Drone-based 3D Reconstruction of Plants in Field Conditions Using Neural Radiance Fields (NeRFs)

Shambhavi Joshi, Ashlyn Rairdin, Elizabeth Tranel, Talukder Jubery, Soumik Sarkar, Arti Singh, Baskar Ganapathysubramanian, Adarsh Krishnamurthy *

Iowa State University, Ames, IA 50011 USA
adarsh@iastate.edu

Abstract

The performance of 3D reconstruction using Neural Radiance Fields (NeRFs) for outdoor phenotyping of plants is strongly influenced by the imaging modality used for data collection. We compare drone, handheld, and 360° ground robot datasets collected over soybean and mungbean plots, and evaluate reconstruction quality using 2D metrics PSNR, SSIM, LPIPS, and 3D geometric metrics precision, recall, and F1 score. Drone imagery produced the highest geometric fidelity, handheld captures achieved the strongest 2D appearance quality, and the 360° captures lagged in both metrics due to spherical distortion and motion artifacts. The consistency of the drone-based reconstructions highlights its suitability for field-scale 3D modeling and positions it as a practical foundation for future phenotyping applications.

Code — https://idealab-isu.github.io/Drone-based-3D-Reconstruction-of-Plants-AAAI26/

Datasets —
https://huggingface.co/datasets/ShambhaviJoshi/Drone_based_3D_Reconstruction_of_Plants_AAAI26

Introduction

Three-dimensional (3D) reconstruction of plant structure has become essential for precision agriculture and plant phenotyping. Quantifying morphological traits such as leaf area, branching architecture, and canopy volume requires capturing the spatial complexity of plants, information that conventional two-dimensional imaging cannot adequately provide (Fiorani and Schurr 2013). Traditional 3D methods, including manual measurements and stereo-based photogrammetry, are labor-intensive and often inadequate for high-throughput phenotyping workflows (Li et al. 2020, 2025).

Recent advances in neural scene representation have introduced new possibilities for dense point cloud 3D reconstruction. Neural Radiance Fields (NeRF) (Mildenhall et al. 2021), in particular, model volumetric radiance as a continuous function of spatial position and viewing direction, enabling both photorealistic view synthesis and detailed geometry recovery from multi-view images. Unlike traditional photogrammetric pipelines that rely on discrete feature matching, NeRF captures subtle variations in reflectance and depth through learned volumetric representations, offering potential advantages for reconstructing complex plant structures (Remondino et al. 2023).

Despite its success in controlled indoor environments and computer graphics applications, applying NeRF to outdoor agricultural settings presents distinct challenges (Hu et al. 2024; Arshad et al. 2024). Field environments introduce dynamic illumination, wind-induced plant motion, and complex occlusions from overlapping foliage. Equally important, the choice of imaging device, including sensor characteristics, lens properties, and capture trajectory, fundamentally affects reconstruction quality (Choi, Park, and Lee 2024; Li et al. 2025). Understanding how different imaging platforms influence NeRF performance under field conditions is essential to develop practical, scalable reconstruction pipelines for agricultural research. Yet systematic comparisons of device-specific reconstruction fidelity in outdoor settings remain limited.

Agricultural data acquisition employs multiple imaging platforms, each offering distinct perspectives and trade-offs. Aerial drones provide consistent orbit-based trajectories and canopy-level coverage suitable for plot-scale monitoring (Hunt et al. 2010; D'Odorico, Rulli, and Dell'Angelo 2020). Handheld devices capture high-resolution, close-range data with flexible viewpoints but limited spatial coverage. Ground-based robots equipped with omnidirectional cameras extend field coverage with low-angle perspectives near the soil surface (Zhang and Kovacs 2012; Thenkabail 2019). While each platform has been evaluated independently in agricultural contexts, systematic cross-platform comparisons using NeRF reconstructions under matched field conditions are lacking. Differences in sensor geometry, capture trajectory, and viewpoint distribution may significantly affect reconstruction completeness and geometric fidelity, yet these effects have not been quantified in controlled comparative studies.

In this study, we present a comparative evaluation of NeRF reconstructions across three imaging platforms operating under similar field conditions: an aerial drone (DJI Inspire 2 with Zenmuse X5S camera), a handheld device (iPhone 16 Pro), and a ground robot equipped with a 360° camera (TerraSentia with Insta360 X4). We evalu-

*Corresponding author.
Accepted at the First International Workshop on AI in Agriculture (Agri AI), co-located with AAAI 2026.

ate reconstruction performance using complementary metrics that capture both geometric accuracy (Precision, Recall, F1 score) and photometric fidelity (PSNR, SSIM, LPIPS). This dual assessment is essential for agricultural phenotyping, where both structural measurements and visual realism under novel viewpoints are required for trait extraction.

Our findings will provide practical guidance for selecting appropriate imaging strategies in multi-device phenotyping workflows and highlight platform-specific strengths and limitations for field-scale 3D plant modeling.

Related Work

Since its introduction, Neural Radiance Fields (NeRF) (Mildenhall et al. 2021) has established a new paradigm for 3D reconstruction by learning continuous volumetric representations from multi-view images. Subsequent developments have addressed key practical limitations: Instant-NGP (Müller et al. 2022) accelerated training through hash-based encodings, Zip-NeRF (Barron et al. 2023) improved anti-aliasing and detail preservation, and Nerfacto (Tancik et al. 2023) integrated these advances into a stable, general-purpose framework. To handle challenging outdoor conditions with limited viewpoint diversity, recent work has explored hybrid approaches that incorporate LiDAR depth priors or geometric regularization (Niemeyer et al. 2022; Remondino et al. 2023). These technical improvements have enabled NeRF deployment in uncontrolled natural environments, including agricultural settings.

Three-dimensional imaging has become integral to crop phenotyping, providing quantitative insights into canopy architecture, leaf orientation, and biomass (Fiorani and Schurr 2013; Li et al. 2020). Early efforts relied on stereo vision and structure-from-motion (SfM) pipelines, often combined with UAV imagery or ground-based sensors (Tsouros, Bibi, and Sarigiannidis 2019). These classical methods, however, remain sensitive to variations in illumination, vegetation texture, and occlusion. The advent of NeRF has introduced new opportunities for generating smooth, continuous 3D reconstructions that better capture plant morphology (Hu et al. 2024; Arshad et al. 2024; Choi, Park, and Lee 2024). Studies have demonstrated NeRF's capability for reconstructing greenhouse crops, such as tomatoes (Choi, Park, and Lee 2024), and evaluating field-grown plants under natural lighting conditions (Arshad et al. 2024). Nevertheless, most existing agricultural NeRF studies focus on a single device or capture configuration, which limits the comparative understanding of how hardware and acquisition geometry affect reconstruction fidelity.

The impact of imaging modality on 3D reconstruction accuracy has been widely discussed in photogrammetry and remote sensing. Aerial drones provide high spatial coverage but suffer from perspective distortion and motion blur, whereas handheld devices capture finer detail at a limited scale (Hunt et al. 2010; Li et al. 2025). Ground-based robots equipped with omnidirectional cameras enable low-angle captures that reveal lower canopy structure but introduce parallax and stitching artifacts. Few studies have systematically compared these modalities within the context of NeRF-based reconstruction. Arshad et al. (2024) noted challenges in field NeRF reconstructions caused by unstable orbits and exposure variations. Li et al. (2025) emphasized the lack of standardized datasets for evaluating cross-device NeRF performance in outdoor scenes. Similarly, Remondino et al. (2023) highlighted that camera calibration, optical distortion, and trajectory stability strongly influence NeRF convergence.

Despite these insights, a unified comparison across aerial, handheld, and ground-based capture devices under shared field conditions has not yet been performed. This study addresses that gap by evaluating orbit-based NeRF reconstructions using a standardized drone trajectory and co-located handheld and 360° ground captures, providing quantitative benchmarks for cross-device performance in agricultural phenotyping.

Methods

Neural Radiance Fields (NeRFs)

Neural Radiance Fields (NeRF) represent scenes as continuous volumetric functions that map 3D position and viewing direction to emitted radiance, enabling photorealistic novel-view synthesis from multi-view imagery (Mildenhall et al. 2021; Tancik et al. 2023). By jointly optimizing scene appearance and geometry, NeRF is capable of reconstructing fine plant structures, even under varying outdoor illumination conditions. All reconstructions in this study were performed using Nerfacto, the default NeRF pipeline in Nerfstudio, which incorporates distortion correction, appearance embeddings, and multi-resolution hash encoding for improved stability in field conditions.

Field Preparation

The experiment involved two legume species: soybean (*Glycine max*, cultivar:*IAS19C3* and mungbean (*Vigna radiata* (L.) Wilczek, cultivar:*ISU Mung-G2*, both released from breeding programs at Iowa State University. These lines were selected as representative examples of their respective crops to evaluate the feasibility of multi-device 3D reconstruction in typical agronomic conditions (Fig. 1 b,c). Plants were established in a 3×3 grid with approximately 0.9 m (3 ft) spacing between individuals and 1.8 m (6 ft) spacing between the two crop sections to maintain clear separation of canopies (Fig. 1 a). Seeds were sown in late June 2025, and each hill was initially seeded with three to five seeds, later thinned to one healthy plant per position to ensure uniform growth and visibility during imaging. One soybean plant failed to emerge in the center of the soybean trial, and one mungbean plant emerged poorly. It did not survive, resulting in slightly incomplete grids that still preserved the overall experimental layout.

Ground control points (GCPs) were distributed across the plot to improve camera pose estimation and spatial alignment during reconstruction. The GCP set included multiple ArUco markers and two black-and-white checkerboard targets positioned around the field. These visual fiducials provided fixed reference points visible across all imaging modalities, helping maintain a consistent reconstruction scale among datasets. Table 1 summarizes the three imaging

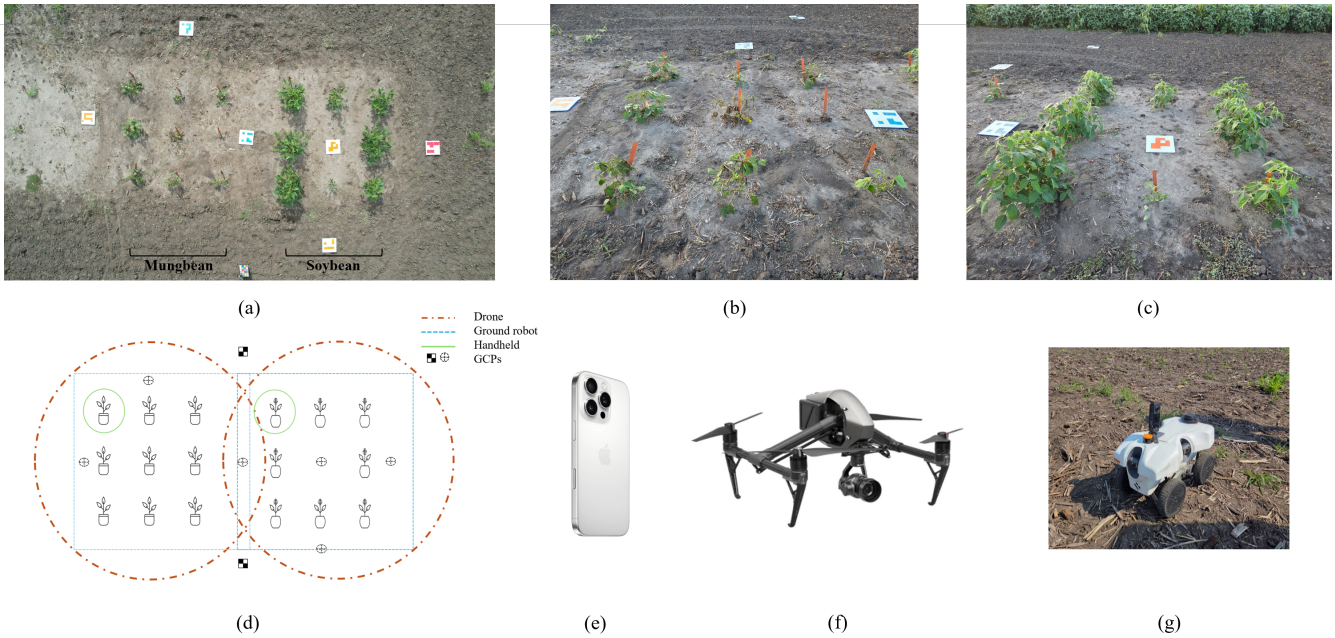

Figure 1: Overview of the experimental setup and data acquisition workflow. (a) Aerial field layout showing mungbean (*ISU Mung-G2*) and soybean (*IAS19C3*) plots with ArUco and checkerboard ground control markers. (b) Ground-level view of the mungbean plants. (c) Ground-level view of the soybean plants. (d) Schematic illustration of capture trajectories for the drone, ground robot, and handheld imaging. (e) Handheld device: iPhone 16 Pro. (f) Aerial platform (Drone): DJI Inspire 2 drone with Zenmuse X5S camera. (g) Ground robot: TerraSentia robot equipped with Insta360 X4.

platforms used in this study, including sensor specifications, image resolution, number of images or frames, viewpoint geometry, and capture distance for each modality.

Although the present experiment was conducted on a small test area to validate workflow feasibility, the same capture strategy can be scaled to larger field experiments. The drone-based orbit approach, in particular, can be automated and standardized using flight-planning applications such as DJI GS Pro , DroneDeploy, or UgCS, enabling consistent data collection across multiple sites and facilitating the integration of NeRF-based reconstruction workflows into broader phenotyping pipelines.

Handheld Dataset

The handheld images collected using the *iPhone 16 Pro* served as the reference reconstruction for this experiment. These images were captured at close range from multiple controlled viewpoints, providing high-resolution and consistent coverage of individual plants (Fig. 1d,e). As no ground-truth LiDAR scans were available, the handheld NeRF models functioned as a practical pseudo–ground truth for evaluating the relative fidelity of the drone and 360° reconstructions. The dataset required no additional preprocessing beyond standard COLMAP pose estimation prior to training with Nerfacto (Schönberger et al. 2016; Schonberger and Frahm 2016).

Drone Dataset

Aerial imagery was acquired using a *DJI Inspire 2* (DJI Technology Co., Shenzhen, China) equipped with a *Zenmuse X5S* camera fitted with an Olympus M.Zuiko 45mm f/1.8 lens (Fig. 1f). The drone was manually piloted to perform a single circular orbit around each soybean and mungbean plot at a height of approximately 4.5 m above ground level, with the camera tilted –20° relative to the horizontal plane to capture oblique canopy views (Fig. 1d). All flights were conducted during early-morning hours under diffuse illumination to minimize wind movement and shadow variation, conditions shown to enhance image stability and reconstruction fidelity (Maes and Steppe 2019).

The aerial dataset consisted of approximately 300 images captured along a circular orbit around each plot. To ensure reconstruction stability, all images were evaluated using an image-quality scoring script that ranked frames based on sharpness, contrast, entropy, and edge density, retaining only the highest-quality frames. The selected images provided consistent oblique viewpoints with sufficient overlap for reliable pose estimation. These images were then used directly for Nerfacto training without additional masking or downsampling.

360° Ground Robot Dataset

Complementary ground-level imagery was obtained using a *TerraSentia* field robot equipped with an *Insta360 X4* camera (Fig. 1g) (McGuire et al. 2021). The robot was man-

Platform	Sensor	Image Resolution	Images	Viewpoint Geometry	Capture Distance
Drone	DJI Inspire 2 with Zenmuse X5S (45 mm)	4K	∼300 images	Circular orbit with oblique viewing angles	∼6 m
Handheld	iPhone 16 Pro	12 MP	∼100 images	Multi-angle close-range viewpoints	∼0.5–1 m
360° Robot	Insta360 X4 mounted on TerraSentia robot	5.7K video	∼300 images	Circular orbit using equirectangular (360°) projection	∼1 m

Table 1: Data collection platforms used to capture complementary viewpoints and spatial scales for field-scale NeRF-based 3D plant reconstruction.

ually driven in orbit-like or square trajectories around each plot, providing low-angle, omnidirectional views of the plant canopies (Fig. 1d).

The recorded 360° video was processed by sampling 8-14 equirectangular frames per orbit. To improve pose estimation, the upper portion of each frame (dominated by sky and horizon) was cropped to remove low-texture regions that negatively affect structure-from-motion. The cropped equirectangular images were then processed using the appropriate camera model in COLMAP before Nerfacto.

Cross-Device Alignment

After NeRF training, dense point clouds were exported from each model. To account for differences in spatial scale, viewpoint, and field of view across the three imaging modalities, all reconstructions were aligned and scaled in CloudCompare using manual correspondences followed by iterative closest point (ICP) refinement. This normalization enabled direct comparison of structural fidelity across the platforms.

Results and Discussion

We evaluate reconstruction quality using both geometric (3D) and appearance-based (2D) metrics. This combination is necessary because agricultural NeRF reconstruction requires both an accurate 3D structure and realistic image synthesis under varying viewpoints.

3D Geometric Metrics

To quantify geometric fidelity, each reconstructed point cloud is compared against the handheld iPhone model, which serves as a pseudo–ground truth due to its high resolution and close-range capture. Following standard practice in multi-view stereo and point cloud evaluation (Schops et al. 2017; Strecha et al. 2008; Guo et al. 2016), we compute Precision, Recall, and F1 score using a distance-threshold protocol.

Precision

$$\text{Precision} = \frac{|\{p \in P_{\text{pred}} : d(p, P_{\text{ref}}) < \tau\}|}{|P_{\text{pred}}|}$$

Recall

$$\text{Recall} = \frac{|\{q \in P_{\text{ref}} : d(q, P_{\text{pred}}) < \tau\}|}{|P_{\text{ref}}|}$$

F1 score

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics capture reconstruction correctness (precision), completeness (recall), and their balance (F1).

2D Appearance Metrics

To assess visual fidelity, we evaluate NeRF-rendered novel views using the PSNR, SSIM, and LPIPS metrics. These metrics capture pixel-level error, perceptual similarity, and deep-feature similarity, respectively.

PSNR Following (Hore and Ziou 2010), Peak Signal-to-Noise Ratio is defined as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{L^2}{\text{MSE}(I, \hat{I})} \right)$$

Where L is the maximum pixel value.

SSIM We use the original formulation of the Structural Similarity Index from (Wang et al. 2004). Given two image patches x and y, SSIM is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x, μ_y are mean intensities, σ_x^2, σ_y^2 are variances, σ_{xy} is covariance, and C_1, C_2 are stability constants.

LPIPS LPIPS (Zhang et al. 2018) computes perceptual similarity using deep feature embeddings from a pretrained neural network. Given features $f_l(\cdot)$ at layer l, LPIPS is:

$$\text{LPIPS}(x, y) = \sum_l w_l \left\| f_l(x) - f_l(y) \right\|_2^2$$

where w_l are learned weights that align the metric with human perceptual judgments. Lower LPIPS values indicate higher perceptual similarity.

Across both crops, the drone reconstructions achieved higher geometric accuracy than the 360° camera, as reflected in the precision, recall, and F1 values in Table 2, with the strongest performance observed in soybean (Fig. 2 e,g). Soybean plants were larger and more visually distinctive

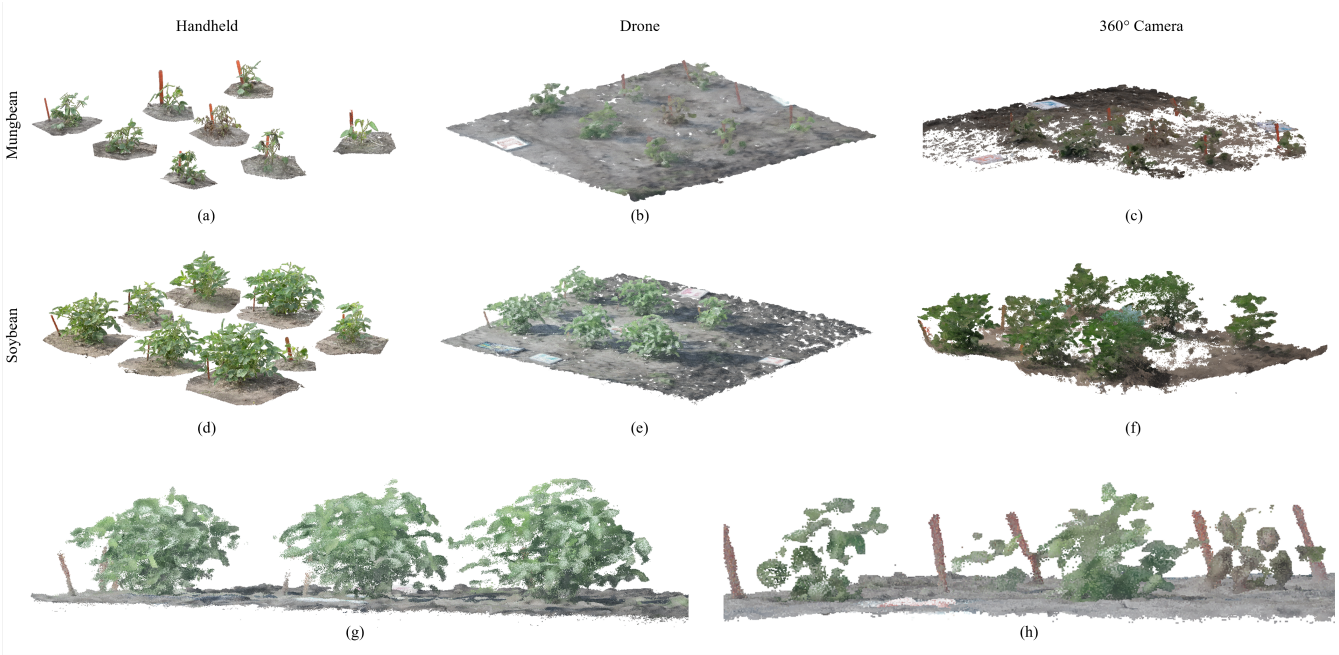

Figure 2: Results from the reconstruction across devices. Handheld iPhone (a,d) Captures provide the highest detail, drone imagery (b,e) preserves global structure, and 360° camera reconstructions (c,f) show increased noise and loss of canopy detail. Close-up views of soybean (g) and mungbean (h) drone reconstructions highlight local canopy structure.

Dataset	**Prec.**	**Rec.**	**F1**	
Drone vs Handheld	Mung	0.389	0.580	0.466
	Soy	0.498	0.670	0.571
360° vs Handheld	Mung	0.221	0.305	0.256
	Soy	0.395	0.468	0.428

Table 2: 3D geometric evaluation of drone and 360° reconstructions relative to the iPhone reference model.

Device	**Crop**	**PSNR**	**SSIM**	**LPIPS**
Drone	Soy	18.07	0.157	0.526
	Mung	20.80	0.239	0.570
Insta360	Soy	17.02	0.482	0.590
	Mung	25.02	0.747	0.281
iPhone	Soy	21.66	0.542	0.243
	Mung	21.00	0.520	0.198

Table 3: 2D appearance evaluation across devices using PSNR, SSIM, and LPIPS. Higher PSNR and SSIM, along with lower LPIPS, indicate better visual fidelity.

at the time of imaging, providing more explicit geometric cues for COLMAP-based reconstruction. In contrast, mungbean canopies were smaller and more compact, introducing greater self-occlusion and reducing the effectiveness of aerial viewpoints, as visible in the drone-based mungbean model (Fig. 2 b).

The 360° camera produced the lowest geometric fidelity for both crops (Fig. 2 c,f), largely due to the sky and soil dominating the equirectangular frames, motion-induced vibrations during maneuvering, and the limited robustness of COLMAP for spherical imagery. Because no frame selection or additional filtering was applied, these distortions led to noisier reconstructions and loss of fine canopy structure, consistent with the lower precision, recall, and F1 scores reported in Table 2. These trends are further supported by the 2D appearance metrics in Table 3, where the handheld iPhone consistently achieves the highest PSNR and SSIM, and the lowest LPIPS, reinforcing its role as the pseudo–ground truth reference.

These results reflect the capture geometry and imaging conditions of each device. The drone's stable circular trajectory yields consistent baselines and strong global structure for both crops (Fig. 2 b,e). The handheld iPhone captures, collected at close range with high image overlap, preserve the highest level of fine-scale detail (Fig. 2 a,d), which explains their superior performance across both geometric metrics (Table 2) and appearance metrics (Table 3). The 360° camera operates close to the ground but does not provide meaningful under-canopy coverage at this growth stage; instead, its horizontal viewpoint combined with spherical distortion and motion blur limits its geometric completeness (Fig. 2 c,f). Overall, the results indicate that drone imagery is most reliable for structured plot-level reconstruction, with the iPhone providing the highest detail and stability. The 360° camera can serve as an auxiliary viewpoint but requires improved preprocessing to achieve consistent re-

sults. The comparisons in Fig. 2 a–f, together with the quantitative summaries in Tables 2–3, highlight clear differences in completeness, noise characteristics, and detail retention across devices.

Conclusions and Future Work

This study systematically compared NeRF-based 3D plant reconstructions across three imaging platforms under matched field conditions. We quantified the platform-specific strengths and limitations for field phenotyping applications by evaluating both geometric accuracy and photometric fidelity across soybean and mungbean plots. Our results reveal distinct tradeoffs across platforms. Drone imagery produced the most geometrically consistent reconstructions, achieving the highest precision, recall, and F1 scores due to stable orbital trajectories and comprehensive plot-level coverage. This geometric reliability makes drone-based acquisition well-suited for measuring canopy structure and spatial plant arrangement. Handheld iPhone captures delivered superior photometric quality, with the highest PSNR and SSIM values and lowest perceptual error (LPIPS), reflecting the benefits of close-range imaging with controlled viewpoints and minimal motion blur. The 360° ground camera, while providing complementary low-angle perspectives, exhibited reduced fidelity in both geometric and visual metrics due to spherical distortion, sky-soil imbalance in equirectangular frames, and motion artifacts during maneuvering.

These findings provide practical guidance for platform selection in agricultural NeRF workflows. For applications requiring accurate spatial measurements and plot-scale coverage, such as canopy height estimation or plant spacing analysis, drone-based acquisition offers the most reliable performance. For tasks demanding high visual fidelity or detailed plant-level feature extraction, handheld devices provide superior image quality at the cost of reduced spatial coverage. Ground-based 360° cameras may serve as auxiliary viewpoints but require improved preprocessing and capture protocols to achieve comparable reconstruction quality.

Several limitations warrant consideration. Our evaluation was conducted on small experimental plots with two crop species at specific growth stages. Reconstruction performance may vary with canopy density, plant height, and phenological stage. The use of handheld iPhone captures as a pseudo-ground truth, while practical in the absence of LiDAR scans, introduces potential reference bias. Additionally, the 360° camera dataset lacked systematic frame selection and preprocessing, which may have contributed to its lower performance. Future work will address these limitations through several extensions. First, we will scale the evaluation to larger field plots with diverse crop types and canopy architectures to assess generalization across agricultural contexts. Second, incorporating GPS metadata for drone and 360° imagery will improve camera initialization and cross-platform spatial alignment. Finally, exploring hybrid capture strategies that combine drone coverage with targeted handheld close-ups may leverage the complementary strengths of each platform.

Overall, this work establishes drone-based imaging as a reliable foundation for field-scale NeRF reconstruction, with demonstrated advantages in geometric consistency and operational scalability. As NeRF methods continue to mature and computational costs decrease, multi-platform 3D reconstruction workflows offer promising pathways toward high-throughput phenotyping systems capable of capturing both plot-level spatial patterns and plant-level morphological detail in operational agricultural settings.

Acknowledgments

This work is supported by the AI Research Institutes program [*AI Institute for Resilient Agriculture (AIIRA)*, Award No. 2021-67021-35329] from the National Science Foundation and U.S. Department of Agriculture's National Institute of Food and Agriculture. We also acknowledge the support from the Plant Science Institute.

References

Arshad, M. A.; Jubery, T.; Afful, J.; Sarkar, S.; and Singh, A. K. 2024. Evaluating Neural Radiance Fields for 3D Plant Geometry Reconstruction in Field Conditions. *Frontiers in Plant Science*, 15: 1458.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2023. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19697–19705.

Choi, H.-B.; Park, J.-K.; and Lee, T.-S. 2024. NeRF-based 3D Reconstruction Pipeline for Acquisition and Analysis of Tomato Crop Morphology. *Frontiers in Plant Science*, 15: 10235.

D'Odorico, P.; Rulli, M. C.; and Dell'Angelo, J. 2020. Global Land and Water Grabbing. *Annual Review of Environment and Resources*, 45: 113–144.

Fiorani, F.; and Schurr, U. 2013. Future Scenarios for Plant Phenotyping. *Functional Plant Biology*, 40(6): 539–552.

Guo, Y.; Bennamoun, M.; Sohel, F.; Lu, M.; Wan, J.; and Kwok, N. M. 2016. A comprehensive performance evaluation of 3D local feature descriptors. *International Journal of Computer Vision*, 116(1): 66–89.

Hore, A.; and Ziou, D. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*, 2366–2369. IEEE.

Hu, K.; Ying, W.; Pan, Y.; Kang, H.; and Chen, C. 2024. High-fidelity 3D reconstruction of plants using Neural Radiance Fields. *Computers and Electronics in Agriculture*, 220: 108848.

Hunt, E. R.; Doraiswamy, P. C.; McMurtrey, J. E.; and Daughtry, C. S. T. 2010. A Visible Band Index for Remote Sensing Leaf Chlorophyll Content at the Canopy Scale. *International Journal of Applied Earth Observation and Geoinformation*, 12(3): 173–182.

Li, J.; Qi, X.; Nabaei, S.; Zhang, X.; Yin, X.; and Li, Z. 2025. A Survey on 3D Reconstruction Techniques in Plant Phenotyping: From Classical Methods to NeRF and Beyond. *Computers and Electronics in Agriculture*. In Press.

Li, L.; Zhang, Q.; Huang, Y.; and Wang, Y. 2020. High-throughput Plant Phenotyping: A Review of Current Capabilities and Future Directions. *Trends in Plant Science*, 25(4): 347–360.

Maes, W. H.; and Steppe, K. 2019. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends in plant science*, 24(2): 152–164.

McGuire, M.; Soman, C.; Diers, B.; and Chowdhary, G. 2021. High throughput soybean pod-counting with in-field robotic data collection and machine-vision based data analysis. *arXiv preprint arXiv:2105.10568*.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15.

Niemeyer, M.; Barron, J. T.; Mildenhall, B.; Sajjadi, M. S.; Geiger, A.; and Radwan, N. 2022. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5480–5490.

Remondino, F.; Karami, A.; Yan, Z.; Mazzacca, G.; and Qin, R. 2023. A Critical Analysis of NeRF-Based 3D Reconstruction. *Remote Sensing*, 15(14): 3585.

Schonberger, J. L.; and Frahm, J.-M. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.

Schönberger, J. L.; Zheng, E.; Frahm, J.-M.; and Pollefeys, M. 2016. Pixelwise view selection for unstructured multi-view stereo. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, 501–518. Springer.

Schops, T.; Schonberger, J. L.; Galliani, S.; Sattler, T.; Schindler, K.; Pollefeys, M.; and Geiger, A. 2017. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3260–3269.

Strecha, C.; Von Hansen, W.; Van Gool, L.; Fua, P.; and Thoennessen, U. 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *2008 IEEE conference on computer vision and pattern recognition*, 1–8. Ieee.

Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A.; et al. 2023. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 conference proceedings*, 1–12.

Thenkabail, P. S. 2019. *Remote Sensing Handbook: Volume II—Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*. CRC Press.

Tsouros, D. C.; Bibi, S.; and Sarigiannidis, P. G. 2019. A review on UAV-based applications for precision agriculture. *Information*, 10(11): 349.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Zhang, C.; and Kovacs, J. M. 2012. The Application of Small Unmanned Aerial Systems for Precision Agriculture: A Review. *Precision Agriculture*, 13(6): 693–712.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.