Eliciting Challenges and User Needs Associated with Annotation Software for Plant Phenotyping

Xiaolei Guo
Department of ECE
University of Florida
Gainesville, Florida, USA
suninth@ufl.edu

Qing Li
Department of CISE
University of Florida
Gainesville, Florida, USA
li.qing@ufl.edu

Sarah Morrison-Smith
Computer Science Department
Hamilton College
Clinton, New York, USA
smorriso@hamilton.edu

Lisa Anthony
Department of CISE
University of Florida
Gainesville, Florida, USA
lanthony@cise.ufl.edu

Alina Zare
Department of ECE
University of Florida
Gainesville, Florida, USA
azare@eng.ufl.edu

Yangyang Song
Agronomy Department
University of Florida
Gainesville, Florida, USA
yangyangsong@ufl.edu

ABSTRACT

Artificial Intelligence (AI) has been enhancing data analysis efficiency and accuracy during plant phenotyping, which is vital for tackling global agricultural and environmental challenges. Designing a reliable AI system to assist precise plant phenotyping begins with high-quality phenotypic feature annotation, which usually involves collaboration between plant scientists and AI specialists. However, due to the high level of diversity in these researchers' backgrounds, it is likely that they have differing user needs from a fine-grained plant feature annotation system. We conducted semi-structured interviews with eight experienced annotators from diverse backgrounds, and observed how they interact with their preferred annotation system, to elucidate the challenges faced when annotating plant features and identify user needs. We collected qualitative responses to the interview questions, and conducted a quantitative evaluation of the agreement of their annotations on the given images. By analyzing the participants' behaviors and the collected data, we identified common user needs and derived implications for the design of an AI-assisted annotation system, including providing a range of annotation options, the flexibility to adapt annotations, and functions to help addressing uncertainty. Our research contributes to the design of systems that make annotations efficient and reliable, not only benefiting plant phenotyping, but also other interdisciplinary fields that rely on user-driven annotations.

KEYWORDS

Image Annotation, Plant Phenotyping Tracing, Human Computer Interaction, Artificial Intelligence Interface, eXplanation User Interface, Interactive Machine Learning

IUI '24, March 18–21, 2024, Greenville, SC, USA

ACM ISBN 979-8-4007-0508-3/24/03
https://doi.org/10.1145/3640543.3645178

ACM Reference Format:
Xiaolei Guo, Qing Li, Sarah Morrison-Smith, Lisa Anthony, Alina Zare, and Yangyang Song. 2024. Eliciting Challenges and User Needs Associated with Annotation Software for Plant Phenotyping. In *29th International Conference on Intelligent User Interfaces (IUI '24), March 18–21, 2024, Greenville, SC, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3640543.3645178

1 INTRODUCTION

Plant phenotyping, the science of measuring and analyzing plant traits or features, has become increasingly crucial in addressing global agricultural and environmental challenges. To understand plant growth, breeding, and how plants respond to their environment, researchers face the challenge of managing and analyzing vast amounts of plant data. Artificial intelligence (AI) techniques have become essential tools with the capacity to automatically identify and analyze intricate plant structures [31, 39, 40, 42, 45]. These AI advancements play a pivotal role in enhancing the precision and efficiency of plant phenotyping, leading to practical applications like weed and disease detection, as well as monitoring plant moisture and nutrient cycling. Yet, a reliable AI-assisted plant phenotyping system depends on high quality phenotypic trait annotations.

Phenotypic traits annotation is an essential step to support AI-assisted agricultural applications. As an interdisciplinary subject, phenotypic traits annotation involves annotators with diverse backgrounds, spanning both plant scientists [21, 23, 26, 30, 38, 43] and AI specialists [4, 12, 46]. This diversity results in varying mental models, leading to distinct user behaviors and needs [38, 40]. For instance, plant scientists specialize in identifying phenotypic traits. However, they may lack experience in integrating AI-assisted features to improve efficient annotation [20, 38]. Additionally, they might be unfamiliar with how AI algorithms will use their annotations and what the different annotation requirements are for distinct AI algorithms. On the other hand, AI experts excel at addressing annotation challenges but may not possess the domain-specific knowledge to know what to annotate [13, 40, 44]. In addition, Designer of AI system may unintentionally overlook the perspectives of non-technical users. This can lead to a lack of awareness when designing explainable AI-assisted features for effective annotation of phenotypic traits, particularly for those who are unfamiliar with

AI concepts. Consequently, designing a system that facilitates efficient, reliable, and collaborative annotation across these diverse user groups remains a significant challenge.

To tackle this challenge, our research focused on how user behaviors are affected and adapted by factors including annotators' prior knowledge, image quality, annotation methods, and system features. We conducted a semi-structured interview study and observed annotators walking through their preferred annotation system, to understand their mental models during interactions with annotation software. Our research was centered on annotating fine, hair-like phenotypic traits [33, 41], primarily because these traits are prevalent in a wide range of plant organs from roots to grass leaves and stems, and naturally occur in diverse environments, including soil and the surrounding vegetation. Moreover, these traits hold significant real-world relevance, with applications ranging from studying the ecological role of plant roots to detecting weeds in agriculture and optimizing turf grass management. In addition to collecting qualitative data (i.e., responses to the interview questions as well as the observations of system use), we conducted a quantitative measurements to assess annotation consistency made among participants.

2 RELATED WORK

This section illustrates how users from various backgrounds select their preferred annotation systems based on distinct user needs observed in previous studies. It also discusses the essential features of current annotation systems, categorized by how the annotator interacts with the systems. Finally, we elucidated the limitations in existing annotation systems, which often overlook diverse user needs.

2.1 User Preferences in Annotation System Selection

Generally, in the domain of fine-grained level plant phenotypic traits annotation, there exist three commonly employed approaches that are driven by the diverse needs of annotators. First, some plant scientists may first seek specialized platforms that support both annotation and phenotypic trait analysis. For instance, root annotators [36, 38] utilize software such as WinRhizo Tron [25], RootNav [23], GLO-Roots [26], and RootSnap [19] to trace roots while measuring their length, surface area, tip angle, child count, etc. However, annotators usually face significant time expenditures when dealing with poorly designed annotation tools. For example, when employing an annotation tool which restricts users to tracing roots with a specific diameter, Xu et al. [38] encountered the need to frequently alter the direction and diameters of annotations to accommodate curly roots with varying orientations and sizes.

When specialized software for phenotyping analysis isn't readily available, annotators often seek flexible and user-friendly alternatives. For instance, Wang et al. [35] employed GIMP [34] for root annotation, while Guo et al. [13] used ImageJ [1] for annotating avocado injuries. Additionally, in studies involving computer engineering students and researchers, some annotators prefer dedicated image annotation systems like the Computer Vision Annotation Tool (CVAT [29]), VGG Image Annotator (VIA [8, 9]), and Segment Anything Model (SAM [17]), as these offer various efficient annotation tools. However, these popular systems can pose challenges in plant phenotyping because they are primarily designed for common objects such as humans and vehicles, which have different characteristics from plant traits (e.g., fine roots mixed with complex surroundings).

In another scenario, when AI scientists take on the role of annotators, some prefer coding annotations from scratch, leveraging techniques and algorithms based on their expertise; for example, Yu et al. [40] and Xu et al. [39] employed superpixel selection to annotate roots, and Biswas, S. and Barma, S. [4] utilized thresholding and morphological operations to annotate potato cells. Besides coding from scratch, open-source code packages supporting phenotypic trait annotation, such as PlantCV [11] and RootPainter [31], have been developed for annotators with programming knowledge. However, installing and running these packages often requires some computer science proficiency, which can be a hurdle for researchers without an engineering background.

Annotators with different backgrounds likely have different preferences on the annotation systems that meet their goals and fit their mental models. Plant scientists might focus more on utility for plant analysis while a computer engineer might consider the annotation from the algorithm perspective; non-expert annotators might find a platform with efficient annotation features to be easier to learn and use. Therefore, to build an explainable annotation system, we need to take users' needs from various perspectives into account and make features transparent among varied target users.

2.2 An Overview of Annotation System Features

In general, existing annotation systems can be categorized by how the users interact with the systems: (i) fully-manual annotation, (ii) automatic annotation, and (iii) semi-automatic annotation. In fully-manual annotation systems, annotators have to manually make all the annotations [7, 10, 18, 38], while the system provides the least assistance. Conversely, with the automatic annotation system, human annotators provide the least instruction where the systems can generate annotations automatically [25, 39, 42]. A fully-manual system typically demands substantial human effort, while a fully-automatic system runs the risk of inaccuracy without human supervision. The semi-automatic annotation system combines human input and system output, where the human annotator can complete the annotation task with AI-assisted features [5, 6, 17, 22, 27, 31, 37]. Semi-automatic annotation methods involve interactive annotation, where annotators guide the AI models to refine their predictions iteratively. Alternatively, some systems may offer tentative annotations based on AI models, which may require manual adjustments without real-time feedback.

Regarding how annotations are made, two common approaches are used to select objects of interest. The first involves covering regions, where closed shapes, scribbles, or supervised clicks are used to cover and select target regions [19, 25, 27]. Another option is to trace the object's boundary, which is achieved by placing vertices connecting lines [6, 29] or curves [22] along the boundary, or by manually outlining it with a mouse or touch pen [2].

Systems typically offer two methods for providing feedback to annotators. The first is an annotation mask, which is typically a translucent shape overlay featuring pixels that potentially belong to the target or a line outlining the target boundary atop the image. In semi-automatic annotation systems, the annotation mask is predicted through an AI model, allowing annotators to refine these suggested annotations to obtain the final annotation [16, 27, 31, 42]. Additionally, annotators might be interested in understanding why the system suggests a specific region for annotation. Therefore, the second type of system feedback involves providing explanations for the predictions; such as visual heatmaps that indicate regions highly correlated with human annotations [14].

2.3 Limitations in Existed Annotation Systems

Although there are numerous systems and approaches supporting phenotypic traits annotation in prior work, previous research has primarily concentrated on technical aspects. For example, Zeng et al. [42] used automatic detection to extract preliminary roots, allowing annotators to make corrections. Smith et al. [31] integrated advanced AI techniques into a semi-automatic root annotation system. However, these works mainly focused on algorithm development, neglecting the diverse mental models of users and their needs.

Plant scientists interested in using AI techniques for automatic plant phenotyping may encounter challenges due to their limited technical background, such as efficiently creating reliable annotations. For example, in weed detection, common concerns include whether a simple bounding box can locate weeds or if every individual leaf needs manual delineation, along with concerns about the time required for annotating numerous leaves in a grassland. On the other hand, AI technicians are well-versed in how AI algorithms work and can design systems tailored for precise pixel-level annotations or less precise bounding box annotations. However, they may require additional training to distinguish specific plant traits, such as weed species or diseased plant organs.

Annotation plays a crucial role in AI algorithms. High-quality annotation is essential for improving AI model quality, but it requires considerable human effort. Plant annotators may question the trade-off between AI model quality and the use of inaccurate yet easily obtainable annotations. For instance, the YOLO algorithm [24] is an effective method that can potentially be applied to weed detection, which requires simple bounding box annotation. However, in a grassland where weed and turf grass are mixed, the accuracy of the bounding box may impact the model's effectiveness. Novice annotators may question the precision level of the bounding box. They may wonder if including turf grass in the bounding box or excluding weeds from the scene will affect the quality of weed detection. On the other hand, it is important for AI technicians to understand how users utilize AI-assisted features [6, 17, 31] and their expectations. This understanding will enable AI techniques to be more helpful in annotating.

These open questions give rise to diverse challenges in building annotation systems to meet varying user needs. Therefore, our study aimed to fill this gap by interviewing annotators to understand their mental models and the actions they take to overcome challenges during the annotation process. These insights are valuable for enhancing annotation systems, shedding light on what constitutes an efficient and reliable system. In general, our study contributes to a more comprehensive understanding of the factors influencing plant phenotyping annotation.

3 METHOD

3.1 Participants Recruitment

We collected data via recorded, semi-structured interviews with eight participants (M = 29.50 years old, S.D. = 5.34 years, three female, five male). Participants were recruited through email invitations and referrals. In-person interviews were conducted with P03, P05, and P08, while virtual interviews were conducted with P01, P02, P04, P06, and P07. The recordings were automatically transcribed via Zoom and were used as the primary data source for analysis. Prior to the interviews, all participants provided informed consent through an anonymous consent form. Compensation at a rate of 20 USD per hour was provided. Our procedures were approved by our Institutional Review Board (IRB).

Our study focused on expert users with a diverse background in terms of degree, major, experience, and software. All participants in our study had experience annotating fine-grained plant traits, focusing on various targets: plant roots (P02-P05, P07, P08), weeds (P01), and plant stems (P06). P02, P03, P05, and P07 annotated minirhizotron roots from different crops. In particular, P05 and P07 worked with roots in a time series. P04 and P08 handled roots grown in transparent plastic boxes. The diverse background of participants and their experience using various annotation systems are summarized in Table 1.

3.2 Interview Conduction and Analysis

The interviews, which spanned approximately 1-2 hours (M = 71.63, S.D. = 14.86 minutes), were structured into three main sections. Initially, participants were asked about their background and previous experience with annotation. We also inquired about their specific targets of interest and how annotated data contribute to their research, leading to a discussion on the significance of annotating plant phenotypes within their respective domains.

Following the initial discussion, participants engaged in the annotation process using their preferred annotation system, working with their own datasets (if available) and our provided images. Throughout this process, participants were encouraged to think aloud, enabling us to capture valuable insights. We closely observed and documented their interactions with the system, paying particular attention to the mutual influence between user input and system feedback. Specifically, user input referred to participants' actions and decisions, while system feedback encompassed visual cues, display settings, and suggestions that guided participants' annotations. Additionally, we observed participants' behaviors and spoken thoughts on how the plant phenotypes impact the bidirectional interaction between human and system.

Towards the end of our interviews, participants who worked with systems that required fully manual annotation watched a short video demonstrating an interactive annotation example. The example utilized the "GrabCut [27]" algorithm to assist annotation. In this video, annotator made simple freehand scribbles to cover part of the target, and the system generated suggested annotations based on this input. The annotator further refined the suggested

Table 1: Participants Background

ID	Degree	Major	Experience	System
P01	Postdoc Researcher	Agricultural Life Sciences	6 months	LabelMe, CVAT
P02	PhD	Electrical & Computer Engineering	2-3 year	Winrhizo Tron, Image Annotator
P03	Master	Electrical & Computer Engineering	3-4 months	LabelMe, Matlab Image Labeler App
P04	PhD	Electrical & Computer Engineering	<1 month	VIA
P05	Master	Agricultural Ecology	1 year	Winrhizo Tron
P06	PhD	Agricultural & Biological Engineering	5 year	Matlab Image Labeler App
P07	Postdoc Researcher	Plant Physiology	7 year	RootFly
P08	PhD	Electrical & Computer Engineering	9 month	Photoshop

annotations by drawing additional scribbles. The purpose of providing the example video was to collect participants' opinions on the interactive framework that employed AI techniques to assist human annotation.

We employed a bottom-up analysis of the interview data by utilizing an affinity diagramming [15] [32] technique to discern emergent themes and patterns. This method aligns with the qualitative analysis framework as delineated by Auerbach and Silverstein [3]. Unlike traditional qualitative coding that involves calculating inter-rater reliability (IRR), our approach involved two researchers collaboratively reaching a consensus on the significant ideas and concepts represented on sticky notes.

During the affinity diagramming process, interview data and observations recorded on sticky notes were organized into coherent clusters through iterative discussions and refinements, eliminating redundancy and enhancing clarity. Subsequently, overarching themes were extracted from these clusters, serving as the basis for a more nuanced content analysis. This allowed us to delve into the intricacies within each theme, identifying commonalities, variations, and key insights. Finally, we explored inter-theme relationships to uncover potential associations or implications, thereby enriching our understanding of the data. This approach is particularly suited for semi-structured interviews, as it helps getting a comprehensive interpretation of the data about participants' interaction process with different annotation systems and also mitigates the risk of applying the same code to different sections of the interview data.

3.3 Quantitative Measurements

To evaluate the consistency of annotations made by different participants, we assigned four minirhizotron root image annotation tasks from PRMI [38] to all participants. (See Fig. 1 for details.) Then, we collected the binary mask, a visual representation indicating the specific regions of interest as marked and annotated by participants. Examples are shown in Fig. 2. To evaluate annotation quality, we computed the Intersection of Union (IoU) score, as suggested by well-known image datasets [7, 18]. The IoU score measures agreement among annotators for each image task and serves as an indication of annotation reliability, with a higher IoU suggesting greater agreement.

4 RESULTS

4.1 User Annotation Experience

Participants commonly mentioned that target characteristics significantly affected the annotation process, citing factors like color contrast, shape, and size. Challenges arose when roots shared colors with the background, making it hard to determine the boundary, as observed by P08. P01 encountered similar difficulties when annotating weeds in turf grass due to visual similarities. Target shape also played a role, with regular, thin, straight roots (P02) being easier to annotate compared to irregular or winding roots with branches (P02, P04). Size mattered, with broader leaves being simpler to annotate (P01), while P06 noted difficulties with tiny stems and needle leaf occlusion. Artifacts like mud, dirt, sands, or stones in the soil further complicated root identification, especially when mixed or covered (P03).

Moreover, participants P05 and P07, familiar with minirhizotron imaging, emphasized image quality's importance. P05 pointed out that blurry, out-of-focus images make root labeling challenging due to unclear boundaries. P07 noted issues with tube installations causing gaps between soil and the tube, degrading image quality and increasing confusion during annotation. These insights underscore collecting high-quality image is an essential initial step in accurate and efficient plant phenotyping annotations.

Annotating solely from images, especially underground minirhizotron root images, presented another challenge. Despite years of studying plants, P05 expressed uncertainty about certain aspects of the roots' condition underground, such as distinguishing between root hairs and fungi. Conversely, P08, who specializes in computer engineering, mentioned that referring back and forth between the actual roots cultivated in transparent plastic box and the root image enhances annotation justification and accuracy.

Regarding the objective of annotation, we collected insights from two perspectives. From the view of computer vision, P01-P04, P06, P08 mentioned that the annotation will be used to generate the training masks (i.e. groundtruth) for a machine learning image segmentation model. From the view of plant phenotype study, P01 and P06 further explained that the purpose of the image segmentation model will be applied to weed (P01) or disease (P06) detection. The extraction of root will be applied to various studies, including root traits such as moisture (P04, P08), length and density (P07), and nutrient cycling (P05).

Figure 1: An overview of the assigned images with varieties of roots. Images are sourced from PRMI [38] dataset. (a) Sesame roots with tiny root hairs. (b) Papaya roots that are straight and thick, with several artifacts such as bricks and stones (c) Cotton roots with a lot of background noise. (d) Switchgrass roots which are thin, but with a clear boundary.

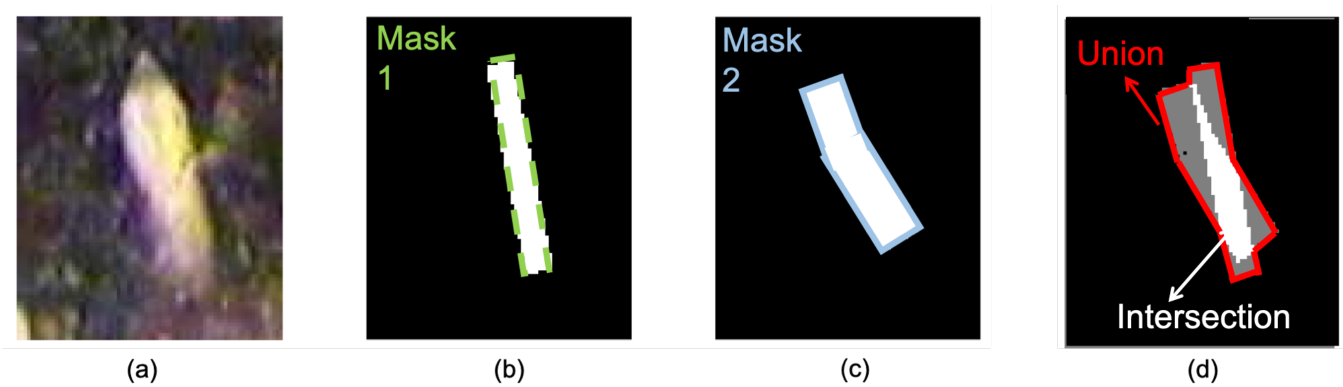

Figure 2: Example masks annotated by two annotators and their intersection of union

4.2 An Overview of the Investigated Systems

As described in Table 1, we investigated diverse annotation tools, encompassing a broad spectrum of systems. These included root tracking software explicitly designed for studying root phenotypes, such as Winrhizo Tron [25] and RootFly [42]. Additionally, we explored software specifically developed for computer vision applications such as CVAT [29], LabelMe [28], and VIA [8, 9]. In addition, we incorporated general image editing software like Photoshop [2]. Furthermore, Image Annotator was a simple GUI developed by the participants' (P02) lab to meet their demands of annotating roots for their research project. Among the systems we studied, the phenotypic software (Winrhizo Tron and RootFly) were powerful in integrated plant trait analysis, such as root length and diameter measurements. While systems designed for computer vision and image processing (CVAT, LabelMe, VIA, and Photoshop) benefited annotation in providing various options of annotation tools, smooth image file import and export with intuitive interface. Our focus primarily revolved around fully manual and semi-automated annotation systems, wherein human input is necessary. Below, we provided details of how annotation works in these systems, as well as the system features and function related to annotation.

Table 2: Interactive Methods for Investigated Systems

Interactive Methods	System	Participants
Fully-manual	CVAT	P01
	Label Me	P03
	VIA	P04
	Winrhizo Tron	P05
	Photoshop	P08
Semi-automatic	Image Annotator	P02
	Matlab Labeler	P03, P06
	RootFly	P07

4.2.1 Interactive methods. How users interact with the system was crucial in the annotation process. Table 2 categorizes the interaction methods of our investigated systems and the corresponding participants. Based on whether the system provides built-in model to assist annotation or not, the systems were categorized as: (i) fully-manual annotation (P01, P03-P05, P08); (ii) semi-automatic annotation (P02, P03, P06, P07).

Notably, the role of the built-in model differed. Image Annotator and RootFly both provided automatic root detection as a suggested annotation. This allowed the annotator to save time by refining the fixed pre-predicted annotation instead of manually tracing everything. The Matlab annotation toolbox (P03) offered a more flexible approach to assist annotation. Instead of a fixed pre-trained model, they provided a tool called "floodfill" that generated suggested annotation that was dynamically updated based on the annotator's input.

4.2.2 Annotation tools. In addition to the interactive method, participants primarily employed two approaches for making annotations: drawing polygons and freehand scribbles. (See Table 3 for details.)

Regarding the polygon tool, participants usually had the option to choose between regular (rectangle, circle, etc.) or irregular shapes. Some systems offered both options, while root tracing systems like Winrhizo Tron and RootFly only provided the rectangle option. The process of drawing polygons was similar across popular annotation systems (e.g., CVAT, Image Annotator, LabelMe, VIA). Users clicked with the mouse to place vertices along the target boundary. Operations in systems designed for minirhizotron root annotation (Winrhizo Tron, RootFly) differed, as users were required to click along the root spine line to place nodes indicating the upper and lower boundaries of the rectangle and then adjust the width to indicate the root diameter. Regarding freehand drawing, users may employ scribbles to overlap the target in the Matlab system (P04, P06) or draw a freehand line to trace the boundary of the selected region in Photoshop (P08). In both approaches, users clicked to indicate the starting point of their "painter brush" and then dragged the mouse while keeping the button pressed to draw with the brush.

4.3 Usability Features of Annotation Systems

In addition to interactive methods and annotation tools, we also reviewed system features intended to aid in the annotation process. We focused on shape editing, annotation mask display, and information editing functionalities used during annotation. This comprehensive examination helped us understand the intersection of system functionality and user interaction in the annotation process.

One key requirement for annotation systems was ease of editing, particularly in terms of error recovery and efficient copy/paste functionality. Editing functions varied depending on the annotation tool. For irregular polygons, most systems allowed the deletion of the entire shape, and some, like CVAT, LabelMe, and VIA, enabled vertex manipulation and more. Winrhizo Tron and RootFly offered similar shape editing options, including node addition and deletion, as well as diameter adjustment. They also supported copy completed annotation from one image to the new one. For freehand annotations without editable vertices or edges, users like P06 and P08 typically undid the last freehand selection or used an eraser brush to remove unwanted regions.

Another significant feature was how the annotation mask was presented. Since annotating mainly relies on visual clues, how the created annotation was displayed over the image impacted the user's perception. Systems like CVAT and Matlab Image Labeler allowed users to customize colors for annotation components (e.g., mask, boundary line, vertices), while RootFly and Winrhizo Tron offered limited color options based on root status. Matlab Image Labeler also enabled opacity adjustment. In Photoshop, user can create a separate mask layer with a selected region and transparent background. Besides manipulating the annotation components, Photoshop offered basic image editing functions like color, brightness, and contrast adjustment.

When it comes to editing annotation information, different systems offered varying degrees of flexibility. General annotation tools such as CVAT, LabelMe, Matlab Image Labeler App, and VIA allowed users to define their own label names. In contrast, systems tailored for root application (Winrhizo Tron, RootFly) only supported default label category, such as live and dead root. In addition to label information, system like VIA also allowed users to add comments associated with specific annotation masks.

4.4 Challenges and Actions during Annotation

In this section, we focused on the annotation behavior of the participants, exploring instances where annotators encountered challenges and their subsequent actions taken in response (Table 4). Additionally, we investigated how system features could influence and assist users in making decisions during the annotation process. One key observation is that functions facilitating the editing of image display settings played a crucial role in assisting annotators in justifying regions of ambiguity. We also note that the chosen annotation tool could impact how users labeled the data, subsequently affecting the overall annotation quality.

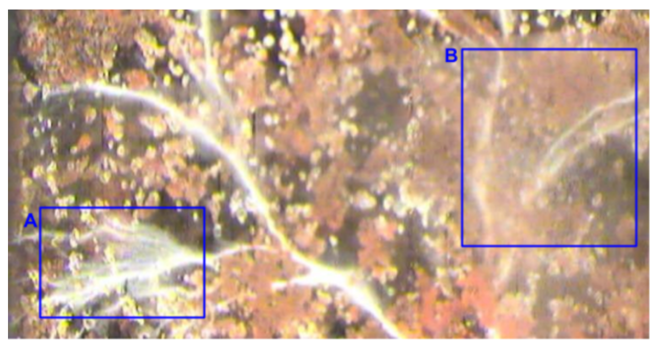

Figure 3: Cotton roots image showing color confusion. Roots are mixed with surroundings in the box A. Roots are covered by dirt in the box B.

4.4.1 Color confusion. One major challenge causing confusion among participants was color. This confusion had two aspects: target-background distinguishability, and label-image distinguishability. Out of eight participants, five (P01, P03, P05, P06, P08) of them found that clear color contrast between roots and soil made annotation easier. When roots blend with the background color (Fig. 3, region A), participants made educated guesses to define the root boundary while maintaining a smooth, thin-line shape. System features can help here. For example, P01 mentioned that zooming in on high-resolution images offered more precise visual clues. When using systems with image editing capabilities, P08 adjusted brightness and contrast to distinguish the root structure from bright surroundings.

Table 3: Annotation Tools for Investigated Systems

Annotation tool		System	Participants
Polygons	Irregular polygons	CVAT, Image Annotator, Label Me, VIA	P01, P02, P03, P04
	Rectangles	Winrhizo Tron, RootFly	P02, P05, P07
Freehand scribbles	Overlap with target	Matlab Labeler	P03, P06
	Trace the target boundary	Photoshop	P08

Table 4: Challenges and User Actions

Challenges	Actions	Participants
Color confusion between target and surroundings	Zoomed in a high-resolution picture to examine the boundary.	P01
	Utilized the system feedback to automatically identify the confused root region.	P03
	preferred manually annotation instead of automatic floodfill when the target cannot simply be distinguished by color.	P06
	Adjusted the image brightness and contrast.	P08
	Annotated all roots despite of varied root color.	P01-P08
Color confusion caused by annotation masks	Picked label color (pink) that is contrastive with target (green).	P01
	Changed mask opacity to examine label quality.	P03
	Switched between the mask layer and the image layer to examine label quality.	P08
Irregular boundary	Used the irregular polygon shape to trace the boundary.	P01-P04
	Edited polygons after closing them.	
	Used freehand drawing to cover the root in a single direction.	P03
	Adjusted the shape and size of freehand scribbles to align with the target shape and boundary.	
	Traced the boundary with freehand drawings.	P08
Tiny roots	preferred not to annotate the fluffy structure to avoid mislabeling background elements.	P05
	Labeled the tiny roots to prevent any missing target.	P07
	Zoomed in on the images for a clearer view of the tiny roots.	P06, P08
	Observed the texture of a broader region.	P08
Roots covered by soil	Annotated the root as "broken" when it was entirely covered by black soil.	P01-P04, P06, P08
	Annotated the root as continuous even if it was entirely covered by soil.	P05, P07
	Annotated the root as continuous if it was partially covered by dirt with some level of transparency.	P01-P08
Branches of root	Include all interconnected roots and their branches within a single polygon.	P01
	Drew multiple polygons to cover main root and branches.	P02, P04
	preferred polyline tool which has more clear display settings and reduce confusion of roots with complex branches.	P04
General ambiguity and uncertain	Examined root images from different time frames to understand the growth of the roots over time.	P05
	Used previous experience as guidance.	P06
	Conducted an initial annotation session followed by revisiting the annotations later to correct any errors.	
	Took a preliminary overview of the dataset before annotation.	P07
Intensive annotation time	Used interactive annotation functions to save time	P02, P03
	Developed custom code to remove background	P06

When the interactive annotation was available, P03 used the floodfill tool and began by selecting the most visually distinguishable bright root, allowing the system to automatically identify root in the blurring region. P03 then only needed to erase the apparently misclassified regions, such as granular bright stones. However, there was concern that the interactive method might be not reliable enough due to the algorithm (P06). Although P06 was aware of the automatic floodfill tool, he preferred manual annotation since his target (stem) and background (needle leaf) cannot be distinguished by color, while the floodfill algorithm relies on color. Another color-related confusion, expressed by P04, pertained to whether roots of different colors should be labeled separately. However, our observations revealed that all participants, including P04, annotated all roots in the image regardless of color. P04 suggested that future tools allow users to categorize the annotations by color.

In addition to the color confusion between target and background, we observed that annotation colors similar to the image can result in ambiguous justification. For example, the yellow line of polygon in Label Me was difficult to see when P04 was annotating a bright image where the cotton roots were mostly white (Fig. 3). Conversely, when CVAT allowed user-defined annotation colors, P01 mentioned that she usually chose pink, as it stood out from her dataset where most pictures depicted green grass. Both P03 and P04 expressed a desire to change the default annotation color setting. In addition to color, we noticed that P03 frequently adjusted the opacity of the annotation surface to assess mask quality. Higher opacity allowed more visual clues from the original image to be visible. Lower opacity, with a clearly visible mask overlaid on the image, helped in identifying mis-annotated targets. The layer switching function in Photoshop aided P08 in similar ways to capture missing targets.

4.4.2 Structure confusion. Participants often remarked the sophisticated structure of the roots they annotated posed challenges that could be categorized into several specific cases: irregular boundary, tiny roots, roots covered by soil, and roots branches. We found that the challenge of winding and irregular boundary tracing is highly related to the participants' choice of annotation tool. For instance, out of six participants having experience with polygons, four (P01-P04) preferred using irregular polygons because of their personalized nature, allowing custom shapes matching root boundaries. P02, experienced with Winrhizo Tron (rectangle) and Image Annotator (irregular polygon), favor the latter due to varying root directions and diameters. However, P07, who used RootFly, mentioned that her minirhizotron root images featured consistent root diameters that were appropriately captured by rectangular masks.

P03 found that using scribbles in Matlab annotation was faster than polygons in LabelMe because it covered the entire root in one direction, eliminating the need to click multiple times, forward and backward. P08 emphasized the advantages of yielding finer and more accurate boundaries with freehand drawings compared to the *"clunky polygon"*. P06, who also preferred scribbles, mentioned that a touch screen and touch pen were even more helpful and efficient. Although the systems used by some participants did not provide a freehand option, we demonstrated this functionality in a video and gathered their opinions. In manual annotation, P04 expressed concern that drawing scribbles overlapping with the target might take long time to get accurate and sharp boundary. P02 and P03 suggested adjusting the brush shape and size to align better with the target. As for the case of interactive annotation, P02 found scribbles efficient for highlighting specific target regions.

Utilizing interactive methods also had potential to relieve the boundary tracing challenge. While P04 worked with a fully-manual system, he proposed an interactive method to expedite his annotation procedure. Instead of tracing the precise boundary of the target, he suggested an automatic adjustment of the rough boundary based on color gradient along boundary.

Another often-discussed question related to the root structure was determining a size cutoff for inclusion in annotation. P05 mentioned being advised to annotate root hairs since they are functional parts of the root. However, it often proved challenging to determine whether the fluffy structure surrounding the root (Fig. 4, sesame, region A) was fungus or root hair. Although there was no strict criterion to determine the diameter at which a "tiny root" should be annotated, we observed that all participants considered thin and bright lines with clear and sharp boundaries as roots (Fig. 4, sesame, region B), while six (P01-P05, P07) out of eight tended to disregard blurred lines (Fig. 4, sesame, region C) to avoid false positive errors. P08 used texture cues in a broader region to justify annotating small roots. For instance, in Fig. 4, switchgrass, region A exhibits a texture that implies the presence of tiny roots compared to the plain soil in region B. Additionally, zooming in on images was also found to be helpful for precise annotations on tiny roots (P06, P08).

Another challenging aspect related to the root structure was when the root was partially or completely covered by soil, giving it a "broken" appearance (Fig. 4, papaya, region A and B). Participants wondered whether they should annotate the hidden part or just the visible fragmented section, as they weren't sure if the roots were covered or broken. Our observations revealed that six out of eight participants chose to annotate the root as "broken" when it was entirely covered by black soil (Fig. 4, papaya, region A and B), except for P05 and P07. However, when the root was covered by dirt with some level of transparency, allowing the root structure to be partially visible (Fig. 2, region B), all participants tended to annotate the root as continuous.

The last challenge caused by root structure arose when annotating roots with branches. This challenge was particularly evident when using the polygon tool, and different annotators made varying decisions. For instance, P01 tended to include interconnected roots and branches within one polygon, adjusting the path to accommodate convex holes. However, this may unintentionally exclude some roots (when the convex hole was large and the roots were small, Fig. 4, sesame, region E, Fig. 5a) or include background areas (when the convex hole was small and the roots were thick, Fig. 4, sesame, region D, Fig. 5b). Others, like P02 and P04, annotated one main root at a time but sometimes encountered issues with disconnected (missing targets) between adjacent polygons (P04, Fig. 5c). Additionally, the display settings of annotation tools, like VIA, can cause confusion on multiple root branches, when the system's default connection line appeared similar to the drawn line (P04).

4.4.3 General solution to uncertainty and ambiguity. Besides the above specific challenges, we also collected general solutions towards uncertainty and ambiguity from experienced annotators. P06,

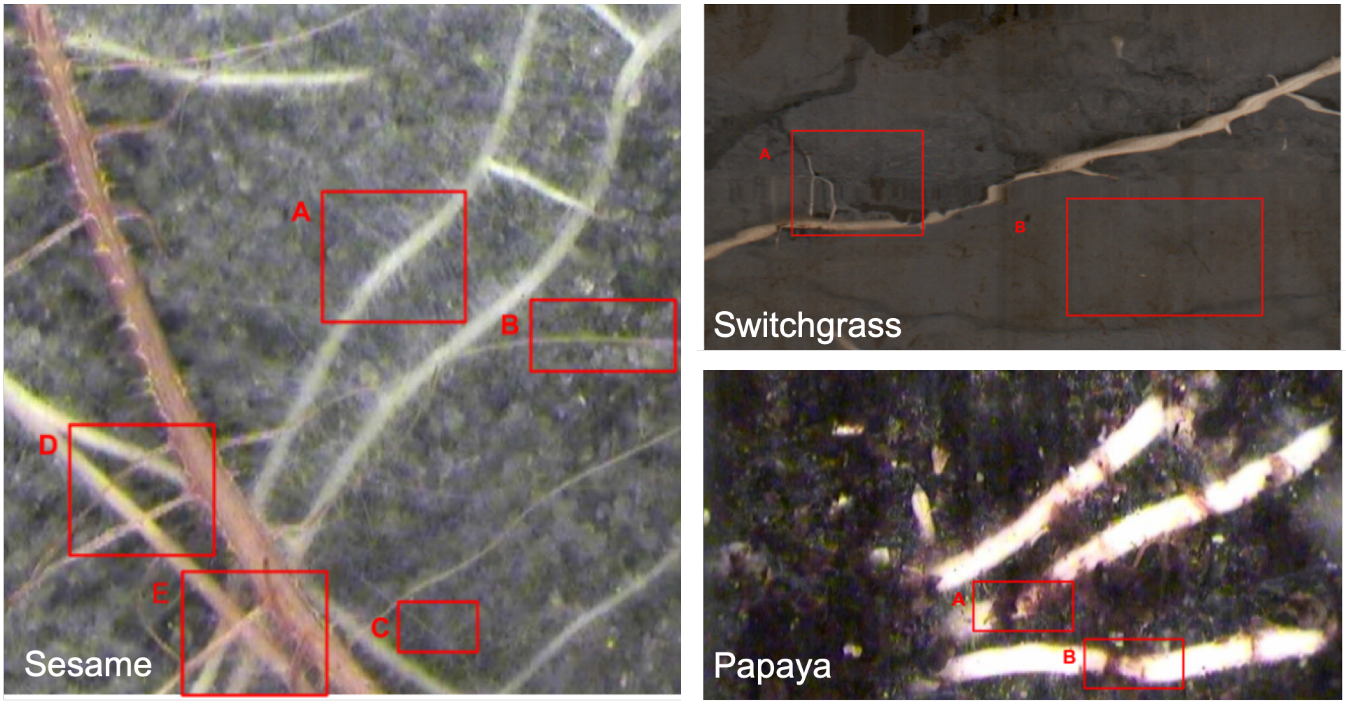

Figure 4: Sesame, switchgrass, and papaya roots images showing structure features of tiny roots (Boxes A, B and C in the sesame roots image), tangled roots (Boxes D and E in the sesame roots image, Box A in the switchgrass roots image), plain soil (Box B in the switchgrass roots image), and "broken" roots (Boxes A and B in the papaya roots image).

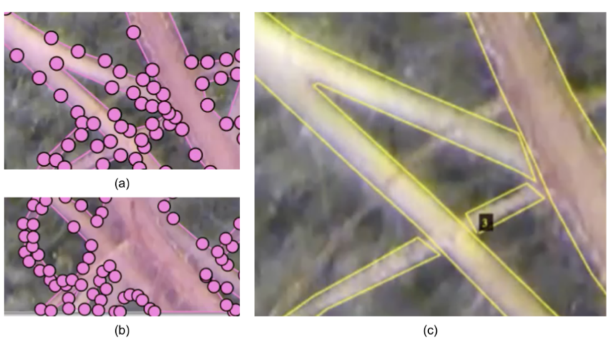

Figure 5: Sesame roots images overlapping with annotation mask on tangled roots structure. (a) and (b) showcase tiny roots mistakenly excluded and background mistakenly included, respectively. (c) showcases broken between multiple polygons.

P07, and P05 possessed extensive annotation experience spanning several years or a significant number of annotated images. P07 mentioned that she adopted a preliminary overview of the entire dataset before initiating the annotation process, allowing her to gain insights into the distribution patterns of the roots. Similarly, P05 examined root images from different time frames to understand the growth of the roots over time. P06, whose focus was on plant stems, emphasized the utilization of prior knowledge, stating that he distinguished between stems and needle leaves based on previous experience (e.g., *"stems has some vertical structures, while needles branch out"*). Additionally, P06 shared a way he used to enhance annotation accuracy: conducting an initial annotation session followed by revisiting the annotations later to correct any errors. This approach allowed him to approach the task with a fresh perspective.

4.4.4 Challenge of time-intensive annotation. In addition to the challenges posed by uncertainty and ambiguity, the time-intensive nature of annotation was a common issue highlighted by P01-P04, and P08. The time required for annotation varied significantly depending on factors such as the complexity of the image, the level of precision needed, and the usability of the annotation tool. For instance, P01 and P08 reported spending over an hour to annotate complex images with intricate leaf (P01) or root (P08) structures. Conversely, P06 took approximately one minute per image to annotate simple, straight plant stems. Participants (P01, P02, P04, P06, P08) noted that achieving highly detailed annotations was a time-consuming process, since they needed to clicking numerous points to trace the precise boundary (utilizing the polygon tool by P04), and zooming in the image to label tiny leaves (P01) or roots (P08). Annotation tools also played an important role. For example, both P06 and P08 used a touch screen for faster and more accurate labeling compared to a mouse.

To address the challenge of time-intensive annotation, several participants employed semi-automatic methods, including P02, P03, and P06. P02's system had build-in automatic root detection, while P06 coded from scratch to remove the background and then label plant stems on the segmented seedling. P03, experienced with both LabelMe (fully-manual) and Matlab Labeler (semi-automatic), preferred the latter due to its efficiency. Using the interactive method in Matlab, the time required per image was reduced to approximately 1-5 minutes, compared to 5-10 minutes per image of fully-manual annotation with LabelMe.

We showed participants using fully-manual systems a video demonstrating the concept of interactive annotation, aiming to accelerate the annotation process. We found that participants highlighted the time-saving benefits of system suggestions. All participants expressed their belief that the interactive approach can save time and facilitate efficient annotation. *"This is definitely our hope. The idea of not having to trace the entire root makes sense," (P05).* However, participants also raised concerns about the quality of the suggested annotation, which was usually determined by the underlying algorithm, particularly in complex scenarios. For instance, when targets were mixed with the background and share similar colors (P01, P03, P04, P06), when the target exhibited color variation (P04), or when targets overlaped with each other (P05). Another concern arising from the quality of suggested annotations was that annotators may spend more time correcting mis-annotated areas (P01, P02, P08). The observation that P07 rarely used automatic detection in RootFly due to poor-quality results suggested that the accuracy of predictions is a crucial factor influencing annotators' preference for labeling methods.

In summary, four (P02, P03, P06, P07) out of eight participants had experience with the AI-assisted annotation feature. Of these, P02 and P03 frequently used the feature to save annotation time. While P06 and P07 were aware of the feature, they hardly used it due to poor detection of their object of interest, which required additional effort to modify the system prediction.

4.5 Labeling Consistency Among Participants

We measured the IoU among different participants to evaluate their labeling consistency. According to our findings, labeling consistency was highly influenced by image quality. Comparing annotation methods, the flexibility and personalization provided by polygon and freehand scribble contributed to more labeling agreement among participants with varied background knowledge. Furthermore, when comparing the labels made within and between annotation methods, our result implied that using different annotation methods impacted the agreement among annotators.

We conducted a comparison of the IoU scores among all the participants (P01-P08) as shown in Fig. 6b. Greater IoU implied more agreement and consistent among annotations from different participants. Specifically, we analyzed IoU scores for participants based on their annotation tools: tracing spine line with diameters in plant phenotyping software (Fig. 6c), polygons (Fig. 6d), and freehand (Fig. 6e). The shaded red, green, blue and cyan color indicate the intersection area, the gray color indicates the union but not intersect area. It is important to note that the switchgrass root example, being a non-regular image not directly captured by the minirhizotron camera, could not be recognized by the plant phenotyping software and, therefore, was unavailable for inclusion in Fig. 6c.

Our results showed that the labeling consistency were strongly dependent on the visibility of roots. According to P05 and P07, who had experience of in-field image collection, issues such as improper tube installations and lack of proper camera focus can lead to poor visibility of roots. The cotton root which was considered as low quality among all the participants resulted in the lowest labeling consistency. In contrast, the switchgrass and papaya image, where the root can be easily identified, achieved more agreement. In the context of annotation tools, our findings indicated that despite the varied background of participants using freehand scribbles and polygons, there was a slightly higher level of labeling consistency observed with freehand scribbles (Fig. 6e) compared to polygons (Fig. 6d). Additionally, both polygons and freehand scribbles exhibited improved consistency when compared to root tracing (Fig. 6c) by annotators with the same agronomy background. One possible explanation was that polygon annotation provided users with the ability to place multiple vertices along the root boundary, offering increased flexibility compared to the regular spine line tracing with fixed diameter to form a rectangle. Additionally, freehand scribbles provided an even greater level of personalization. Therefore, allowing users more freedom in drawing led to a more accurate representation of the targeted root structure, and more agreement even among the annotators with diverse background.

Furthermore, a significant decrease in labeling consistency was observed when comparing the annotation masks collected from all participants (Fig. 6b). To ascertain whether this decline in performance was attributable to variations in annotation tools or variations among participants, we conducted a more detailed investigation of the labeling consistency within the same annotation tools as well as between different tools. Specifically, we calculated the pairwise consistency between each pair of participants and categorized all pairwise consistencies as either "within" or "between" annotation tools. The box chart in Fig. 6f illustrates that in most tasks (Sesame, Papaya, Switchgrass), the median pairwise consistencies within annotation tools are higher than those between different tools. These disparities suggested that the mental models of annotators might be affected by the annotation tools, leading to divergent labeling behaviors and decisions. To construct reliable annotation for plant trait dataset, it is important to take a comprehensive consideration of annotators' expertise and the influence of their chosen annotation tools.

5 DISCUSSION

According to the mapping of challenges and corresponding actions discussed above, we summarized common user needs. These needs were based not only on the actions taken on current systems but also on open challenges that need resolution. Then, we derived several implications for designing an efficient and reliable annotation design to support fine-grained phenotypic traits annotation.

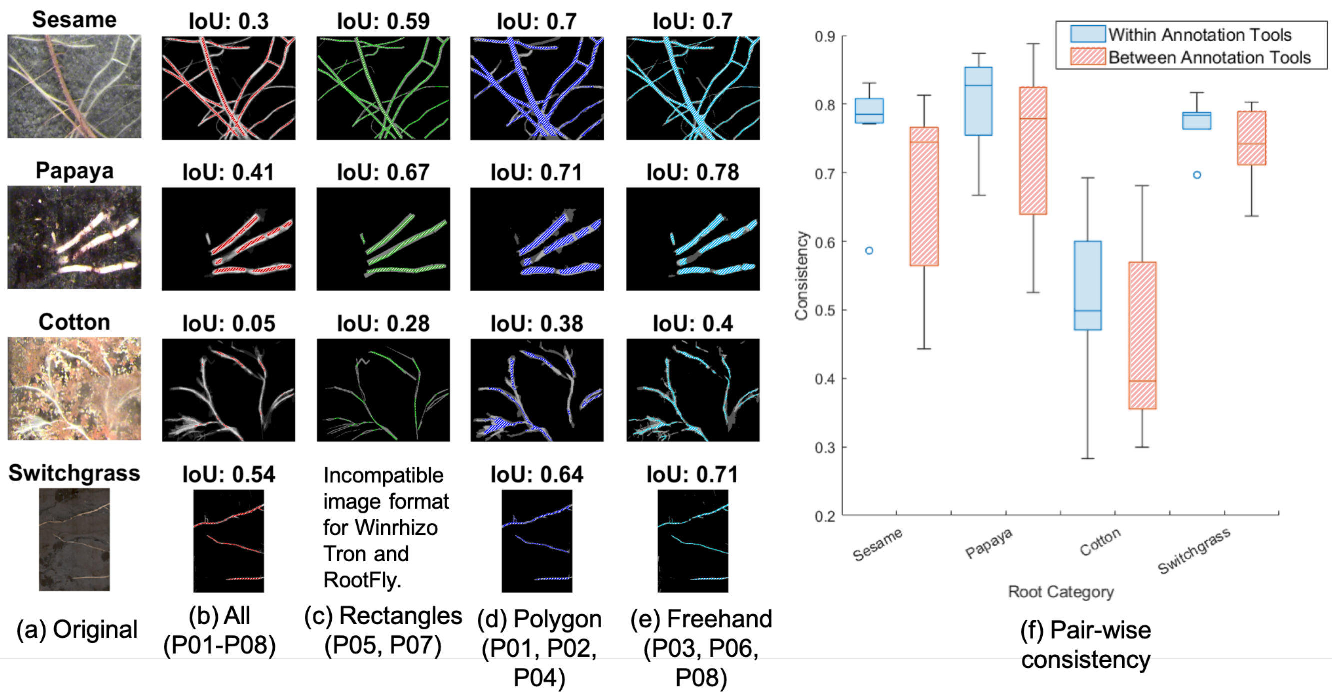

Figure 6: The consistency between masks annotated by varied participants in (a-e), and bar chart for pairwise consistency within and between annotation tools in (f).

5.1 Clear image and annotation view

Our users' need for image clarity arose from the reliance on visual cues, as distinguishing the target area in images was greatly dependent on their visual appearance. Basic functionalities like zooming in and out, adjusting brightness and contrast were valuable for examining the image. Additionally, editable display settings for the annotation components, including color and opacity, can help avoid confusion caused by mask color.

5.2 Flexible annotation tools

Users need a certain level of flexibility to create personalized shapes that accurately represent the root boundaries efficiently. Features like adjustable settings, such as brush shape and size in freehand scribbles and editable vertex points in polygon annotations, provided users with the flexibility to prevent errors, particularly in cases involving intricate and tangled root structures. Besides, semi-automatic interactive annotation with reliable root detection can enhance efficiency by saving annotation time. The incorporation of multi-modal interaction methods, including keyboard, mouse, and touchscreen inputs, offered users diverse and intuitive ways to trace targets. Therefore, providing a variety options of flexible annotation tools can significantly accommodate various preferences and ultimately facilitating more precise and efficient annotations.

5.3 Uncertainty resolution

Participants often expressed uncertainty during annotating. Their decisions were based on their experience and habits, leading to potentially divergent actions. For instance, P01, preferred labeling tiny roots to avoid missing any targets, while P05 chose to ignore extremely small roots to prevent mislabeling the background. To address annotator uncertainty, the system can offer two solutions. First, systems can facilitate adding explanatory notes when they are uncertain, such as distinguishing between root hair and fungus. These notes help annotators review and revise their annotations as needed. Second, reliable automatic detection from system can serve as guidance to address confusion. For instance, an automatic detection system can identify tiny roots, aiding annotators in making quicker decisions. These solutions can be combined for greater flexibility. For example, the system can predict all small roots, categorize them by different sizes, and let users add confidence notes for each group.

5.4 AI-assisted annotation

The AI-based feature had the potential to save annotation time. However, users expected automatic root detection to be accurate, explainable, and adaptable to different needs. Inaccurate detection risked extra effort to correct the mis-annotated regions and may cause users to opt out of the semi-automatic option. Explainable feedback helped users to easily handle the AI feature. For example, P03 received system feedback based on color, which influenced his behavior to select the most visually distinguishable bright root. Users wanted to understand the mechanism of automatic detection to make better use of the AI-assisted feature efficiently. Additionally, flexible automatic detection adaptable to various scenarios would be beneficial. For instance, adjustable parameters can help users group detected roots by color and shape, based on their research focus.

5.5 Efficient image navigation

Integrating features that provided a comprehensive view of the entire image dataset is essential. These functions, combined with seamless image navigation, empowered users to understand how targeted phenotypic traits appear across diverse backgrounds. This capability helped resolve uncertainties that may occur when working with specific images.

6 LIMITATIONS AND FUTURE WORK

This study had several limitations, including its focus on a specific type of phenotypic trait, a relatively small sample size, and inherent challenges associated with relying on self-reported interview data. We focused on annotating specific fine-grained traits found in various plant organs, such as roots, grass leaves, and stems. However, the challenges associated with other types of phenotypic traits may differ significantly. Additionally, while our participant pool included individuals from diverse backgrounds, including agronomy and computer engineering, with varying levels of education, a broader and more experienced participant pool would have provided further insights. Besides, the reliance on self-reported information about system features, particularly regarding the absence of certain features, may introduce imprecision. Although participants were encouraged to clarify whether the system genuinely lacked these features or if they were simply unaware of how to use them, the study's findings could be further strengthened by validating system features with a larger and more diverse user base. Moreover, we acknowledged the challenges of implementing solid quantitative analytics in our small-scale study with participants using diverse software with varying features. In the current study, we concentrated the quantitative analytics on a fundamental feature, the annotation tools, and primarily investigated its impact on annotation agreement. However, it is also important to identify the effect of other key features.

To address the limitations, we plan to conduct additional interviews that encompass a broader range of phenotypic traits, including those across florets and fruits. Moreover, expanding the participant pool to include more individuals will allow us to generalize our conclusions more effectively. Additionally, we intend to conduct follow-up interviews with the participants from this study to ensure the accuracy of our interpretations of the qualitative data. On the quantitative data side, a more comprehensive analysis that includes additional system features would improve our understanding of how these features are used by users and the extent to which they contribute to annotation efficiency.

7 CONCLUSION

Our interview study identified the fundamental functions and features of systems employed by experienced annotators for plant trait annotation. We identified challenges in annotating fine-grained plant traits across diverse annotator backgrounds. These challenges were categorized into difficulties related to color and difficulties related to structural complexities. We analyzed how the participants addressed the challenges with existing system features, and thereby identified common user needs and implications of system design: i) functionalities that provide a clear view for accurate assessment of image color; ii) flexible annotation tools such as personalized polygon and freehand scribbles; iii) functionalities to help manage uncertainty; iv) reliable, explainable and adaptive automatic detection mechanisms that assist users in making informed decisions efficiently; v) easy image set scanning and review capabilities. We believe that these insights will significantly contribute to the development of future robust plant trait annotation systems. Beyond software design, we are also contributing to an instance of studying user needs for an AI-based annotation system, which is a crucial step in supporting more AI-based agricultural applications.

REFERENCES

[1] Michael D Abràmoff, Paulo J Magalhães, and Sunanda J Ram. 2004. Image processing with ImageJ. *Biophotonics international* 11, 7 (2004), 36–42.
[2] Adobe Inc. 2019. *Adobe Photoshop*. Adobe Inc. https://www.adobe.com/products/photoshop.html
[3] Carl Auerbach and Louise B Silverstein. 2003. *Qualitative data: An introduction to coding and analysis*. Vol. 21. NYU press.
[4] Sumona Biswas and Shovan Barma. 2020. A large-scale optical microscopy image dataset of potato tuber for deep learning based plant cell assessment. *Scientific Data* 7, 1 (2020), 1–11.
[5] Yuri Y Boykov and M-P Jolly. 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, Vol. 1. IEEE, 105–112.
[6] Lluis Castrejon, Kaustav Kundu, Raquel Urtasun, and Sanja Fidler. 2017. Annotating object instances with a polygon-rnn. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5230–5238.
[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2015. The cityscapes dataset. In *CVPR Workshop on the Future of Datasets in Vision*, Vol. 2. sn.
[8] Abhishek Dutta, Ankush Gupta, and Andrew Zissermann. 2016. VGG image annotator (VIA). *URL: http://www. robots. ox. ac. uk/~ vgg/software/via* (2016).
[9] Abhishek Dutta and Andrew Zisserman. 2019. The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3343031.3350535
[10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 2 (2010), 303–338.
[11] Malia A Gehan, Noah Fahlgren, Arash Abbasi, Jeffrey C Berry, Steven T Callen, Leonardo Chavez, Andrew N Doust, Max J Feldman, Kerrigan B Gilbert, John G Hodge, et al. 2017. PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ* 5 (2017), e4088.
[12] Xiaolei Guo, Yogesh K Ahlawat, Tie Liu, and Alina Zare. 2022. Evaluation of Postharvest Senescence of Broccoli via Hyperspectral Imaging. *Plant Phenomics* 2022 (2022).
[13] Xiaolei Guo, Chiwah Tseung, Alina Zare, and Tie Liu. 2023. Hyperspectral image analysis for the evaluation of chilling injury in avocado fruit during cold storage. *Postharvest Biology and Technology* 206 (2023), 112548.
[14] Huimin Han, Ritvik Prabhu, Timothy Smith, Kshitiz Dhakal, Xing Wei, Song Li, and Chris North. 2022. Interactive Deep Learning for Exploratory Sorting of Plant Images by Visual Phenotypes. (2022).
[15] Karen Holtzblatt and Hugh Beyer. 1997. *Contextual design: defining customer-centered systems*. Elsevier.
[16] Suyog Dutt Jain and Kristen Grauman. 2019. Click carving: Interactive object segmentation in images and videos with point clicks. *International Journal of Computer Vision* 127, 9 (2019), 1321–1344.
[17] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643* (2023).
[18] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.
[19] Guillaume Lobet. 2017. Image analysis in plant sciences: publish then perish. *Trends in plant science* 22, 7 (2017), 559–566.
[20] Hooshang Majdi. 1996. Root sampling methods-applications and limitations of the minirhizotron technique. *Plant and soil* 185 (1996), 255–258.

[21] Sergio Marconi, Ben G Weinstein, Sheng Zou, Stephanie A Bohlman, Alina Zare, Aditya Singh, Dylan Stewart, Ira Harmon, Ashley Steinkraus, and Ethan P White. 2022. Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network. *Remote Sensing of Environment* 282 (2022), 113264.

[22] Eric N Mortensen and William A Barrett. 1995. Intelligent scissors for image composition. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 191–198.

[23] Michael P Pound, Andrew P French, Jonathan A Atkinson, Darren M Wells, Malcolm J Bennett, and Tony Pridmore. 2013. RootNav: navigating images of complex root architectures. *Plant physiology* 162, 4 (2013), 1802–1814.

[24] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[25] Inc. Regent Instruments. 2022. *WinRHIZO Tron*. Quebec, Canada. https://regentinstruments.com/assets/winrhizotron_about2.html

[26] Rubén Rellán-Álvarez, Guillaume Lobet, Heike Lindner, Pierre-Luc Pradier, Jose Sebastian, Muh-Ching Yee, Yu Geng, Charlotte Trontin, Therese LaRue, Amanda Schrager-Lavelle, et al. 2015. GLO-Roots: an imaging platform enabling multidimensional characterization of soil-grown root systems. *elife* 4 (2015), e07597.

[27] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, Vol. 23. ACM, 309–314.

[28] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision* 77, 1 (2008), 157–173.

[29] Boris Sekachev, N Manovich, and A Zhavoronkov. 2019. Computer vision annotation tool: a universal approach to data annotation. *Intel [Internet]* 1 (2019).

[30] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. 2020. PlantDoc: a dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 249–253.

[31] Abraham George Smith, Eusun Han, Jens Petersen, Niels Alvin Faircloth Olsen, Christian Giese, Miriam Athmann, Dorte Bodin Dresbøll, and Kristian Thorup-Kristensen. 2022. RootPainter: deep learning segmentation of biological images with corrective annotation. *New Phytologist* 236, 2 (2022), 774–791.

[32] Hariharan Subramonyam, Steven M Drucker, and Eytan Adar. 2019. Affinity lens: data-assisted affinity diagramming with augmented reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[33] Milan Šulc and Jiří Matas. 2017. Fine-grained recognition of plants from images. *Plant Methods* 13 (2017), 1–14.

[34] The GIMP Development Team. [n. d.]. *GIMP*. https://www.gimp.org

[35] Tao Wang, Mina Rostamza, Zhihang Song, Liangju Wang, G McNickle, Anjali S Iyer-Pascuzzi, Zhengjun Qiu, and Jian Jin. 2019. SegRoot: a high throughput segmentation method for root image analysis. *Computers and Electronics in Agriculture* 162 (2019), 845–854.

[36] Jennifer M Withington, Marc Goebel, Bartosz Bułaj, Jacek Oleksyn, Peter B Reich, and David M Eissenstat. 2021. Remarkable similarity in timing of absorptive fine-root production across 11 diverse temperate tree species in a common garden. *Frontiers in plant science* (2021), 2238.

[37] Ning Xu, Brian Price, Scott Cohen, Jimei Yang, and Thomas S Huang. 2016. Deep interactive object selection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 373–381.

[38] Weihuang Xu, Guohao Yu, Yiming Cui, Romain Gloaguen, Alina Zare, Jason Bonnette, Joel Reyes-Cabrera, Ashish B Rajurkar, Diane Rowland, Roser Matamala, et al. 2021. PRMI: A Dataset of Minirhizotron Images for Diverse Plant Root Studies. In *AI for Agriculture and Food Systems*.

[39] Weihuang Xu, Guohao Yu, Alina Zare, Brendan Zurweller, Diane L Rowland, Joel Reyes-Cabrera, Felix B Fritschi, Roser Matamala, and Thomas E Juenger. 2020. Overcoming small minirhizotron datasets using transfer learning. *Computers and Electronics in Agriculture* 175 (2020), 105466.

[40] Guohao Yu, Alina Zare, Hudanyun Sheng, Roser Matamala, Joel Reyes-Cabrera, Felix B Fritschi, and Thomas E Juenger. 2020. Root identification in minirhizotron imagery with multiple instance learning. *Machine Vision and Applications* 31, 6 (2020), 1–13.

[41] Dan Zeng, Mao Li, Ni Jiang, Yiwen Ju, Hannah Schreiber, Erin Chambers, David Letscher, Tao Ju, and Christopher N Topp. 2021. TopoRoot: a method for computing hierarchy and fine-grained traits of maize roots from 3D imaging. *Plant methods* 17 (2021), 1–17.

[42] Guang Zeng, Stanley T Birchfield, and Christina E Wells. 2008. Automatic discrimination of fine roots in minirhizotron images. *New Phytologist* 177, 2 (2008), 549–557.

[43] Sheng Zou, Paul Gader, and Alina Zare. 2019. Hyperspectral tree crown classification using the multiple instance adaptive cosine estimator. *PeerJ* 7 (2019), e6405.

[44] Sheng Zou, Hao Sun, and Alina Zare. 2017. Hyperspectral Unmixing with Endmember Variability using Semi-supervised Partial Membership Latent Dirichlet Allocation. *arXiv preprint arXiv:1703.06151* (2017).

[45] Sheng Zou, Yu-Chien Tseng, Alina Zare, Diane L Rowland, Barry L Tillman, and Seung-Chul Yoon. 2019. Peanut maturity classification using hyperspectral imagery. *biosystems engineering* 188 (2019), 165–177.

[46] Sheng Zou and Alina Zare. 2017. Hyperspectral unmixing with endmember variability using partial membership latent dirichlet allocation. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6200–6204.