



Title	Toward Reliable Vision-Guided Suction-and-Cut for Robotic Leaf Sampling in Arabidopsis Plants
Author(s)	Chang, Jiarui; Wan, Weiwei; Sato, Ryoichi et al.
Citation	IEEE Robotics and Automation Letters. 2026, 11(7), p. 8880-8887
Version Type	AM
URL	https://hdl.handle.net/11094/105207
rights	© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Towards Reliable Vision-Guided Suction-and-Cut for Robotic Leaf Sampling in Arabidopsis Plants

Jiarui Chang¹², Weiwei Wan¹², Ryoichi Sato², Miki Fujita², Nobuyuki Tanaka³,
Masami Hirai², Koichi Takahashi³, Kensuke Harada¹

Abstract—Automated leaf sampling is essential for high-throughput genotype-phenotype mapping but remains challenging for small plants, such as *Arabidopsis thaliana*, that have dense canopies, tiny leaves (6~10 mm) and slender petioles (0.8~1 mm). Such morphological traits demand high positional accuracy and specialized manipulation to reliably isolate individual leaves. To address these challenges, we present a fully automated robotic sampling system that combines a suction-cutting end-effector with a two-stage vision framework for coarse-to-fine cutting-point alignment. The end-effector is driven by a non-interfering four-bar linkage, enabling compact motion and minimal disturbance to surrounding foliage. The control strategy iteratively replans motions based on local camera feedback until high-confidence alignment is achieved. Experiments across varied growth stages of *Arabidopsis* demonstrate high success rates with minimal disturbance, while ablation studies confirm the effectiveness of the selected perception methods. The system offers a reliable and precise automatic solution for plant genotype-phenotype mapping and broader laboratory automation workflows.

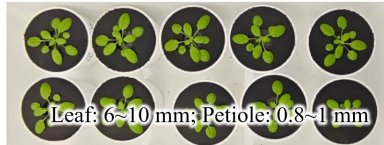
Index Terms—Laboratory Automation, End-effector Design

I. INTRODUCTION

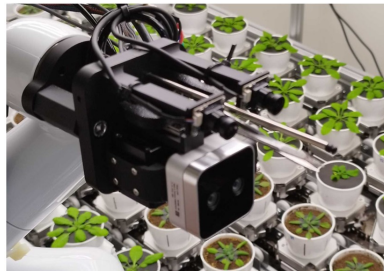
ARABIDOPSIS *thaliana* has long served as a model organism in plant biology due to its small genome, short life cycle, and well-established genetic resources [1]. It is extensively used in phenotyping and molecular studies, many of which require the collection of plant tissues, particularly leaves, for downstream analyses such as gene expression profiling and metabolite quantification. As the scale of these experiments continues to grow, manual leaf sampling becomes increasingly impractical due to its labor-intensive nature and susceptibility to human error, creating a pressing need for precise and automated sampling systems [2][3]. Robotic sampling in this context can accelerate data acquisition and ensure consistent leaf collection, thereby enhancing repeatability and scalability compared with manual approaches.

Despite recent advances in agricultural robotics [4][5], automated leaf sampling for *Arabidopsis* remains particularly challenging. Fig. 1(a) shows several *Arabidopsis* plants. They have small rosette morphology, characterized by a flat, ground-hugging arrangement of densely packed leaves, placing individual leaves close to the soil surface and in frequent

(a) Tiny and difficult to sample



(b) End-effector with vision guidance



(c) Suction-cut sampling



Fig. 1: (a) *Arabidopsis* has tiny leaves with slender petioles. They are difficult to sample. (b) Proposed system design. (c) Suction-cut sampling process: (c.1) Suction lifting the target leaf; (c.2) Cutting operation.

contact with neighboring foliage. In addition, the leaves are tiny (6~10 mm), supported by thin petioles (0.8~1 mm), and highly susceptible to damage, making them difficult to cut and grasp. Sampling *Arabidopsis* leaves demands high positional accuracy, careful end-effector design, and reliable vision modules to detect and localize slender petioles. However, existing robotic systems for plant manipulation [6][7] are more suitable for larger plants and struggle with tiny leaves with thin petioles. In addition, their vision modules often lack sufficient precision for such fine-scale targets [8].

In this paper, we propose a fully automated robotic leaf sampling system for *Arabidopsis*. Fig. 1(b) presents an overview of the system. It integrates a suction-cutting end-effector actuated by a non-interfering four-bar linkage, enabling compact motion and minimizing disturbance to neighboring foliage. The suction cup lifts the target leaf away from the soil surface, while the scissor mechanism protrudes only during cutting, avoiding obstruction of the suction operation at other times. Fig. 1(c.1) and (c.2) show the lifting and cutting process. To address the detection and alignment challenges posed by small leaves with slender petioles, we adopt a two-stage hierarchical vision framework that combines a global camera for coarse target estimation with two local cameras for fine alignment. This hierarchical design improves robustness to partial occlusion and enables accurate positioning of the cutting tool. At the

control level, the system repeatedly adjusts scissors positions based on local camera feedback, and executes the suction and cutting actions only when a high-confidence alignment is achieved. This combination of mechanical design, perception, and control enables reliable and precise leaf sampling across varying plant configurations.

We validate the system through experiments conducted across a range of plant growth conditions and configurations. Our results demonstrate high success rate and low damage to sampled plants, indicating the potential of this method for deployment in automated phenotyping pipelines and laboratory workflows. Additionally, we conducted ablation studies comparing learning-based and analytical methods in leaf segmentation and plant center prediction to confirm the effectiveness of the selected methods.

The main contributions of this work are: (1) A novel suction-cutting end-effector designed with a compact, non-colliding four-bar linkage for precise and gentle leaf detachment. (2) A two-stage hierarchical vision framework that performs global estimation and local refinement of the cutting target, enabling precise alignment with leaves and petioles. These contributions collectively enable reliable robotic leaf sampling for *Arabidopsis* under diverse growth conditions.

II. RELATED WORK

A. Leaf Manipulation and Sampling

Leaf manipulation and sampling have been widely studied in the context of plant phenotyping robots, encompassing tasks from pose estimation for grasping to the precise detachment of target leaves [3]. Traditional computer vision and planning techniques play a crucial role in many studies. For example, Alenyà et al. [9] performed plant probing using depth and infrared-intensity images from a ToF camera and modeled the target leaf through an analytical approach. Foix et al. [10] constructed a multi-layer 3D occupancy grid map based on the task requirements and iteratively selected the Next-Best-View (NBV) to maximize the information gain of the target leaf until a suitable probing point was obtained. This data-efficient strategy avoids collecting 3D point cloud data for leaf reconstruction. Chaudhury et al. [11] attached a 3D laser scanner to a 7-DoF robot arm to capture and construct the 3D mesh of *Arabidopsis*, which was then used to compute the volume and surface area for growth pattern analysis. The authors implemented a two-step alignment method using the scans from 12 angles to obtain a fine result.

More recently, learning-based methods have been explored to improve leaf recognition, segmentation, and manipulation performance. Notably, Ahlin et al. [12] applied convolutional neural networks in image processing and proposed Monoscopic Depth Analysis (MDA) for visual-servoing to grasp the target leaf. Bao et al. [13] performed K-means clustering to locate individual plants from a batch and used the region growing method to segment the leaves. For probing, they proposed a motion planning framework based on Probabilistic Roadmap to avoid collision between the robot and the environment. Chen et al. [14] employed traditional machine vision methods to segment the target leaf of the soybean plant and planned the

robot's motion to avoid robot singularities and to accommodate the leaf's slope. Mortazavi et al. [4] applied an instance segmentation network to mask and reconstruct the 6D pose of leaves for robotic grasping. Selvam et al. [15] employed YOLOv8 for leaf detection, point cloud modeling for 3D shape reconstruction, and heuristic metrics for selecting the target leaf and grasping point. Building on this, they proposed a hybrid autonomous grasping method that integrates geometric algorithms with neural networks using self-supervised learning [16]. Campbell et al. [17] first acquired a 3D point cloud of avocado leaves, then estimated their poses to create bounding boxes, and finally sampled them with a customized end-effector. Deb et al. [18] designed deep learning-based methods to estimate poses of corn and sorghum leaves for grasping and sampling. Giang et al. [19] generated semantic 2D segmentations and 3D point clouds of sweet pepper plants to plan robotic motions for leaf pruning.

Many previous studies focus on larger leaves, where centimeter-level accuracy is often sufficient for grasping with a parallel gripper. In contrast, our work targets *Arabidopsis*, which has small leaves with slender petioles and a dense arrangement. It requires higher detection accuracy, a delicate end-effector design, and closed-loop control strategies capable of refining the target position. To meet these requirements, we develop a suction-cutting end-effector with a compact four-bar linkage, combined with a hierarchical vision framework that performs global estimation and local refinement for precise leaf targeting. The control layer executes iterative policy replanning based on local camera feedback, ensuring that suction and cutting actions are triggered only when the target alignment reaches a high level of confidence. Since end-effector capability plays an essential role, we next review relevant designs for sampling and harvesting in plant robotics.

B. Sampling and Harvesting End-Effector Design

Various end-effectors have been designed and fabricated for different types of target crops, especially fruits and vegetables. For example, Lehnert et al. [20] designed a sweet pepper harvesting system that can decouple the suction cup and the cutting blade to enable independent motion, and the two parts can recouple through gravity and magnetic force for the next target. Silwal et al. [21] designed a 3-finger underactuated end-effector to pull and twist apples off the vine. Sobol et al. [22] proposed an elongated multi-layer physical system that includes a gripping layer, a clamping layer, and a final cutting layer to perform steady and accurate strawberry harvesting. Ochoa et al. [23] designed a pneumatic-powered cutter that translates along the shaft. It cuts and pinches the stem of strawberries with a 3D-printed hook during extension and drops the fruit during retraction. Birrell et al. [24] designed a caging system that has two pneumatic actuators. The first one powers a soft finger to hold the lettuce inside the cage and the second powers a blade to cut. Navas et al. [25] created a three-finger soft gripper system that can apply a gentle but firm grip on the fruit, and the rotating base aids the detachment of the target from the stem. Yu et al. [26] proposed a deformable blade system that can adjust its hyperbola shape

and height to fit the Oolong tea bush to harvest tea leaves. Besides academic studies, Alexander et al. [27] from Iron Ox developed an automated plant-handling system for indoor agricultural facilities, focusing on scanning and transplanting with robotic manipulators. Boudreau et al. [28] patented an automated harvesting system for mushrooms utilizing a SCARA arm and a specialized end-effector with a suction cup to hold the pileus and a neck structure with graduated elasticity to facilitate damage-free picking while mimicking human “twist-and-pull” motion. More recently, Wisdom et al. [29] developed an end-effector in which pivoting doors and a cutting mechanism cooperate to enclose, sever, and retain an edible portion of a plant within a receptacle-like housing.

The main difference between our system and the above designs is that we target individual *Arabidopsis* leaves that are tiny, fragile, and have slender petioles. Our end-effector utilizes a miniature suction cup to grip the target leaf gently. A built-in air-blowing function enables reliable leaf release after cutting. The cutting tool consists of surgical scissors mounted on a retractable crank-slider four-bar linkage. This mechanism allows the blades to extend only during cutting and remain retracted otherwise, thereby avoiding obstruction of the suction motion. The cutting mechanism is actuated by a single, lightweight motor, making it suitable for low-payload desktop robotic arms that manipulate small plants. The end-effector also integrates a global camera for coarse localization and two local cameras for fine alignment, enabling precise tool positioning.

III. END-EFFECTOR DESIGN

Our end-effector is designed to first secure a target leaf with a suction cup and then cut its petiole using scissors. A major challenge in realizing this design is to prevent interference between the scissor blades and the leaf during suction. To address this issue, we employ a crank-slider four-bar linkage to control the scissor blades, as illustrated in Fig. 2. In the mechanism, a motor-driven crank rotates one handle of the scissors, while the other handle is constrained by a slider that translates linearly. Depending on the crank angle, the system operates in two modes: fold mode and cut mode. During suction, the blades are folded to avoid collision with the target leaf. Once a leaf is secured, the crank drives the scissors to cut the petiole. The slider translation is designed to be 15 mm. The motor rotates 48.3° to fully close the scissors. The distance between the motor and the scissors is 17.76 mm. To integrate the suction and cutting actions spatially, the suction cup is mounted in parallel with the scissors at a 7 mm offset. When the cup grips the center of a leaf, the cutting location corresponds to a point on the petiole with the offset away from the suction center. This configuration leverages the phyllotaxis of *Arabidopsis*, where leaves are arranged symmetrically along the stem, and ensures that a small section of the petiole remains on the sampled leaf, which, as advised by our collaborating botanists, facilitates downstream biological analyses. The sampled leaf will be lifted and transferred by a robot arm.

With this layout ensuring interference-free and biologically meaningful cutting, the suction module is further engineered

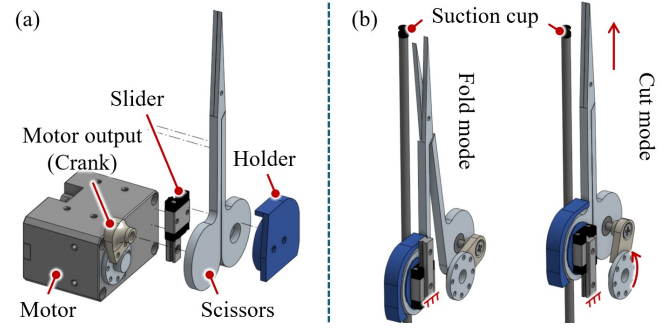


Fig. 2: (a) Exploded view of the scissor driving mechanism. (b) The fold and cut modes of the scissors.

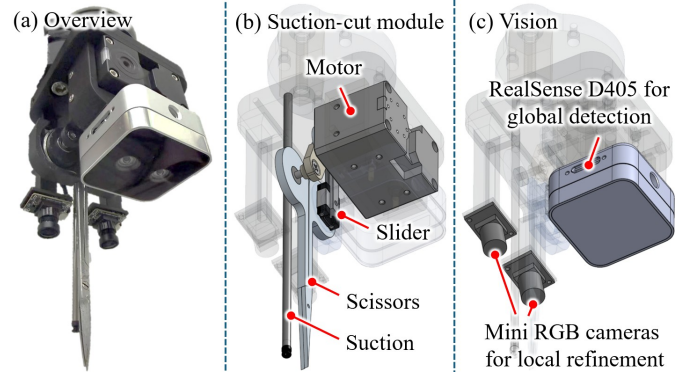


Fig. 3: (a) Overview of the end-effector. (b) Scissors-suction module. (c) Global and local vision systems. The design enables the scissors to fold after cutting and facilitates leaf handling in crowded spaces.

to provide reliable attachment. A stainless steel suction tube with a diameter of 3.5 mm was selected to match the typical width (around 8 mm) of mature *Arabidopsis* leaves. At its tip, an insertable suction cup (outer diameter 4 mm, insertion diameter 3.2 mm) is press-fitted into the tube (inner diameter 3.1 mm, outer diameter 3.5 mm) and connected via a polyurethane line (inner diameter 2.99 mm, outer diameter 4.76 mm) to a vacuum generator. The generator supports both suction and blow-off modes depending on the input voltage: suction is used to secure the leaf, while blow-off is used to release a sample into test tubes for liquid-nitrogen storage. The blow-off is necessary because a leaf may not be heavy enough to fall off from the suction cup naturally.

In addition to the mechanism, the end-effector is equipped with two complementary vision systems. Global localization is provided by an Intel RealSense D405 depth camera, which segments 2D images to identify the plant center, the target leaf, and the petiole-blade junction point. These features are then transformed into the robot base frame using depth sensing and hand-eye calibration to obtain the working pose of the end-effector. For local refinement, two miniature web cameras are mounted adjacent to the suction-cutting module. They provide online visual feedback to compensate residual errors from the global estimate, ensuring that the suction cup attaches accurately to the leaf surface and that the scissors cut the petiole precisely. All vision systems are mounted on the end-

effector to make the robotic manipulator self-contained, reduce occlusion, and avoid reliance on external setups.

IV. GLOBAL VISION FOR ROUGH ALIGNMENT

The D405 RGB-D camera mounted on the end-effector is used to identify the target leaf and determine the required cutting pose of the end-effector. The procedure consists of two steps. First, the 2D RGB image from the D405 is used to segment all leaves, from which we extract the petiole-blade junction points and identify candidate targets. We focus on detecting the junction point rather than the petiole itself because the petiole is slender and has a similar color to the blade of the Arabidopsis leaf, making direct segmentation or detection difficult. Second, the corresponding 3D positions of both the leaf center and the petiole-blade junction point are obtained from the depth point cloud captured by the D405. These 3D positions are then used to compute the goal position and orientation of the end-effector for cutting. The whole process is illustrated as a flow chart in Fig. 4(a). The boxes with dark blue and green shadows are the starting and ending points of the workflow. The blue boxes are the key algorithmic components. The gray boxes indicate intermediate data.

For the 2D step, we first adopt a Mask R-CNN model [30] to segment leaves. Mask R-CNN is used because it provides a balance of precision, reliability, and speed over others like YOLO or SAM2 [31]. More details on this can be found in Appendix B. We fine-tune a Mask R-CNN based on the plant phenotyping dataset [32][33] and segment leaves as binary masks. Fig. 4(b.1) shows an example. To identify the petiole-blade junction point, we erode the segmented masks to suppress the protrusion introduced by the petiole region and employ PCA to obtain the principal axis of each mask. The axes correspond to midrib lines extending from the petiole through the leaf center to the tip, as shown by the white segments of Fig. 4(b.2). We then calculate the two intersections of each principal axis and the contour of an eroded mask. One intersection corresponds to a petiole-blade junction point. The other corresponds to the leaf tip. Their midpoint is taken as the leaf center. The yellow and blue points of Fig. 4(b.3) are the two intersections. The red points are the leaf centers. To further determine which intersection is the junction, we estimate the plant center by applying MeanShift [34] to all intersections, as shown by Fig. 4(b.4), and regard the intersections spatially close to the estimated plant center as the junction points.

After detecting the leaf centers and petiole-blade junction points, we select the most isolated leaf and obtain the 3D positions of its center and junction point through RGB-to-point-cloud mapping. The end-effector's goal pose is computed using the obtained 3D leaf center and junction point positions. Selecting the most isolated leaf improves sampling feasibility and reduces the risk of interfering with neighboring leaves. Specifically, for each segmented leaf, we compute the minimum Euclidean distance from its center to the centers of all other leaves, and choose the leaf with the largest minimum distance. During the process, leaves that are too large or too small are filtered based on their distance to the pot center: leaves with a distance between the plant center and leaf center

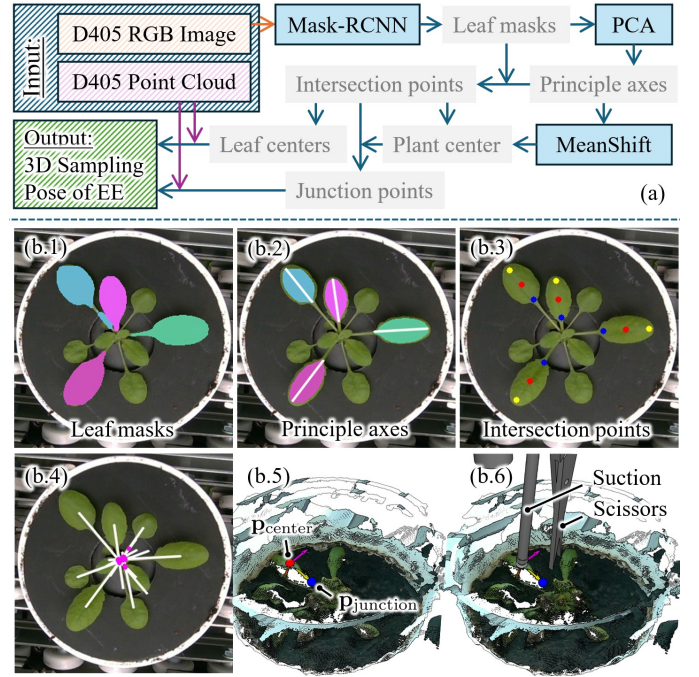


Fig. 4: (a) Workflow of global vision. (b.1~3) Junction point (blue) detection and leaf center (red) estimation. (b.4) Intersections (magenta) of principal axes are used to evaluate the plant center. (b.5, 6) Detected junction point and leaf center visualized on the 3D point cloud of a plant.

smaller than two-fifths of the pot radius are considered too small. Those exceeding the pot radius are considered too large.

Fig. 4(b.5) illustrates the leaf center $\mathbf{p}_{\text{center}}$ and junction point $\mathbf{p}_{\text{junction}}$ of a target leaf in the plant point cloud. To apply firm suction to the leaf center, the suction cup must be oriented so that it is properly aligned with the leaf blade surface. To achieve this pose of the end-effector, we first compute the unit vector in the radial direction as $\mathbf{v}_{\text{radial}} = (\mathbf{p}_{\text{center}} - \mathbf{p}_{\text{junction}}) / \|\mathbf{p}_{\text{center}} - \mathbf{p}_{\text{junction}}\|$. Next, a unit vector in the tangential direction within the horizontal plane is obtained as $\mathbf{v}_{\text{tangential}} = \mathbf{v}_{\text{radial}} \times \mathbf{z}$, where $\mathbf{z}$ is the unit vector in the vertical direction. The surface normal vector of the leaf is then calculated by $\mathbf{v}_{\text{normal}} = \mathbf{v}_{\text{tangential}} \times \mathbf{v}_{\text{radial}}$ considering the Arabidopsis leaf has little torsion along the petiole axis. Note that due to the limited depth resolution of the camera and the thinness of the petiole, the depth of a junction point may be inaccurate or unavailable. In such cases, it is approximated as 1 mm below the leaf center according to the typical geometry of Arabidopsis plants.

The target end-effector pose comprises position $\mathbf{p}_{\text{center}}$ and rotation $\mathbf{R}_{\text{rot}} = [\mathbf{v}_{\text{tangential}} \ \mathbf{v}_{\text{radial}} \ \mathbf{v}_{\text{normal}}]$. The pose aligns the line connecting the suction cup and the scissors' cutting center with the target leaf's central axis, as shown in Fig. 4(b.6). The corresponding robot configuration can then be solved by inverse kinematics. For practical purposes, the robot is first commanded to a pre-action pose located 5 mm along $\mathbf{v}_{\text{normal}}$ off the estimated $\mathbf{p}_{\text{center}}$ to prevent accidental contact before heading downward for suction and cutting.

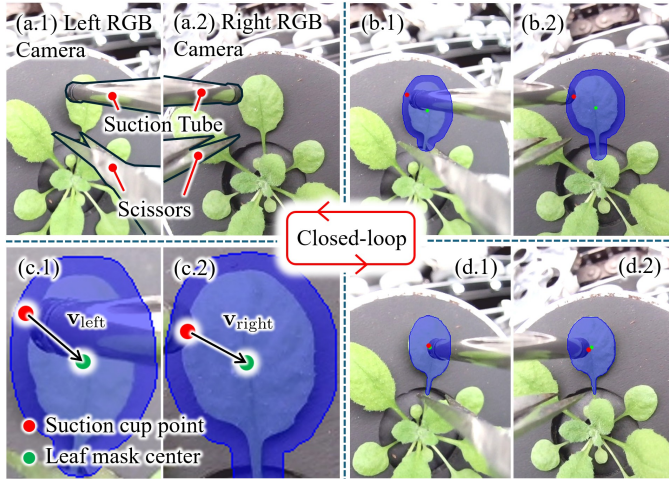


Fig. 5: (a) Left and right local RGB views. (b) Target-leaf masks segmented by a Mask R-CNN model fine-tuned considering tube occlusion. (c) Alignment vectors v_{left} and v_{right} between the suction cup and mask center. (d) Suction tube correction based on the averaged vectors. Steps (a)–(d) are repeated until convergence or until a limit is reached.

V. CLOSED-LOOP LOCAL REFINEMENT

Global vision alone is insufficient for reliable sampling because sensing noise and calibration errors may leave residual misalignment between the end-effector and the target leaf. To compensate for this error, we introduce a closed-loop local refinement strategy using two miniature RGB cameras mounted on both sides of the suction tube. As shown in Fig. 5(a.1) and (a.2), the metallic cylindrical component marked with a black contour is the suction tube, with the scissors located below it and the target leaf underneath. The robot iteratively estimates the residual alignment error from the local views and corrects the end-effector, enabling the suction cup to attach near the leaf center and the scissors to capture the petiole.

In detail, we first re-detect the target leaf from new images acquired using the local cameras. The re-detection is again performed through Mask R-CNN segmentation. However, the suction cup causes severe occlusion in the local-view images. To ensure that the leaf can still be accurately segmented under such occlusions, we fine-tune the Mask R-CNN model using leaf images that exhibit varying degrees of occlusion caused by the suction tube. The fine-tuning details are provided in Appendix D. Using the fine-tuned model, the target leaf is segmented from the local-camera images, and its pixel-level mask center is compared with the pixel location of the suction cup to compute the refinement offset, as shown in Fig. 5(b) and (c). Specifically, for both the left and right cameras, we compute the vector from the suction cup to the mask center, and get two pixel-space displacement vectors. These two vectors are averaged and the magnitude of the resulting vector is multiplied by an empirically tuned coefficient to obtain the direction and scale of the refinement motion. The end-effector is then commanded to move according to this motion. The refinement process is executed iteratively in closed loops: After each adjustment, the local cameras

provide updated feedback, and the estimation-and-correction cycle is repeated until the refinement offset falls below a preset threshold or when the maximum number of refinement iterations is reached, as denoted by the looping arrows in the center of Fig. 5. Iteration is required because the suction cup and the leaf surface are not in the same plane, making a single-time correction based on pixel offsets insufficiently accurate. Notably, the refinement process does not require transforming the 3D position of the leaf point into the robot base frame, thereby avoiding additional calibration.

After convergence, the end-effector descends along v_{normal} to secure the leaf. Because the relative position between the suction cup and scissors is fixed, aligning the suction cup with p_{center} alone may cause the scissors to cut the leaf blade instead of the petiole when leaves are large. Therefore, after alignment, an additional translation is applied along the direction from p_{center} to $p_{junction}$ when $p_{junction}$ is observed lower than the scissor tip pixels in either local-view image. The p_{center} and $p_{junction}$ of the target leaf are then re-detected from the two local cameras for following refinement iterations.

Finally, once the suction cup secures the leaf, the gap between the leaf and the black soil-shading cover is evaluated to determine whether the leaf should be lifted before cutting. The cover is commonly used in Arabidopsis phenotyping systems to suppress reflected light from the soil during spectral measurements. If the estimated gap at the leaf center is less than 4.5 mm, the leaf is elevated to provide clearance for the scissors and reduce disturbance to surrounding foliage; otherwise, the end-effector performs a direct cut. After cutting, the suction tube transfers the leaf above a test tube at a known position. The robot then moves downward and releases the leaf into the tube using a short air burst from the vacuum generator.

VI. EXPERIMENTS AND ANALYSIS

We evaluated the developed methods by two sets of experiments: (1) comparative studies against other methods, and (2) real-world deployment besides a plant phenotyping system.

A. Comparative Studies

We compared the PCA-based analytical method for junction point estimation with a learning-based one. The analytical method computes the intersection of the principal axis with the contour of eroded mask. In contrast, the learning-based approach directly predicts the junction points from the leaf image. We trained a U-Net [35] on the public plant phenotyping dataset for 40 epochs with a 90%-10% train-validation split. Each leaf of a plant is hand labeled with a junction point and there are in total 120 plants and 1157 leaves. Table I shows the performance of different junction point prediction methods. In total, 10 plants were used for comparison. Junction points are estimated for each segmented leaf and are compared with the hand-labeled ground truth. The mean and standard deviation of the errors are reported. The results show that the PCA method predicts the junctions with higher accuracy. The U-Net showed larger error variations, indicating that additional control or rejection strategies would be necessary to safely use its predictions for robotic execution. Robust generalization across

TABLE I: Quantitative junction points prediction comparison in terms of errors compared to ground truth.

ID	P-mean	P-std	U-mean	U-std	ID	P-mean	P-std	U-mean	U-std
1	1.79	0.82	3.77	2.33	6	4.11	2.84	5.18	8.36
2	2.46	0.90	3.87	2.64	7	1.63	0.91	4.93	2.66
3	4.08	1.80	14.96	11.74	8	2.54	0.58	4.80	2.17
4	3.87	3.56	7.53	9.19	9	1.64	0.63	4.24	1.18
5	3.39	2.24	3.33	1.47	10	2.52	1.71	3.10	1.47

¹ Values are differences in pixels from hand-labeled ground truth. 1 px is about 0.24 mm at the given camera view.

² Average difference: P-mean: 2.80, P-std: 1.60, U-mean: 5.57, U-std: 4.32.

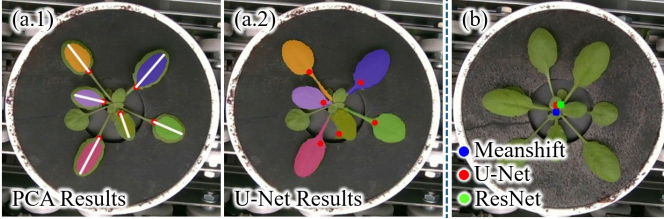


Fig. 6: (a.1) Junction points detected as the intersections between principal axes and mask contours. (a.2) Junction points predicted by U-Net. (b) Estimated plant centers.

different leaf sizes, plant structures, and lighting conditions would also require a larger annotated dataset and extensive data augmentation. Therefore, we did not adopt this learning-based method in the current system. Fig. 6(a) provides a visual comparison of the two methods of one plant.

For plant center estimation, we compared our MeanShift method with U-Net- and ResNet-based alternatives. Fig. 6(b) illustrates one result. Although the approaches may show pixel-level differences, these differences do not fundamentally affect the disambiguation between the junction point and the leaf tip, because the tip is always located at the distal end of the leaf principal axis and is farther from the plant center.

B. Real-World Experiments and Analysis

The developed robot system was deployed in an automated plant phenotyping system (as shown in Fig. 7) to collect leaf samples for 70 different plants. Table II shows the results of all trials. Here, a trial is defined as successful if the robot can cut the target leaf, pick it up with the suction cup, and deposit it into the test tube using the proposed method. In total, 61 out of 70 trials were successful, yielding a success rate of 87.1%. Among the nine failures, four were due to global localization errors, and the other four were due to local refinement failures. The global localization failures were primarily caused by occlusions from overlapping leaves, which led to inaccurate estimation of the leaf center and junction point. Besides, there are cases in which the hard leaf selection criterion rules out all detected leaves, leaving the system to terminate at a plant. The failures in local refinement resulted from the inability to perform rotational correction. When the principal axis estimated during global localization is inaccurate, translational refinement alone may be insufficient. Incorporating rotational correction into the local refinement process is therefore an important direction for future work. We encourage readers to

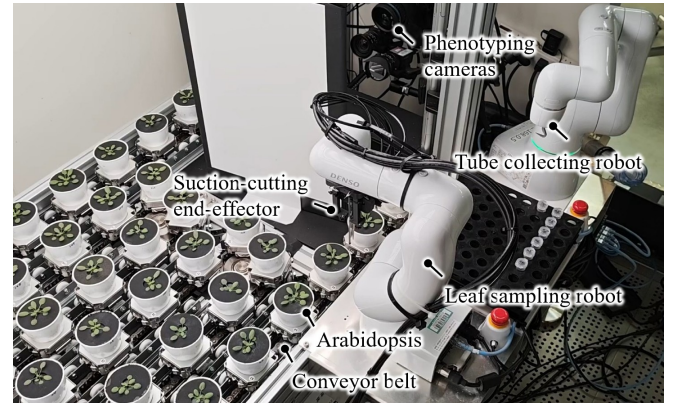


Fig. 7: The leaf-sampling robot samples Arabidopsis leaves from plants carried on the conveyor belt and deposits sampled leaves into test tubes. The tube-collecting robot transfers the test tubes into liquid nitrogen for instant preservation. The complete workflow can be found in a supplementary video.

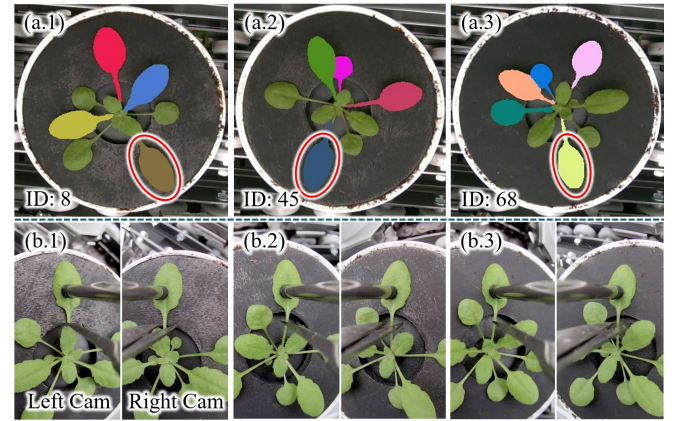


Fig. 8: (a) Global segmentation results of three samples. (b) Final refinement results after suction tube descent.

refer to the supplementary video released with this paper for execution details and success and failure inspections. Fig. 8 visualizes three representative successful trials. The first row shows the global recognition results, the second row shows the local refinement results, and the last row shows the cut leaves collected in test tubes.

Regarding time efficiency, the system took an average of 9.23 s to globally select and approach a target leaf, and 15.68 s to complete the local refinement and cutting process. Among the 65 trials involving local refinement, 64 trials took five iterations (60 converged, 4 failed). The remaining one converged in four iterations. Each refinement iteration took approximately 0.6 s. In addition, before the first refinement iteration, the system spent about 1.8 s warming up and loading the inference model, resulting in an average refinement time of 4.5 s. In total, the system spent 137.7 min sampling 70 plants, including 31.8 min of idle time associated with robot setup and human intervention. The net operation time was 105.9 min, corresponding to 1.51 min per plant, or 0.66 picks per min. This duration is longer than the sum of t^{gl} and t^{loc} in Table II, because it also included coordination time with the

TABLE II: Performance of the developed robot on 70 plants rotated in an automated plant phenotyping system.

ID	Result	t^{gl} (s)	t^{loc} (s)	#Iter	ID	Result	t^{gl} (s)	t^{loc} (s)	#Iter
1	Succ.	8.83	16.07	5	36	Succ.	8.05	15.76	5
2	Succ.	8.12	15.8	5	37	Succ.	8.93	15.79	5
3	Succ.	9.91	16.02	5	38	Succ.	8.55	13.89	4
4	Fail.(G)	-	-	-	39	Succ.	10.03	15.72	5
5	Succ.	8.19	14.78	5	40	Succ.	8.23	14.64	5
6	Succ.	11.97	16.39	5	41	Succ.	7.83	19.09	5
7	Succ.	8.36	15.79	5	42	Succ.	8.35	15.78	5
8	Succ.	8.97	15.91	5	43	Succ.	8.82	15.70	5
9	Succ.	9.10	15.83	5	44	Succ.	8.88	15.65	5
10	Succ.	9.05	15.95	5	45	Succ.	12.75	16.10	5
11	Succ.	8.24	15.78	5	46	Succ.	9.68	15.70	5
12	Fail.(L)	7.91	13.44	5	47	Fail.(G)	-	-	-
13	Succ.	16.29	16.00	5	48	Succ.	8.46	14.52	5
14	Fail.(L)	8.93	15.62	5	49	Succ.	9.31	15.56	5
15	Fail.(L)	10.22	14.40	5	50	Succ.	8.91	15.94	5
16	Succ.	8.69	15.64	5	51	Fail.(G)	-	-	-
17	Succ.	9.21	15.90	5	52	Succ.	8.61	15.74	5
18	Succ.	8.64	15.91	5	53	Succ.	11.62	15.50	5
19	-	-	-	-	54	Succ.	9.66	15.66	5
20	Succ.	8.03	15.71	5	55	Succ.	9.11	15.84	5
21	Succ.	9.16	15.92	5	56	Succ.	9.05	15.86	5
22	Succ.	10.44	15.78	5	57	Succ.	8.91	15.74	5
23	Fail.(G)	-	-	-	58	Succ.	9.20	15.77	5
24	Fail.(L)	9.15	16.4	5	59	Succ.	9.05	15.74	5
25	Succ.	9.51	15.85	5	60	Succ.	9.27	15.76	5
26	Succ.	8.74	15.85	5	61	Succ.	9.17	15.57	5
27	Succ.	8.17	15.58	5	62	Succ.	11.94	15.86	5
28	Succ.	9.13	15.92	5	63	Succ.	8.28	15.50	5
29	Succ.	8.31	15.32	5	64	Succ.	8.56	15.66	5
30	Succ.	8.31	14.61	5	65	Succ.	9.80	15.66	5
31	Succ.	9.12	14.85	5	66	Succ.	9.11	15.81	5
32	Succ.	9.85	15.70	5	67	Succ.	9.40	15.78	5
33	Succ.	8.85	16.02	5	68	Succ.	9.67	15.69	5
34	Succ.	8.76	15.84	5	69	Succ.	8.74	15.52	5
35	Succ.	8.70	16.06	5	70	Succ.	9.36	15.76	5

¹ Fail.(G): Global Recognition Failure; Fail.(L): Local Refinement Failure.

² t^{gl} and t^{loc} : Time costs of global and local processes.

³ #Iter: Number of local refinement iterations. The maximum number is set to 5, and successful trials often reach this limit due to the strict termination threshold used to reduce false-positive convergence.

⁴ ID19 is an empty pot.

tube collecting robot.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a fully automated robotic system for leaf sampling in Arabidopsis. Comparative studies with learning-based junction point and plant center prediction methods confirmed the robustness of the analytical approaches adopted in the workflow. Real-world experiments demonstrated the system's effectiveness across diverse plant phenotypes and environmental conditions.

Despite the successful implementation and deployment, several limitations remain in the current system. First, the current implementation prioritizes accuracy over speed by using a strict stopping threshold and redundant refinement iterations, which slows down the sampling throughput. A better balance between accuracy and efficiency should be explored according to task requirements. Second, although the vacuum generator provides pressure sensor readout, this signal is not used as active feedback in the current implementation, because the local refinement stage was assumed to provide sufficiently reliable suction in most trials. Incorporating pressure feedback

could provide an additional safeguard when higher reliability is required. Third, although lifting the leaf can reduce the risk of interference between the scissors and the soil-shading cover, the strategy is not yet completely reliable. Cover damage may still occur due to depth perception errors and the use of a fixed 4.5 mm threshold. Therefore, the relative configuration between the scissors and the suction cup may need to be adjusted according to specific plant and cover conditions.

Several broader directions also remain for future work: (i) extending perception from planar rosettes to non-planar, multi-layer leaf arrangements through multi-view depth fusion and 3D petiole modeling; (ii) generalizing target selection from 2D separability to 3D spacing and occlusion metrics; and (iii) adapting the end-effector pose and cutting trajectory to curved leaf surfaces and non-coplanar stems, potentially with active viewpoint selection to reduce occlusion. These extensions would make the system more adaptable to complex plant architectures and broader phenotyping applications.

APPENDIX

A. Empirical parameters

There are several empirical parameters in the system. Although these parameters are not theoretically derived, they are indispensable for successful implementation. The parameters include (i) a default junction-point height offset of -1 mm relative to the detected leaf center when the junction point is difficult to perceive, (ii) a refinement displacement factor of 0.061 mm/px for converting local pixel offsets to corrective motion, and (iii) a radial refinement displacement factor of 0.036 mm/px for large leaves.

B. Leaf mask extraction

We adopted Mask R-CNN as the leaf mask extraction method in the proposed system. There are many other choices like the YOLO series, SAM series, and their combinations. We compared them in this appendix for reference. Specifically, we tested the Mask R-CNN only, YOLOv11 only, and YOLOv11-prompted SAM2^{Note 1} on 45 artificial plant images. The ground truth masks are annotated by hand-prompted SAM2. The results are shown in Table III.

TABLE III: Segmentation Metrics Comparison

Method	Precision	Recall	F1-Score	Mean IoU	Time(s)
Mask R-CNN	0.64	0.90	0.75	0.82	0.06
YOLOv11	0.41	0.79	0.53	0.68	0.03
YOLOv11-SAM2	0.51	0.97	0.66	0.88	0.21

We can see from the results that the Mask R-CNN model efficiently identifies true positive leaves while maintaining the reliability of our positive predictions. Although YOLOv11-prompted SAM2 achieved the highest mean IoU, correct instance-wise prediction of each individual leaf is more critical for stable sampling than mask overlap alone. This is an important reason why Mask R-CNN was adopted in this study.

^{Note 1} SAM2 must be prompted by YOLOv11 to generate meaningful masks. Or else, every part of the image will be segmented.

When selecting target masks after Mask R-CNN segmentation in global vision phase, we eliminate the leaves whose singular values are subject to $\sigma_1/\sigma_2 < 1.5$. This constraint guarantees that circular leaves are filtered out and only leaves with a principal axis in the petiole direction remain valid.

C. MeanShift for plant center estimation

When applying MeanShift, we captured multiple frames for Mask R-CNN segmentation and computed all principal axes of the recognized masks. This strategy increases the number of intersections for robust convergence to the real center of the plant because the wrong intersections scatter to multiple locations and will not add to density at the real center.

D. Leaf detection under suction tube occlusion

We fine-tuned the Mask R-CNN model used in the local refinement stage to handle leaf detection under suction tube occlusion. This process adopted a training dataset consisting of both real and synthetic images. For the real images, the contours of the leaf masks are manually annotated. For the synthetic images, the suction tube is digitally overlaid onto the target leaf at different levels of occlusion, while the original (non-occluded) mask is retained as the ground truth. The overlaid images are then placed on randomly generated backgrounds (e.g., white noise, polygons, or stripes). In total, 458 images (122 real and 366 synthetic) are used for fine-tuning. Examples of the dataset are shown in Fig. 9.

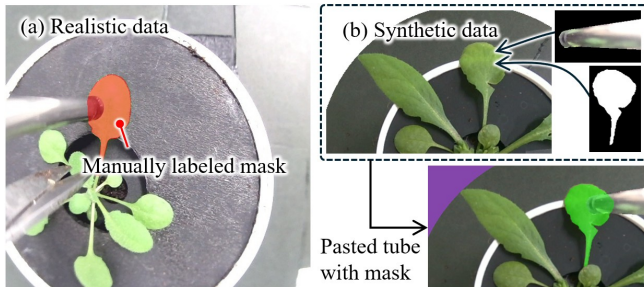


Fig. 9: (a) Real occluded leaf with a manually labeled mask. (b) Synthesized occluded leaf and mask.

REFERENCES

- [1] D. W. Meinke *et al.*, “Arabidopsis thaliana: A Model Plant for Genome Analysis,” *Science*, vol. 282, no. 5389, pp. 662–682, 1998.
- [2] T. Gill *et al.*, “A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping,” *Phenomics*, vol. 2, no. 3, pp. 156–183, 2022.
- [3] A. Atefi *et al.*, “Robotic technologies for high-throughput plant phenotyping: Contemporary reviews and future perspectives,” *Frontiers in Plant Science*, vol. 12, p. 611940, 2021.
- [4] M. Mortazavi, D. J. Cappelleri, and R. Ehsani, “Romu4o: A robotic manipulation unit for orchard operations automating proximal hyperspectral leaf sensing,” *arXiv:2501.10621*, 2025.
- [5] T. Mueller-Sim *et al.*, “The robotanist: A ground-based agricultural robot for high-throughput crop phenotyping,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3634–3639.
- [6] S. Vijayarangan *et al.*, “High-throughput robotic phenotyping of energy sorghum crops,” in *Field and Service Robotics*, 2018, pp. 99–113.
- [7] E. Navas *et al.*, “Soft grippers for automatic crop harvesting: A review,” *Sensors*, vol. 21, no. 8, p. 2689, 2021.
- [8] A. Kaleem *et al.*, “Development challenges of fruit-harvesting robotic arms: A critical review,” *AgriEngineering*, vol. 5, no. 4, pp. 2216–2237, 2023.
- [9] G. Alenya *et al.*, “Robotized plant probing: Leaf segmentation utilizing time-of-flight data,” *IEEE Robotics & Automation Magazine*, vol. 20, no. 3, pp. 50–59, 2013.
- [10] S. Foix, G. Alenya, and C. Torras, “3d sensor planning framework for leaf probing,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 6501–6506.
- [11] A. Chaudhury *et al.*, “Machine vision system for 3d plant phenotyping,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 6, pp. 2009–2022, 2019.
- [12] K. Ahlin *et al.*, “Autonomous leaf picking using deep learning and visual-servoing,” *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 177–183, 2016.
- [13] Y. Bao, L. Tang, and D. Shah, “Robotic 3D Plant Perception and Leaf Probing with Collision-Free Motion Planning for Automated Indoor Plant Phenotyping,” in *ASABE Annual International Meeting*, 2017.
- [14] Z. Chen, J. Wang, and J. Jin, “Fully automated proximal hyperspectral imaging system for high-resolution and high-quality in vivo soybean phenotyping,” *Precision Agriculture*, vol. 24, pp. 2395–2415, 2023.
- [15] S. Selvam, A. Silwal, and G. Kantor, “T-rex: Vision-based system for autonomous leaf detection and grasp estimation,” *arXiv preprint arXiv:2505.01654*, 2025.
- [16] —, “Self-supervised learning for robotic leaf manipulation: A hybrid geometric-neural approach,” *arXiv preprint arXiv:2505.03702*, 2025.
- [17] M. Campbell, A. Dechemi, and K. Karydis, “An integrated actuation-perception framework for robotic leaf retrieval: Detection, localization, and cutting,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 9210–9216.
- [18] A. Deb, K. Kim, and D. J. Cappelleri, “Deep learning-based leaf detection for robotic physical sampling with p-agbot,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 8291–8297.
- [19] T. T. H. Giang and Y.-J. Ryoo, “Autonomous robotic system to prune sweet pepper leaves using semantic segmentation with deep learning and articulated manipulator,” *Biomimetics*, vol. 9, no. 3, p. 161, 2024.
- [20] C. Lehnert *et al.*, “Autonomous sweet pepper harvesting for protected cropping systems,” *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 872–879, 2017.
- [21] A. Silwal *et al.*, “Design, integration, and field evaluation of a robotic apple harvester,” *Journal of Field Robotics*, vol. 34, no. 6, pp. 1140–1159, 2017.
- [22] Z. Sobol *et al.*, “Prototype of a new head grabber for robotic strawberry harvesting with a vision system,” *Sensors*, vol. 24, no. 20, p. 6628, 2024.
- [23] E. Ochoa and C. Mo, “Design and field evaluation of an end effector for robotic strawberry harvesting,” *Actuators*, vol. 14, no. 2, p. 42, 2025.
- [24] S. Birrell *et al.*, “A field-tested robotic harvesting system for iceberg lettuce,” *Journal of Field Robotics*, vol. 37, no. 2, pp. 225–245, 2020.
- [25] E. Navas *et al.*, “Soft gripper for small fruits harvesting and pick and place operations,” *Frontiers in Robotics and AI*, vol. 10, 2024.
- [26] R. Yu *et al.*, “Development and experiment of adaptive oolong tea harvesting robot based on visual localization,” *Agriculture*, vol. 14, no. 12, p. 2213, 2024.
- [27] B. A. Alexander and J. Binney, “Method for automating transfer of plants within an agricultural facility,” US Patent US10225993B2, 2019, assignee: Iron Ox, Inc.
- [28] M. H. Boudreau *et al.*, “Automated mushroom harvesting system,” U.S. Patent Application US20230044563A1, 2023.
- [29] A. Wisdom and I. Mintz, “End effector for harvesting plants,” U.S. Patent Application US20240147903A1, 2024.
- [30] K. He *et al.*, “Mask r-cnn,” in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [31] N. Ravi *et al.*, “Sam 2: Segment anything in images and videos,” *arXiv:2408.00714*, 2024.
- [32] M. Minervini *et al.*, “Finely-grained annotated datasets for image-based plant phenotyping,” *Pattern Recognition Letters*, vol. 68, pp. 231–241, 2015.
- [33] International Plant Phenotyping Network (IPPN), “Datasets,” <https://www.plant-phenotyping.org/datasets>, 2016.
- [34] D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [35] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab *et al.*, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.