



Plot2Phenome: A UAV-Based Deep Learning Framework for Automated Micro-Plot Segmentation and Plot Level Phenotyping

Chuanjing Jin¹, Yan Li¹, Xiaohu Zhang^{1, 2, 3, *}

¹National Engineering and Technology Center for Information Agriculture, Nanjing Agricultural University, Nanjing 211800, China

²Key Laboratory for Crop System Analysis and Decision Making, Ministry of Agriculture and Rural Affairs, Nanjing 211800, China

³Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing 211800, China

**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.

Automated micro-plot segmentation is a foundational requirement for plot-level phenotyping from UAV orthomosaics in field breeding trials. However, reliable delineation of individual plots remains difficult in realistic agronomic settings, particularly under canopy closure that erodes inter-plot gaps and under irregular or degraded plot boundaries caused by lodging, variable emergence, and field operations. These conditions reduce boundary contrast, increase instance adjacency, and introduce shape ambiguity, which frequently degrades the robustness of conventional computer vision pipelines and limits the accuracy and consistency of downstream plot-level phenotypic measurements. Approaches that rely solely on RGB texture often fail under canopy closure, because plot boundaries become visually indistinguishable, which hinders robust digital management of breeding plots and reliable phenotypic interpretation at the plot level.

To address this bottleneck, this study aims to develop an end-to-end automated framework, termed Plot2Phenome, for the phenotyping of field breeding trials. The primary objectives are to: (1) create a robust deep learning model capable of precisely segmenting micro breeding plots from UAV-captured orthomosaics under varying field conditions; (2) establish a plot-level digital identification and management framework that assigns unique identities to individual breeding plots, thereby supporting longitudinal phenotypic data integration and time-series analysis; and (3) demonstrate that the resulting segmentation masks support reliable extraction of plot-level phenotypic traits.

High-resolution RGB imagery and the corresponding digital surface models (DSM), derived from low-altitude UAV flights, were acquired over breeding trial fields that contained diverse genetic varieties across multiple crops and sites. To address the segmentation challenge, we propose RGBD-PlotFormer, a dual-stream instance segmentation framework that integrates RGB orthomosaics and DSMs via cross-modal attention fusion, enabling the delineation of plot boundaries that are otherwise indiscernible from spectral and textural cues alone. The integration of spectral and height information improves boundary discrimination under canopy closure and enables consistent extraction of plot-level instances with geometric regularity. Following segmentation, each plot is automatically georeferenced and assigned a unique digital identifier, which enables the extraction of key phenotypic traits such as canopy coverage, excess green index (ExG), and canopy height.

The proposed framework was evaluated on six UAV field experiments across three crops and four sites in China, comprising 412 micro-plots acquired from 2020 to 2025. RGBD-PlotFormer achieved a mean intersection over union (mIoU) of 93.64% and ranked first across all evaluated metrics among five baseline models. Moreover, the plot-level traits extracted from the predicted masks exhibited strong agreement with the traits derived from manually annotated masks, with coefficients of determination ranging from 0.637 to 0.853, which confirms the reliability of the extracted data for plot-level phenotypic analysis. In conclusion, these results demonstrate that the Plot2Phenome framework effectively bridges raw UAV imagery and agronomic decision-making by enabling accurate, standardized, and temporally consistent plot-level phenotyping. By integrating deep learning-based plot segmentation with digital plot identification and management, the framework supports multi-environment breeding trials, facilitates cross-site comparison, and enhances the practicality of UAV-based phenotyping in routine breeding workflows. The source code of Plot2Phenome is available at <https://github.com/Little-King2022/Plot2Phenome>.

Keywords.

UAV-based Phenotyping, Micro-plot Segmentation, Crop Breeding Trials, Instance Segmentation, Deep Learning.

Introduction

High-throughput field phenotyping requires repeated and objective observation of breeding populations at the plot level, and UAV imagery has become a widely adopted data source because it provides high-resolution observations at comparatively low cost (Yang et al., 2017). Canopy and height traits can be summarized accurately for each genotype only when the boundary of every micro-plot is delineated precisely, yet breeding plots are small, densely arranged, and frequently obscured by canopy closure, lodging, and weed encroachment, so rule-based and threshold-based extraction remains highly sensitive to field geometry and operator intervention (Khan & Miklavcic, 2019). Recent deep-learning methods have improved plot extraction (Han et al., 2024; Zhang et al., 2025; Li et al., 2025), but most rely primarily on RGB appearance and degrade when canopy closure removes the visible inter-plot gaps. Digital surface models (DSM), generated routinely alongside orthomosaics during photogrammetric processing, provide complementary structural information, because height gradients and local surface discontinuities remain discriminative when color contrast diminishes. A dual-stream workflow termed Plot2Phenome was therefore developed to integrate RGB and DSM inputs for instance-level plot delineation and subsequent plot-level trait extraction.

Materials and Methods

Six sub-datasets were acquired from breeding trial fields in China between 2020 and 2025, encompassing three crop species across four locations and comprising 412 micro-plots, with phenological stages ranging from early-season open canopies to late-season closed canopies. Co-registered RGB orthomosaics and DSM products were generated with OpenDroneMap and partitioned into overlapping 512 × 512 pixel tiles. RGBD-PlotFormer is a dual-stream instance segmentation network (Fig. 1) in which two parameter-independent ResNet-50 encoders (He et al., 2016) process the RGB and DSM streams separately. A hierarchical bimodal fusion module (HBFM), a multi-scale spatial-channel attention module (MSCAM), and a boundary attention module (BAM) progressively integrate the two modalities and strengthen plot borders before a Mask2Former-style query decoder with 100 queries (Cheng et al., 2022) predicts the plot instances. Predicted masks were converted to georeferenced polygons, assigned unique identifiers through row-column coding in a field-oriented coordinate system, and used to extract plot-level traits, with canopy height derived from the DSM through robust quantile statistics (Bendig et al., 2014).

Results

RGBD-PlotFormer achieved a mean intersection over union (mIoU) of 93.64%, an AP@50 of 93.84%, an AP@75 of 87.84%, a panoptic quality (PQ) of 86.92%, and a boundary F1 (BF1) of 78.15%, and it ranked first across all reported metrics among five baseline models (Table 1, Fig. 2) on the validation set. The closest competitor, YOLOv8x-seg, attained a comparable mIoU of 93.61% but a lower boundary F1 of 77.16%, whereas Mask R-CNN reached a competitive AP@50 of 92.27% but a substantially weaker boundary F1 of 68.47%.

An ablation study comparing the full model with an RGB-only variant and module-removal variants demonstrated that the DSM modality was the dominant source of improvement (Fig. 3). Incorporating DSM information, even though simple fusion, increased mIoU by 2.00 percentage points (from 88.31% to 90.31%) and boundary F1 by 8.39 percentage points (from 73.44% to 81.83%) relative to the RGB-only model. The addition of HBFM, MSCAM, BAM, and the alignment loss collectively contributed a further 0.77 and 1.39 percentage points over the basic RGB-DSM baseline, respectively. Among the individual modules, removal of the boundary attention module resulted in the largest mIoU reduction.

To assess phenotyping reliability, plot-level traits derived from predicted and manually annotated masks were compared on the 66 maize micro-plots at Dengzhou on 31 July 2025, the most

challenging case for boundary delineation. The two sets agreed closely across all three traits. Canopy coverage showed the strongest agreement, with R^2 of 0.853 and RMSE of 0.0122; canopy height followed with R^2 of 0.759 and RMSE of 0.13 m; and canopy volume reached R^2 of 0.637 and RMSE of 1.93 m³. Biases remained near zero across the full value range, indicating no appreciable systematic deviation from manual annotation.

Discussion

Accurate micro-plot segmentation is fundamentally a problem of boundary delineation, and the visual distinctiveness of plot boundaries degrades as the canopy closes. Under canopy closure the spectral and textural contrast between adjacent plots is lost, which makes models that rely solely on RGB prone to instance merging. As shown by the maize examples in [Fig. 3](#), the RGB-only model frequently merged adjacent plots in canopy-closure regions, whereas the model that incorporated DSM separated them correctly, because canopy height drops and local surface discontinuities at plot edges persist even when color and texture become indistinguishable. An ablation analysis confirmed that the DSM modality was the dominant source of improvement, because the addition of DSM information to the RGB-only model raised the mIoU from 88.31% to 90.31% and the boundary F1 from 73.44% to 81.83%. Because UAV-derived DSM products are affected by wind-induced canopy motion, reconstruction artifacts over texture-poor regions, and local voids, the hierarchical fusion formed by HBFM, MSCAM, and BAM was adopted to suppress modality-specific noise and to base boundary delineation on relative structural transitions rather than absolute elevation, which further improved the boundary F1 under challenging conditions.

Conclusion

This study presented Plot2Phenome, a UAV-based workflow that couples RGB-DSM instance segmentation with plot-level trait extraction for breeding micro-plots. RGBD-PlotFormer delineated dense micro-plots with high spatial agreement (mIoU 93.64%, AP@50 93.84%, PQ 86.92%, boundary F1 78.15%), outperforming all five baseline models including U-Net, Mask R-CNN, and YOLO-seg family. The ablation study indicated that DSM height information was the dominant source of improvement over RGB-only segmentation. Traits extracted from the predicted masks agreed closely with traits obtained from manually annotated masks, with coefficients of determination ranging from 0.637 to 0.853, which demonstrates that accurate micro-plot segmentation provides a reliable spatial foundation for plot-level phenotyping. Future work will extend the framework to independent multi-site evaluation and to field validation of the derived traits against in-situ measurements.

References

- Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., & Bareth, G. (2014). Estimating biomass of barley using crop surface models derived from UAV-based RGB imaging. *Remote Sensing*, 6(11), 10395-10412.
- Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1290-1299).
- Chen, C. J., & Zhang, Z. (2020). GRID: A Python package for field plot phenotyping using aerial images. *Remote Sensing*, 12(11), 1697.
- Han, X., Zhou, M., Guo, C., Ai, H., Li, T., Li, W., et al. (2024). A fully convolutional neural network model combined with a Hough transform to extract crop breeding field plots from UAV images. *International Journal of Applied Earth Observation and Geoinformation*, 132, 104057.
- Hazirbas, C., Ma, L., Domokos, C., & Cremers, D. (2017). FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture. In *Computer Vision - ACCV 2016* (pp. 213-228). Cham, Switzerland: Springer.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).
- Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., & Hawkesford, M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sensing*, 8(12), 1031.

- Khan, Z., & Miklavcic, S. J. (2019). An automatic field plot extraction method from aerial orthomosaic images. *Frontiers in Plant Science*, 10, 683.
- Li, S., Song, Y., Wang, K., Liu, Y., Xian, J., Wu, H., et al. (2025). STANet-TLA: Leveraging deep learning and prior knowledge for large-scale soybean breeding plot segmentation and high-yielding variety screening from UAV time-series data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 227, 397-414.
- Meyer, G. E., & Camargo Neto, J. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63(2), 282-293.
- 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.
- Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1), 259-269.
- Zhang, R., Zhang, Y., Jin, S., Zang, J., Zhao, R., Yao, J., et al. (2025). OSNet: An oriented instance segmentation network of breeding plot extraction from UAV RGB imagery. *Computers and Electronics in Agriculture*, 236, 110436.

Appendix

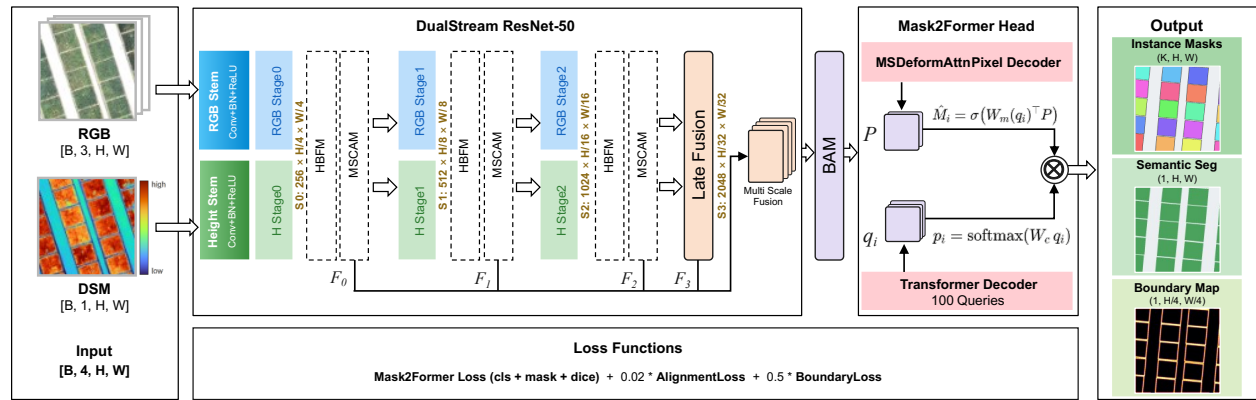


Fig. 1 RGBD-PlotFormer architecture.

Table 1. Comparison of RGBD-PlotFormer with baseline models. All models were trained and evaluated on the same dataset under identical settings. Best results are highlighted in bold.

Model	Modality	AP@50	AP@75	PQ	SQ	RQ	BF1	CountMAE↓	mIoU
U-Net (s5-d16+FCN)	RGB	-	-	-	-	-	-	-	92.53
Mask R-CNN R50-FPN	RGB	92.27	85.16	83.40	88.31	94.45	68.47	0.529	91.81
YOLOv8x-seg	RGB	92.18	86.91	86.48	90.48	95.58	77.16	0.548	93.61
YOLO11x-seg	RGB	89.87	83.90	84.57	89.93	94.05	75.24	0.697	92.86
YOLO26x-seg	RGB	89.43	84.04	84.73	90.28	93.86	75.95	0.816	92.19
RGBD-PlotFormer (Ours)	RGB+DSM	93.84	87.84	86.92	90.52	96.24	78.15	0.364	93.64

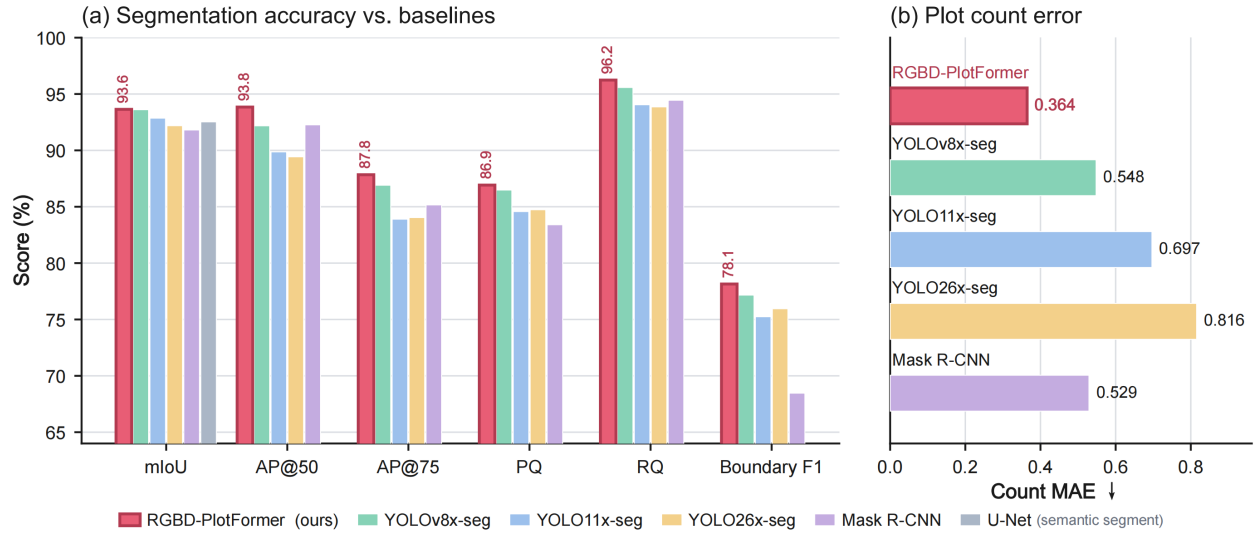


Fig. 2. Quantitative comparison of RGBD-PlotFormer (ours) with five baselines trained and evaluated under identical settings: (a) grouped bars for mIoU, AP@50, AP@75, PQ, RQ and Boundary F1; (b) plot-count error (Count MAE, lower is better). Our model ranks first on every metric.

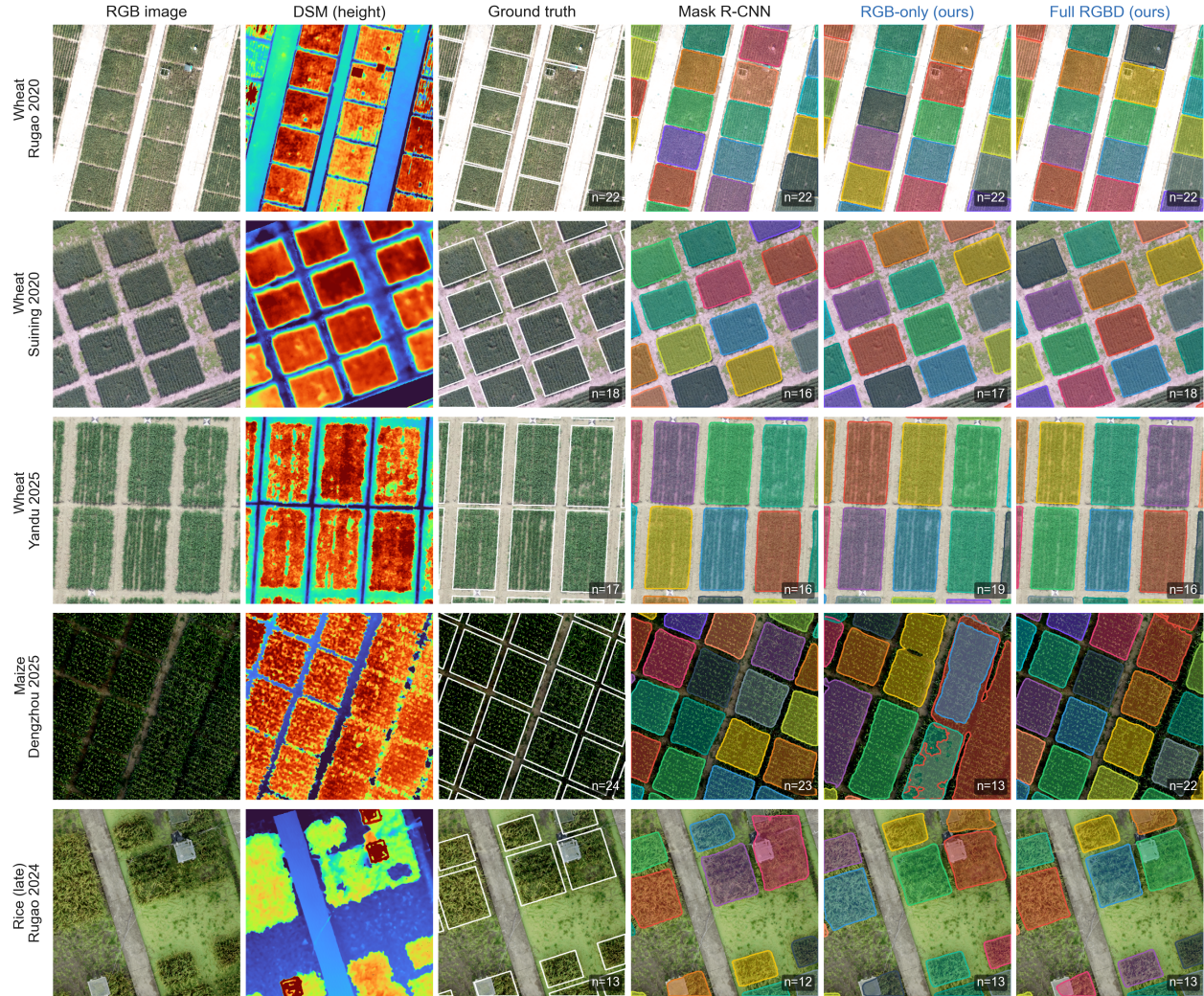


Fig. 3. Qualitative segmentation results. Columns: RGB, DSM, ground truth, Mask R-CNN, RGB-only, and full RGBD model; insets give the number of detected plot instances. The red box marks the case of segmentation failure, where the RGB-only models merge adjacent plots, while the DSM-assists model separates them.