Agronomist-in-the-Loop Mask-Guided 3D Cotton Boll Phenotyping from UAV Imagery

Harshitha Manjunatha[1], Shekhar Suman Borah[2], Aryan Anand[2], Theoren Stroud[2], Lakshman Tamil[1], Andrew Price[3], H. Allen Torbert[3] and Prabha Sundaravadivel[2]

[1]The University of Texas at Dallas, Richardson, Texas, USA.
[2]Center for Robotics and Intelligent Systems, The University of Texas at Tyler, Tyler, TX, USA
[3] National Soil Dynamics Lab, USDA ARS, Auburn, Alabama, USA.

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

Cotton boll phenotyping from UAV RGB imagery is limited by the small apparent size of exposed lint, partial canopy occlusion, soil and residue confusion, and illumination variation across field acquisitions. Most UAV-based cotton phenotyping studies emphasize boll counts, canopy indices, yield maps, or plot-scale point clouds, while fewer approaches connect aerial scouting to boll-level morphology under pre- and post-defoliation conditions. This study presents a mask-guided phenotyping framework in which defoliation is treated as a visibility intervention for exposing boll-bearing structures. The framework resolves acquisition phase, detects candidate boll regions using contrast-normalized morphology and color filtering, refines lint-like regions through detector-guided mask extraction, and maps retained pixels into an image-derived 2.5D review space for proxy trait estimation. In the current dataset run, the framework indexed 1,549 UAV images, including 680 pre-defoliation and 869 post-defoliation frames, and produced 4.52 million raw candidate regions before balanced candidate analysis. Across 16,000 analyzed candidates, post-defoliation imagery showed an 8.84% higher measurement-readiness score and a 22.98% lower green-pixel fraction than pre-defoliation imagery, indicating improved lint exposure after defoliation. Proxy diameter and ellipsoid-volume summaries increased by 39.17% and 50.15%, respectively; these values are reported as image-derived proxies, not calibrated physical measurements. The framework provides a practical bridge between UAV cotton scouting and future calibrated boll-scale 3D phenotyping for precision agriculture.

Keywords
cotton boll phenotyping, UAV remote sensing, mask-guided segmentation, defoliation visibility analysis, image-derived 2.5D reconstruction, precision agriculture

Agronomist-in-the-Loop Mask-Guided 3D Cotton Boll Phenotyping from UAV Imagery

1 Introduction

Cotton yield assessment and harvest management depend on the development, exposure, and spatial distribution of bolls. UAV imagery has made it possible to observe large field areas at high temporal frequency, but most image-based boll analyses remain closer to counting than to measurement. A field can contain thousands of bright lint structures in a single frame, and those structures vary in visibility because foliage, branches, shadows, soil, and neighboring bolls obscure the organ boundary. A count can therefore be useful and still incomplete: it says little about which bolls were measurable, which were partially occluded, and whether organ-scale traits were stable enough to support agronomic decisions.

The pre- and post-defoliation setting provides an unusually informative view of this problem. Defoliation is not merely a change in image appearance; it is an agronomic intervention that removes part of the canopy and changes the visibility of the reproductive structures. In a pre-defoliation image, a boll may be detectable only as a small white region between leaves. In a post-defoliation image, the same region may become more separable from the canopy, allowing a better estimate of size and location. Treating these phases as a controlled visibility contrast makes the study more precise than a generic pre/post-split. The difficulty is that organ-scale 3D measurement is not automatically obtained from UAV imagery. Classical structure-from-motion pipelines such as COLMAP rely on stable local texture and sufficient view overlap, conditions that are not guaranteed for repeated white cotton lint (Schoenberger and Frahm 2016; Schoenberger et al. 2016). Neural view synthesis and 3D Gaussian Splatting can produce visually compelling scene representations, but photorealistic rendering alone does not prove accurate boll length or volume (Mildenhall et al. 2020; Kerbl et al. 2023). Recent foundation models for segmentation and visual geometry, including Segment Anything, SAM 2, DUSt3R, MASt3R, and VGGT, suggest useful components for masks and geometry-aware reconstruction, but they must be adapted carefully to agricultural imagery (Kirillov et al. 2023; Ravi et al. 2024; Wang S. et al. 2024; Leroy et al. 2024; Wang J. et al. 2025).

This research states the contribution as a measurement-ready pipeline rather than a solved metrology system. Candidate bolls are detected from UAV frames, refined into lint masks, projected into a proxy 3D review space, and aggregated into plot-level summaries. The system is designed to expose where measurements are reliable and where calibration or additional views are still required.

- A pre/post-defoliation formulation that treats defoliation as a visibility intervention for cotton boll phenotyping.
- A detector mask extraction stage that turns raw boll detections into measurement-ready lint candidates.
- A mask-to-3D review module that exports a highlighted boll-mask point cloud and supports visual inspection of organ-level evidence.
- Explicit proxy trait definitions for count, visibility, mask length, mask width, diameter, ellipsoid volume, extraction confidence, and plot-cell aggregation.

2 Related Work

2.1 UAV and field-based cotton phenotyping

UAV phenotyping has become central in precision agriculture because it offers field-scale coverage with repeatable image acquisition. UAV remote sensing has become a practical field-phenotyping tool because it supports repeated, high-resolution observation of crop structure over large plots (Yang et al. 2017; Xie and Yang 2020; Feng et al. 2021; Gano et al. 2024). In cotton, UAV and close-range imagery have been used for stand assessment, canopy characterization, boll detection, and yield-related traits. The closest field-scale works for this manuscript are the cotton boll point-cloud studies by Sun et al. (2020) and Xiao et al. (2024), which demonstrate that 3D reconstruction can support boll counting, spatial distribution, volume analysis, and yield estimation when capture geometry is sufficiently stable. The current research position its contribution around paired defoliation, mask-guided measurement readiness, and explicit proxy boundaries rather than claiming novelty for cotton 3D reconstruction itself.

2.2 Cotton boll detection and counting

Cotton boll counting has been studied with classical image processing, supervised object detectors, weak supervision, region-based semantic segmentation, and density-map learning (Li et al. 2016; Adke et al. 2022). Image-based plant phenotyping increasingly uses automated visual analysis to convert large image collections into quantitative crop traits (Jiang and Li 2020). The current work inherits a phase-aware detector based on contrast enhancement, multi-scale top-hat morphology, thresholding, contour filtering, and color gates. This detector is useful because it produces candidate regions without dense manual annotation, but a detector box is not a physical measurement. Image-based plant phenotyping increasingly uses automated visual analysis to convert large image collections into quantitative traits (Jiang and Li 2020). The gap addressed here is the conversion of detection evidence into masks, proxy traits, and 3D review objects. Recent cotton yield studies have also combined UAV imagery with object detection and SAM-based segmentation, but their emphasis is mainly boll segmentation or yield prediction rather than proxy 2.5D measurement-readiness review (Reddy et al. 2024; Tan et al. 2025).

2.3 3D reconstruction and geometry

Structure-from-motion and multi-view stereo remain standard tools for image-based 3D reconstruction (Schoenberger and Frahm 2016; Schoenberger et al. 2016). Learned local features and matchers, including SuperPoint and SuperGlue, improve correspondence in many settings but still require texture and viewpoint consistency (DeTone et al. 2018; Sarlin et al. 2020). Newer geometry models such as DUSt3R, MASt3R, and VGGT reduce some of the engineering burden by predicting geometric relationships more directly, but their use in dense cotton scenes must be validated rather than assumed (Wang S. et al. 2024; Leroy et al. 2024; Wang J. et al. 2025).

2.4 Segmentation foundation models

Segment Anything introduced interactive or spatially conditioned segmentation at large scale and made point-, box-, and mask-conditioned extraction a practical design pattern for downstream systems (Kirillov et al. 2023). SAM 2 extended this idea to images and video, making temporal or sequential mask propagation more accessible (Ravi et al. 2024). In this paper, the term SAM-style refers to the box-conditioned segmentation design a detector box identifies a candidate, and a mask stage isolates lint-like pixels. The current implementation does not claim to run official SAM/SAM 2 unless those models are integrated and evaluated.

2.5 Neural rendering and Gaussian Splatting

NeRF and 3D Gaussian Splatting have changed how scenes are represented for novel-view synthesis (Mildenhall et al. 2020; Kerbl et al. 2023). Cotton3DGaussians is particularly relevant because it connects Gaussian scene representation with cotton boll phenotyping, including the mapping of segmentation masks into a multi-view 3DGS representation (Jiang et al. 2025). The distinction for this manuscript is scale and evidence: the present work starts from UAV field imagery and paired defoliation, and it treats mask-to-3D review as a measurement-support layer rather than a rendering-only objective.

Field-scale cotton boll phenotyping sits at the intersection of image-based counting, segmentation, and three-dimensional reconstruction. Earlier cotton studies showed that 3D point clouds can support boll counting, spatial mapping, and organ-level characterization when the image capture geometry is strong enough: Sun et al. (2020) reconstructed cotton plot point clouds from multi-view imagery, while Xiao et al. (2024) also proposed UAV-based 3D reconstruction for in situ cotton boll characterization. These works establish that 3D cotton phenotyping is feasible, but they also make clear that reliable geometry depends on stable scale, pose, overlap, and validation. In parallel, cotton boll detection and counting have advanced through region-based segmentation, supervised or weakly supervised learning, density-map estimation, and recent SAM-assisted aerial-image workflows (Li et al. 2016; Adke et al. 2022; Tan et al. 2025). These methods are useful for finding or counting visible lint, but a detected bright region is still not the same as a measurable boll. Broader computer-vision work provides additional tools: SfM/MVS methods estimate camera pose and dense structure from overlapping images (Schoenberger and Frahm 2016; Schoenberger et al. 2016), Segment Anything and SAM 2 make box- or point-based masks practical at scale (Kirillov et al. 2023; Ravi et al. 2024), and neural scene representations such as NeRF and 3D Gaussian Splatting can produce detailed visual reconstructions (Mildenhall et al. 2020; Kerbl et al. 2023). More recent cotton-specific 3D Gaussian work connects segmentation masks with multi-view boll mapping and plant architecture analysis (Jiang et al. 2025). The gap addressed in the present study is therefore not simply cotton detection or 3D reconstruction alone. Instead, this research focuses on the missing middle step: converting pre- and post-defoliation UAV evidence into detector lint masks, measurement-readiness scores, proxy traits, proxy 2.5D review artifacts, and image-coordinate plot summaries while keeping calibrated metric 3D claims separate until scale, camera geometry, ground control, and physical boll measurements are available.

3 Dataset and Study Setting

This study uses near-nadir UAV RGB imagery acquired over cotton plots before and after defoliation. The dataset was organized into two agronomic phases: a pre-defoliation condition, in which green canopy, shadows, stems, and partially hidden lint create substantial occlusion, and a post-defoliation condition, in which leaf removal improves the visual exposure of boll-bearing structures. Rather than treating these phases as only two image folders, the study uses them as a visibility contrast: pre-defoliation imagery represents a more occluded sensing condition, whereas post-defoliation imagery represents a more exposed condition for boll-level phenotyping.The current dataset audit includes 1,549 UAV RGB images, with 680 pre-defoliation images and 869 post-defoliation images. Image-header inspection showed that both phases have a consistent resolution of 5280 × 3956 pixels, corresponding to approximately 20.88 megapixels per image. No low-resolution or high-resolution outliers were identified in either phase. This consistency is important because it allows the same preprocessing scale, tiling strategy, and candidate-detection parameters to be applied across the full dataset without introducing phase-specific resolution bias. The imagery is visually dense: a single frame may contain thousands of small bright lint-like regions distributed across cotton rows. Some of these regions correspond to exposed cotton bolls, while others may arise from partially merged lint clusters, bright soil, residue, glare, or occluded plant material. Dense instance-level annotation at this spatial scale is therefore expensive and time-consuming. For this reason, the proposed framework uses a high-confidence candidate strategy. Raw detections are retained for count auditing, but only measurement-ready candidates are carried forward for proxy trait analysis. This subset is selected using lint fraction, visibility, brightness, shape regularity, size consistency, green-canopy penalty, and phase evidence. The present implementation reports image-derived proxy traits rather than calibrated physical measurements. Length, width, diameter, and ellipsoid volume are therefore interpreted as proxy quantities until they can be validated with ground sampling distance, camera calibration, ground control points, RTK/GPS metadata, orthomosaic geometry, multi-view reconstruction, or direct physical boll measurements. Under this claim boundary, the dataset supports pre/post visibility analysis, measurement-readiness ranking, proxy 2.5D review, and image-coordinate plot summaries, while calibrated boll-scale 3D metrology remains a future validation step.

4 Method

4.1 Overview

The proposed framework treats pre- and post-defoliation UAV imagery as a visibility intervention for cotton boll phenotyping. RGB frames are first assigned to an acquisition phase and processed through a phase-aware candidate detector. Candidate boxes then guide lint-mask extraction, after which each candidate is ranked by measurement readiness using visibility, lint fraction, brightness, shape, size, green-canopy penalty, and phase evidence. Retained candidates are mapped into a proxy 2.5D review space and summarized as structured phenotyping records, including visibility, mask axes, diameter proxy, volume proxy, readiness score, phase, and plot-cell summaries. Solid arrows indicate the implemented proxy pipeline, while dashed arrows indicate calibration-dependent validation steps required before metric 3D trait claims as shown in the figure 1.

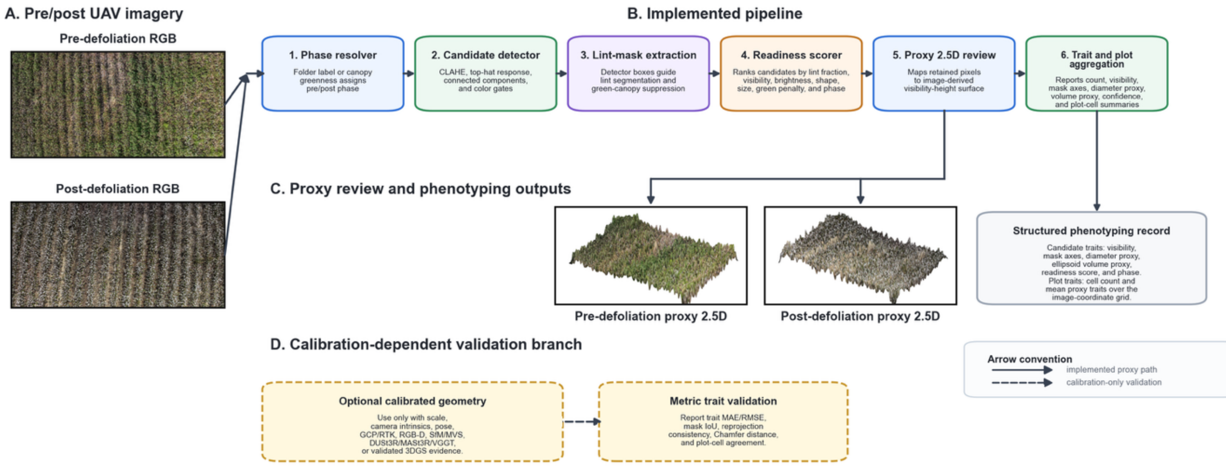

Figure1: Mask-Guided Cotton Boll Phenotyping Pipeline

4.2 Cotton boll candidate detection

The candidate detector is designed to be transparent, repeatable, and suitable for dense UAV imagery where exposed lint appears as many small bright structures within cotton rows. Each RGB frame is first resized to a stable working scale and contrast-normalized using contrast-limited adaptive histogram equalization. Multi-scale top-hat filtering is then applied to enhance bright lint-like regions against darker canopy, soil, and residue backgrounds. Otsu thresholding provides an initial binary response, from which connected components are extracted as candidate boll regions. These components are filtered using area, aspect ratio, saturation, value, luminance, and green-canopy constraints so that obvious background structures and highly vegetative regions are removed before trait analysis. when phase metadata are unavailable, a greenness heuristic is used to separate pre- and post-defoliation imagery. The detector returns candidate boxes, an adjusted image-level count, and an annotated audit image for downstream inspection.

4.3 SAM-style boll mask extraction

Each retained candidate box defines a local crop window around a potential cotton boll. Within this crop, the image is converted to HSV color space so that lint-like pixels can be separated using low saturation, high value, and reduced green response. Morphological opening and closing are applied to remove isolated noise and fill small gaps, and the dominant connected lint component is retained as the candidate mask. This step converts a coarse detector response into a pixel-level lint region that can be used for visibility scoring, mask-axis estimation, proxy trait computation, and proxy 2.5D review. The design is compatible with box-conditioned segmentation models such as SAM or SAM 2, but the current implementation should be described as detector-guided deterministic mask extraction unless official SAM/SAM 2 inference is integrated and evaluated on expert-labeled cotton examples.

4.4 Proxy 3D reconstruction and mask-to-3D review

The current implementation estimates a morphology-aware height proxy from image evidence rather than reconstructing calibrated metric geometry. The proxy surface is derived from row position, brightness, lint likelihood, canopy greenness, and local texture, and is used to provide a reviewable 2.5D representation of the scene. Pixels belonging to measurement-ready lint masks are projected into this same image-derived coordinate frame, producing a boll-mask review layer that highlights where the selected candidate evidence occurs within the field image. This representation is useful for quality control because it links the original RGB evidence, the extracted lint mask, and the proxy spatial surface in a single inspection loop. It should not be presented as calibrated 3D reconstruction. A metric version would require validated scale, camera intrinsics, camera pose, ground control, RGB-D sensing, multi-view reconstruction, or physical boll measurements. Until those data are available, the output is best described as a proxy 2.5D review artifact supporting measurement-readiness analysis and candidate selection for future calibrated reconstruction.

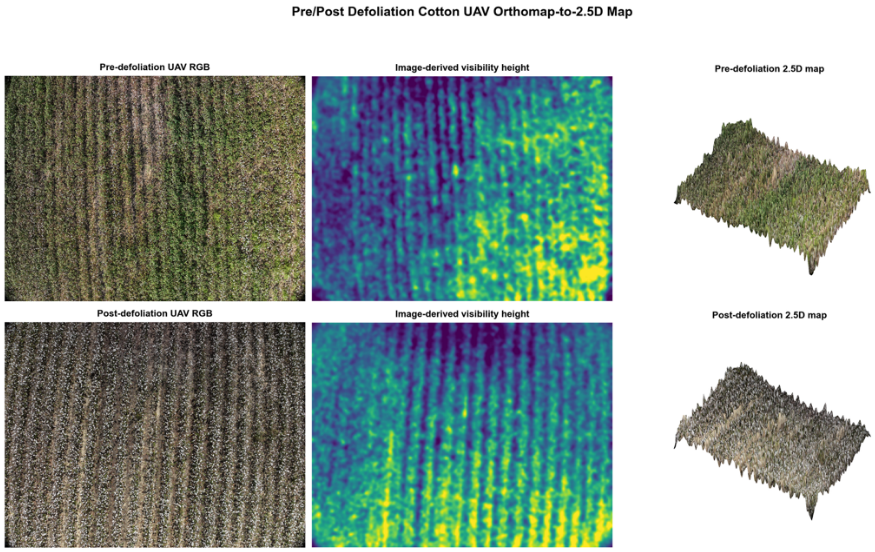

Figure 2: The left column shows real UAV RGB crops under pre- and post-defoliation conditions. The middle column shows an image-derived visibility-height proxy computed from canopy/lint appearance, row structure, and brightness cues. The right column renders the same evidence as a proxy 2.5D surface for visual review. These views are used for measurement-readiness analysis. The figure 2 provides a visual example of this proxy review step. The pre- and post-defoliation RGB crops show how defoliation changes the visibility of lint-like structures, while the corresponding visibility-height maps expose the image evidence used to form the 2.5D review surfaces. The figure is included to clarify the intermediate representation used for measurement-readiness analysis.

4.5 Trait estimation

Let $s$denote the image scale in millimeters per pixel. For candidate i, let $M_i$denote the extracted lint mask, A_i^{mask}its foreground area, and $A_i^{\text{box}}$the area of the corresponding detector box. The major and minor mask axes are denoted by $a_i$and b_i, while $w_i$and $h_i$denote the detector-box width and height in pixels. The framework reports mask length L_i, mask width W_i, coarse diameter proxy D_i, ellipsoid volume proxy U_i, visibility V_i^{vis}, measurement-readiness score q_i, and plot-cell count C_{rc}. These quantities are designed for ranking, comparison, and inspection under the current image-derived scale. They should not be interpreted as calibrated physical measurements until validated with ground sampling distance, camera geometry, ground control, multi-view reconstruction, or direct boll measurements.

The mask visibility ratio is defined as

$$V_i^{\text{vis}} = \frac{A_i^{\text{mask}}}{A_i^{\text{box}} + \epsilon} \qquad (1)$$

where ϵ prevents division by zero. The mask-derived length and width proxies are

$$L_i = s\,a_i, W_i = s\,b_i \qquad (2)$$

and the coarse diameter proxy is

$$D_i = \frac{1}{2}(L_i + W_i). \qquad (3)$$

Assuming a simple ellipsoid approximation with one major axis and two equal minor axes, the volume proxy is

$$U_i = \frac{4\pi}{3}\left(\frac{L_i}{2}\right)\left(\frac{W_i}{2}\right)\left(\frac{W_i}{2}\right). \qquad (4)$$

To penalize elongated or merged detections, the detector-box aspect ratio is computed as

$$\rho_i = \frac{\max(w_i, h_i)}{\min(w_i, h_i) + \epsilon}. \qquad (5)$$

The corresponding shape-regularity term is

$$R_i = \max\left(0{,}1 - \frac{\rho_i - 1}{3.5}\right). \qquad (6)$$

A bounded size prior is computed from the foreground mask area:

$$S_i = \min\left(\frac{\sqrt{A_i^{\text{mask}}}}{20}, 1\right). \qquad (7)$$

The measurement-readiness score combines lint fraction, visibility, brightness, shape regularity, size, green-canopy penalty, and phase evidence:

$$q_i = 0.32\ell_i + 0.20V_i^{\text{vis}} + 0.16b_i^{\text{bright}} + 0.14R_i + 0.12S_i + 0.06(1 - g_i) + 0.081[p_i = \text{post}]. \qquad (8)$$

Here, $\ell_i$is the normalized lint fraction, $b_i^{\text{bright}}$is normalized brightness, $g_i$is the green-canopy fraction, and $\mathbf{1}[p_i = \text{post}]$indicates post-defoliation imagery.
For phase-level comparison, the absolute difference in a trait $x$is

$$\Delta_\mu(x) = \mu_{\text{post}}(x) - \mu_{\text{pre}}(x), \qquad (9)$$

and the relative percent change is

$$\Delta_{\%}(x) = 100 \cdot \frac{\mu_{\text{post}}(x) - \mu_{\text{pre}}(x)}{\max(|\,\mu_{\text{pre}}(x)\,|, \epsilon)}. \qquad (10)$$

For proxy volume-mutation analysis, extreme volumes are capped at the 99th percentile before computing sorted first differences:

$$U_i^{99} = \min(U_i, \text{percentile}_{99}(U)), D_{\text{thr}} = 5\,\text{mean}\big(\text{diff}(\text{sort}(U^{99}))\big). \qquad (11)$$

Finally, each measurement-ready candidate is assigned to an image-coordinate plot cell G_{rc}, and the cell count is

$$C_{rc} = \sum_i \mathbf{1}\,[\,\text{center}_i \in G_{rc}\,]. \qquad (12)$$

The ellipsoid volume proxy in Eq. (4) is a pragmatic approximation for ranking and phase comparison, not a direct physical volume measurement. Once physical boll measurements and calibrated geometry are available, the same trait definitions can be evaluated using mean absolute error, relative volume error, correlation against measured dimensions, and plot-level agreement.

4.6 Plot-level mapping

To make dense UAV detections interpretable at the field scale, measurement-ready candidates are aggregated into an image-coordinate plot grid. In the current implementation, a 4×43grid is overlaid on the central study area, and each retained candidate is assigned to a grid cell according to its mask or detector-box center. For each occupied cell, the framework stores candidate count, mean readiness score, mean visibility, mean diameter proxy, and mean ellipsoid-volume proxy. This aggregation converts thousands of local detections into row-column summaries that can be inspected for spatial density, measurement quality, and failure regions. At this stage, the grid should be interpreted as an image-coordinate plot proxy rather than a georeferenced field map. Metric plot mapping will require plot boundaries, orthomosaic coordinates, camera pose, ground control points, or other field-registration metadata.

4.7 Agronomist-in-the-loop reporting

The agronomist-in-the-loop layer is treated as a downstream reporting component rather than part of the detection, segmentation, or reconstruction pipeline. Its role is to translate structured phenotyping records into concise agronomic summaries that can be reviewed by a domain expert. These records may include candidate counts, measurement-readiness scores, visibility summaries, proxy diameter and volume statistics, plot-cell patterns, and calibration warnings. If a language-model-based reporting layer is used, it should be evaluated for schema validity, consistency, expert agreement, latency, and unsupported agronomic claims. In this study, the reporting layer is therefore positioned as an interpretive aid for human review, not as an autonomous source of agronomic decisions.

5 Algorithm

Mask-guided cotton boll phenotyping with proxy 2.5D review

Input: UAV image I, phase label $p \in \{\mathrm{pre}, \mathrm{post}, \mathrm{unknown}\}$, scale s or calibration metadata, detector parameters Θ_d, mask parameters Θ_m, proxy-depth parameters Θ_z, plot grid G, readiness thresholdτ_q.

Output: Measurement-ready candidate set R, trait table T, proxy boll-mask point cloud P, plot-cell summaries C.

if $p$is unknown then
    infer $p$from canopy greenness
end if
$I_n \leftarrow \mathrm{CLAHE}(I)$
$R_d \leftarrow \mathrm{MultiScaleTopHat}(I_n)$
$Y_d \leftarrow \mathrm{OtsuThreshold}(R_d)$
$B_0 \leftarrow \mathrm{ConnectedComponents}(Y_d)$
$B \leftarrow \mathrm{FilterCandidates}(B_0$;area, aspect ratio, saturation, value, luminance, green-canopy gates)
$R \leftarrow \emptyset$
$T \leftarrow \emptyset$
$P \leftarrow \emptyset$
for each candidate box $b_i \in B$do
    $\Omega_i \leftarrow$ padded crop around b_i
    convert $\Omega_i$from RGB to HSV
    $M_i^0 \leftarrow$ low-saturation, high-value, low-green binary lint map
    $M_i \leftarrow$ largest connected component after morphological opening and closing
    compute lint fraction ℓ_i
    compute visibility $V_i^{\mathrm{vis}} = A_i^{\mathrm{mask}} / (A_i^{\mathrm{box}} + \epsilon)$
    compute brightness b_i^{bright}, green fraction g_i, and aspect ratio ρ_i
    compute shape regularity $R_i$and size prior S_i
    $q_i \leftarrow 0.32\ell_i + 0.20V_i^{\mathrm{vis}} + 0.16b_i^{\mathrm{bright}} + 0.14R_i + 0.12S_i + 0.06(1 - g_i) + 0.081[p = \mathrm{post}]$

if $q_i < \tau_q$ or M_i violates mask size/shape constraints then
continue
end if
estimate mask major axis a_i and minor axis b_i
$L_i \leftarrow s a_i$
$W_i \leftarrow s b_i$
$D_i \leftarrow (L_i + W_i)/2$
$U_i \leftarrow (4\pi/3)(L_i/2)(W_i/2)(W_i/2)$
$Z_i \leftarrow \mathrm{ProxyDepth}(I, M_i; \Theta_z)$
for each foreground pixel $(u, v) \in M_i$ do
$P_i(u, v) \leftarrow [s(u - u_0), s(v - v_0), Z_i(u, v)]^\top$
add $P_i(u, v)$ to P
end for
$x_i \leftarrow [\mathrm{center}_i, A_i^{\mathrm{mask}}, V_i^{\mathrm{vis}}, L_i, W_i, D_i, U_i, q_i, p]$
$R \leftarrow R \cup \{i\}$
append x_i to T
end for
for each candidate $i \in R$ do
assign center_i to image-coordinate grid cell G_{rc}
update cell count C_{rc}
update mean readiness, visibility, diameter proxy, and volume proxy for G_{rc}
end for
return R, T, P, C

6.1 Dataset-scale audit

The first audit was performed at the image-set level to quantify the amount of visual evidence processed before measurement-ready filtering as showed in table 1. Pre- and post-defoliation images were separated by acquisition phase, and the same phase-aware detector was applied across the full dataset. This audit reports both the number of images available in each phase and the detector load created by dense cotton imagery. Because each UAV frame contains many small lint-like structures, the raw candidate count is treated as detection evidence for downstream filtering rather than as a validated boll count. Across 1,549 UAV images, the detector produced millions of candidate regions. The pre-defoliation subset had a higher mean adjusted count per image, whereas the post-defoliation subset produced more raw retained candidates. This difference should be interpreted cautiously because canopy density, overlapping lint, and phase-specific detector adjustment can influence candidate counts. Manual validation is therefore required before these values are treated as ground-truth boll counts.

Table 1. Phase-level detector audit on the current UAV folders.

Phase	Images	Adjusted detector count	Raw candidate regions	Mean adjusted count per image
Pre-defoliation	680	2,914,620	1,821,639	4,286.2
Post-defoliation	869	2,698,919	2,698,919	3,105.8

6.2 Measurement-ready candidate analysis

For organ-scale analysis, the raw detections were reduced to a balanced set of high-confidence candidates from each phase as shown in table 2. The measurement-readiness score summarizes whether a candidate is bright, lint-like, compact, sufficiently visible, and minimally contaminated by green canopy. It is used as a ranking and filtering signal, not as a direct probability of correctness. The score distributions in figure 3 show a rightward shift for post-defoliation candidates, indicating that defoliation increased the proportion of candidates suitable for mask-based inspection and proxy trait estimation.

Table 2. Measurement-ready candidate statistics. Diameter and volume are image-derived proxy values under the current scale assumption.

Phase	Candidates	Readiness score	Lint fraction	Green fraction	Visibility	Diameter proxy (mm)	Volume proxy $((10^3\ \mathrm{mm}^3))$
Post-defoliation	8,000	0.636	0.491	0.058	0.459	211.35	16,105.3
Pre-defoliation	8,000	0.584	0.525	0.075	0.509	151.86	10,726.1

Table 3. Phase contrast in proxy morphology and visibility.

Metric	Pre mean	Post mean	Post − pre	Relative change
Measurement-readiness score	0.584	0.636	0.052	+8.84%
Lint fraction	0.525	0.491	−0.034	−6.49%
Green fraction	0.075	0.058	−0.017	−22.98%
Visibility	0.509	0.459	−0.049	−9.72%
Diameter proxy (mm)	151.86	211.35	59.49	+39.17%
Volume proxy $((10^3\ \mathrm{mm}^3))$	10,726.1	16,105.3	5,379.2	+50.15%

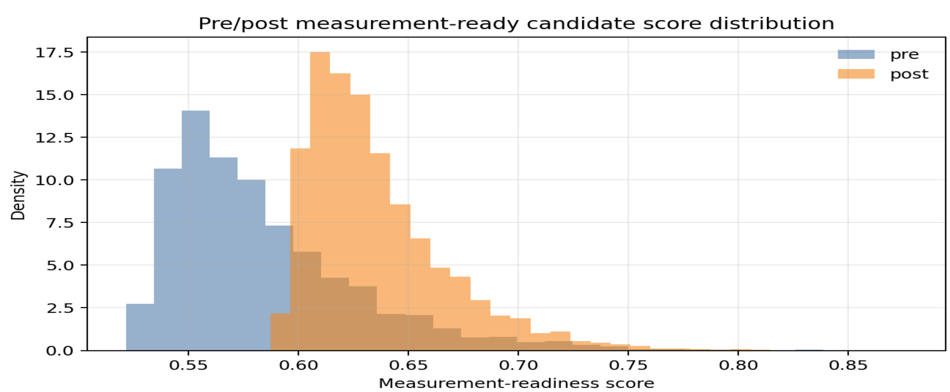

Figure 3: Measurement-readiness distribution across pre- and post-defoliation imagery. The histogram shows that post-defoliation candidates shift toward higher readiness scores, indicating improved suitability for mask-based inspection and proxy trait estimation.

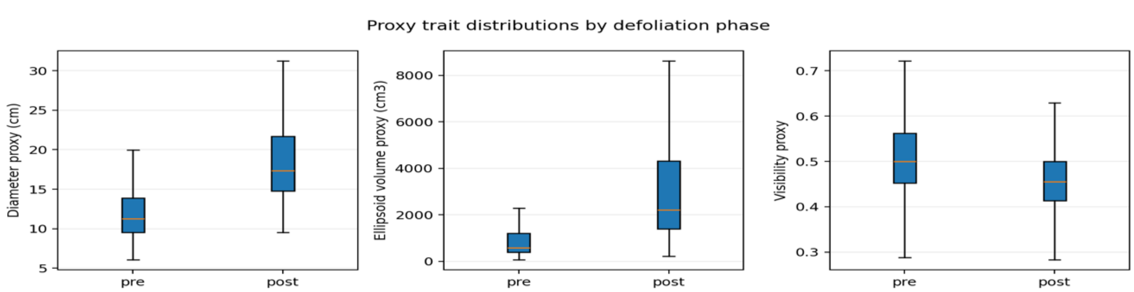

Figure 4: Proxy trait distributions for visibility, diameter, and ellipsoid volume. The boxplots summarize phase-level differences among retained candidates. Diameter and volume are image-derived proxy traits and require physical or geometric calibration before being interpreted as true boll measurements.

Post-defoliation candidates had a higher mean readiness score and a lower green fraction, supporting the interpretation of defoliation as a visibility intervention. The lower green fraction indicates reduced canopy contamination, while the higher readiness score suggests that more candidates were suitable for downstream proxy-trait analysis. The phase contrast shown in table 3 says post-defoliation imagery improved measurement readiness and reduced green-pixel contamination. The proxy trait distributions in figure 4 further show that retained post-defoliation candidates had larger diameter and ellipsoid-volume proxies than pre-defoliation candidates. These increases should be interpreted as image-derived morphology differences among retained candidates, not as calibrated physical growth or yield measurements. Because volume scales cubically with mask dimensions, it is especially sensitive to merged lint clusters, mask extent, and scale assumptions. Physical boll measurements or calibrated geometry are needed before these proxy values can be treated as true organ-scale traits.

6.3 Confidence intervals and phase interpretation

The confidence intervals in table 4 show that the measurement-readiness score is tightly estimated for the sampled candidates, while the proxy volume interval is much wider. This pattern is expected because ellipsoid volume scales cubically with mask dimensions and is sensitive to merged lint clusters, partial occlusion, and mask extent. The result supports the main interpretation of the pipeline: readiness can be used as a stable ranking signal, but proxy volume should remain a review and comparison metric until physical boll measurements or calibrated geometry are available.

Table 4. Ninety-five percent confidence intervals for selected proxy traits.

Phase	Metric	CI low	CI high
post	Measurement-readiness score	0.635	0.636
post	Visibility	0.458	0.461
post	Diameter proxy (mm)	208.22	214.48
post	Volume proxy $(10^3 mm3)$	10198.1	22012.5
pre	Measurement-readiness score	0.583	0.585
pre	Visibility	0.507	0.511
pre	Diameter proxy (mm)	148.82	154.91
pre	Volume proxy $(10^3 mm3)$	8466.4	12985.9

6.4 Ablation of measurement-readiness terms

The ablation results in table 5 and figure 5 test whether the readiness score depends on meaningful visual evidence rather than a single arbitrary term. Removing lint fraction or visibility changes candidate selection most strongly, showing that the method favors candidates that are not only bright, but also separable and measurable. The phase-bonus term affects ranking order because it encodes the visibility shift introduced by defoliation. Overall, the ablation supports the design of the score as a practical candidate-ranking function for proxy 2.5D review, not as a ground-truth accuracy measure.

Table 5. Candidate-ranking ablation. Higher Spearman correlation and top five overlap indicate closer agreement with the full score.

Variant	Mean score	Spearman	Top-5 overlap
full	0.690	1.000	1.000
minus_lint	0.528	0.901	0.684
minus_visibility	0.593	0.939	0.805
minus_brightness	0.600	0.939	0.791
minus_shape	0.565	0.950	0.833
minus_size	0.570	1.000	1.000
minus_green_penalty	0.634	0.995	0.965
minus_phase_bonus	0.650	0.531	0.516

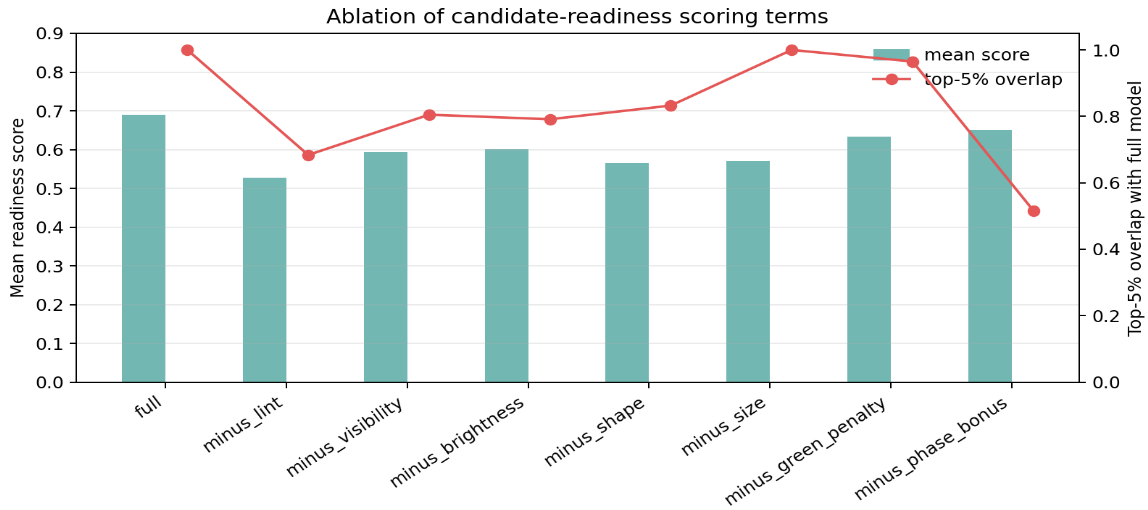

Figure 5: Ablation of the measurement-readiness score. Removing lint fraction or visibility most strongly changes candidate selection, supporting the method design: the system prioritizes candidates that are bright, compact, separable, and plausibly measurable.

6.5 Plot-grid mapping

The plot-grid analysis converts dense candidate detections into an image-coordinate spatial summary. In the current implementation, candidates are assigned to a 4×43 grid using their mask or detector-box centers, and each cell stores count and mean proxy-trait values. As shown in figure 6, this representation makes row-column patterns easier to inspect and helps identify regions with high candidate density or possible failure modes. The grid is not yet a georeferenced field map; it is a spatial accounting layer that can later be registered to field plots when orthomosaic coordinates, plot boundaries, or GCPs are available.

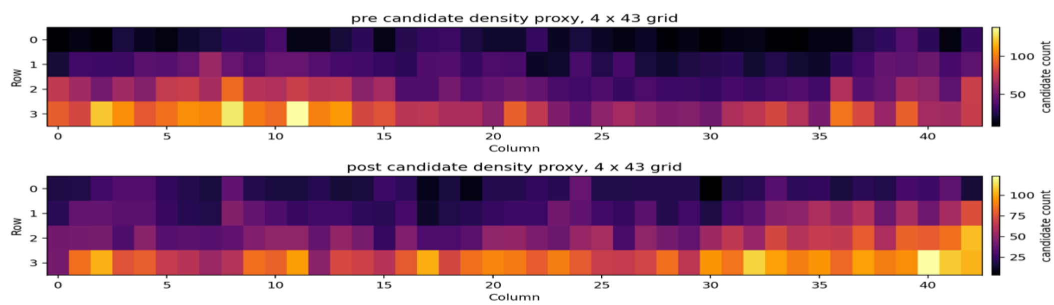

Figure 6: Image-coordinate plot-grid maps for candidate density and proxy traits. The grid summarizes dense UAV detections as row-column patterns that can later be registered to field plots when orthomosaic coordinates and plot boundaries are available.

6.6 Proxy volume mutation analysis

Proxy volume was further examined by sorting candidate volumes within each phase and inspecting large first-difference changes in the upper tail of the distribution. Before computing the mutation diagnostic, volumes were capped at the 99th percentile to reduce the influence of extreme merged components. As shown in Figure 7, most candidates occupy a stable low-volume range, while a small tail shows abrupt increases. These jumps are treated as quality-control flags for merged, adherent, or over-extended lint masks rather than as biological thresholds. The diagnostic therefore helps identify candidates that should be inspected before proxy volume is used in yield-related interpretation.

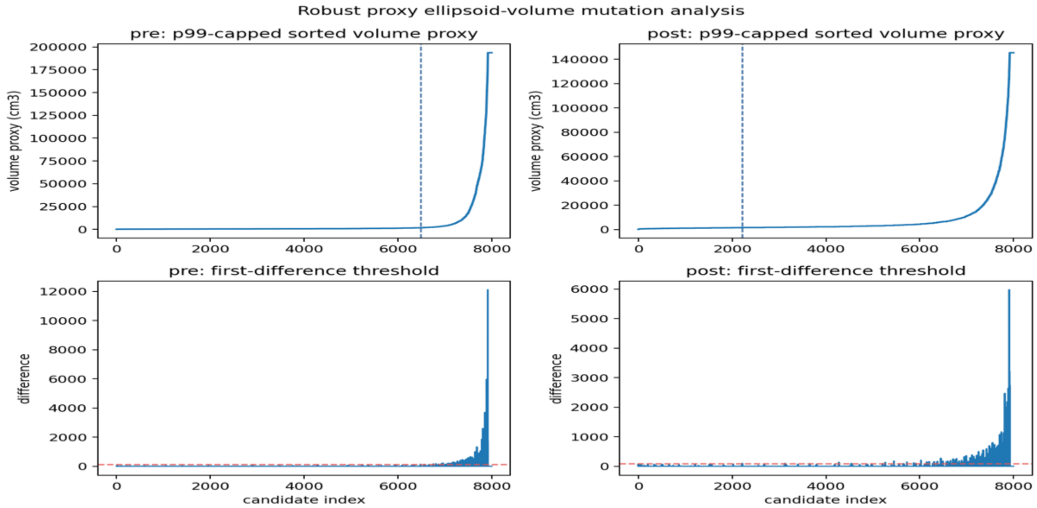

Figure 7: Sorted proxy-volume curves and first-difference diagnostics after 99th-percentile capping. Abrupt tail increases indicate candidates that may contain merged or adherent lint regions and should be reviewed before entering yield-related analysis.

6.7 Local 2.5D target selection

The local 2.5D module selects the highest-ranked candidates for visual inspection and crop-level export. These targets are useful because they identify where calibrated reconstruction, manual mask review, or multi-view follow-up should be concentrated. They are not treated as final 3D cotton boll geometry. Instead, they serve as a priority list for the next validation stage, where scale, camera geometry, or physical boll measurements can be added.

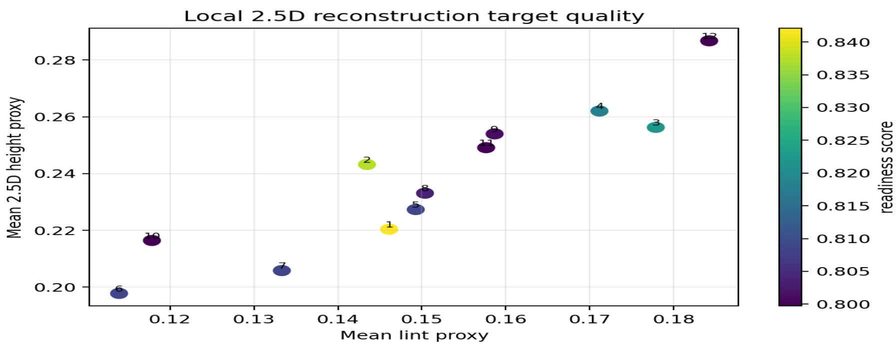

Figure 8: Local 2.5D target selection. High-readiness candidates are selected as inspection targets for future calibrated reconstruction. These points indicate where follow-up should be focused, not final metric 3D boll measurements.

6.8 Validation required before final accuracy claims

The present results should be interpreted as evidence for measurement readiness and proxy 2.5D review. The validation step is to add ground truth at three levels. First, detection should be evaluated on a stratified subset of pre- and post-defoliation images using manual boll counts, with precision, recall, F1, MAE, and RMSE reported. Second, lint-mask quality should be compared with expert or SAM-assisted masks using mask IoU, boundary F1, and mask-area error. Third, proxy diameter and volume should be checked against physical boll measurements or scale-calibrated orthomosaic/GCP information before being interpreted as metric traits. For calibrated geometry, selected high-visibility crops should be reconstructed with SfM/MVS, DUSt3R/MASt3R/VGGT, RGB-D, also validated Gaussian Splatting and evaluated using reprojection consistency, Chamfer distance, point density, and view-rendering quality. Any agronomist-in-the-loop reporting layer should be tested separately as a structured summarization aid, with attention to schema validity, expert agreement, latency, consistency, and unsupported agronomic claims.

7 Discussion and Limitations

The results support the central argument of this study: defoliation changes not only the number of visible lint-like regions, but also the measurement readiness of the detected candidates. Post-defoliation imagery produced a higher mean readiness score and lower green-canopy contamination, indicating that defoliation can be treated as a visibility intervention for organ-scale cotton boll phenotyping. This is different from reporting a simple pre/post count difference. It links an agronomic management step to the quality of the visual evidence available for measurement. The main technical contribution is the separation of detection, lint-mask extraction, proxy 2.5D review, and plot-level aggregation. A conventional counting pipeline would usually stop after candidate detection. In the proposed framework, each candidate box is refined into a lint mask, scored for measurement readiness, projected into a reviewable proxy 2.5D space, and summarized at the plot-grid level. This design makes failure modes visible before they become trait or yield claims. Merged lint clusters, unusually large proxy volumes, low-visibility masks, and green-contaminated candidates can be identified and reviewed rather than silently entering the final summaries. The most important limitation is scale. A single near-nadir UAV image or uncalibrated orthomosaic cannot, by itself, guarantee organ-scale metric 3D geometry. True physical diameter and volume require ground sampling distance, camera intrinsics, ground control, multi-view pose, RGB-D sensing, or direct boll measurements. For that reason, the present study reports diameter and volume as proxy traits and uses volume-mutation diagnostics as a quality-control tool. This conservative framing is essential because recent cotton 3D phenotyping studies show that more complete boll characterization depends on validated multi-view geometry and physical scale information (Sun et al. 2020; Xiao et al. 2024; Jiang et al. 2025).

The present system has limitations that define the boundary of its claims. First, the proxy 2.5D output should not be described as calibrated Gaussian Splatting, SfM/MVS, or metric 3D reconstruction until camera geometry, scale, and field control are validated. Second, the detector may confuse cotton lint with bright soil, residue, glare, or exposed background objects, so manual labels are still needed to estimate detection precision, recall, F1, MAE, and RMSE. Third, the ellipsoid volume proxy assumes a simplified boll shape and remains sensitive to merged or adherent lint masks. Finally, the plot-grid summaries are image-coordinate summaries rather than georeferenced field maps until orthomosaic coordinates, plot boundaries, or GCPs are added. The agronomist-in-the-loop reporting layer should therefore be treated as a structured summarization aid and should remain subject to expert review.

8 Conclusion

This paper formulates pre- and post-defoliation cotton boll phenotyping as a measurement-readiness problem for precision agriculture. The proposed framework combines phase-aware candidate detection, detector-guided lint-mask extraction, proxy 2.5D review, plot-grid aggregation, and conservative trait reporting. Across 1,549 UAV images, the method summarized 4.52 million raw candidate regions and analyzed 16,000 balanced candidates from pre- and post-defoliation imagery. Post-defoliation candidates showed higher measurement readiness and lower green-canopy contamination, supporting the interpretation of defoliation as a visibility intervention rather than only a change in image appearance. The framework is useful for precision agriculture because it converts dense UAV imagery into auditable candidate records, proxy morphology summaries, and plot-level spatial patterns that can guide scouting, inspection, and future calibrated reconstruction. At the same time, diameter and volume are reported as image-derived proxy traits, not final physical measurements. The next step is to add manual labels, calibrated scale, camera geometry, and physical boll measurements so that the same architecture can mature from proxy review into validated organ-scale 3D cotton phenotyping.

Acknowledgement

The materials are based upon work supported by the USDA-ARS NACA 58-6010-4-007. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of USDA-ARS.

References

Adke, S., Li, C., Rasheed, K. M., and Maier, F. W. (2022). Supervised and weakly supervised deep learning for segmentation and counting of cotton bolls using proximal imagery. *Sensors*, 22(10), 3688.

DeTone, D., Malisiewicz, T., and Rabinovich, A. (2018). SuperPoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Feng, L., Chen, S., Zhang, C., Zhang, Y., and 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.

Gano, B., Bhadra, S., Vilbig, J. M., Ahmed, N., Sagan, V., and Shakoor, N. (2024). Drone-based imaging sensors, techniques, and applications in plant phenotyping for crop breeding: A comprehensive review. *The Plant Phenome Journal*, 7(1), e20100.

Jiang, L., Sun, J., Chee, P. W., Li, C., and Fu, L. (2025). Cotton3DGaussians: Multiview 3D Gaussian Splatting for boll mapping and plant architecture analysis. *Computers and Electronics in Agriculture*, 234, 110293.

Jiang, Y., and Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics*, 2020, 4152816.

Kerbl, B., Kopanas, G., Leimkuehler, T., and Drettakis, G. (2023). 3D Gaussian Splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 1-14.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Leroy, V., Cabon, Y., and Revaud, J. (2024). Grounding image matching in 3D with MASt3R. In *Proceedings of the European Conference on Computer Vision*, 71-91.

Li, Y., Cao, Z., Lu, H., Xiao, Y., Zhu, Y., and Cremers, A. B. (2016). In-field cotton detection via region-based semantic image segmentation. *Computers and Electronics in Agriculture*, 127, 475-486.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision*.

Ravi, N., Gabeur, V., Hu, Y. T., Hu, R., Ryali, C., Ma, T., et al. (2024). SAM 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714*.

Reddy, J., Yu, J., Thomasson, J. A., Huang, Y., and Rains, G. C. (2024). Cotton yield prediction via UAV-based cotton boll image segmentation using YOLO model and Segment Anything Model (SAM). *Remote Sensing*, 16(23), 4346.

Sarlin, P. E., DeTone, D., Malisiewicz, T., and Rabinovich, A. (2020). SuperGlue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Schoenberger, J. L., and Frahm, J. M. (2016). Structure-from-Motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Schoenberger, J. L., Zheng, E., Frahm, J. M., and Pollefeys, M. (2016). Pixelwise view selection for unstructured multi-view stereo. In *Proceedings of the European Conference on Computer Vision*.

Sun, S., Li, C., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., et al. (2020). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 195-207.

Tan, C., Sun, J., Song, H., and Li, C. (2025). A customized density map model and Segment Anything model for cotton boll number, size, and yield prediction in aerial images. *Computers and Electronics in Agriculture*, 232, 110065.

Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., and Novotny, D. (2025). VGGT: Visual Geometry Grounded Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., and Revaud, J. (2024). DUSt3R: Geometric 3D vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Xiao, S., Fei, S., Ye, Y., Xu, D., Xie, Z., Bi, K., et al. (2024). 3D reconstruction and characterization of cotton bolls in situ based on UAV technology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 209, 101-116.

Xie, C., and Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Computers and Electronics in Agriculture*, 178, 105731.

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Frontiers in Plant Science*, 8, 1111.