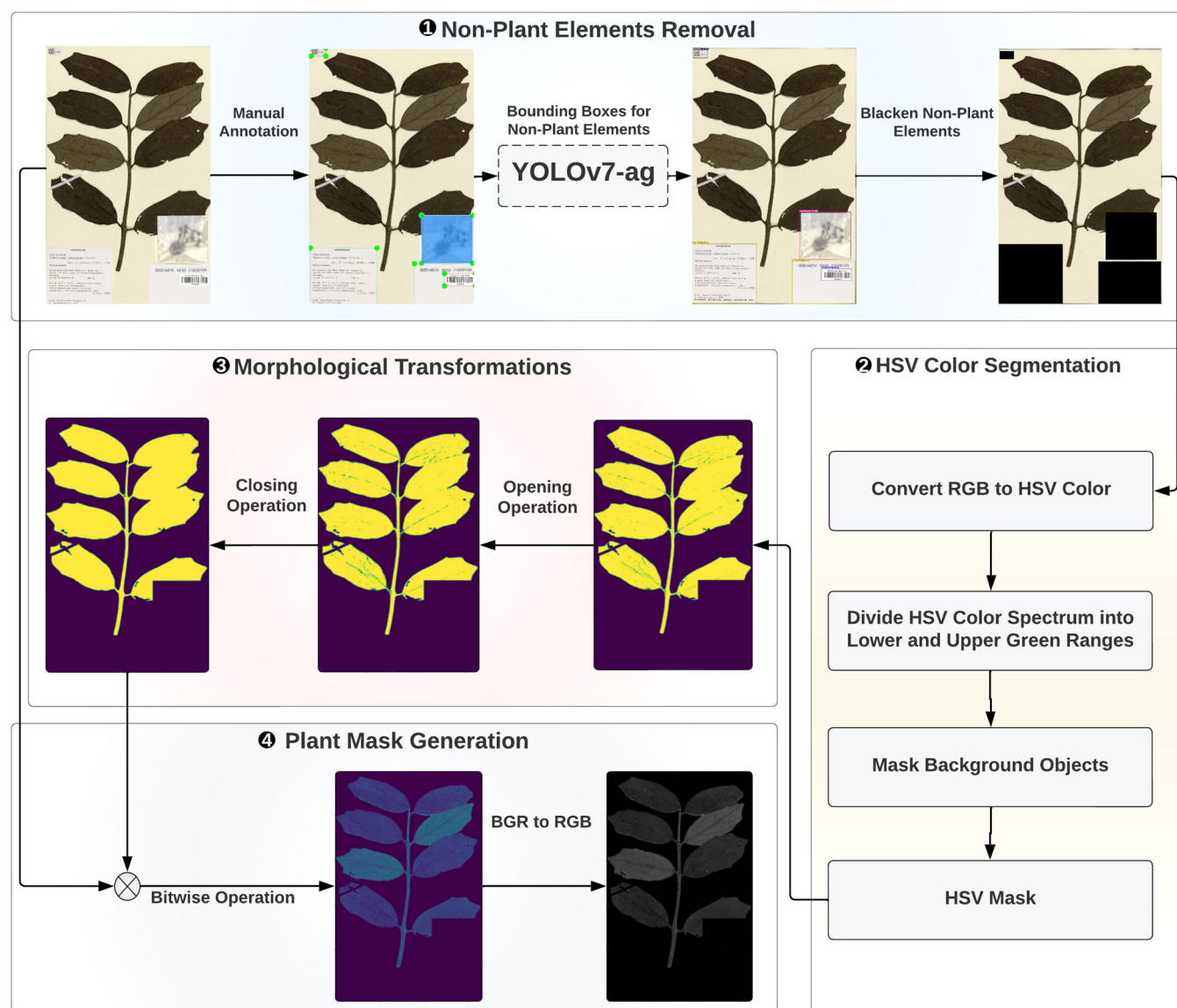

FIGURE 4 The mask generation pipeline. This pipeline defines an automated approach for preparing an initial segmented image dataset to train models for deep learning segmentation. We were able to build a first dataset of 5000 images, from which we selected 2227 images to train the deep learning model. Of these, we had to manually enhance 1500 images.

train a YOLO-based model (YOLOv7-ag; Ariouat et al., 2024) for identifying non-plant elements. The YOLOv7-ag model is distinguished by the integration of an attention mechanism, which significantly improves the model's focus on relevant features (Ariouat et al., 2024). Following the detection process, the outputs from YOLOv7-ag undergo post-processing, which is instrumental in distilling the relevant information regarding non-plant elements, thereby preparing the images for more detailed segmentation in the next phase of the pipeline. We removed all non-plant elements identified in the object detection phase by converting their pixel color to black or white (for black or white background).

- *HSV color segmentation:* To isolate the plant region, we utilize the HSV color space, which is better aligned with human color perception than the RGB color space. Rather than limiting our scope to green hues alone, we define specific ranges that capture a broader spectrum of colors. Specifically, we set the lower HSV bounds to [0,18,20] and the upper bounds to [88,200,160]. These intervals are chosen not only to target the green hues typically associated with plant tissues but also to include a wider range of colors present in the plant.
- *Morphological transformations:* We apply morphological operations, primarily targeting binary images, which involve the original image and a structuring element. Our focus is on three operations: opening, closing, and bitwise. The opening operation, similar to erosion but less destructive, preserves regions that match the structuring element's shape. The closing operation, akin to dilation, maintains the original boundary shape while enlarging foreground regions. Both operations involve a sequence of erosion and dilation with the same structuring element.

- *Plant mask generation:* Finally, we apply bitwise operations on images and their corresponding binary masks. This technique outlines the plant by assigning a value of 1 to the plant pixels and 0 to the background pixels. We applied a bitwise 'AND' operation between the image and the mask, retaining only the pixels where both the original image and the mask had a value of 1, corresponding to the plant. All other pixels, representing the background or non-plant elements, were set to 0, rendering the background black. The result is a segmented image in which only the plant is visible, distinctly isolated from its surroundings. Next, we convert the segmented image, initially in blue, green, red (BGR) format, to RGB format, which is most commonly used in computer vision.

Deep learning segmentation

Our deep learning–based segmentation approach (Figure 5) is based on the U-Net architecture, which was originally introduced by Ronneberger et al. (2015) for precise segmentation of medical images but has also demonstrated significant effectiveness in domains such as botanical image processing. Its effectiveness stems from an upsampling mechanism and the ability to capture multi-scale features, allowing it to segment plant elements of varying sizes and shapes. To further enhance the model's performance, we employed a strategy using small image patches rather than entire images. This approach allows the model to better represent the intricate details and varied characteristics present in plant images, such as fine granularity and diverse shapes and sizes. This refinement ensures more detailed and accurate segmentation, accommodating the unique textural and structural complexities of diverse plant species.
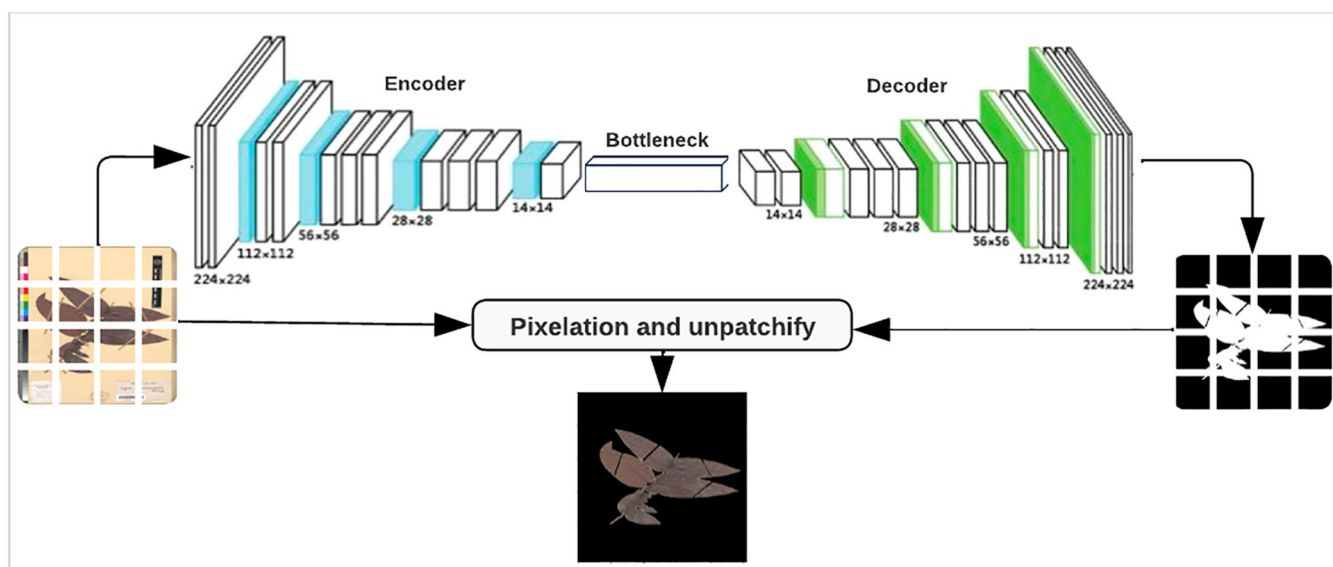

FIGURE 5 Segmentation approach using the U-Net architecture with small image patches to enhance the model's ability to accurately identify and segment plant elements in diverse and challenging botanical datasets.

For training, we used binary annotations for the masks. However, this binary setup presents challenges, particularly because the number of background pixels often far exceeds the number of plant pixels, resulting in an imbalance in the binary classification task. This disproportion skews the model's learning, as it tends to overwhelmingly encounter and learn from background pixels rather than the relatively fewer plant pixels. Such an imbalance can potentially affect the model's ability to accurately identify and segment plant features, requiring strategies to mitigate this issue for effective training and accurate segmentation outcomes. To address this, we combined the binary cross-entropy (BCE) loss with the Jaccard loss to guide the learning process. This combination helps balance the focus between plant and background pixels. BCE loss ensures accuracy, while Jaccard loss enhances the overall shape similarity between the predicted segmentation and the ground truth. By combining these two losses, the model is better able to accurately predict object pixels while minimizing overall errors.

The result of the segmentation process is a set of small binary masks, which are subsequently combined using the *patchify* library (https://github.com/dovahcrow/patchify.py). The combined masks are then overlaid onto the original image to achieve the pixelation, which involves running through the acquired masks and substituting zero-value pixels with zero in the original image. This approach is critical for preserving the plant's original pixels, which is of paramount importance in the context of our study.

RESULTS

Experimental setup

We evaluated our approach by comparing the pixel label predictions from U-Net (using both the WB and BB) with the manually curated ground truth labels for 358 validation images, covering 11 distinct plant families and genera. The plant phenotyping dataset used is detailed in the Methods section (under "Dataset collection and preprocessing"). Initially, our experimental analysis focused on plant segmentation, comparing the performance of both the WB and BB models on the test set. Subsequently, we assessed the performance improvements achieved by using these models' outputs in two image classification networks, namely ResNet101 and ViT.

Implementation details and evaluation protocol

The images in our dataset were divided into two separate groups: 80% were allocated to the training set and 20% for validation. Our experiments were conducted on the TensorFlow platform (Google Brain, Mountain View, California, USA), leveraging the computational power of four NVIDIA A100 GPUs, each equipped with 80 GB of memory. To ensure efficient and stable convergence during training, we structured our mini-batches by randomly selecting samples

of a batch size of 16. The adaptive moment estimation (Adam) optimizer (Kingma and Ba, 2014) was chosen for its rapid convergence capabilities, and we trained the model for 300 epochs. This duration was chosen as we observed that beyond this point, the model's performance plateaued with no further improvements. The learning rate was maintained at 0.0001 throughout the training process.

Metrics for segmentation

We used the IoU metric as our primary evaluation benchmark, which is commonly used in segmentation tasks. IoU quantifies the overlap between the segmented plant region and its actual ground truth. However, given the class imbalance between our two classes (background and plants), relying solely on accuracy as an evaluation metric can be misleading. High accuracy might not fully capture the model's performance when one class is significantly more prevalent than the other.

To provide a more comprehensive evaluation, we also computed precision and recall. Precision measures the proportion of correctly identified plant pixels among all pixels predicted as plants, while recall assesses the model's ability to capture all relevant plant pixels from the ground truth. Together, these metrics offer a balanced view of the model's performance, highlighting both false positives and false negatives.

To evaluate the impact of segmentation on the performance of classification models, we conducted experiments using ResNet101 and ViT, using another dataset comprising 4005 images that were annotated for the presence or absence of five distinct plant traits by botanical experts using the Les herbonautes platform (http://lesherbonautes.mnhn.fr). In total, the experts annotated 27 traits, but in this paper, we illustrate the results using only five of these traits: thorns (sharp structures on branches or stems), fruit, leaves with an acuminate tip, infructescence (arrangements of elementary fruits, e.g., grapes), and leaves with an acute base (Figure 6). These images underwent a segmentation process, resulting in a dataset of 4005 segmented images. A significant challenge arose due to the substantial imbalance in trait annotations. To address this, we created a series of balanced datasets for each trait, each comprising segmented and unsegmented images (Table 1).

Performance of deep learning segmentation

Our segmentation model achieved an IoU coefficient of 0.82% for the predicted masks across the entire test set of 358 images, regardless of the background used (WB or BB). This evaluation extended to each of the 11 plant families and genera represented in our test data (Table 2), where we observed that the performance of the two models was generally similar across these groups. For most taxa, the IoU coefficient was above 0.90%, indicating high segmentation

FIGURE 6 Herbarium specimen examples (left column) showcasing traits (thorns, fruits, leaves with acuminate tips, infructescence, leaves with an acute base) and their corresponding segmented images with black background (middle column) and white background (right column).

TABLE 1 The distribution of images in the balanced training and validation datasets for the five studied traits, derived from the initial dataset of 4005 images annotated by expert botanists.

Trait	Number of images	
	Train	Validation
Thorns	2426	500
Fruits	2776	580
Leaves with acuminate tips	2360	500
Infructescence	2340	480
Leaves with an acute base	2308	500

accuracy. However, for some taxa, such as Convolvulaceae and *Desmodium*, the IoU coefficients were slightly lower, with values around 0.82% and 0.87%, respectively. This slight reduction in IoU might be attributed to the complex morphology of the plants within these taxa, which include small leaves, flowers, and slender stems that challenge the segmentation process.

Overall, while both models performed well, the choice of background color significantly impacted the segmentation outcomes. In some cases, the BB model provided better recall (Table 2), indicating its effectiveness in capturing a higher number of true positives, especially for lighter or

TABLE 2 Evaluation of the segmentation model in two variants (with white background and black background) across 11 plant families and genera, using a validation dataset of 358 images to assess four performance metrics: intersection over union (IoU) coefficient, precision, recall, and F1 score.

Family	No. of images	Models	Metrics			
			IoU coefficient	Precision	Recall	F1 score
Amborella	5	White background	0.94	0.98	0.95	0.97
		Black background	0.93	0.94	0.98	0.96
Castanea	61	White background	0.94	0.97	0.96	0.97
		Black background	0.94	0.95	0.99	0.97
Convolvulaceae	25	White background	0.82	0.94	0.87	0.90
		Black background	0.84	0.86	0.97	0.91
Desmodium	11	White background	0.87	0.98	0.88	0.93
		Black background	0.88	0.89	0.98	0.93
Eugenia	40	White background	0.90	0.95	0.94	0.94
		Black background	0.90	0.91	0.99	0.95
Laurus	53	White background	0.94	0.97	0.97	0.97
		Black background	0.94	0.95	0.99	0.97
Litsea	17	White background	0.91	0.94	0.96	0.95
		Black background	0.91	0.92	0.99	0.95
Magnolia	34	White background	0.94	0.97	0.97	0.97
		Black background	0.95	0.95	0.99	0.97
Monimiaceae	37	White background	0.92	0.95	0.96	0.96
		Black background	0.92	0.93	0.99	0.96
Rubus	22	White background	0.92	0.97	0.95	0.96
		Black background	0.94	0.95	0.99	0.97
Ulmus	53	White background	0.91	0.98	0.93	0.95
		Black background	0.92	0.93	0.98	0.96

TABLE 3 Performance evaluation of the ViT model on unsegmented images and two segmented image variants (segmented with black backgrounds [BB] and segmented with white backgrounds [WB]) across five botanical traits.[a]

Trait	Unsegmented		Segmented BB		Delta[b]		Segmented WB		Delta[b]	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Thorns	97.00	92.66	95.60	92.59	−01.40	−00.07	94.80	94.26	−2.20	+1.60
Fruits	64.31	63.86	64.66	66.86	+00.35	+03.00	62.93	66.01	−1.38	+2.15
Leaves with acuminate tips	82.00	77.52	83.00	80.23	+01.00	+02.71	80.20	75.32	−1.80	−2.20
Infructescence	70.63	64.98	71.46	72.65	+00.83	+07.67	69.17	70.51	−1.46	+5.53
Leaves with an acute base	71.80	68.22	75.20	75.40	+03.40	+07.18	69.80	70.85	−2.00	+2.63

[a]The metrics used for comparison are accuracy (Acc) and F1 score (F1).

[b]The delta values represent the performance difference between segmented and unsegmented images for both variants.

more intricate plant features. The WB model, on the other hand, exhibited superior precision, particularly with darker specimens, suggesting it was better at minimizing false positives. For example, in *Magnolia* and *Eugenia*, the BB model showed higher recall, which is crucial for ensuring that all relevant plant features are detected. Conversely, the WB model showed higher precision, which is essential for correctly identifying plant features without including non-plant pixels. The segmentation results for different images from the two models are shown in Figure 6.

DISCUSSION

In our study, we developed an automated method to extract herbarium masks using semantic segmentation models, with the goal of removing the background from herbarium images and retaining only the plant. This step is crucial for cleaner preprocessing in subsequent classification models.

Our segmentation process captures various plant components, including leaves, stems, flowers, fruits, and seeds, and the models we employed achieved IoU scores surpassing 96%. The choice between the WB and BB models depends on the specific characteristics of the plant elements, with the WB model being more effective for plants with darker features and the BB model excelling with lighter and finer

TABLE 4 Performance evaluation of the ResNet101 model on unsegmented images and two segmented image variants (segmented with black backgrounds [BB] and segmented with white backgrounds [WB]) across five botanical traits.[a]

Trait	Unsegmented		Segmented BB		Delta[b]		Segmented WB		Delta[b]	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Thorns	94.20	92.86	95.20	93.47	−01.00	−00.60	82.00	80.08	−12.20	−12.78
Fruits	60.17	67.28	64.66	67.22	−04.49	+00.06	52.51	66.67	−07.66	−00.61
Leaves with acuminate tips	79.40	78.88	77.00	73.92	+02.24	+04.96	71.40	69.47	−08.00	−09.41
Infructescence	70.83	72.87	69.37	69.10	+01.46	+03.77	56.04	66.67	−14.79	−06.20
Leaves with an acute base	73.00	72.66	71.40	69.42	+01.60	+03.24	58.20	65.03	−14.80	−07.63

[a]The metrics used for comparison are accuracy (Acc) and F1 score (F1).

[b]The delta values represent the performance difference between segmented and unsegmented images for both variants.
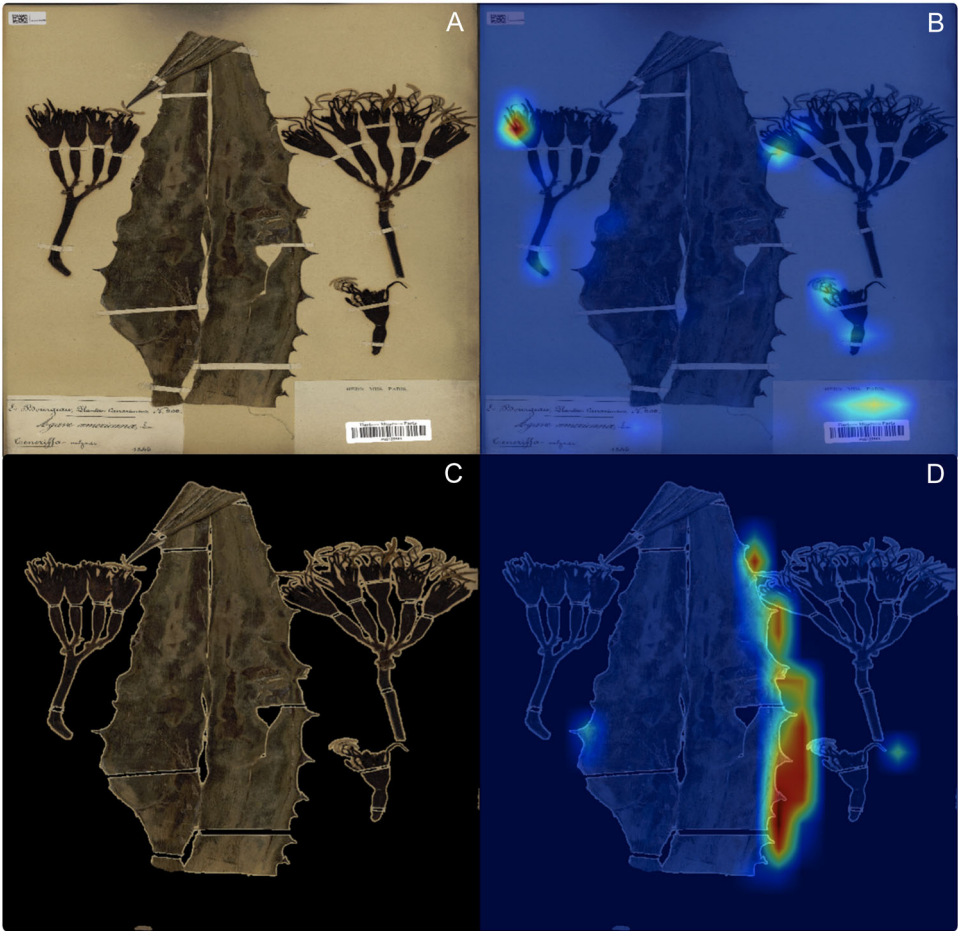

FIGURE 7 Comparison of the ViT model's performance in two distinct scenarios. (A) The model incorrectly predicted the absence of thorns, likely due to excessive focus on the background. (B) Corresponding heatmap showing inadequate attention to relevant plant regions. (C) The model accurately identified the presence of thorns after background removal. (D) Heatmap demonstrating appropriately directed attention on relevant plant features, leading to correct classification.

plant details. Our research highlights the effectiveness of the models on previously unseen images while acknowledging certain limitations, particularly when dealing with unfamiliar backgrounds.

The evaluation of the impact of segmentation on the performance of classification models demonstrates an improvement in the performance of ViT for four out of five traits, with increases up to 3% in accuracy and 7% in F1 score (Table 3) when using segmented images with the BB model. A similar trend is observed for ResNet101 (Table 4). Conversely, when using segmented images with the WB model, ViT shows lower accuracy compared to using non-segmented images, but the F1 score improves for four out of five traits. As for ResNet101 with the WB model, we observed a drastic decrease in performance across all traits and both metrics. This decline can be attributed to the nature of the images used for classification, which are predominantly lighter, thus the WB model did not perform well, as it is better suited for images with darker plant specimens, as previously discussed.

Another important aspect is illustrated in the key regions influencing the predictions made by ViT on both non-segmented and segmented images (Figure 7, with BB). In the first example, ViT inaccurately predicted the absence of thorns on a plant that, in fact, had thorns (Figure 7A, B). This error may be attributed to the model's undue focus on the background rather than on the plant itself. Conversely, in a second example, ViT correctly identified the presence of thorns, attributed to the model's appropriate focus on the plant, leading to an accurate prediction (Figure 7C, D). This comparison highlights the influence of background removal on the model's ability to correctly identify essential plant features, emphasizing the significance of targeted focus for precise classification in deep learning models. Our study underlines the critical role of dataset preprocessing in enhancing the generalization capabilities of models within the domain of botanical image classification.

This study clearly demonstrates the potential of deep learning semantic segmentation as a preprocessing tool to reduce visual noise in herbarium images before applying classification models. Looking ahead, our goal is to expand this research to develop a comprehensive herbarium feature recognition system, which will include digital reconstruction of lost or damaged plant parts. Such advances promise not only to improve the accuracy of plant identification, but also to make a significant contribution to the preservation and study of botanical specimens.

AUTHOR CONTRIBUTIONS

H.A. and Y.S. planned and designed the data preprocessing pipeline and the experimental strategies. H.A. designed the model architecture and wrote the model code. Y.S. prepared the working environment on the high-peformance computing cluster, and planned and supervised data acquisition and preparation. H.A. and Y.S. conducted the training and wrote the initial manuscript. E.C., J.-D.Z., and E.P. contributed to the e-Col+ project planning and design. All authors contributed to revising and editing the text and approved the final version of the manuscript.

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The source code for segmentation, examples, and trained models for networks with black and white backgrounds are available at https://github.com/IA-E-Col/Herbarium-Image-Segmentation. The used dataset is available at https://doi.org/10.6084/m9.figshare.27685914.v1 (Sklab et al., 2024b).

ORCID

Hanane Ariouat 🄳 http://orcid.org/0009-0002-7066-4358

REFERENCES

Ariouat, H., Y. Sklab, M. Pignal, R. V. Lebbe, J.-D. Zucker, E. Prifti, and E. Chenin. 2023. Extracting masks from herbarium specimen images based on object detection and image segmentation techniques. *Biodiversity Information Science and Standards* 7: e112161. https://doi.org/10.3897/biss.7.112161

Ariouat, H., Y. Sklab, M. Pignal, F. Jabbour, R. V. Lebbe, E. Prifti, and E. Chenin. 2024. Enhancing YOLOv7 for plant organs detection using attention-gate mechanism. *In* D. N. Yang, X. Xie, V. S. Tseng, J. Pei, J. W. Huang, and J. C. W. Lin [eds.], Advances in knowledge discovery and data mining, PAKDD 2024. Springer, Singapore.

Borhani, Y., J. Khoramdel, and E. Najafi. 2022. A deep learning-based approach for automated plant disease classification using vision transformer. *Scientific Reports* 12: e11554. https://doi.org/10.1038/s41598-022-15163-0

Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv 1802.02611 [preprint]. Available at https://arxiv.org/abs/1802.02611 [posted 7 February 2018; accessed 28 January 2025].

Corney, D., J. Clark, H. Tang, and P. Wilkin. 2012. Automatic extraction of leaf characters from herbarium specimens. *Taxon* 61: 231–244. https://doi.org/10.1002/tax.611016

Dhaka, V. S., S. V. Meena, G. Rani, D. Sinwar, Kavita, M. F. Ijaz, and M. Woźniak. 2021. A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. *Sensors* 21(14): e4749. https://www.mdpi.com/1424-8220/21/14/4749

Dice, L. R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26(3): 297–302.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, et al. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations. Available at: https://openreview.net/forum?id=YicbFdNTTy

Fan, X., R. Zhou, T. Tjahjadi, S. Das Choudhury, and Q. Ye. 2022. A segmentation-guided deep learning framework for leaf counting. *Frontiers in Plant Science* 13: e44522 https://doi.org/10.3389/fpls.2022.844522

He, K., X. Zhang, S. Ren, and J. Sun. 2015. Deep residual learning for image recognition. ArXiv 1512.03385 [preprint]. Available at: http://arxiv.org/abs/1512.03385 [posted 10 December 2015; accessed 10 January 2025].

Hussein, B. R., O. A. Malik, W.-H. Ong, and J. W. F. Slik. 2020. Semantic segmentation of herbarium specimens using deep learning techniques. *In* R. Alfred, Y. Lim, H. Haviluddin, and C. K. On [eds.], Computational Science and Technology, 321–330. Springer, Singapore.

Kingma, D., and J. Ba. 2014. Adam: A Method for Stochastic Optimization. arXiv 1412.6980 [preprint]. Available at https://arxiv.org/abs/1412.6980 [posted 22 December 2014; accessed 28 January 2025].

Leksut, J. T., J. Zhao, and L. Jetti. 2020. Learning visual variation for object recognition. *Image and Vision Computing* 98: e103912. https://doi.org/10.1016/j.imavis.2020.103912

Lorieul, T., K. D. Pearson, E. R. Ellwood, H. Goeau, J.-F. Molino, P. W. Sweeney, and A. Joly. 2019. Toward a large-scale and deep phenological stage annotation of herbarium specimens: Case studies from temperate, tropical, and equatorial floras. *Applications in Plant Sciences* 7: e1233. https://doi.org/10.1002/aps3.1233

Moayeri, M., P. Pope, Y. Balaji, and S. Feizi. 2022. A comprehensive study of image classification model sensitivity to foregrounds, backgrounds, and visual attributes. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 19065–19075. Los Alamitos, California, USA. https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01850

Otsu, N. 1979. A threshold selection method from Gray-Level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9: 62–66.

Pohlen, T., A. Hermans, M. Mathias, and B. Leibe. 2017. Full-resolution residual networks for semantic segmentation in street scenes. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4151–4160, 21–26 July 2017, Honolulu, Hawaii, USA.

Ronneberger, O., P. Fischer, and T. Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). Springer, Cham, Switzerland.

Sahraoui, M., Y. Sklab, M. Pignal, R. V. Lebbe, and V. Guigue. 2023. Leveraging multimodality for biodiversity data: Exploring joint representations of species descriptions and specimen images using CLIP. *Biodiversity Information Science and Standards* 7: e112666. https://doi.org/10.3897/biss.7.112666

Simonyan, K., and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556 [preprint]. Available at https://arxiv.org/abs/1409.1556 [posted 4 September 2014; accessed 28 January 2025].

Sklab, Y., H. Ariouat, Y. Boujydah, Y. Qacami, E. Prifti, J. D. Zucker, and E. Chenin. 2024a. Towards a deep learning-powered herbarium image analysis platform. *Biodiversity Information Science and Standards* 8: e135629. https://doi.org/10.3897/biss.8.135629

Sklab, Y., H. Sklab, E. Prifti, E. Chenin, and J. D. Zucker. 2024b. Herbarium image segmentation dataset with plant masks for enhanced morphological trait analysis. figshare [dataset]. Available at: https://doi.org/10.6084/m9.figshare.27685914.v1 [posted 17 November 2024; accessed 15 January 2025].

Sklab, Y., H. Ariouat, E. Prifti, E. Chenin, and J. D. Zucker. In press. Identification of non-plant elements in herbarium images using YOLO. In Proceedings of the Conférence Africaine sur la Recherche en Informatique et en Mathématiques (CARI).

Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 5(4): 1–34

Thompson, K. M., R. Turnbull, E. Fitzgerald, and J. L. Birch. 2023. Identification of herbarium specimen sheet components from high-resolution images using deep learning. *Ecology and Evolution* 13(8): e10395. https://doi.org/10.1002/ece3.10395

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

Triki, A., A. Bouaziz, and W. Mahdi. 2022a. A deep learning-based approach for detecting plant organs from digitized herbarium specimen images. *Ecological Informatics* 69: e101590. https://doi.org/10.1016/j.ecoinf.2022.101590

Triki, A., B. Bouaziz, W. Mahdi, H. Hamed, and J. Gaikwad. 2022b. Deep learning-based approach for digitized herbarium specimen segmentation. *Multimedia Tools and Applications* 81: 28689–28707. https://doi.org/10.1007/s11042-022-12935-8

Wang, C.-Y., A. Bochkovskiy, and H.-Y. M. Liao. 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 17–24 June 2023, Vancouver, Canada.

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

White, A. E., R. B. Dikow, M. Baugh, A. Jenkins, and P. B. Frandsen. 2020. Generating segmentation masks of herbarium specimens and a data set for training segmentation models using deep learning. *Applications in Plant Sciences* 8(6): e11352.

Wilde, B. C., J. G. Bragg, and W. Cornwell. 2023. Analyzing trait-climate relationships within and among taxa using machine learning and herbarium specimens. *American Journal of Botany* 110(5): e16167. https://doi.org/10.1002/ajb2.16167

Younis, S., C. Weiland, R. Hoehndorf, S. Dressler, T. Hickler, B. Seeger, and M. Schmidt. 2018. Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Botany Letters* 165(3–4): 377–383. https://doi.org/10.1080/23818107.2018.1446357

Younis, S., M. Schmidt, C. Weiland, S. Dressler, B. Seeger, and H. Thomas. 2022. 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

Yue, Z., T. Durand, E. Chenin, M. Pignal, P. Gallinari, and R. Vignes-Lebbe. 2017. Using a deep convolutional neural network for extracting morphological traits from herbarium images. *Biodiversity Information Science and Standards* 1: e20400. https://doi.org/10.3897/tdwgproceedings.1.20400

Zhou, Z., M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. 2018. Unet++: A nested U-Net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA ML-CDS 2018). Springer, Cham, Switzerland.