Received: 15 March 2025 | Accepted: 30 September 2025

DOI: 10.1002/aps3.70034

Applications in Plant Sciences

APPLICATION ARTICLE

PlantSAM: An object detection-driven segmentation pipeline for herbarium specimens

Youcef Sklab[1] | Florian Castanet[1] | Hanane Ariouat[1] | Souhila Arib[2] | Jean-Daniel Zucker[1,3] | Eric Chenin[1] | Edi Prifti[1,3]

[1]Institut de Recherche pour le Développement (IRD), Sorbonne Université, UMMISCO, Paris, France

[2]Laboratoire Etis, UMR 8051 CNRS, CY Cergy Paris Université, Cergy-Pontoise, France

[3]Sorbonne Université, INSERM, Nutrition et Obesities; systemic approaches, NutriOmique, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France

Correspondence

Youcef Sklab, Institut de Recherche pour le Développement (IRD), Sorbonne Université, UMMISCO, Paris, France.
Email: Youcef.sklab@ird.fr

Abstract

Premise: Deep learning–based classification of herbarium images is hampered by background heterogeneity, which introduces noise and artifacts that can potentially mislead models and degrade their accuracy. Addressing these effects is essential to enhance overall performance.

Methods: We introduce PlantSAM, an automated segmentation pipeline that integrates YOLOv10 for plant region detection and the Segment Anything Model (SAM2) for segmentation. YOLOv10 generates bounding box prompts to guide SAM2, enhancing segmentation accuracy. Both models were fine-tuned on herbarium images and evaluated using intersection over union (IoU) and Sørensen–Dice coefficient metrics.

Results: PlantSAM achieved state-of-the-art segmentation performance, with an IoU of 0.94 and a Sørensen–Dice coefficient of 0.97. Incorporating segmented images into classification models led to consistent performance improvements across five tested botanical traits, with accuracy gains of up to 4.36% and F1 score improvements of 4.15%.

Conclusions: Our findings highlight the importance of background removal in herbarium image analysis, as it significantly enhances classification performance by enabling models to focus more effectively on the foreground plant structures.

KEYWORDS

botanical analysis, herbarium specimens, Segment Anything Model (SAM), semantic segmentation, YOLOv10

Plants are a cornerstone of Earth's biodiversity, playing a critical role in ecosystem stability, carbon cycling, and the regulation of atmospheric gases (Raven, 2019). However, this biodiversity is under severe threat from accelerating climate change and anthropogenic pressures (Soltis, 2017; Besnard et al., 2018). Natural history collections, particularly herbarium specimens, preserve centuries of plant data and serve as valuable archives for tracking morphological traits, species distribution, and long-term environmental changes (Meredith, 1996; Besnard et al., 2018; Younis et al., 2020; Abdelaziz and Walid, 2022; Ariouat et al., 2023).

Large-scale digitization initiatives have significantly expanded access to these collections (Sweeney et al., 2018). Institutions such as the National Museum of Natural History (Muséum National d'Histoire Naturelle [MNHN]) in Paris now make millions of specimens publicly available via platforms like ReColNat (https://www.recolnat.org/en/). Additionally, broader aggregators such as GBIF (https://www.gbif.org) provide centralized access to global biodiversity data contributed by institutions worldwide. These digital resources enable high-throughput analyses of morphological traits such as leaf size, shape, and organ presence (Zhang et al., 2022), while also supporting research in crop quality (Jiang and Li, 2020), disease classification (Borhani et al., 2022), and soil evolution (Grasso, 2024). However, fully exploiting these vast datasets requires robust automatic tools to extract descriptive morpho-anatomical information, including specimen-specific traits such as organ presence or conservation state (Dhaka et al., 2021).

In recent years, deep learning has emerged as a powerful solution for such automated analyses (Dhaka et al., 2021;

Sahraoui et al., 2023; Ariouat et al., 2023, 2025; Sklab et al., 2024a, 2025a, 2025b). Leveraging advances in convolutional neural networks (CNNs) and Vision Transformers (ViTs) (Lecun et al., 1998; Simonyan and Zisserman, 2015; Szegedy et al., 2015; He et al., 2016; Krizhevsky et al., 2017; Dosovitskiy et al., 2021), deep learning models have achieved strong performance in image classification, object detection, and segmentation (Krizhevsky et al., 2017; Dosovitskiy et al., 2021). In botanical research, these methods have been successfully applied to identify plant organs in herbarium images (Mochida et al., 2018; Younis et al., 2020; Ariouat et al., 2025), particularly for segmentation-based tasks such as leaf counting, organ detection, and trait extraction (Triki et al., 2021; Fan et al., 2022; Weaver and Smith, 2023; Wilde et al., 2023; Sklab et al., 2025a).

Nonetheless, herbarium images present unique challenges for deep learning–based models due to complex, heterogeneous backgrounds and visual artifacts, including labels, scale bars, handwritten annotations, overlapping structures, and aging effects (Mochida et al., 2018; Sklab et al., 2024a, 2025a, 2025b; Ariouat et al., 2025). Such noise can lead models to learn spurious features, thereby reducing their effectiveness in downstream tasks like species identification and trait classification (Hussein et al., 2020; White et al., 2020; Triki et al., 2021; Fan et al., 2022). Isolating the foreground plant structures (e.g., leaves, stems, flowers, and fruits) from non-plant elements is thus essential for improving model accuracy and robustness (Ariouat et al., 2025; Sklab et al., 2025a).

Various approaches have been proposed to address these issues. Fan et al. (2022) applied UNet++ (Zhou et al., 2018) with a ResNet50 backbone (He et al., 2016) to segment *Arabidopsis* and tobacco leaves, showing good performance on natural backgrounds but limited generalization to herbarium images. Triki et al. (2022) developed a segmentation pipeline based on coarse annotation masks assuming uniform transparent backgrounds, which restricts its use on complex specimens. White et al. (2020) employed Otsu's thresholding to segment ferns, but their method focused on fronds and was not extended to other organs. Hussein et al. (2020) and Lee et al. (2025) introduced models targeting specific organs, such as roots and leaves, without addressing full-specimen segmentation. In parallel, object detection pipelines such as GinJinn (Ott et al., 2020) and GinJinn2 (Ott and Lautenschlager, 2022) have been designed to detect intact leaves or non-plant components using bounding boxes. Thompson et al. (2023) trained a YOLOv5-based model to detect annotations and scale bars, achieving good generalization with minimal fine-tuning. However, these approaches generally lack pixel-level precision and remain limited in scope.

A recent effort by Sklab et al. (2025a) introduced SIM-Net, a multimodal network that combines CNN-based appearance features with 3D geometric cues inferred from segmentation masks. While the approach improves robustness in regard to background noise and occlusion, it relies heavily on the availability of high-quality masks and does not solve the problem of generating them automatically at scale.

Traditional segmentation models (Hussein et al., 2020; White et al., 2020; Triki et al., 2021, 2022; Fan et al., 2022; Weaver and Smith, 2023; Wilde et al., 2023) also face difficulties when confronted with the structural complexity and visual variability of herbarium images. Recently, foundation models such as the Segment Anything Model (SAM) (Kirillov et al., 2023) and its enhanced version SAM2 (Ravi et al., 2024) have emerged as powerful alternatives. These models leverage large-scale pretraining and prompt-based segmentation to generalize across domains (Zhang et al., 2023; Zhao et al., 2023; Chen et al., 2024; Yin et al., 2024). SAM2, in particular, incorporates architectural refinements and improved prompt integration, making it more robust to the ambiguous boundaries and intricate structures often seen in herbarium sheets. Designed for promptable segmentation, SAM2 generates masks based on user- or algorithm-provided prompts such as points, boxes, or coarse regions.

Despite their versatility, SAM and SAM2 face limitations in botanical contexts. They require high-quality prompts, struggle with degraded contours and overlapping organs, and lack scalability without domain-specific fine-tuning. To overcome these challenges, Ariouat et al. (2025) proposed a UNet-based model with a ResNet101 backbone (He et al., 2016), achieving high intersection over union (IoU) scores on simplified binary background masks. Yet, this approach proved less effective for specimens with fine stems or densely packed structures.

In this work, we propose an automatic segmentation pipeline, PlantSAM, which integrates YOLOv10 (Wang et al., 2024) for prompt generation with SAM2 for high-resolution plant segmentation. YOLOv10 provides bounding box prompts that guide SAM2 in isolating plant structures from the background. This design eliminates the need for manual annotation and scales effectively across large herbarium datasets. Our main contributions are as follows: (i) we introduce a novel segmentation pipeline tailored for herbarium specimens, combining object detection and promptable segmentation; (ii) we fine-tune SAM2 on a curated dataset to handle domain-specific challenges such as background heterogeneity and complex plant morphologies; (iii) we evaluate segmentation performance using IoU and Sørensen–Dice metrics, showing significant improvements over baseline models like UNet (Ariouat et al., 2025) and standard SAM2; and (iv) we demonstrate the impact of segmentation on trait classification, achieving accuracy gains of up to 4.36% and F1 score improvements of 4.15%. Finally, we incorporate our pipeline into a semi-automatic annotation tool, enabling expert-guided mask refinement to further enhance dataset quality and usability.

METHODS

Our proposed PlantSAM pipeline, as illustrated in Figure 1, comprises four phases: patching, plant region detection with YOLOv10, segmentation with SAM, and unpatching to recombine the segmented patches into a complete image mask.

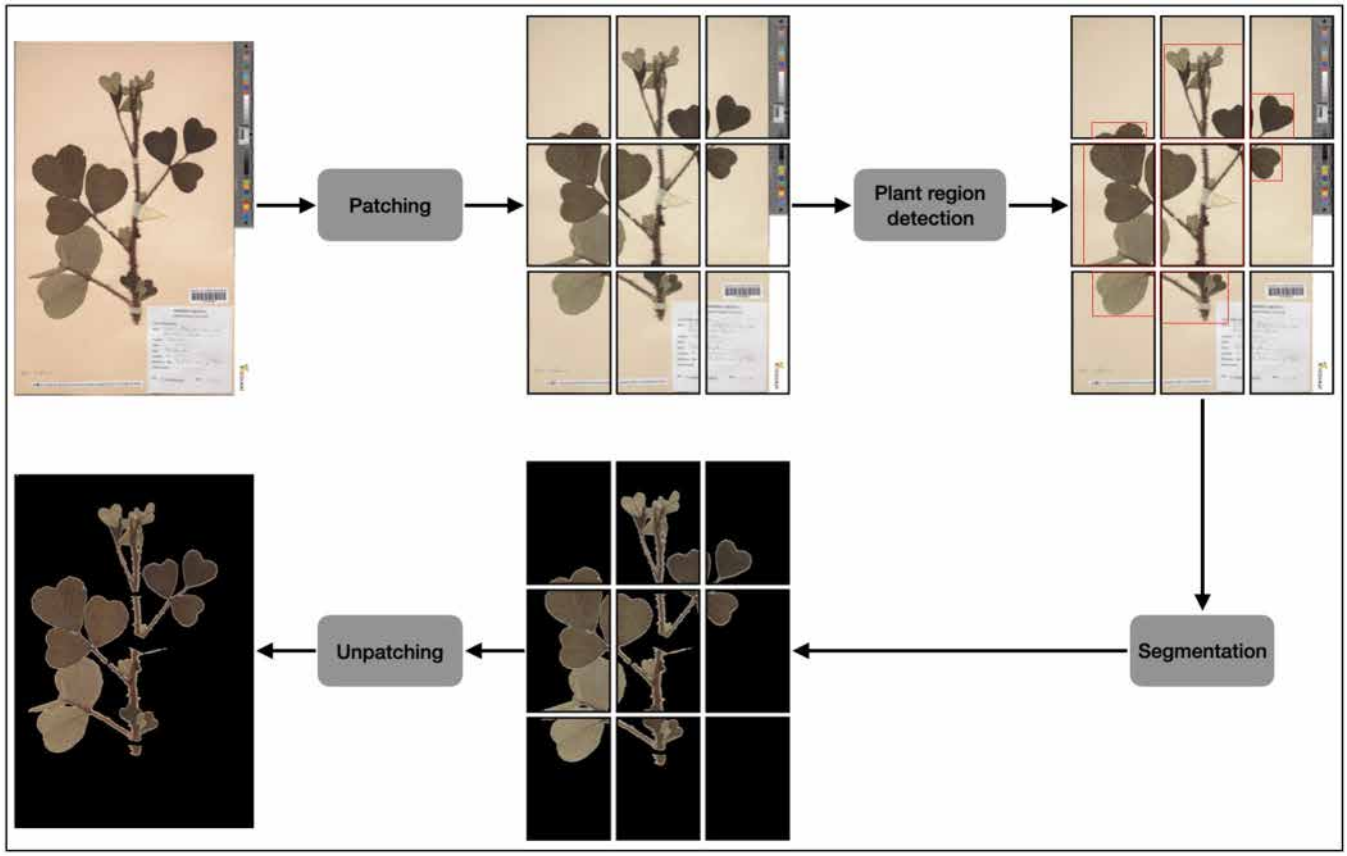

FIGURE 1 The PlantSAM plant region detection and segmentation pipeline for herbarium images. Each image is divided into patches, processed by YOLOv10 for plant region detection, and segmented by SAM. The segmented patches are then recombined to create a complete plant mask.

We evaluated both versions of SAM integrated into our pipeline: the first version (Kirillov et al., 2023), which we refer to here as SAM1, and SAM2. We used both SAM1 and SAM2 in our experiments to compare their performance with that of UNet (Ariouat et al., 2025). Although SAM2 consistently outperformed SAM1 across all benchmarks, we retained SAM1 in our experiments for comparative purposes. We denote the pipeline using SAM1 as PlantSAM1, and the one using SAM2 as PlantSAM2. When referring to the method generically, we use PlantSAM, which encompasses both variants based on the integration of YOLOv10 with SAM.

Image preprocessing and patching

Each herbarium image, typically around 4000 × 6000 pixels, is divided into smaller patches to preserve fine structural details and ensure efficient processing prior to segmentation. This patching strategy mitigates resolution loss that occurs when handling the entire image and reduces computational load. The image patch generation procedure is shown in Box 1. We begin by loading the image and applying a sequence of morphological operations (specifically, erosion followed by dilation) to refine the visual quality of the mask. This preprocessing step addresses noise artifacts present in our initial dataset, which included small, isolated pixel clusters along the mask boundaries. These artifacts originated from a previous segmentation pipeline that combined a UNet-based model with manual corrections (Ariouat et al., 2025). We first use erosion to remove these spurious pixels while preserving the core structure of the plant. We then apply dilation to recover the slightly eroded edges, enhancing both the clarity and completeness of the specimen contours.

Once preprocessing is complete, we use the image width to determine the optimal patch size: 1024, 512, or 256 pixels. This adaptive strategy allows us to balance the need for contextual information (with larger patches) against the risk of including irrelevant regions or losing resolution (with smaller patches). Selecting the appropriate patch size is crucial, as smaller patches help reduce background noise but may exclude important context, whereas larger patches may include too much irrelevant information, which can hinder models like SAM that operate with a fixed resolution of 256 × 256 pixels. After patch extraction, we normalize each patch to have zero mean and unit variance, ensuring that the input distribution aligns with what the pre-trained SAM model expects. Finally, we apply standard data augmentations, such as random flips and rotations, to improve the model's generalization capacity and the robustness of the segmentation.

BOX 1 Image processing pipeline overview.

Objective

✓ Split a high-resolution image into smaller patches after basic morphological processing (erosion and dilation).

Inputs and Outputs

✓ Input: File path of the image to be processed.
✓ Output: A set of image patches extracted based on image width.

Processing Steps

✓ Load the image from the specified path.
✓ Apply erosion to the image to reduce noise.
✓ Apply dilation to restore the main structures.
✓ Determine the width of the image.
✓ If the width is larger than 3 × 1024 pixels, split it into 1024-pixel-wide patches.
✓ If the width is larger than 3 × 512 pixels, split it into 512-pixel-wide patches.
✓ Otherwise, split it into 256-pixel-wide patches.

Return

✓ A list of patches corresponding to the size-based splitting strategy.

Plant region detection with YOLOv10

In our pipeline, we use YOLOv10 to detect plant regions within herbarium image patches, identifying component regions such as leaves, stems, flowers, and fruits. After the patching step, YOLOv10 processes each patch to produce bounding boxes around the plant structures. The detection process handles various challenges, such as overlapping plant parts, background noise, and artifacts like labels, pins, or aged paper textures. YOLOv10's bounding boxes effectively capture the spatial extent of the plant regions while minimizing false positives caused by background elements. This capability is essential for ensuring better segmentation in the subsequent step with SAM.

As illustrated in Figure 1, each image patch is processed to generate bounding boxes that localize plant structures. The number and size of these boxes vary depending on the morphological complexity of the specimen. In patches containing multiple distinct plant areas, YOLOv10 generates several bounding boxes, enabling finer-grained segmentation. This multi-region detection strategy facilitates the isolation of individual plant components. To train YOLOv10, we used the Plant Region Detection Dataset (Castanet et al., 2025b), which we constructed specifically for this task based on a previously published segmented herbarium image dataset (Ariouat et al., 2025). To build this dataset, we first split the segmented herbarium images into smaller patches, then generated bounding boxes around the plant regions by detecting contiguous non-black pixel areas, as the background in the segmented images is uniformly black. This process enables automated and consistent annotation. The final dataset contains 19,078 image patches annotated with bounding boxes, which are split into groups of 14,307 for training, 3815 for validation, and 956 for testing.

Segmentation with SAM

We explored two types of prompts for guiding the segmentation process: point prompts and bounding box prompts. Initial experiments with point prompts proved ineffective, primarily due to the absence of an automated method for accurately placing them. In particular, distinguishing between positive points (placed on the plant) and negative points (placed on the background) was challenging. Additionally, maintaining a proper balance between positive and negative points introduced further complexity. For these reasons, we chose to focus on bounding box prompts.

The generated bounding boxes serve as input prompts for the SAM models, which then generate segmentation masks for each patch. The segmented patches are then resized and reassembled to reconstruct the full-resolution segmentation of the original image. In our study, we investigated two bounding box prompting strategies:

- Single-box strategy: A single bounding box is generated to enclose the entire plant structure within each patch, encompassing all visible plant regions. The box is created by identifying the outermost non-background pixels in the mask, forming a rectangle around the complete specimen (Figure 2A). This strategy is computationally efficient and requires only one prompt per patch. However, it often captures surrounding background areas, which may introduce noise and reduce segmentation precision.
- Multi-region strategy: This approach generates multiple bounding boxes, each corresponding to a distinct connected component within the patch. It isolates clusters of connected foreground pixels more precisely, allowing finer control over the segmentation (Figure 2B). By focusing on localized regions, this strategy minimizes background inclusion and improves segmentation accuracy. However, it requires multiple prompts per patch, increasing computational cost and processing time.

Morphological operations, such as erosion and dilation, were applied to enhance mask quality by removing problematic residual pixels and ensuring more accurate detection of plant structures (Ariouat et al., 2025). We conducted a comparative study on a set of 19 images selected as representative of our dataset, and found that the multi-region strategy reduces background inclusion within bounding boxes by approximately 10% compared to the single-box strategy (Figure 3). Therefore, we incorporated it into our

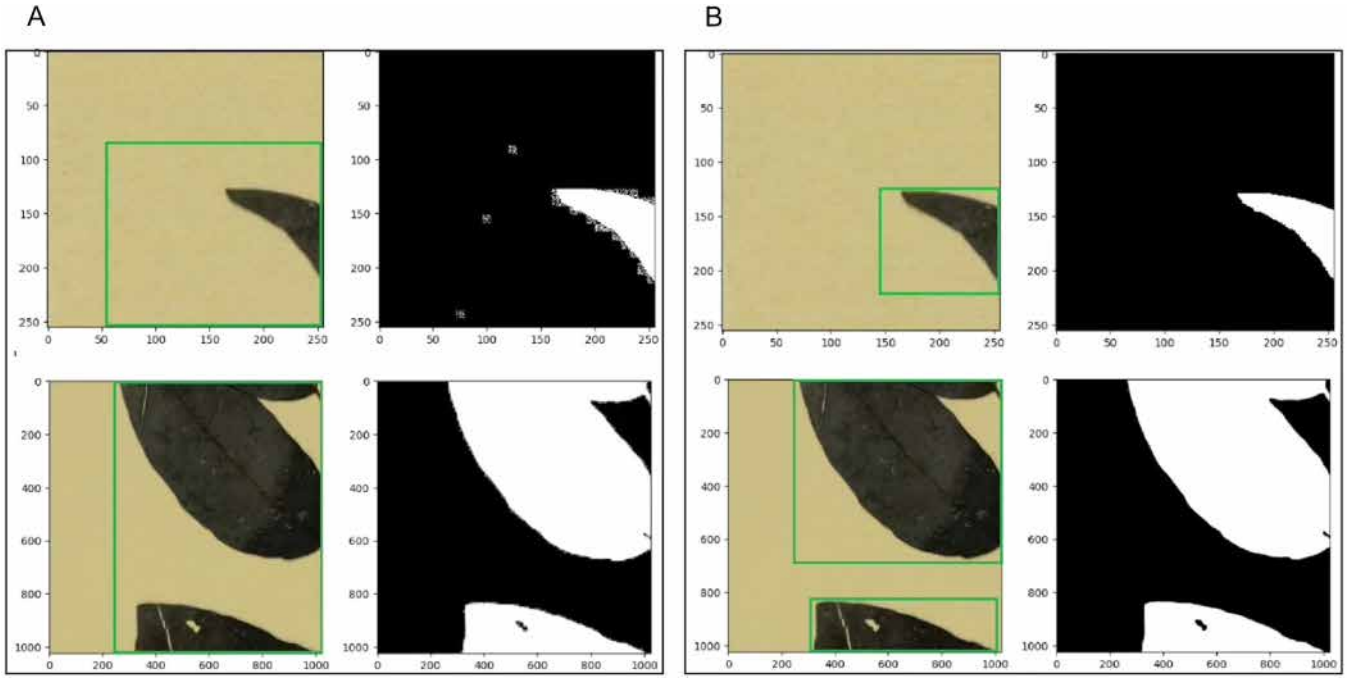

FIGURE 2 Comparison of the boxing strategies used to generate prompts for segmentation. (A) The single-box strategy encloses all plant parts in one bounding box, potentially capturing noise. (B) The multi-region strategy isolates separate plant structures with multiple boxes, leading to cleaner segmentation masks.

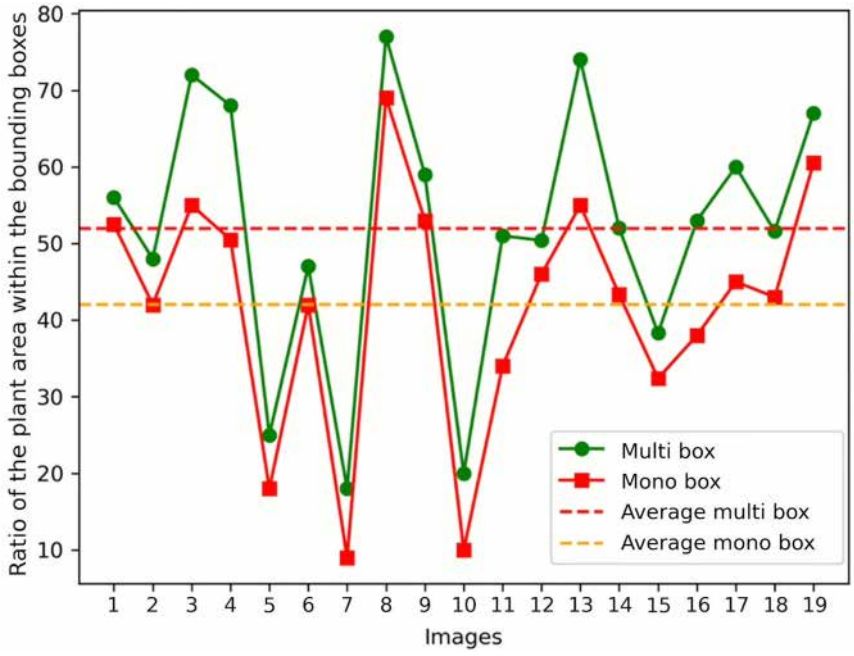

FIGURE 3 Graph illustrating the comparison of segmentation between the multi-region strategy and the single-box strategy, based on the ratio of plant area to bounding box area. Each value represents the average ratio across all bounding boxes identified within a given image, using a total of 19 images. The multi-region strategy achieves a higher average ratio (51.96%) compared to the single-box strategy (42.01%), indicating improved segmentation performance by reducing background noise.

pipeline for better segmentation, particularly for complex or noisy herbarium images.

To fine-tune the SAM models, we used a curated subset (Castanet et al., 2025b) of the segmentation dataset (Sklab et al., 2024b) published in our previous work (Ariouat et al., 2025); this original dataset was constructed using a semi-automatic pipeline combining morphology and deep learning. The subset included 1476 herbarium specimen images from nine

genera and two families, chosen for their high segmentation quality based on visual inspection. We then generated heat maps to reveal distinctive spatial distribution patterns of plant structures on a herbarium sheet for each taxon. These heat maps highlighted taxon-specific traits such as density and surface coverage (Figure 4). Each visualization was centered around the largest specimen representative of the corresponding taxon. By aggregating and normalizing the segmentation masks across all images within a taxon, we produced color gradients revealing the typical spatial positioning of plant structures on the herbarium sheet. For instance, the heat map for *Laurus* exhibits a centralized and dense spatial distribution, suggesting a compact structure with plant elements predominantly clustered near the center of the sheet. The heat map for *Magnolia* reveals a similarly centralized and dense configuration, reflecting the compact arrangement and large, broad leaves typical of this genus. In contrast, *Desmodium* shows a broader and more dispersed distribution on the herbarium sheet, consistent with its branching morphology and scattered, smaller leaves.

The heat-map patterns are corroborated by quantitative, per-taxon measures of plant coverage on herbarium sheets, computed as the ratio of plant-pixel count to total image-pixel count. We observed significant variation in coverage across taxa, reflecting differences in plant morphological traits and the arrangement of the specimen on the herbarium sheet (Table 1). For instance, *Magnolia* exhibits the highest average coverage (19.68%), consistent with its large, broad leaves and compact structure. In contrast, Convolvulaceae has the lowest coverage (7.12%), likely due to its more dispersed and slender growth habit. Other taxa, such as *Castanea* and Monimiaceae, show intermediate coverage levels, with values of 17.27% and 14.80%, respectively, illustrating the diversity in plant architecture present in the dataset. In some specimens, over 90% of the image consists of background rather than plant material. This finding quantitatively illustrates the challenge posed by high background content in herbarium specimen images and highlights the need for segmentation as a preprocessing step. Removing the background ensures that downstream models focus on relevant plant features.

Mask reconstruction

After processing all patches of an image, the segmented masks were reassembled to create a complete plant mask. This step, called "unpatching," involves recombining the segmented patches to restore the original spatial context of the specimen. Reconstructed masks, adjusted for patching-related padding and refined using multi-region prompts, improve segmentation quality by minimizing background noise and preserving plant structures for downstream analysis.

RESULTS

We assessed the performance of the segmentation pipeline under both ideal and non-ideal conditions, focusing on the comparison between UNet, SAM1, and SAM2. This

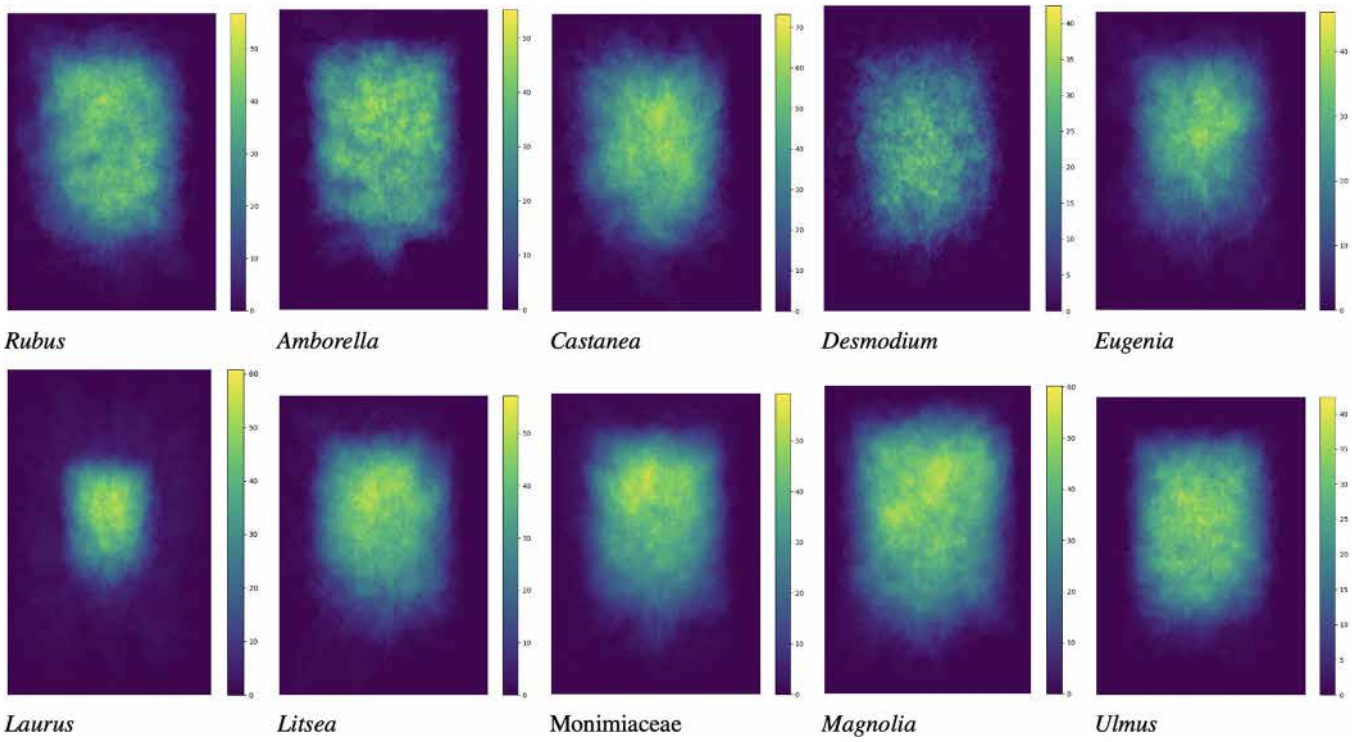

FIGURE 4 Heat maps of the plant taxa, generated by combining and normalizing their masks. These visualizations highlight spatial distribution patterns, density, and surface coverage for each taxon.

TABLE 1 Comparison of segmentation performance across 11 plant taxa. For each taxon, we report the average plant and background coverage, IoU and Sørensen–Dice scores obtained with the three models (UNet, PlantSAM1, and PlantSAM2), and their performance improvements (Δ_1, Δ_2) over UNet. PlantSAM2 consistently achieves the best results across most taxa, especially for challenging cases with low plant coverage and complex backgrounds.

Taxon	No. of images	Coverage (%)		IoU					Sørensen–Dice score				
		Plant	Background	UNet	PlantSAM1	Δ_1	PlantSAM2	Δ_2	UNet	PlantSAM1	Δ_1	PlantSAM2	Δ_2
Magnolia	34	19.68	80.32	0.9497	0.9625	+1.28	0.9656	+1.59	0.9822	0.9769	−0.53	0.9836	+0.14
Castanea	36	17.27	82.73	0.9325	0.9394	+0.69	0.9425	+1.00	0.9772	0.9706	+0.66	0.9788	+0.16
Amborella	5	15.97	84.03	0.9005	0.9387	+3.82	0.9432	+4.27	0.9619	0.9662	+0.43	0.975	+1.31
Rubus	22	15.05	84.95	0.9185	0.9485	+3.00	0.9504	+3.19	0.9714	0.9711	−0.03	0.9798	+0.84
Litsea	17	14.17	85.83	0.9343	0.9299	−0.44	0.9357	+0.14	0.9766	0.9603	−1.63	0.9692	−0.74
Eugenia	40	10.88	89.12	0.9078	0.9275	+1.97	0.9324	+2.46	0.9668	0.9566	−1.02	0.9682	+0.14
Ulmus	53	10.87	89.13	0.8995	0.9360	+3.65	0.9386	+3.91	0.9654	0.96	−0.54	0.9729	+0.75
Desmodium	11	8.09	91.91	0.8337	0.9007	+6.70	0.9018	+6.81	0.941	0.9344	−0.66	0.9528	+1.18
Laurus	53	7.37	92.63	0.9420	0.9558	+1.38	0.9569	+1.49	0.9795	0.9752	−0.43	0.9817	+0.22
Convolvulaceae	25	7.12	92.88	0.8222	0.8698	+4.76	0.8789	+5.67	0.9381	0.9181	−2.00	0.9447	+0.66
Monimiaceae	37	14.8	85.20	0.9356	0.9506	+1.50	0.9531	+1.75	0.9775	0.9698	−0.77	0.9775	+0.00

Note: IoU = intersection over union.

evaluation also considered the impact of combining YOLOv10 with SAM models using two datasets: a curated and annotated dataset representing ideal conditions and an out-of-distribution (OOD) dataset (Castanet et al., 2025a) featuring noisy and complex backgrounds.

Experimental setup

We fine-tuned both SAM models (SAM1 and SAM2), as well as UNet (Ariouat et al., 2025), on herbarium images using the curated segmentation dataset (Castanet et al., 2025b) and identical training conditions. To mitigate class imbalance during segmentation, we employed the Sørensen–Dice loss as the objective function for SAM1, SAM2, and UNet. All models were optimized using the Adam optimizer with an initial learning rate of 10^{-5}, reduced by a factor of 10^{-1} every 10 epochs, following a cosine annealing schedule. Training was conducted on the curated dataset of 1476 herbarium segmentation masks (Castanet et al., 2025a), with an 80/20 split between the training and validation groups. SAM1 and SAM2 were trained for up to 80 epochs with a batch size of 1, while UNet (Ariouat et al., 2025) was trained for a maximum of 100 epochs, with early stopping based on validation Sørensen–Dice scores. YOLOv10 was trained for 250 epochs on the plant region detection dataset (Castanet et al., 2025b), which was split into 14,307 training images, 3815 for validation, and 956 for testing. All input images were resized to 640 × 640 pixels. The best-performing YOLOv10 model achieved a mean average precision (mAP) of 0.959 at an IoU threshold of 0.5. All training and evaluation were conducted on NVIDIA A100 GPUs.

Segmentation performance

We evaluated the performance of our pipeline by comparing PlantSAM1, PlantSAM2, and UNet through three complementary analyses: quantitative benchmarking with standard metrics, generalization testing under challenging conditions, and measuring the impact of segmentation on trait classification.

Quantitative evaluation

Two primary metrics were used to quantitatively compare the models across multiple plant taxa: IoU and Sørensen–Dice coefficient.

- Intersection over union (IoU) measures the overlap between the predicted segmentation mask and the ground truth mask. It is defined as the ratio of the intersection of these two masks to their union. A higher IoU score indicates a greater degree of overlap, reflecting more accurate segmentation.

$$IoU = \frac{\text{Intersection}(\text{Predicted Mask} \cap \text{Ground Truth Mask})}{\text{Union}(\text{Predicted Mask} \cup \text{Ground Truth Mask})}$$

- Sørensen–Dice coefficient quantifies the overlap between the predicted segmentation mask and the ground truth mask. It is computed as twice the intersection divided by the sum of the areas of both masks. The Sørensen–Dice coefficient ranges from 0 to 1, where a higher value indicates better segmentation performance.

$$\text{Sørensen–Dice coefficient} = \frac{2 \times \text{Intersection}}{\text{Area of predicted mask} + \text{Area of ground truth mask}}$$

We conducted the quantitative evaluation on a separate test dataset of 333 herbarium images (Castanet et al., 2025a). PlantSAM2 consistently outperformed both PlantSAM1 and UNet, achieving an average IoU of 0.94 and a Sørensen–Dice coefficient of 0.97 (Table 1). Significant improvements were observed for taxa such as Convolvulaceae and *Desmodium*, with PlantSAM2 yielding IoU gains of 5.67% and 6.81%, respectively, over UNet. In contrast, performance gains were minimal for taxa like *Litsea*, suggesting variability in boundary-sensitive segmentation performance. Notably, PlantSAM1 was outperformed by UNet in nearly all cases with respect to the Sørensen–Dice coefficient (Table 1). We observed a better alignment between the segmentation masks generated by our pipeline and the ground truth (Figure 5).

Generalization to challenging conditions

We evaluated the generalization capacity of the models under challenging conditions using a distinct OOD dataset (Castanet et al., 2025a) comprising 171 images. These images included specimens with noisy or colored backgrounds, intricate plant structures, and artifacts such as pins or overlapping components. The dataset featured a variety of background textures and colors, including mosaic patterns and yellow, pink, dark gray, and newspaper-like backgrounds (Figure 6). PlantSAM2 generated usable masks in over 75% of cases (Table 2), even in the presence of thin armatures (e.g., spines, prickles, or thorns) or colored backgrounds. Usable masks are defined as outputs that, based on visual evaluation, accurately capture plant structures with minimal background inclusion or artifacts, preserving the fine details necessary for downstream tasks such as species identification or morphological analysis (see Figure 5 for an example). In contrast, UNet struggled significantly under these conditions, producing a higher proportion of unusable masks. These failed segmentations were often caused by excessive background inclusion, missed plant components, or errors introduced by overlapping structures or visual noise such as pins (Figure 7A, B).

We further analyzed the proportion of best-performing masks for each condition. PlantSAM2 consistently outperformed UNet in nearly all scenarios, with particularly strong results in complex cases such as thin armatures (73.08%) and long thin armatures (75%). For difficult background colors such as yellow and gray, it achieved best-mask rates of 89.66% and 69.09%, respectively. In contrast, UNet's performance remained limited, rarely exceeding 25% across these conditions. PlantSAM2 also demonstrated high robustness when handling artifacts such as pins, reaching a success rate of 90.91%. The column labeled "None" (Table 2) indicates instances where neither model succeeded, highlighting the persistent challenges in cases with highly complex backgrounds, such as brown or gray paper. These difficult cases often featured low contrast and poorly defined plant contours that impaired segmentation (Figure 7A, B). Nevertheless, PlantSAM2 reduced the proportion of unusable masks by more than 50% compared to UNet, confirming its superior generalization and resilience across diverse and noisy scenarios.

These results should be interpreted with caution, as the datasets used to compute IoU and Sørensen–Dice scores were not perfect representations of ideal segmentation masks. Although they closely approximated the ground truth, we occasionally observed segmentation outputs from PlantSAM2 that could be considered improvements over the original annotations. Paradoxically, such enhancements sometimes led to lower IoU scores, highlighting a limitation of current evaluation metrics. To better account for these discrepancies and gain a deeper understanding of PlantSAM's segmentation performance, we extended our analysis to include classification-based evaluation.

Impact of segmentation on classification

To investigate the impact of segmentation on deep learning classification performance, we designed an experiment using a trait-annotated dataset comprising herbarium images labeled for five botanical traits: armatures, fruits, acuminate leaf tips, infructescence, and acute leaf bases. Each image was processed using PlantSAM to generate segmented versions, and we subsequently trained a ResNet101 model on three variations of the dataset: (i) unsegmented images (raw, with background), (ii) segmented images (background removed), and (iii) segmented cropped images (plant regions only).

A key challenge in classifying herbarium images lies in the dominance of background over plant structures (Figure 4). Non-plant elements accounted for more than 90% of some images, increasing the risk of spurious correlations (Table 1). Segmentation helped isolate the relevant botanical content to mitigate this effect. Table 3 details the distribution of images across the training and validation sets for each trait. A further limitation of standard pipelines is the compression of both plant structures and background when resizing high-resolution images to fixed input dimensions (e.g., 224 × 224), which often leads to a loss of fine morphological details. Cropping segmented images helped preserve higher-resolution features while eliminating irrelevant regions. This preprocessing led to notable gains in

FIGURE 5 Illustration of the bounding box pipeline, showing the progression from the original image to the ground truth mask (GT Mask), bounding box–based prompts (Box prompt), and the final segmentation output (Prediction).

classification accuracy and F1 score across all traits (Table 4). The largest improvements were seen for fine-scale features such as armatures, with increases of 4.36% in accuracy and 2.38% in F1 score. Cropped segmented images consistently outperformed both raw and unprocessed segmented images, confirming that background removal improved the model's focus on informative regions. Overall, segmentation enhanced both accuracy and robustness by

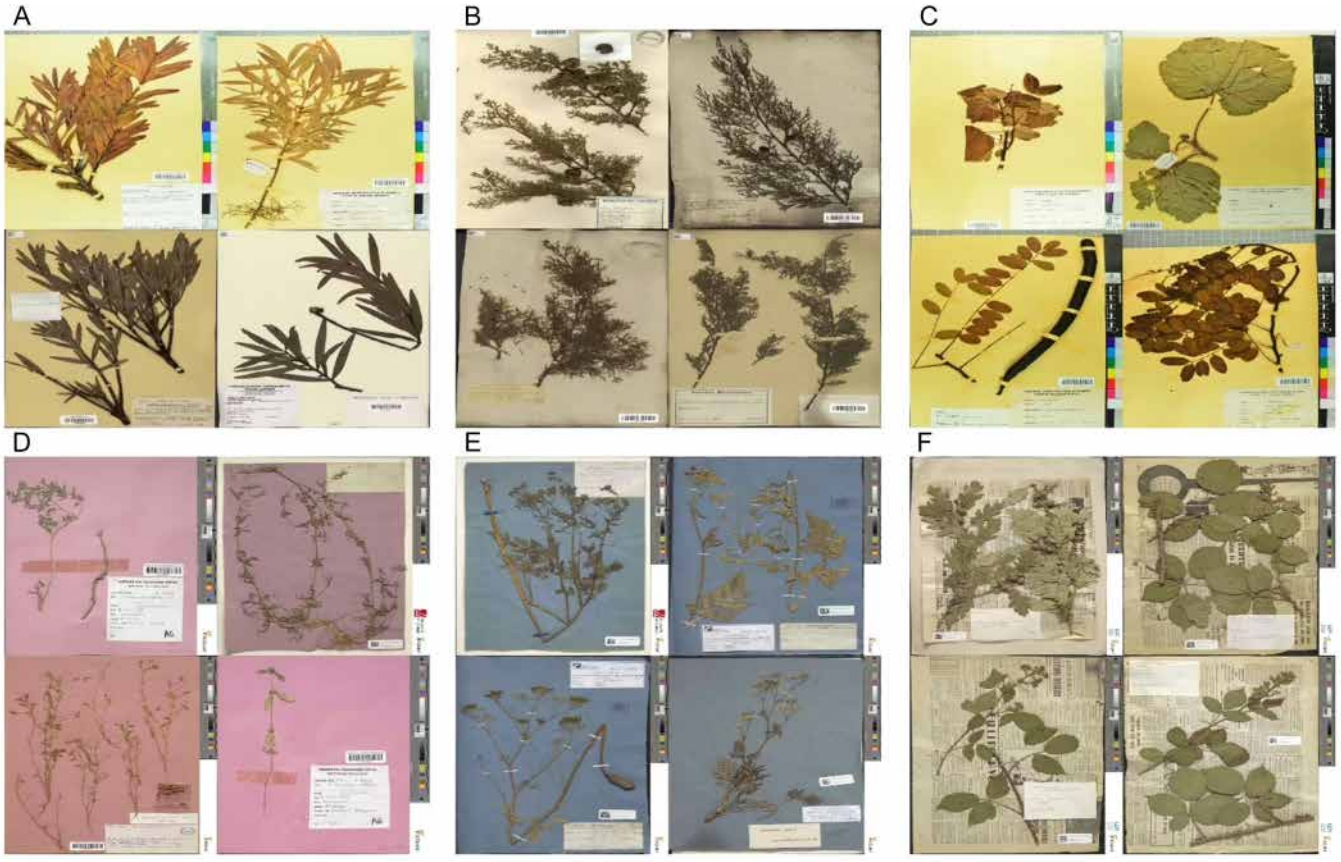

FIGURE 6 Examples of herbarium images with various background types, illustrating the diversity in texture and color found across the dataset. The backgrounds include (A) slender leaves on a pale background, (B) aged paper, (C) yellow background, (D) pink background, (E) dark gray background, and (F) Newspaper. This variability in background characteristics poses challenges for segmentation and highlights the importance of robust preprocessing and model adaptability.

TABLE 2 Comparison of UNet and PlantSAM2 performance across various challenging conditions, highlighting PlantSAM2's overall superior robustness in producing usable masks under visually complex scenarios. For each condition, we report (i) the number of analyzed images, (ii) the percentage of usable masks, (iii) the percentage of unusable masks, and (iv) the percentage of best-performing masks (categorized as best predicted by UNet, PlantSAM2, or none).

		Usable masks (%)		Unusable masks (%)		Best-performing masks (%)		
Challenging condition	**No. of images**	**UNet**	**PlantSAM2**	**UNet**	**PlantSAM2**	**UNet**	**PlantSAM2**	**None**
Blue background	11	0.00	9.09	90.91	54.55	9.09	72.73	18.18
Gray background	55	5.45	34.55	67.27	29.09	7.27	69.09	23.64
Yellow background	29	10.34	65.52	48.28	0.00	10.34	89.66	0.00
Brown background	15	0.00	26.67	100.0	53.33	20.0	46.67	33.33
Orange background	3	33.33	33.33	33.33	0.00	33.33	66.67	0.00
Pink background	5	0.00	0.00	40.0	40.0	60.0	40.0	0.00
Thin armatures	26	19.23	61.54	7.69	3.85	23.08	73.08	3.85
Long thin armatures	12	25.0	58.33	41.67	8.33	25.0	75.0	0.00
Long leaves	4	25.0	75.0	50.0	0.00	25.0	75.0	0.00
Pins	11	9.09	90.91	18.18	0.00	9.09	90.91	0.00

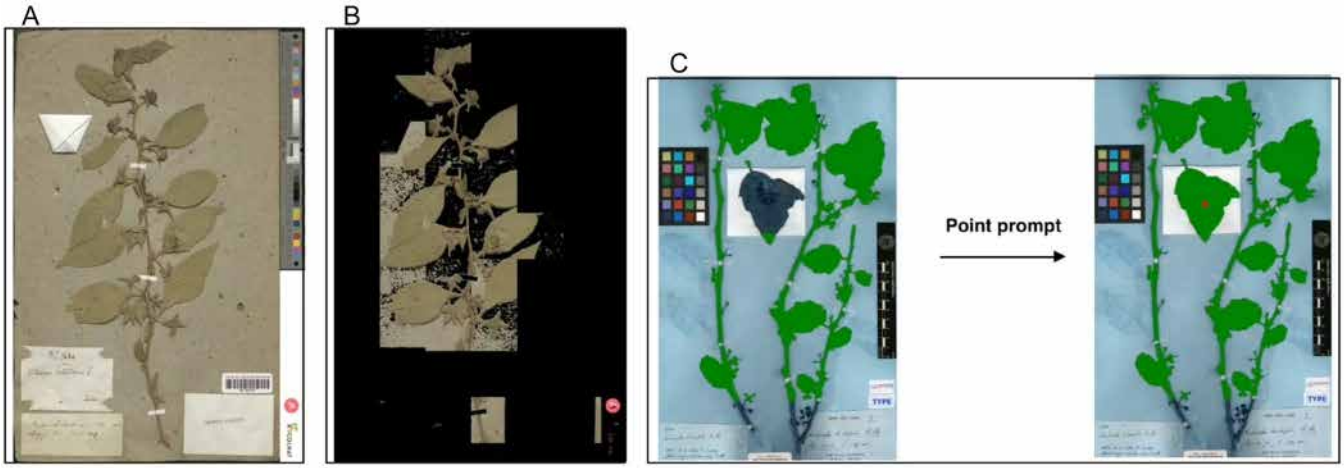

FIGURE 7 Examples illustrating segmentation limitations and refinements. (A) Original herbarium image with challenging conditions. (B) Segmentation result showing an unusable mask due to poorly defined contours, low contrast, or inclusion of non-plant elements. (C) Enhanced segmentation using a point prompt in the developed application, where user interaction improves the quality of the mask by focusing on relevant regions.

TABLE 3 Distribution of images in the training and validation datasets for the five studied traits for the classification downstream task.

Trait	Training	Validation
Armatures	1607	344
Fruits	1623	335
Leaves with acuminate tips	1479	305
Infructescence	1394	302
Leaves with an acute base	1219	256

enabling the model to extract more discriminative features and reducing the influence of irrelevant visual noise.

Semi-automatic annotation tool for refining segmentation masks

Our pipeline demonstrates strong potential for integration into a semi-automatic annotation tool. While fully automatic segmentation remains challenging for some cases, this limitation can be addressed through a semi-automatic approach where users refine the masks by providing precise point prompts to SAM. We developed an application that incorporated the PlantSAM2 pipeline, enabling the semi-automatic correction of masks initially deemed unusable. This process relies on direct interaction with expert users who guide SAM2 to produce more accurate segmentation results (Figure 7C). The application serves two main purposes: correcting poor segmentations produced by the PlantSAM2 pipeline and enhancing the segmentation of complex cases to expand the training dataset for fine-tuning the SAM2 model. Users can upload their images, which are initially processed by PlantSAM2. The images are then presented one by one, allowing users to interact with specific regions by providing point prompts to SAM2. This interaction transforms unusable masks into usable ones.

DISCUSSION

In this work, we developed an automatic segmentation pipeline tailored for herbarium specimens, leveraging object detection techniques to integrate the strengths of the Segment Anything Model (SAM) while addressing its limitations. By combining SAM with YOLOv10, we created a streamlined pipeline that significantly outperformed traditional models like UNet. The results demonstrated PlantSAM2's consistent superiority over both UNet and PlantSAM1 in terms of segmentation accuracy (IoU and Sørensen–Dice scores) and its ability to generalize to challenging scenarios, such as colored backgrounds and intricate plant regions.

Beyond segmentation, we demonstrated the substantial impact of background removal on classification performance. By isolating plant structures and eliminating irrelevant background elements, segmented images improved the accuracy and F1 score of classification models across all studied botanical traits. Additionally, the segmentation-based cropping strategy allowed for better utilization of image resolution by focusing on relevant plant regions, preventing the loss of fine morphological details during preprocessing. These improvements reinforce the importance of segmentation as a preprocessing step for deep learning applications in herbarium image analysis.

This pipeline represents a technical improvement while simultaneously providing practical advancements in herbarium image processing. The integration of a

TABLE 4 Performance evaluation of the ResNet101 model on unsegmented and segmented images across five botanical traits. The metrics used for comparison are accuracy (Acc) and F1 score (F1). The delta (Δ) values represent the performance difference between segmented and unsegmented images.

	Unsegmented		Segmented		Δ		Segmented cropped		Δ	
Trait	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Armatures	90.41	89.18	92.73	89.77	+2.32	+0.59	94.77	91.56	+4.36	+2.38
Fruits	57.01	62.02	58.21	64.32	+1.2	+2.3	60.00	64.41	+2.99	+2.39
Leaves with acuminate tips	69.84	70.24	70.16	71.47	+0.32	+1.23	72.46	73.43	+2.62	+3.19
Infructescence	58.94	59.56	60.26	62.98	+1.32	+3.42	59.93	63.71	+0.99	+4.15
Leaves with an acute base	68.75	68.24	69.18	69.57	+0.43	+1.33	69.53	69.75	+0.78	+1.51

semi-automatic annotation application empowered users to efficiently refine masks, converting otherwise unusable masks into usable ones with minimal manual intervention. This approach could accelerate the image annotation process, enabling the rapid generation of new datasets while substantially reducing the time and labor typically required for manual annotation.

In conclusion, the combination of SAM2 with an object detection model and semi-automatic refinement tools has expanded the scope of treatable herbarium images. Future extensions of this pipeline could include adaptive prompts to further enhance segmentation in challenging cases or multi-modal techniques that integrate textual and morphological data. These advancements would support the development of increasingly sophisticated tools for herbarium data analysis, fostering broader applications in taxonomy, conservation biology, and biodiversity research.

ACKNOWLEDGMENTS

This work was partly funded by the French National Research Agency in the context of the e-Col+ project (ANR-21-ESRE-0053). This project was provided with computing HPC and storage resources through the 2023-A0150114385 grant by GENCI at IDRIS.

AUTHOR CONTRIBUTIONS

Y.S. led the design of the segmentation pipeline, co-developed the experimental strategy, co-designed the model architecture, and supervised the entire data acquisition and preparation process. He was also responsible for setting up the HPC environment, managing computational resources (GENCI/IDRIS), designing the prompt strategy, conducting heatmap and classification analyses, and drafting the initial version of the manuscript. F.C. co-designed the pipeline and model architecture, developed the preprocessing code, implemented model integration, and fine-tuned the segmentation models. He also curated the datasets, contributed significantly to model training and evaluation, and co-authored the initial draft. H.A. contributed to the pipeline and model design, and participated in dataset construction and fine-tuning. S.A. assisted with model training and figure preparation. E.C., J.-D.Z., Y.S., and E.P. were involved in the strategic planning of the e-Col+ project, with E.C. and Y.S. jointly overseeing the scientific coordination of Work Package 4 (WP4). All authors participated in the manuscript's revision and approved its final version.

DATA AVAILABILITY STATEMENT

The source code for segmentation, including examples and trained models, is available at: https://github.com/IA-E-Col/PlantSAM. The source code of the segmentation application is available at: https://github.com/IA-E-Col/plantsam-app. The datasets used in this study are the following: the segmentation dataset used to train the UNet model is available at https://doi.org/10.6084/m9.figshare.27685914 (Sklab et al., 2024b); the object detection dataset used to train YOLOv10 for plant region detection is available at https://doi.org/10.6084/m9.figshare.29528882 (Castanet et al., 2025b); the dataset used to fine-tune SAM (a subset of the segmentation images from Sklab et al. [2024b]) and the out-of-distribution dataset, used for evaluating segmentation robustness under challenging conditions, are available at https://doi.org/10.6084/m9.figshare.29538065 (Castanet et al., 2025a).

ORCID

Youcef Sklab https://orcid.org/0000-0002-7045-9402

REFERENCES

Abdelaziz, B., and M. Walid. 2022. A deep learning-based approach for detecting plant organs from digitized herbarium specimen images. *Ecological Informatics* 69: 101590. https://doi.org/10.1016/j.ecoinf.2022.101590

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, E. Prifti, J.-D. Zucker, and E. Chenin. 2025. Enhancing plant morphological trait identification in herbarium collections through deep learning–based segmentation. *Applications in Plant Sciences* 13(2): e70000. https://doi.org/10.1002/aps3.70000

Besnard, G., M. Gaudeul, S. Lavergne, S. Muller, G. Rouhan, A. P. Sukhorukov, A. Vanderpoorten, and F. Jabbour. 2018. Herbarium-based science in the twenty-first century. *Botany Letters* 165(3–4): 323–327. https://doi.org/10.1080/23818107.2018.1482783

Borhani, Y., J. Khoramdel, and E. Najafi. 2022. A deep learning based approach for automated plant disease classification using Vision

Transformer. *Scientific Reports* 12: 11554. https://doi.org/10.1038/s41598-022-15163-0

Castanet, F., H. Ariouat Sklab, E. Chenin, and Y. Sklab. 2025a. Digitised herbarium image segmentation dataset. *figshare* https://doi.org/10.6084/m9.figshare.29538065 [posted 28 July 2025; accessed 1 September 2025].

Castanet, F., H. Ariouat Sklab, E. Chenin, and Y. Sklab. 2025b. Plant region detection in digitised herbarium specimens. *figshare* https://doi.org/10.6084/m9.figshare.29528882.v1 [posted 10 July 2025; accessed 1 September 2025].

Chen, W.-T., Y.-J. Vong, S.-Y. Kuo, S. Ma, and J. Wang. 2024. RobustSAM: Segment anything robustly on degraded images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024)*, pp. 4081–4091. https://doi.org/10.1109/CVPR52733.2024.00391

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): 4749. https://doi.org/10.3390/s21144749

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, et al. 2021. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv 11929 [preprint]. Available at https://doi.org/10.48550/arXiv.2010.11929 [revision posted 3 June 2021; accessed 6 November 2025].

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: 844522. https://doi.org/10.3389/fpls.2022.844522

Grasso, G. 2024. Soils associated with herbarium plants: A resource to address the temporal evolution of plant-associated microbiomes. PhD thesis, Museum national d'histoire naturelle, Paris, France, and Università degli studi di Torino, Turin, Italy. https://theses.hal.science/tel-05061486v1

He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

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, 603. Springer, Singapore. https://doi.org/10.1007/978-981-15-0058-9_31

Jiang, Y., and C. Li. 2020. Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics* 2020: 4152816. https://doi.org/10.34133/2020/4152816

Kirillov, A., E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, et al. 2023. Segment anything. arXiv 2304.02643v1 [preprint]. Available at https://doi.org/10.48550/arXiv.2304.02643 [posted 5 April 2023; accessed 1 September 2025].

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60(6): 84–90. https://doi.org/10.1145/3065386

Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11): 2278–2324. https://doi.org/10.1109/5.726791

Lee, S., H. Moon, S. Kim, and J. Lee. 2025. Multiple kernel-enhanced encoder for effective herbarium image segmentation. *Electronics Letters* 61: e70155. https://doi.org/10.1049/ell2.70155

Meredith, L. 1996. Roles of natural history collections. *Annals of the Missouri Botanical Garden* 83(4): 536–545. https://www.jstor.org/stable/2399994

Mochida, K., S. Koda, K. Inoue, T. Hirayama, S. Tanaka, R. Nishii, and F. Melgani. 2018. Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective. *GigaScience* 8(1): giy153. https://doi.org/10.1093/gigascience/giy153

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

Raven, P. H. 2019. Saving plants, saving ourselves. *Plants, People, Planet* 1(1): 8–13. https://doi.org/10.1002/ppp3.3

Ravi, N., V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, et al. 2024. SAM2: Segment anything in images and videos. arXiv 2408.00714 [preprint]. Available at https://arxiv.org/abs/2408.00714 [posted 28 October 2024; accessed 1 September 2025].

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. 2015. 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 1 September 2025].

Sklab, Y., H. Ariouat, Y. Boudjydah, Y. Qacami, E. Prifti, J.-D. Zucker, R. V. Lebbe, 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* Available at https://doi.org/10.6084/m9.figshare.27685914.v1 [posted 17 November 2024; accessed 22 September 2025].

Sklab, Y., H. Ariouat, E. Chenin, E. Prifti, and J.-D. Zucker. 2025a. SimNet: A multimodal fusion network using inferred 3D object shape point clouds from RGB images for 2D classification. *IET Computer Vision* 19(1): e70036. https://doi.org/10.1049/cvi2.70036

Sklab, Y., H. Ariouat, E. Prifti, J.-D. Zucker, and E. Chenin. 2025b. Identification of non-plant elements in herbarium images using YOLO. In D. Aissani, K. Barkaoui, and M. Roche [eds.], Research in Computer Science. CARI 2024. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-031-88226-5_10

Soltis, P. 2017. Digitization of herbaria enables novel research. *American Journal of Botany* 104: 1281–1284. https://doi.org/10.3732/ajb.1700281

Sweeney, P. W., B. Sajan, P. J. Morris, Y. Xia, A. Jirasek, R. Srinivasan, C. J. Grassa, and C. C. Davis. 2018. Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system. *Taxon* 67(1): 165–178. https://doi.org/10.12705/671.10

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, et al. 2015. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

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., B. Bouaziz, W. Mahdi, H. Hamed, and J. Gaikwad. 2022. 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, A., H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding. 2024. YOLOv10: Real-time end-to-end object detection. arXiv 2405.14458 [preprint]. Available at https://arxiv.org/abs/2405.14458 [posted 22 May 2024; accessed 1 September 2025].

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. https://doi.org/10.1002/aps3.11352

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

Yin, F., J. Li, Y. Wei, W. Zhang, and C. Xu. 2024. SAMS: One-shot learning for the segment anything model using similar images. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. https://doi.org/10.1109/IJCNN60899.2024.10651474

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, R., Z. Jiang, Z. Guo, S. Yan, J. Pan, H. Dong, P. Gao, and H. Li. 2023. Personalize Segment Anything Model with one shot. arXiv 2305.03048 [preprint]. Available at https://arxiv.org/abs/2305.03048 [posted 5 May 2023; accessed 1 September 2025].

Zhang, W., J. Wang, Y. Liu, K. Chen, H. Liu, Y. Dai, W. Wang, et al. 2022. Deep-learning-based in-field citrus fruit detection and tracking. *Horticulture Research* 9: uhac003. https://doi.org/10.1093/hr/uhac003

Zhao, X., W.-Y. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang. 2023. Fast Segment Anything. arXiv 2306.12156 [preprint]. Available at https://arxiv.org/abs/2306.12156 [posted 22 June 2023; accessed 1 September 2025].

Zhou, Z., M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. 2018. UNet++: A nested U-Net architecture for medical image segmentation. In D. Stoyanov et al. [eds.], Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA ML-CDS 2018). Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-00889-5_1

How to cite this article: Sklab, Y., F. Castanet, H. Ariouat, S. Arib, J.-D. Zucker, E. Chenin, and E. Prifti. 2025. PlantSAM: An object detection-driven segmentation pipeline for herbarium specimens. *Applications in Plant Sciences* 13(6): e70034. https://doi.org/10.1002/aps3.70034