

Developing a segment anything model-based framework for automated plot extraction

Han Sae Kim¹ • Ismail Olaniyi¹ • Anjin Chang² • Jinha Jung¹💿

Accepted: 23 April 2025 © The Author(s) 2025

Abstract

Purpose Automated plot extraction in agronomic research field trials is essential for highthroughput phenotyping and precision agriculture. Accurate delineation of plot boundaries enables reliable crop type classification, yield estimation, and crop health monitoring. However, traditional plot extraction methods rely heavily on manual digitization, which is time-consuming, labor-intensive, and prone to inconsistencies. This study aims to develop a Segment Anything Model (SAM)-based framework that automates plot extraction while maintaining high accuracy across diverse agricultural field conditions.

Methods The proposed framework consists of mask generation, plot orientation estimation, and plot refinement. SAM is leveraged to generate plot masks, which are subsequently filtered and refined to ensure precise boundary delineation. The method is designed to function without the need for model training or fine-tuning, making it highly adaptable across different datasets.

Results The framework was validated on five datasets, demonstrating robust performance under varying field conditions. The pixel-based evaluation yielded an average F1 score of 89.54%. For polygon-based evaluation, the framework achieved 99.71% precision at IoU=50% and an average precision of 68.51% across IoU thresholds from 50 to 95%, confirming its ability to accurately extract plot boundaries. A Canopeo-based regression analysis further demonstrated that the extracted plots provide more reliable phenotypic estimates compared to manually digitized ground reference data.

Conclusions The proposed framework significantly reduces manual effort while ensuring high precision and scalability for large-scale phenotyping applications. By relying solely on RGB imagery and zero-shot segmentation, it enhances accessibility for real-world agricultural research. Future work will focus on extending the framework to irregular plot structures, diverse crop types, and computational optimizations for large-scale implementation.

Keywords Plot boundary · Plot extraction · Segment anything model (SAM) · High-throughput phenotyping (HTP)

Extended author information available on the last page of the article

Introduction

Phenotyping plants on a field scale is crucial for understanding how genetic makeup and environmental factors affect plant growth and its development (Araus & Cairns, 2014; Fiorani & Schurr, 2013). This process involves systematically measuring and analyzing a range of plant characteristics throughout their developmental stages (Cobb et al., 2013). The complexity of plant traits necessitates a great deal of field measurements for comprehensive analysis (Furbank & Tester, 2011). In this manner, high-throughput phenotyping (HTP) has become a critical area of exploration in precision agriculture. HTP in a precise and noninvasive way is paramount for leveraging the advancements in biotechnological fields such as DNA sequencing and gene editing (Furbank & Tester, 2011; Fahlgren et al., 2015) within plant breeding (Li et al., 2020).

Many studies have demonstrated the importance of plot-level analysis for HTP (Herrero-Huerta et al., 2020), plant lodging (Kumar et al., 2024), yield prediction (Fieuzal et al., 2017; Aghighi et al., 2018; Kuwata & Shibasaki, 2016; You et al., 2017), and biomass estimation (Masjedi et al., 2019) as a precursor for agronomic research. Each field trial is often represented by research plots with different treatments or genotypes (Sankaran et al., 2015). Since a wide range of biophysical properties can be associated through correlations with phenotypic information, and phenotypic features are often summarized at the plot level (Lelong et al., 2008), a critical preliminary step in agronomic research involves extracting plot-specific data. Traditional methods to collect the data at the plot level depend on manual digitization of plot boundaries, which is often laborious and impractical. This dependency on manual intervention constrains the scalability and efficiency of the phenotyping process, highlighting the need for more automated solutions to enhance the potential of HTP in agricultural research.

Unoccupied Aerial Vehicle (UAV) offers a viable solution for regular and extensive data collection at plot level by utilizing photogrammetry and remote sensing techniques (Zhang & Kovacs, 2012; Sankaran et al., 2015; Feng et al., 2021). It has demonstrated considerable potential for applications in the agricultural area due to its capability to promptly gather data across expansive areas (Zhang et al., 2021; Fahlgren et al., 2015), leading to a substantial transform in facilitating a more efficient and accurate assessment of plant traits under varying environmental conditions (Li et al., 2014). These technological advancements are attributed to the development of more refined sensors, improved data collection platforms, and advanced data analysis techniques (Weiss et al., 2020; Araus & Cairns, 2014).

Consequently, many studies have utilized UAV remote sensing data to advance agricultural research. In the task of extracting plots for agronomic research field trials, RGB imagery and LiDAR point clouds have been pivotal as they provide clear visual boundaries and height values. Makanza et al. (2018) assessed crop cover and canopy senescence in maize field trials, demonstrating a significant correlation between UAV imagery and canopy senescence. Despite the promising findings, their methodology relies on manual plot digitization, which poses a limitation to fully automated HTP capabilities. Khan and Miklavcic (2019) proposed an image-based optimization algorithm to find the alignment of plots. They proposed the energy function to locate the optimal position of plots and leveraged particle swarm optimization to specify the location of separated field plots robustly and accurately. Yang et al. (2021) proposed a Comb Function Optimization Plot Extraction (COPE) to locate the rows and columns of plots that adhere to a strict grid-like layout. Recent advancements in deep learning-based segmentation models have significantly improved the accuracy of agricultural applications. For instance, Elmessery et al. (2024) employed SegFormer to perform semantic segmentation for microbial alterations in strawberry plants, demonstrating the effectiveness of transformer-based models in agricultural image analysis. Ribera et al. (2017) proposed a convolutional neural network to identify and count each plant from imagery.

On the other hand, Sofonia et al. (2019) examined both the UAV LiDAR and photogrammetry system in monitoring sugarcane growth, suggesting that the UAV LiDAR can provide more consistent information in estimating the biophysical parameters. Lin and Habib (2021) used a UAV LiDAR point cloud to identify rows and columns of the plot automatically. The algorithm is based on the principle that areas with higher point densities or elevations typically indicate the presence of vegetation. Its effectiveness has been proven across different agricultural conditions, including various crops, growth stages, and planting densities.

Although various methods have demonstrated their capability for accurate plot extraction, certain research gaps persist, particularly when dealing with non-ideal field conditions. Their performance can be influenced by multiple factors such as the crop's growth stage, background interference, plot size, and the grid's pattern or orientation. To mitigate such influences, it is crucial to enhance the generalization capabilities of the model or algorithm. Segment Anything Model (SAM), created by the Meta AI research team, has showcased remarkable generalization capabilities in image segmentation tasks, particularly highlighting its exceptional zero-shot performance. Introducing a novel promptable segmentation, the model was designed for zero-shot transfer (Kirillov et al., 2023). Meanwhile, there have been attempts to deploy SAM in the remote sensing domain (Osco et al., 2023; Gui et al., 2024), there are only a few studies conducted in agricultural applications (Chen et al., 2024). Considering the generalization capability and its competitive performance to fully supervised state-of-the-art models, SAM can effectively mitigate the influence of various factors in plot extraction scenarios.

This paper presents a framework for automated plot extraction by integrating SAM to address the challenges under diverse and complex field conditions. By taking advantage of the zero-shot capabilities of SAM, the framework can better accommodate irregular grid patterns, variations in crop growth stages, and different crop types while minimizing the impact of background interference. Furthermore, the proposed framework solely relies on RGB images and does not necessitate any training process. This feature can offer significant advantages: (i) it can be implemented using only UAV RGB camera systems with no multi-spectral or LiDAR sensors, and (ii) the performance of the model is not contingent on specific training data, ensuring more consistent results across different environments. This characteristic enhances the framework's accessibility and cost-effectiveness, making it a practical solution for widespread agricultural applications. In the segmentation stage, a preprocessed orthomosaic UAV RGB image is utilized to generate masks using SAM. It then estimates the orientation of the field trials to appropriately rotate the image orthogonally, enhancing the quality of the segmentation results. The segmented masks are subsequently converted into polygons, which undergo a series of refining processes. Finally, the refined plots are projected onto the corresponding coordinate systems.

The objectives of this study are to (i) explore the feasibility of SAM, (ii) establish a framework for automated plot extraction that incorporates SAM, and (iii) evaluate the performance of the proposed framework across various plot extraction scenarios. The expected outcomes of this study aim to diminish the need for manual labor in precise plot extraction and offer a more robust and versatile solution for plot extraction in precision agriculture, thereby ultimately contributing to improving high-throughput phenotyping and precision agriculture practices. The remainder of this paper is structured as follows. Section "Materials and methods" details the datasets and the study area used in the experiment, along with the methodology employed in the proposed framework. Section "Results and discussion" presents the results of the study and provides an in-depth discussion of the findings. Finally, Sect. "Conclusions" offers conclusions drawn from the research work.

Materials and methods

Datasets description

The study sites are located across four distinct states within the United States. Table 1 shows the summary of the datasets, showing the five datasets that are collected across the US continent. Figure 1 illustrates the locations where the UAV imagery was captured over wheat crop field trials, alongside the orthophotos generated using photogrammetry software. The selection of these sites was based on their distinct characteristics, which have the potential to influence plot extraction performance. Key factors considered include plot dimensions, background variations, grid patterns, and growth stages at the time of data collection. As shown in Fig. 1, the planting orientation of Datasets W-I and W-IV aligns closely with the cardinal directions, whereas Datasets W-II and W-III exhibit a slight tilt, and Dataset W-V shows moderate tiltness. The datasets also feature variations in plot dimensions, reflecting the diversity of field trial designs. The regions of interest (ROI) were selected to focus on specific treatment groups, with row and column numbering customized to align with each study's design. The number of plots within each dataset's ROI was substantial, with counts of 569, 500, 803, 656, and 260 for Datasets W-I, W-II, W-III, W-IV, and W-V, respectively. Additionally, differences in crop growth stages across datasets contribute to variations in plot appearance and grid patterns. These factors underscore the complexity of accurately capturing and analyzing agricultural data across diverse field trials. A robust and adaptable plot extraction approach is essential to accommodate these variations and ensure reliable performance under different field conditions.

Dataset	Crop type	Location	Collection date	Day after planting	Reso- lution (cm)	Plot dimension $(w \times h)$	Num- ber of plots
W-I	Wheat	Yolo County, California	05/09/2022	159	1.4	1.2 m×3.3 m	569
W-II		Elder County, Utah	06/28/2022	238	1.1	1.2 m×3.25 m	500
W-III		Ellis County, Kansas	05/19/2022	226	1.1	2.43 m×1.2 m	803
W-IV		Potter County, Texas	05/11/2022	204	0.8	$3 \text{ m} \times 1 \text{ m}$	656
W-V		Tompkins County, New York	05/11/2022	210	1.8	3.7 m×1 m	260

Table 1 Summary of datasets used in this study

The table includes details on crop type, location, collection date, days after planting, image resolution, plot dimensions, and the number of plots for each dataset



Fig. 1 Locations of the study sites and corresponding orthophotos used in this study. The UAV imagery was captured over wheat crop field trials across five distinct datasets, highlighting variations in plot dimensions, orientations, and grid patterns

Computational environment

All experiments were performed on a machine with the following hardware and software configurations: a 13 th Gen Intel(R) Core(TM) i9-13900 K CPU @ 5.80 GHz with 24 cores (Intel Corporation, Santa Clara, CA, USA), an NVIDIA GeForce RTX 4090 GPU with 24 GB VRAM running CUDA Version 12.2 (NVIDIA Corporation, Santa Clara, CA, USA), and 128 GB of DDR4 RAM. The system operated on Ubuntu 22.04.3 LTS with Kernel Version 6.8.0–40-generic. The experiments were conducted in a Python 3.9.18 environment with PyTorch 2.4.0, leveraging CUDA 12.4 for GPU-accelerated computations.

Methodology for plot extraction

The proposed framework consists of mask generation, plot orientation estimation, and plot refinement tasks. A flowchart of the proposed plot extraction framework is shown in Fig. 2. Section "SAM segmentation and mask filtering" details the SAM segmentation process and the subsequent filtering of the produced masks. Section "Plot orientation estimation and the image rotation" explains the methodology for estimating plot orientation and performing image rotation to enhance segmentation accuracy. Section "Plot generation and refinement" outlines the techniques for plot generation and refinement employed in this study. Finally, Sect. "Experiments description" presents the evaluation methods and metrics used to assess the framework's performance.

SAM segmentation and mask filtering

A key distinction between plot extraction and boundary delineation is that accurately delineating the exact perimeters of a plot's canopy may not be essential for quantifying phenotypic traits (Khan & Miklavcic, 2019). Instead, phenotypic assessments can be more effectively conducted by calculating canopy coverage within a predefined constant plot dimension. In this context, the mechanized research field trials are presumed to be systematically organized with predetermined specifications, such as the number of rows and columns and the plot's dimensions. Assuming that this information has already been identified, the hyperparameters of the SAM automatic mask generator can be configured for each specific plot extraction scenario. Several hyperparameters can be configured, namely points per side, predicted IoU threshold, and stability score threshold. The mask generation process is controlled by several hyperparameters, including points per side, predicted IoU threshold, and stability score threshold. It begins by establishing a regular grid that serves as a prompt for zero-shot transfer. This grid is defined by the 'points per side' parameter. Increasing this



Fig. 2 Flowchart of the proposed plot extraction framework. The methodology consists of three main steps: mask generation, plot orientation estimation, and plot refinement

parameter can yield masks at a finer scale but comes at the cost of greater computational expense. Once the mask is generated, a filtering process follows, which involves evaluating the predicted IoU and stability scores. By setting thresholds for these parameters, only masks that are confident and stable are retained as final segmented masks. It is anticipated that increasing the number of points per side will enhance the accuracy of the segmentation results; however, this improvement comes at the expense of increased computational costs. The size of the image array is adjusted to streamline processing efficiency while maintaining the quality of the segmentation results. The model architecture remained unaltered, and the standard SAM was employed to infer segmentation masks using the pre-trained ViT-H model. Unlike conventional deep learning models that require dataset-specific training and fine-tuning, SAM operates in a zero-shot manner, meaning it can generalize to new images without additional training. By leveraging its pre-trained capabilities, SAM enables robust segmentation across diverse field conditions, eliminating the need to collect training samples or manually annotate large-scale datasets. This makes it particularly suitable for agricultural field trials, where variability in plot structure, crop type, and environmental conditions often makes manual labeling impractical. Comprehensive information on the model architecture along with experimental setup is available in the relevant documentation (Kirillov et al., 2023). Following mask generation, the initial output often includes unwanted masks that do not correspond to actual plots. At the mask filtering stage, these masks can be filtered out based on the dimensions of the plots. This step allows us to identify approximately the initial plot candidates. The sample outputs of the SAM segmentation and mask filtering stage are presented in Fig. 3.

Plot orientation estimation and the image rotation

Since the plot generation is based on the centroid of the mask, width, and height of the plot cell, planting orientation should be estimated to get accurate boundaries. Furthermore, the performance of segmentation can decline when the field trials are not aligned with a strict grid pattern due to the nature of SAM's prompt segmentation. To tackle this challenge, this section outlines a method for estimating the plot orientation of the entire research field trials. Figure 4 illustrates an example of the plot orientation estimation and the image rotation process. The method begins with edge detection applied to the initial mask generated during the segmentation stage. These detected edges are then processed using the Hough Trans-



Fig. 3 Sample segmentation results produced by the SAM model. (Left) Original RGB image, (Center) Random color-coded output masks, and (Right) Filtered masks retained based on plot dimensions (Color figure online)



Fig. 4 Illustration of the plot orientation estimation and image rotation process. a RGB image, b Segmented masks, c Filtered masks, d Detected edges, e Extracted lines based on the Hough Transform, and f Rotated image after alignment correction

form to extract dominant lines, which are identified based on their frequency of occurrence. The dominant lines identified based on the highest frequency of occurrence are utilized to estimate the plot orientation. Finally, the image is rotated perpendicularly to the x- and y-axes according to the estimated orientation. This iterative process continues until the rotation angle converges to 0 or 90 degrees, ensuring proper alignment of the field trials for optimized segmentation and plot extraction

Plot generation and refinement

The plot refinement stage consists of polygon generation, grid filling, and grid removal. Masks filtered based on plot dimensions during segmentation and mask filtering serve as the foundation for defining plot entities. The centroid of these masks is used to generate a fixed-size plot polygon. Initially, the centroid of each plot is determined, followed by the creation of plot polygons, defined by the predetermined width and height of the plot. An illustration of the plot generation process is shown in Fig. 5.

To increase the overall accuracy of plot extraction results, the refinement process is followed by conducting grid filling and grid removal. In this process, with the initial plots generated, the horizontal and vertical spacings between plots are approximated by analyzing the spacing between adjacent plot centroids, with the most commonly occurring distances adopted as the nominal horizontal and vertical spacings. This is achieved by organizing the plot centroids in both ascending and descending orders along the x and y axes, respectively. Subsequently, a regular grid is established using the number of rows and columns, vertical and horizontal spacings. A graphical illustration of the plot refinement step is depicted in Fig. 6. Refinement is conducted by calculating the distances between the centroids of nearest neighbor plots (d_1 in Fig. 6) and the grid points (d_2 in Fig. 6).



Fig. 5 Illustration of the plot generation process. (Left) Mask array obtained from segmentation, (Center) Extracted plot boundaries and centroids, and (Right) Final plot polygons with fixed dimensions



Fig. 6 Plot refinement process using grid-based distance calculations. The figure demonstrates how spacing between plots and regular grid structures is used to refine plot positioning

Grid filling occurs when the distance d_1 , the distance from a given regular grid point to the nearest plot centroid, exceeds a predefined threshold. These grid points then serve as"point prompts"that act as the foreground in the SAM segmentation process. From the segmented mask, the initial step involves calculating the area of the mask. If this area falls within the range of a predefined threshold around the fixed area designated for a plot, it is considered valid for a new plot. Subsequently, a plot cell is created based on these dimensions and added to the existing collection of plots. This process ensures that only the objects that match the expected size and dimensions of plots are recognized and incorporated, enhancing the accuracy of the grid filling process. Figure 7a, b, and c illustrates the grid filling process, showcasing how each step contributes to grid filling during the plot refinement stage.

The subsequent phase is grid removal, which involves the deletion of duplicated plots and the commission error. For each plot centroid, if the distance d_2 is smaller than a specified threshold, grid removal is executed. This process is designed to eliminate redundancies by identifying and removing plots that are too close together, as they are considered duplicates. Furthermore, this process can reduce the commission error, removing incorrectly identified objects by examining Canopeo coverage (Patrignani & Ochsner, 2015). Figure 7d, e, and f display examples of the grid removal process, illustrating how duplicated plots can be removed. If the distance d_2 , the distance between the nearest plot centroids, falls within the predefined threshold, both the plot at the centroid and the corresponding plot at distance



Fig. 7 Examples of the plot refinement process. a Initial plot centroids (red) and regular grid (yellow), b Initial plot cells, c Plots after grid filling, d Duplicated plot centroids (random color-coded), e Duplicated plot cells, f Plots after grid removal (Color figure online)

 d_2 are flagged as duplicates and subject to removal. These flagged plots can be removed based on Canopeo coverage, to determine which plot has been extracted more accurately. The underlying rationale for this approach is that Canopeo serves as a representative metric for various vegetation indices and exhibits a strong correlation with a wide spectrum of phenotypic traits. The equation to extract a canopy cover is defined as follows:

$$Canopeo = (R/G < T_1) \land (B/G < T_2) \land (2 \times G - R - B > T_3)$$

$$(1)$$

where R, G, and B are the red, green, and blue bands, and Ts are the threshold. As presented in Patrignani and Ochsner (2015), threshold values were set to default values of 0.95, 0.95, and 20, respectively. Figure 8 shows an example of extracted Canopeo using red, green, and blue bands.

Upon extracting the Canopeo data, the Canopeo coverage rate can be determined by calculating the ratio of Canopeo pixels to the total pixels within each plot. The Canopeo coverages for duplicate plot candidates are computed. Among plots that are deemed duplicates of each other, only the plot exhibiting the highest Canopeo coverage is retained. After the plot refinement, the coordinate system of the final outcome is transformed into map coordinates, resulting in the generation of the final plot boundaries.

The pseudocode for the plot refinement is presented in Algorithm 1. Before initiating plot refinement, two *k*-d trees are constructed to accelerate the *k*-nearest neighbor search: one for the regular grid and plot centroids and another for distances among plot centroids. The ini-

tial phase, grid filling, aims to identify and rectify omissions from the segmentation phase. Subsequently, the second phase, grid removal, aims to identify and eliminate commission errors that may arise from both the segmentation and grid-filling phases.

1:	Definitions:
2:	R: regular grid points
3:	A_p : fixed area of the plot
4:	A_m : area of the mask
5:	C: Canopeo coverage
6:	T : thresholds (e.g., T_{d_1}, T_A, T_{d_2})
7:	N: refined plot set
8:	
9:	Grid Filling
10:	for $P_i \in R$ do
11:	Find the nearest plot centroid
12:	Compute d_1
13:	$\mathbf{if} \ d_1 > T_{d_1} \ \mathbf{then}$
14:	Point prompt segmentation using P_i
15:	Compute A_i and update d_1
16:	end if
17:	$ {\bf if} \ A_p - T_A < A_i < A_p + T_A \ {\bf and} \ d_1 < T_{d_1} \ {\bf then} \\$
18:	Add new plot to N
19:	end if
20:	end for
21:	
22:	Grid removal
23:	for $P_i \in N$ do
24:	Find the k nearest plot centroids
25:	Compute d_2
26:	$\mathbf{if} d_2 < T_{d_2} \mathbf{then}$
27:	Compute C_k
28:	else
29:	$\mathbf{if} C_k < \max(C_i) \mathbf{then}$
30:	Delete plot centroid P_i from N
31:	end if
32:	end if
33:	end for

Algorithm 1 Pseudocode for the plot refinement

Experiments description

All experiments were conducted using consistent parameter settings, with image resizing as the only variable. The SAM parameters were defined as follows: points per side = 100, stability score threshold = 0.92, and prediction IoU threshold = 0.86. Image dimensions were adjusted to 1024×2048 or 2048×1024 , depending on the field's aspect ratio. The distance thresholds T_{d_1}, T_{d_2} and T_A were uniformly set to 1. Table 2 shows parameter settings used in the experiments. The computational efficiency of SAM segmentation was highly dependent on the selected resizing ratio, which directly influenced segmentation accuracy.



Fig. 8 Canopeo-based canopy coverage extraction. (Left) RGB image, (Right) Extracted Canopeo coverage. Green pixels indicate detected canopy, while yellow pixels represent non-canopy areas (Color figure online)

The default ViT-H model was used as the pre-trained checkpoint, with no transfer learning or fine-tuning applied. The inference relied exclusively on the pre-trained weights, demonstrating the generalization capability of SAM across various test sites without requiring training samples for executing the proposed framework.

Evaluation

Two approaches were used for the performance evaluation: pixel-based and polygon-based evaluation. For pixel-based evaluation, the performance of the proposed framework was measured by evaluating the generated plots on manually digitized ground-reference plots. Intersection over Union (IoU) is reported, as this metric is widely used in segmentation tasks, quantifying the extent of overlap between the predicted segmentation and the actual ground reference (Eq. 2). For both evaluation cases, precision, recall, and F-1 score are reported to quantitatively evaluate the plot extraction result (Eqs. 3 to 5). In the case of pixel-based evaluation, true positive (TP) refers to pixels that correctly overlap with the ground reference area. Conversely, a false positive (FP) denotes pixels that are predicted outside the actual ground reference, while a false negative (FN) indicates pixels within the ground reference area that were not covered by the predicted plot. Additionally, a polygonbased precision-recall evaluation was introduced to assess the accuracy of boundary delineation. This evaluation was conducted by calculating precision, recall, and F1-score for the extracted plot polygons against manually digitized ground-reference polygons. A detected plot was considered a true positive if its IoU with the corresponding ground reference polygon exceeded a specified threshold; otherwise, it was classified as a false positive. To provide a more comprehensive evaluation, precision, recall, and F1-score were also computed

Table 2	Experimental	parameter settings	for the SAM-ba	used plot extrac	tion framework
---------	--------------	--------------------	----------------	------------------	----------------

Parameters	Points per side	Stability score threshold	Prediction IoU threshold	$T_{d_1}(\mathbf{m})$	$T_{d_2}(\mathbf{m})$	$T_A(\mathbf{m})$
Value	100	0.92	0.86	1	1	1

53

across multiple IoU thresholds from 50% to 95% (in 5% increments), denoted as P@50-95. This approach enables a more nuanced assessment of segmentation accuracy by considering how well the framework performs at varying levels of overlap. This method provides an interpretable assessment of how well the proposed framework captures plot boundaries, particularly in scenarios where minor segmentation deviations may still result in functionally correct detections.

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F_1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$
(5)

Results and discussion

This section details the performance of the proposed plot extraction framework and includes a discussion of the results. For Dataset W-I and W-IV, the plot orientation was found to be approximately 0 degrees, negating the need for image rotation. In contrast, Datasets W-II, W-III, and W-V exhibited rotations of approximately 1°, 1°, and 25°, respectively. These datasets required image rotation based on the estimated plot orientation before further processing.

Performance of the plot extraction

Figure 9 displays the RGB images, Canopeo data, segmented masks, filtered masks, final plots, Canopeo, and Canopeo coverage for all datasets. The results indicate that SAM effectively segmented individual plots, as evidenced by the segmented and filtered masks. In Dataset W-I, certain background regions were indistinguishable in the RGB image, and this ambiguity persisted even in the Canopeo data. Nonetheless, segmentation performance was outstanding, with the filtered masks successfully identifying almost all the plots in the field. Similarly, Dataset W-III showed that Canopeo failed to detect the plot areas, but the segmentation results were impressively accurate. Notably, these results were achieved using only a pre-trained model, highlighting SAM's remarkable zero-shot capabilities across diverse field conditions. Furthermore, although Canopeo struggled to accurately detect plots with its default parameters, the proposed framework successfully extracted plots, demonstrating its superior performance over vegetation index-based methods for plot extraction.

Utilizing the filtered masks shown in Fig. 9c, a plot refinement process is applied to generate the final plot boundaries. Figure 10 provides a zoomed-in view of the plot extraction results across all datasets. The red boxes highlight the extracted plots, while the yellow

boxes represent manually digitized ground references, offering a comparative assessment of the framework's accuracy and precision. Overall, the plots extracted using the proposed framework closely match the ground reference, affirming that the proposed framework is capable of accurately identifying nearly all plots across different datasets. The minor misalignments observed in Dataset W-III are likely due to inherent limitations of manual digitization, where human error can become significant given the large number of plots requiring digitization. This fact also highlights that given the considerable number of plots that required digitization, the potential for human error can be significant. Despite these challenges, the consistency of results across all datasets validates the effectiveness of the proposed framework, demonstrating the framework could precisely determine individual plot locations and delineate plot boundaries under a wide range of conditions. Furthermore, the framework's robustness in handling variations in plot dimensions, orientations, and diverse field patterns underscores its utility in high-throughput phenotyping, facilitating precise and efficient agricultural research.

Table 3 presents the average processing time required to extract plots across different datasets, providing insights into the computational efficiency of the proposed framework. The results demonstrate the framework's feasibility for large-scale high-throughput phenotyping, though variations in runtime suggest that field size, plot density, and computational resources influence overall efficiency. While the framework performs efficiently under current settings, further optimizations, such as parallel processing or adaptive resolution strate-



Fig. 9 Subset of experimental results from the five datasets. The figure displays RGB images, Canopeo data, segmented masks, filtered masks, final extracted plots, and Canopeo coverage

gies, could enhance computational performance, making it more suitable for large-scale agricultural applications.

For the quantitative assessment, the metrics outlined in Sect. "Experiments description" are reported in Table 4. The accuracy results from both pixel-based and polygon-based evaluations exhibit consistency, indicating that the framework maintains stable performance across varying field conditions. The polygon-based evaluation, in particular, highlights the framework's suitability for plot extraction, as evidenced by its high accuracy. Despite Dataset W-III demonstrating a lower accuracy in the pixel-based evaluation, the polygon-based evaluation underscores the success of the plot extraction process. The observed discrepancies in alignment with the ground reference in Dataset W-III warrant further analysis. The next section examines these alignment issues in detail to better understand the factors affecting plot extraction accuracy and identify potential areas for improvement.

Regression analysis of Canopeo coverage

Despite the overall high accuracy of the proposed framework, Dataset W-III exhibited lower accuracy compared to the other datasets, primarily due to misalignment with the ground reference. Similarly, Dataset W-V showed a slightly higher Canopeo coverage in the ground reference plots compared to the estimated plots. This minor discrepancy is likely due to the tilted nature of the field, where image rotation was not perfectly aligned, introducing small deviations in segmentation accuracy. To further assess the accuracy of the plot extraction



W-IV

W-V



Fig. 10 Comparison of extracted plots and manually digitized ground reference plots. (Red) Extracted plot boundaries, (Yellow) Ground reference plots. The close alignment between the extracted and reference plots validates the accuracy of the proposed framework (Color figure online)

Table 3 Average processing time	Dataset							
(in seconds) required for plot	Metric	W-I	W-II	W-III	W-IV	W-V		
extraction across the five datasets	Processing time (sec)	82.8	85.5	100.9	84.29	48.8		

Table 4 Evaluation of the proposed plot extraction framework using pixel-based and polygon-based metrics

Dataset	Pixel-based				Polygon-based@50			Polygon-based@50-95		
	IoU (%)	Preci- sion (%)	Recall (%)	F-1 (%)	Precision (%)	Recall (%)	F-1 (%)	Precision (%)	Recall (%)	F-1 (%)
W-I	87.68	93.51	93.36	93.43	100	99.82	99.91	80.81	80.67	80.74
W-II	86.47	93.99	90.41	92.16	100	100	100	76.14	76.14	76.14
W-III	65.81	79.62	79.15	79.38	98.51	98.75	98.63	37.76	37.86	37.81
W-IV	90.19	94.84	94.84	94.84	100	100	100	85.55	85.55	85.55
W-V	78.38	90.68	85.24	87.88	100	100	100	62.31	62.31	62.31
Average	81.71	90.53	88.60	89.54	99.70	99.71	99.71	68.51	68.51	68.51

process, a regression analysis was conducted to compare the Canopeo coverage within the estimated plots and the ground reference plots. The regression results, presented in Fig. 11, demonstrate a clear correlation between the estimated and ground reference plot Canopeo coverage.

Notably, the analysis indicated that, in most cases, the Canopeo coverage within the estimated plots was consistently higher than in the ground reference plots. The histogram distribution of the differences in Canopeo coverage supports this observation, suggesting that the proposed framework may provide a more precise estimation of plot coverage than initially inferred from pixel-based evaluation metrics alone. However, for Dataset W-V, the ground reference plots exhibited slightly higher Canopeo coverage, likely due to the field's tilt affecting the rotation process, leading to minor segmentation misalignments.

This outcome is particularly significant considering the close association between phenotypic traits and vegetation indices. The framework's ability to delineate plot boundaries with greater accuracy suggests that it can provide more reliable and valuable information for assessing phenotypic characteristics. Thus, even in cases where pixel-based evaluations suggest lower accuracy, the framework's performance in accurately capturing Canopeo coverage highlights its effectiveness and potential utility in agricultural research, offering insights crucial for the accurate evaluation of plant phenotypes.

Limitations and future work

While the proposed framework has demonstrated robust performance in various field conditions, several limitations remain. First, the current approach has only been tested on fields with regular grid patterns, limiting its applicability to more complex or irregularly structured agricultural plots. Additionally, the framework has primarily been evaluated on wheat crop datasets, and its generalizability to other crop types remains unverified. Furthermore, variations in growth stages, represented by different days after planting, have not been extensively explored, leaving uncertainties regarding its adaptability to diverse phenotypic conditions.



Fig. 11 Regression analysis of Canopeo coverage. (Left) Scatter plot showing the correlation between estimated and ground reference Canopeo coverage. (Right) Histogram illustrating the difference in Canopeo coverage between extracted plots and ground reference plots

Another key limitation is the computational cost associated with SAM-based segmentation. The framework's reliance on SAM increases the demand for processing power, particularly when handling large-scale fields with high-resolution imagery. This dependency may hinder its practical application in real-time or resource-constrained environments. Future research should focus on optimizing the segmentation process to improve computational efficiency without compromising accuracy. Potential directions include reducing memory overhead, implementing adaptive resolution strategies, and integrating lightweight models tailored for agricultural plot extraction.

To address these limitations, future work will explore (i) applying the proposed framework to field trials involving various crop types to assess its versatility, (ii) extending its applicability to fields with irregular layouts, (iii) evaluating performance across different growth stages to ensure robustness over the entire phenotypic cycle, and (iv) developing computational optimizations to enhance efficiency in large-scale agricultural applications. These improvements will contribute to refining the framework's scalability and usability, ultimately advancing high-throughput phenotyping and precision agriculture.

Conclusions

This study presents a Segment Anything Model (SAM)-based automated plot extraction framework, addressing the limitations of manual digitization methods commonly used in agronomic research field trials. By leveraging zero-shot segmentation, the framework eliminates the need for model training while ensuring high accuracy and adaptability across diverse field conditions. The method was validated using five datasets, demonstrating its effectiveness in extracting plot boundaries with an average F1 score of 89.54% in pixel-based evaluation. The polygon-based evaluation further confirmed its robustness, achieving 99.71% precision at IoU=50% and 68.51% precision across IoU thresholds from 50 to 95%. Additionally, a Canopeo-based regression analysis implied that the extracted plots provide more reliable phenotypic estimates than manually digitized ground reference data, underscoring the framework's practical applicability in high-throughput phenotyping.

The key contribution of this study lies in its ability to provide a fully automated and scalable solution for plot extraction, significantly reducing the labor-intensive nature of manual methods. Unlike existing approaches, the proposed framework operates solely on RGB imagery without requiring fine-tuning, making it highly accessible and adaptable to largescale agricultural applications. Its robustness across varying field conditions demonstrates its potential for enhancing the accuracy and efficiency of agronomic research. Furthermore, the integration of Canopeo-based analysis highlights its capability to provide more precise vegetation coverage estimates, which is essential for phenotypic assessments and precision agriculture.

Despite its strong performance, some limitations remain. The framework has been primarily tested on regular grid-patterned wheat fields, and its applicability to irregular layouts and different crop types warrants further exploration. Additionally, computational costs associated with SAM segmentation may pose challenges for real-time applications in largescale field trials. Future work will focus on expanding the framework to irregular plot structures and diverse crop types, evaluating its robustness across different phenotypic growth stages, and optimizing computational efficiency for deployment in resource-constrained environments.

Funding This project was supported by the Agriculture and Food Research Initiative Competitive Grant 2022-68013-36439 (WheatCAP) from the USDA National Institute of Food and Agriculture.

Data availability The datasets generated during and/or analysed during the current study are not publicly available.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H. S., & Radiom, S. (2018). Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 oli. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12), 4563–4577. https://doi.org/10.1109/JSTARS.2018.2823361
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, 19(1), 52–61.
- Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, 126, 867–887.
- Chen, R., Lu, H., Wang, Y., Tian, Q., Zhou, C., Wang, A., Feng, Q., Gong, S., Zhao, Q., & Han, B. (2024). High-throughput uav-based rice panicle detection and genetic mapping of heading-date-related traits. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2024.1327507
- Elmessery, W. M., Maklakov, D. V., El-Messery, T. M., Baranenko, D. A., Gutiérrez, J., Shams, M. Y., El-Hafeez, T. A., Elsayed, S., Alhag, S. K., & Moghanm, F. S. (2024). Semantic segmentation of microbial alterations based on segformer. *Frontiers in Plant Science*, 15, 1352935.
- Feng, L., Chen, S., Zhang, C., Zhang, Y., & He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Computers and Electronics in Agriculture*, 182, 106033.
- Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*, 24, 93–99.
- Fieuzal, R., Marais Sicre, C., & Baup, F. (2017). Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *International Journal of Applied Earth Observation* and Geoinformation, 57, 14–23. https://doi.org/10.1016/j.jag.2016.12.011
- Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. Annual Review of Plant Biology, 64, 267–291.
- Furbank, R. T., & Tester, M. (2011). Phenomics-technologies to relieve the phenotyping bottleneck. Trends in Plant Science, 16(12), 635–644.
- Gui, B., Bhardwaj, A., & Sam, L. (2024). Evaluating the efficacy of segment anything model for delineating agriculture and urban green spaces in multiresolution aerial and spaceborne remote sensing images. *Remote Sensing*, 16(22), 414. https://doi.org/10.3390/rs16020414

- Herrero-Huerta, M., Bucksch, A., Puttonen, E., & Rainey, K. M. (2020). Canopy roughness: A new phenotypic trait to estimate aboveground biomass from unmanned aerial system. *Plant Phenomics*. https://d oi.org/10.34133/2020/6735967
- Kumar, M., Bhattacharya, B. K., Pandya, M. R., & Handique, B. K. (2024). Machine learning based plot level rice lodging assessment using multi-spectral uav remote sensing. *Computers and Electronics in Agriculture*, 219, 108754. https://doi.org/10.1016/j.compag.2024.108754
- Khan, Z., & Miklavcic, S. J. (2019). An automatic field plot extraction method from aerial orthomosaic images. *Frontiers in Plant Science*, 10, 683. https://doi.org/10.3389/fpls.2019.00683
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything (arXiv:2304.02643) [cs]
- Kuwata, K., & Shibasaki, R. (2016). Estimating corn yield in the united states with modis evi and machine learning methods. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, III-8, 131–136. https://doi.org/10.5194/isprs-annals-III-8-131-2016
- Lelong, C. C. D., Burger, P., Jubelin, G., Roux, B., Labbé, S., & Baret, F. (2008). Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. *Sensors*,8(55), 3557– 3585. https://doi.org/10.3390/s8053557
- Li, Z., Guo, R., Li, M., Chen, Y., & Li, G. (2020). A review of computer vision technologies for plant phenotyping. Computers and Electronics in Agriculture, 176, 105672.
- Lin, Y.-C., & Habib, A. (2021). Quality control and crop characterization framework for multi-temporal UAV LiDAR data over mechanized agricultural fields. *Remote Sensing of Environment*, 256, 112299. https:// doi.org/10.1016/j.rse.2021.112299
- Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. Sensors, 14(11), 20078–20111.
- Masjedi, A., Carpenter, N.R., Crawford, M.M., & Tuinstra, M.R. (2019). Prediction of sorghum biomass using uav time series data and recurrent neural networks. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 2695–2702) https://doi.org/10.1109/CVPRW .2019.00327. https://ieeexplore.ieee.org/document/9025597
- Makanza, R., Zaman-Allah, M., Cairns, J., Magorokosho, C., Tarekegne, A., Olsen, M., & Prasanna, B. (2018). High-throughput phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. *Remote Sensing*, 10(2), 330. https://doi.org/10.3390/rs10020330
- Osco, L. P., Wu, Q., De Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Marcato, J. (2023). The segment anything model (sam) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*, 124, 103540. https://doi.org/10.1016/j.jag .2023.103540
- Patrignani, A., & Ochsner, T. E. (2015). Canopeo: A powerful new tool for measuring fractional green canopy cover. Agronomy Journal, 107(6), 2312–2320. https://doi.org/10.2134/agronj15.0150
- Ribera, J., Chen, Y., Boomsma, C., & Delp, E.J. (2017). Counting plants using deep learning. In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp 1344–1348) IEEE. https ://doi.org/10.1109/GlobalSIP.2017.8309180. http://ieeexplore.ieee.org/document/8309180/ Accessed 2023-11-15
- Sankaran, S., Khot, L. R., & Carter, A. H. (2015). Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Computers and Electronics in Agriculture*, 118, 372–379.
- Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., & Knowles, N. R. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy*, 70, 112–123.
- Sofonia, J., Shendryk, Y., Phinn, S., Roelfsema, C., Kendoul, F., & Skocaj, D. (2019). Monitoring sugarcane growth response to varying nitrogen application rates: A comparison of uav slam lidar and photogrammetry. *International Journal of Applied Earth Observation and Geoinformation*, 82, 101878. https://do i.org/10.1016/j.jag.2019.05.011
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. https://doi.org/10.1016/j.rse.2019.111402
- Yang, C., Baireddy, S., Cai, E., Crawford, M., & Delp, E.J. (2021). Field-Based Plot Extraction Using UAV RGB Images. In 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (pp. 1390–1398) IEEE. https://doi.org/10.1109/ICCVW54120.2021.00160. https://ieeexplo re.ieee.org/document/9607827/ Accessed 2023-11-07
- You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep gaussian process for crop yield prediction based on remote sensing data. In *Proceedings of the AAAI Conference on Artificial Intelligence* (vol. 31)
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, 13, 693–712.

Zhang, H., Wang, L., Tian, T., & Yin, J. (2021). A review of unmanned aerial vehicle low-altitude remote sensing (uav-lars) use in agricultural monitoring in china. *Remote Sensing*. https://doi.org/10.3390/rs1 3061221

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Han Sae Kim¹ · Ismail Olaniyi¹ · Anjin Chang² · Jinha Jung¹

☑ Jinha Jung jinha@purdue.edu

> Han Sae Kim kim4012@purdue.edu

Ismail Olaniyi iolaniyi@purdue.edu

Anjin Chang achang@tnstate.edu

- ¹ Lyles School of Civil and Construction Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA
- ² Department of Agricultural Sciences and Engineering, Tennessee State University, 3500 John A. Merritt Boulevard, Nashville, TN 37209, USA