UAV-BASED HIGH-THROUGHPUT PHENOTYPING OF SOYBEAN USING LIGHTWEIGHT POINT DETECTION FOR MULTI-ORGAN TRAIT EXTRACTION

/

基于高通量无人机的轻量化大豆关键点检测模型用于多器官表型获取

Jianing LI[1)], Jinye LU[2)], Luyan LIU[3)], Kai WANG*[4)]

[1)] College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Shandong / China

[2)] Research & Development Affairs Office, Tsinghua University, Beijing / China

[3)] Qingdao Topscomm Communication Co., Ltd, Shandong / China

[4)] College of Electrical Engineering, Weihai Innovation Research Institute, Qingdao University, Shandong / China

Tel: 15863060145; E-mail: wangkai@qdu.edu.cn

DOI: https://doi.org/10.35633/inmateh-78-33

Keywords: *UAV imagery, Multi-organ detection, Lightweight point-based network, Counting and localization, End-to-end framework, Super Dynamic Convolution.*

ABSTRACT

Accurate soybean field phenotyping is increasingly important for breeding. However, traditional measurement methods are labor-intensive and subjective, while UAV-based approaches are challenged by complex backgrounds and densely distributed small targets. This study first develops UAV-ZSAR to transform oblique UAV images into horizontal-view images and reconstruct plant geometry. A lightweight point-based model, Soy-MOPNet, is then proposed for fast and parallel detection of soybean seeds and stem nodes. The model incorporates the proposed SDConv, optimized hierarchical dilated convolution (HDC) principles, and PBOS to enhance adaptive feature fusion, receptive field design, and multi-branch training stability, respectively. Based on the detected keypoints, six phenotypic traits are extracted in parallel, providing comprehensive support for field phenotyping, breeding selection, and precision agricultural management.

摘要

精准的田间大豆表型分析对育种研究日益重要，但传统表型测量方法费时费力且主观性强，而基于无人机的高通量表型采集又面临田间背景复杂和小目标密集等挑战。本文首先构建了 UAV-ZSAR 方法，将倾斜无人机图像转换为水平视角图像，以恢复植株几何形态。随后，构建了轻量化点式模型 Soy-MOPNet，用于大豆豆粒和茎节点的快速并行检测。该模型引入了所提出的 SDConv、优化的 HDC 原则和 PBOS，分别用于提升自适应特征融合能力、感受野设计效果和多分支训练稳定性。基于检测到的关键点，进一步并行提取了 6 个表型性状，从而为田间表型分析、育种筛选和精准农艺管理提供更系统、更全面的支持。

INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is an important pulse and oil crop whose seeds are rich in protein and oil and serve as major sources of vegetable oil and industrial materials *(De Pretto et al., 2018; Hernandez et al., 2023; Zhang et al., 2015)*. Breeding high-yield and high-quality soybean varieties has therefore become a major goal. Deep learning combined with low-altitude remote sensing is enabling high-throughput field phenotyping *(Awais et al., 2025; Maraveas, 2024; Oliveira et al., 2025)*. At maturity, organ traits are directly related to final yield and variety selection, yet most existing studies focus on single organs *(Hu et al., 2024; Wang et al., 2025)*, limiting multi-phenotype extraction and robustness in complex field environments. Therefore, a high-throughput phenotyping approach for UAV field images is proposed.

Soybean yield and grain quality are key indicators of varietal performance. At maturity, pods and stems are the main visible organs, and their traits are closely related to plant architecture, height, branch number, and stem curvature. Accurate detection of these organs enables efficient acquisition of most breeding-relevant traits. Previous studies used edge detection to measure pod length, width, area, and seed count *(He et al., 2022)*, YOLOX to classify pods by seed number *(Zhang et al., 2023)*, and instance segmentation for pod segmentation and per-pod seed estimation *(Yang et al., 2024)*. To characterize whole-plant architecture, stem traits were further incorporated. PodNet was proposed for post-harvest pod counting and localization *(Yu et al., 2024)*, instance segmentation was applied to disassembled stem-pod images for stem length and pod size extraction *(Li et al., 2022)*, and DSBEAN used YOLOv5 to detect nodes and pods for effective stem length estimation *(Zhang et al., 2024)*. A directed search algorithm was further introduced to reconstruct stem spatial configuration *(Guo et al., 2022)*.

Soybean phenotyping research has been largely limited to laboratory conditions, restricting its applicability to non-contact and automated trait extraction. In recent years, UAVs have become low-cost and efficient platforms for non-destructive, high-throughput field phenotyping *(Feng et al., 2021; Kerec et al., 2026; Xie and Yang, 2020)*, and have been applied to crops such as wheat, grape, and rapeseed for extracting traits including spike count, grape cluster morphology, and plant height *(Ariza-Sentís et al., 2023; Shen et al., 2024)*. However, phenotyping accuracy in the field is still affected by complex backgrounds, occlusion, and inconsistent image quality. In soybean fields, small targets, slender plant structures, low contrast with soil, uneven fallen-leaf cover, and motion blur caused by wind further increase detection difficulty. To address these problems, anchor box clustering in YOLOv3 was used to improve sensitivity to pod shape *(Lu et al., 2022)*, transfer learning and synthetic occlusion augmentation were adopted to improve YOLOv7 pod detection under complex conditions *(Kanglei et al., 2025)*, and coordinate attention was introduced to enhance pod recognition and yield estimation under occlusion *(He et al., 2023)*. Despite these efforts, UAV-based multi-trait soybean phenotyping remains challenging because of imaging angle, resolution limits, and plant occlusion, and current research is still exploratory.

In UAV-based field phenotyping, dense targets often cause overlapping bounding boxes and reduce the accuracy of conventional object detection. Recently, point annotations have emerged as a lightweight alternative with lower labeling cost and better performance in densely packed scenes. Their effectiveness has been demonstrated in field crops, such as keypoint localization in complex point clouds *(Chen et al., 2025)*, centroid detection for overlapping wheat ear counting *(Yao et al., 2024b)*, and wheat leaf counting using tip-point annotation *(Li et al., 2023)*. Compared with bounding boxes, point annotations better represent the fine morphology of dense, small, and diverse soybean organs. They have also supported extraction of complex phenotypes, including seed sequence detection, per-plant yield estimation *(He et al., 2024)*, and pod shape and spatial orientation analysis *(Liu et al., 2025)*. Therefore, point-based methods are lightweight, accurate, and robust for automated multi-organ soybean phenotyping.

Because dense targets are closely distributed in pixels, traditional density-map regression often causes point displacement and is mainly limited to counting. P2PNet *(Song et al., 2021)* introduced a one-to-one matching strategy based on the Hungarian algorithm for direct point-level supervision, overcoming the long-standing limitation of counting without localization in dense scenes. This framework has shown superior localization accuracy in multiple tasks. For example, P2PNet-EFF *(Yao et al., 2024a)* reduced localization errors in rice ear counting under leaf occlusion, P2PNet-Soy *(Zhao et al., 2023)* combined attention mechanisms and unsupervised clustering to distinguish overlapping soybean grains and improve per-plant yield estimation, and Transformer-based point matching was further introduced into UAV soybean image analysis to improve pod localization accuracy and counting stability *(Li et al., 2024)*.

Although deep learning has advanced plant phenotyping, field-based soybean phenotype extraction still faces major challenges. Most existing methods focus on single-organ traits, limiting support for high-dimensional phenotypic analysis. Joint extraction of multiple organs and traits requires stronger representation ability and lightweight design. In addition, complex backgrounds, occlusion, dense targets, and blurred features in UAV images further reduce detection accuracy. Meanwhile, point-based annotation methods, despite their advantages in localization and efficiency, remain underutilized.

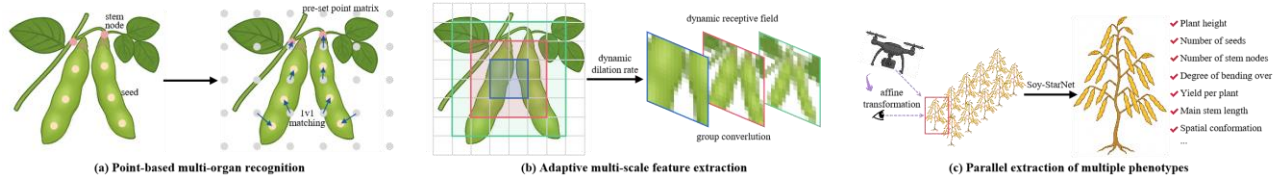


Fig. 1 - Overview of the Proposed Phenotyping Framework for UAV-based Multi-Organ Detection

To address these gaps, this study proposes a lightweight framework for point-based multi-organ phenotyping from UAV field images. An affine transformation is first applied to rectify oblique UAV images into horizontally aligned views (Fig. 1c). Point-level matching is then used to identify soybean pods and stem nodes (Fig. 1a), supporting keypoint detection and multi-phenotype extraction. Based on the detected keypoints, six soybean traits are extracted, including plant height, stem node count, seed count, yield per plant, main stem length, and lodging resistance.

The main contributions are as follows:

1) An affine-transformation-based image rectification method, UAV-ZSAR, is proposed to transform soybean plant regions in oblique UAV images into millimeter-accurate horizontal-view images, restoring the true spatial configuration of soybean plants for downstream phenotypic analysis.

2) A lightweight network, Soy-MOPNet, is developed for parallel point-based detection of seeds and stem nodes in densely planted UAV field images. Based on the detected keypoints, six maturity traits are extracted concurrently: plant height, stem node count, seed count, yield per plant, main stem length, and lodging resistance. The framework is extensible to field phenotyping of other crops.

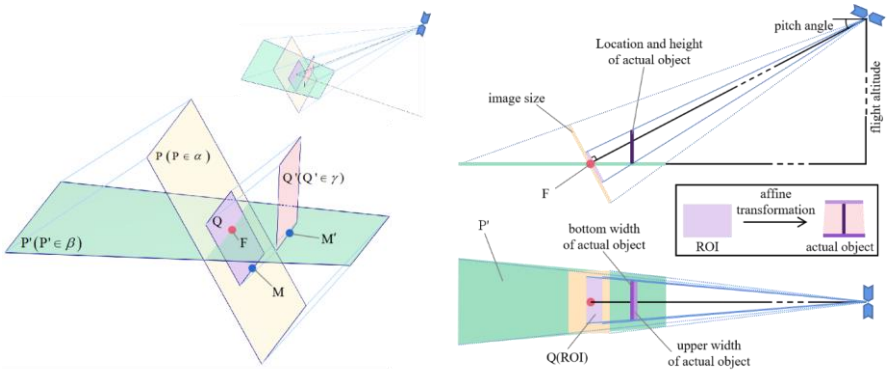
3) A Super Dynamic Convolution module, SDConv, is designed with a five-dimensional attention mechanism and adaptive receptive-field modulation to capture and fuse small-target features of dense seeds and stem nodes with the elongated structure of soybean plants, while suppressing interference from complex field backgrounds.

MATERIALS AND METHODS

UAV-ZSAR for Rectifying Oblique UAV Imagery

Oblique UAV imagery causes geometric distortion and limits precise phenotyping. To address this issue, UAV-ZSAR establishes two geometric mappings: from oblique imaging to a virtual imaging region referenced to vertical projection, and from the plant ROI in the image to its true horizontal-view configuration determined by the actual ground position of the plant.

As shown in Fig. 2 and Fig. 3, the method includes scale determination, image positioning, and ROI correction. First, the ground sampling distance (GSD) is calculated from the flight height, pixel size, and focal length, and the virtual imaging region P, constructed with reference to vertical projection, is determined from the image resolution. Then, P is projected onto the ground to form the actual photographic area P'. Within this mapping, the root pixel M of the plant corresponds to the field coordinate M', which further determines the plant growth plane Q'. Finally, the plant ROI Q in the oblique image is projected onto Q' through an affine transformation matrix A, thereby recovering the true plant configuration in the horizontal view. Lanczos interpolation is then used to generate the corrected image with millimeter-level precision. A Python-based tool was developed for ROI extraction and perspective correction, and it can be combined with an existing detection model for automatic multi-ROI processing in high-throughput applications.



(a) Projection Principle of UAV Oblique Photogrammetry (b) Top and Front Views of the Oblique Image Rectification Model

Fig. 2 - Schematic Illustration of the Modeling Principle for UAV-ZSAR.

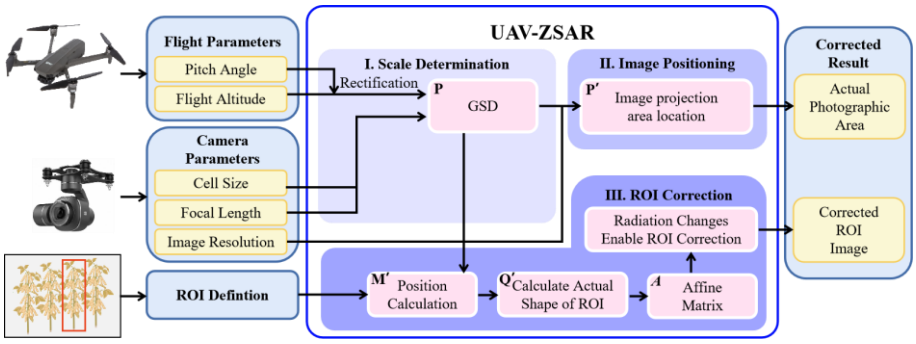


Fig. 3 - The flow diagram of the wheat reduction phase in a milling plant with the capacity of 4.2 t/h

Field Image Acquisition and Dataset Construction

The experiment was conducted in Dongying, Shandong, China (118.65°E, 37.31°N) under local standard soybean cultivation, with an average intra-row spacing of 8.3 cm and row spacing of 0.5 m. Oblique field images were acquired using a MATRICE 300 RTK UAV equipped with a downward-facing Zenmuse H20T gimbal. To ensure complete visibility of individual plants, the flight path was arranged perpendicular to the planting rows. The image resolution was 3840 × 2160 pixels, with a pixel size of 1.43 μm, a focal length of 36.0 mm, a zoom ratio of 10×, a flight altitude of 12 m, and pitch and yaw angles of 30°.

As shown in Fig. 4, field images were extracted from UAV video at five-frame intervals. Complete single-plant ROIs were obtained by either direct cropping or UAV-ZSAR correction, followed by image sharpening and 1.5× lossless upscaling using Video2X. A total of 424 high-quality images were generated, covering 26 soybean cultivars and major maturity-related phenotypic types. Among them, 40 UAV-ZSAR-corrected images were used as the test set, and the remaining images were randomly divided into training and validation sets at a ratio of 8:2. Data augmentation, including flipping, brightness and contrast adjustment, and Gaussian and salt-and-pepper noise, was applied to improve robustness. Finally, LabelMe was used to annotate the centers of soybean seeds and stem nodes; seeds were labeled when more than 10% of their area was visible, and stem nodes were annotated even under occlusion.

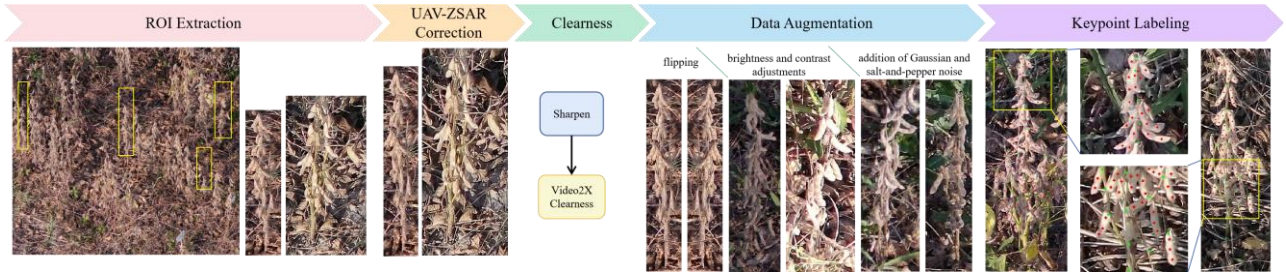


Fig. 4 - Schematic Diagram of Dataset Construction Process

Soy-MOPNet

Soy-MOPNet is a lightweight point-level multi-organ detection model tailored for mature soybean plants, enabling classification, counting, and precise localization of both seeds and stem nodes. As shown in Fig. 5, Soy-MOPNet consists of three components: a feature extraction backbone named SD-StarNet, built upon SDConv and an improved version of StarNet; three parallel point prediction heads responsible for category classification and coordinate regression; and a point matching module based on the Hungarian algorithm, which is used only during training. The model is designed to suppress complex background interference from the early convolutional stages. It expands the receptive field using a lightweight recursive multi-layer dilated convolution structure, which directs attention toward the stem and organ regions of the plant. Instead of relying on traditional dense-object detection, Soy-MOPNet adopts simplified point classification and coordinate regression, achieving a balance among accuracy, efficiency, and architectural adaptability.

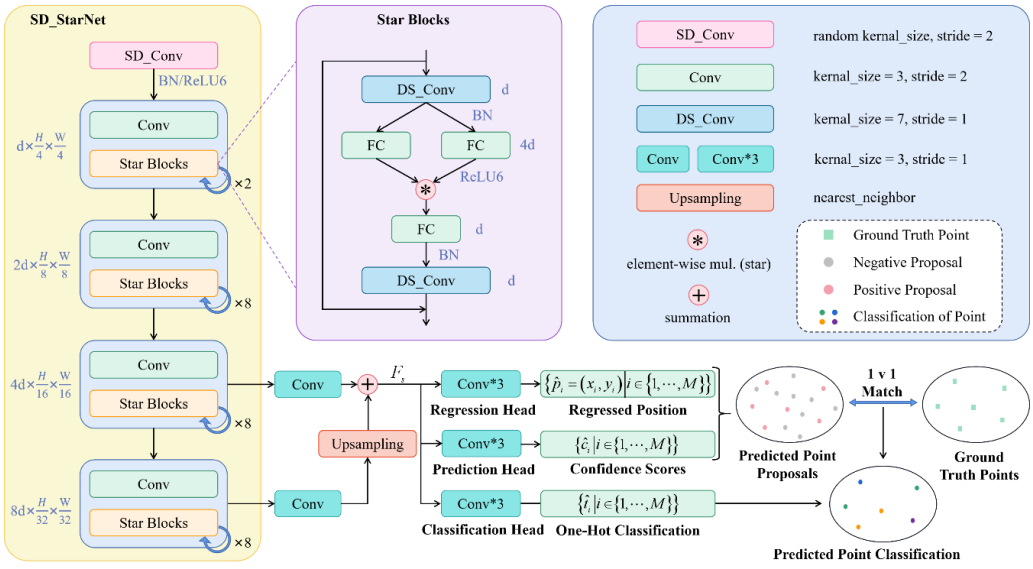
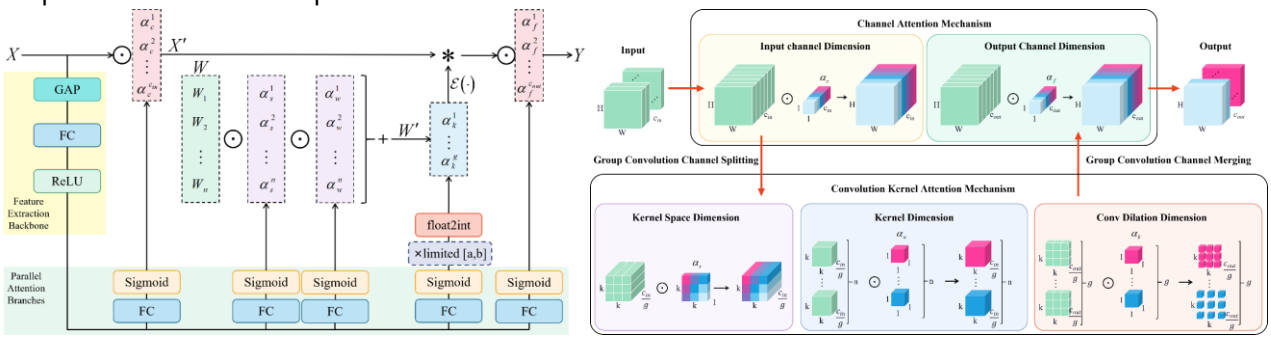


Fig. 5 - Overall Architecture of Soy-MOPNet

SDConv Based on Group Convolution

Dynamic convolution improves parameter efficiency and feature adaptability over static convolution. To enhance multi-scale modeling for dense soybean organs, SDConv is developed on the basis of group convolution. Different from existing dynamic convolutions that mainly modulate kernel weights, SDConv further introduces attention-driven dynamic dilation, assigning each convolution group an adaptive dilation rate to jointly model target scale and structural variation. This design extends dynamic modulation from convolution weights to the convolution operation itself, thereby strengthening sensitivity to dense small targets and elongated plant structures.

As shown in Fig. 6(a), SDConv adopts a shared attention-generation backbone with parallel lightweight branches. The input feature map is first compressed by global average pooling and nonlinear transformation, and then five attention weights are generated to modulate the input channel, output channel, kernel space, kernel strength, and dilation rate, respectively. As shown in Fig. 6(b), The feature map is first weighted by input-channel attention and split into groups for parallel convolution. During convolution, spatial attention assigns positional weights, kernel attention adjusts kernel response, and dilation attention controls the receptive field of each group. The grouped outputs are then concatenated and further refined by output-channel attention. In this way, SDConv achieves coordinated modulation across channels, kernels, space, and receptive fields with limited parameter overhead.



(a). SDConv Structure Diagram (b). SDConv Structure Diagram

Fig. 6 - Overall Structure and Five-Dimensional Attention Composition of SDConv

SDConv is formulated as:

$$\begin{aligned} X' &= \left[\left[\alpha_c^1,\ldots,\alpha_c^m\right]\odot X_1,\ldots,\left[\alpha_c^{c_{in}-m},\ldots,\alpha_c^{c_{in}}\right]\odot X_g\right] \\ W' &= \alpha_w^1\odot\alpha_s^1\odot W_1+\cdots+\alpha_w^n\odot\alpha_s^n\odot W_n \\ Y &= \alpha_f\odot\left(\mathcal{E}\left(\alpha_k,W'\right)*X'\right) \end{aligned} \tag{1}$$

where, $X\in\mathbb{R}^{c_{in}\times h\times w}$ and $Y\in\mathbb{R}^{c_{out}\times h\times w}$ denote the input and output feature maps, respectively; g is the number of convolution groups; W denotes the convolution kernel set; and n is the number of kernels. The attention coefficients α_c and α_f represent input- and output-channel modulation, respectively, whereas α_s, α_w, and α_k denote spatial weighting, kernel weighting, and dilation control. The operator $\mathcal{E}(\alpha_k,W')$ expands the kernel according to the dilation coefficient, $\odot$ denotes element-wise modulation across dimensions, and $*$ denotes group convolution. To maintain structural validity, the predicted dilation rate is constrained and discretized to an integer before convolution. This dynamic dilation strategy is the key improvement of SDConv, enabling more flexible receptive-field adaptation and more effective feature extraction under complex field backgrounds.

SD-StarNet: Lightweight Multiplicative Backbone with HDC

Conventional feature extraction networks usually improve performance by increasing model width and depth, which intensifies the conflict between lightweight design and task complexity. Compression methods such as pruning, knowledge distillation, and low-rank decomposition mainly reduce model size but do not fundamentally improve feature extraction efficiency. Recent studies have shown that element-wise multiplication enhances feature interaction and nonlinear representation. As a representative multiplicative network, StarNet achieves favorable accuracy and latency with a compact four-stage architecture.

Because soybean seeds and stem nodes depend on different receptive fields, a lightweight backbone named SD-StarNet is developed based on StarNet. Seed detection mainly relies on fine-grained local features, whereas stem node detection depends more on the overall posture of the soybean plant. To accommodate these different demands, SD-StarNet replaces the base convolutions with SDConv to enhance general feature

extraction, substitutes conventional convolutions in Star Blocks with depthwise separable convolutions to reduce parameters, and introduces hybrid dilated convolution (HDC) to adaptively regulate receptive field size and attention focus. This design improves adaptability to the slender structure of soybean plants and the scale variation of different organs.

To configure HDC more reasonably, the spatial receptive behavior of repeated modules was further analyzed. A metric termed Pixel Receptive Frequency (PRF) is introduced to measure how many times each input pixel contributes to the receptive field of a given output pixel within a repetitive convolutional module. Based on PRF, a Pixel Receptive Frequency Heatmap (PRFH) is constructed to visualize spatial coverage under different dilation settings, as shown in Fig. 7. Since the fully connected layers and multiplication operations in StarNet mainly modulate channels rather than spatial receptive fields, they are excluded from this analysis. The results show that mixed dilation settings can expand receptive coverage, and the configuration with both wider coverage and stronger central response is more suitable for soybean organ detection. Therefore, it is adopted as the default setting.

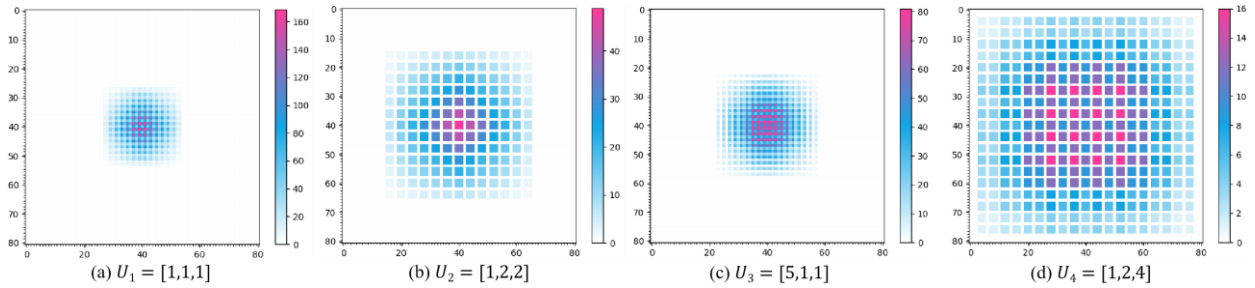


Fig. 7 - Analysis of Receptive Field Structure via PRFH under Hybrid Dilated Convolution

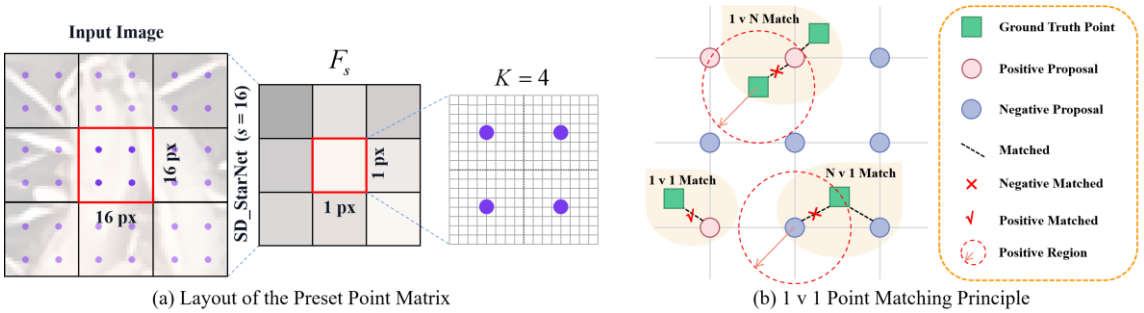
Based on the above analysis, a hybrid dilation configuration strategy for repetitive modules is proposed. It follows three principles: complete spatial coverage to alleviate the gridding effect, stronger central response with weaker boundary response to preserve structural focus, and alignment between high-response regions and target scale to improve feature extraction. Although simple, this strategy provides an intuitive and practical guideline for HDC design.

Purely Point-based Prediction and Matching

Following existing point-based detection frameworks, Soy-MOPNet adopts a purely point-based formulation instead of using intermediate representations such as bounding boxes or density maps. As shown in Fig. 8(a), the input image is projected by the backbone onto a feature map F_s, and K pre-set points are uniformly assigned to each patch according to the feature-map stride, yielding a pre-set point matrix $\mathcal{R}=\{R_k \mid k\in\{1,...,M\}\}$, where $H\times W\times K=M$ and each $R_k=(x_k,y_k)$ denotes the absolute coordinates of a reference point in the input image. Based on this representation, the network predicts point confidence $\hat{\mathcal{C}}=\{\hat{c}_j \mid \hat{c}_j\in[0,1], j\in\{1,...,M\}\}$, coordinate $\hat{\mathcal{P}}=\{\hat{p}_j=(\hat{x}_j,\hat{y}_j) \mid j\in\{1,...,M\}\}$, and category probabilities $\hat{\mathcal{T}}=\{\hat{t}_j \mid j\in\{1,...,M\}\}$ in parallel. The absolute coordinates of the predicted points are obtained by combining the reference points with the regressed offsets, and is defined as:

$$(\hat{x}_j,\ \hat{y}_j)=(x_j+\gamma\cdot\Delta x_j,\ y_j+\gamma\cdot\Delta y_j) \tag{2}$$

where γ is a normalization factor used to standardize the offset magnitude.



(a) Layout of the Preset Point Matrix (b) 1 v 1 Point Matching Principle

Fig. 8 - Analysis of Receptive Field Structure via PRFH under Hybrid Dilated Convolution

During training, a one-to-one matching strategy $\Omega\left(\mathcal{P},\hat{\mathcal{P}},\mathcal{D}\right)$ is adopted to associate predicted points with ground-truth points through the Hungarian algorithm, as shown in Fig. 8(b). The matching cost is defined as:

$$\mathcal{D}\left(\mathcal{P},\hat{\mathcal{P}}\right)=\left(\tau\left\|p_i-\hat{p}_j\right\|_2-\hat{c}_j\right)_{i\in N,j\in M} \tag{3}$$

where $\|\cdot\|_2$ denotes the l_2 distance, and τ is a weight balancing the distance term. In this way, predictions with smaller localization error and higher confidence are preferentially matched; the matched proposals are treated as positive samples, and the remaining proposals are assigned as negative samples for subsequent training.

Purely Point-based Prediction and Matching

The loss function $\mathcal{L}$ consists of three terms: Binary Cross-Entropy loss $\mathcal{L}_{obj}$ for point confidence, Euclidean loss $\mathcal{L}_{loc}$ for coordinate regression, and Cross-Entropy loss $\mathcal{L}_{cls}$ for point classification:

$$\mathcal{L}=\lambda_1\mathcal{L}_{obj}+\lambda_2\mathcal{L}_{loc}+\lambda_3\mathcal{L}_{cls} \tag{4}$$

$$\left(\mathcal{L}_{obj},\mathcal{L}_{loc},\mathcal{L}_{cls}\right)=\left(-\frac{1}{M}\left\{\sum_{i=1}^{N}\log\hat{c}_{\xi(i)}+\lambda_p\sum_{i=N+1}^{M}\log\left(1-\hat{c}_{\xi(i)}\right)\right\},\frac{1}{N}\sum_{i=1}^{N}\left\|p_i-\hat{p}_{\xi(i)}\right\|_2^2,\sum_{i=1}^{N}\left\{-t_i\log\hat{t}_{\xi(i)}-\left(1-t_i\right)\log\left(1-\hat{t}_{\xi(i)}\right)\right\}\right) \tag{5}$$

where $\|\cdot\|_{l2}$ denotes the Euclidean distance, λ_p is the reweighting factor for negative proposals, and $\lambda=[\lambda_1,\lambda_2,\lambda_3]$ balances the confidence, regression, and classification losses.

Evaluation Metrics

Three types of metrics were used to evaluate Soy-MOPNet and the compared models, include regression performance, classification performance, and efficiency. For point detection, counting, and regression, *MAE*, *RMSE*, and R^2 were adopted, and is defined as:

$$\left(MAE,RMSE,R^2\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\left|P_i-\hat{P}_i\right|,\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i-\hat{P}_i\right)^2},1-\left(\sum_i\left(P_i-\hat{P}_i\right)^2\Big/\sum_i\left(P_i-\bar{P}\right)^2\right)\right) \tag{6}$$

where n is the number of test images, P and $\hat{P}$ denote the ground-truth count and its mean, respectively, and $\bar{P}$ is the predicted count. For category classification, the task was simplified as a binary problem, with seeds as the positive class and stem nodes as the negative class.

Accuracy (*Acc*) and Precision (*Pre*) were used, and is defined as:

$$\left(Acc,Pre_{Seed},Pre_{Node}\right)=\left(\frac{TP+TN}{TP+FP+FN+TN},\frac{TP}{TP+FP},\frac{TN}{TN+FN}\right) \tag{7}$$

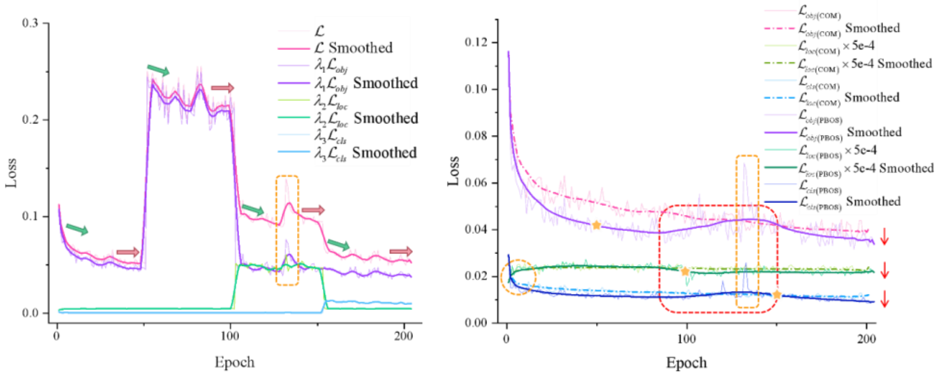
For efficiency evaluation, Parameters (*Param*) and Frames Per Second (*FPS*) were used to assess model compactness and inference speed.

RESULTS

PBOS Training Strategy and Implementation Details

The main hyperparameters were set as follows: $s=16$, $K=4$ to insure $M\gg N$, $U=[5,1,1]$, $\gamma=100$, $\tau=5\times10^{-2}$, $\lambda_p=0.5$, $\lambda=\left[\lambda_1,\lambda_2,\lambda_3\right]=\left[1,1\times10^{-3},0.5\right]$, $batchsize=8$, $lr=3\times10^{-4}$, $lr_{backbone}=5\times10^{-5}$, Adam optimizer, and $epoch=200$ which were equally divided into four stages.

To address the multi-branch and multi-task optimization problem, a Progressive Branch-wise Optimization Strategy (PBOS) was proposed. By sequentially increasing the loss weight of a single branch and training in stages, PBOS promotes both branch-wise and global optimization. In this study, training started with balanced optimization and then proceeded through confidence prediction, position regression, and classification stages. The confidence branch was prioritized because it determines proposal quality and directly affects localization and classification, whereas the classification branch was emphasized in the final stage to reduce interference from earlier optimization and improve downstream phenotypic structure analysis. Although this process may slightly constrain position regression, the densely pre-set candidate points still ensure robust localization performance.



(a) Weighted loss function (STTS) (b) Absolute loss function

Fig. 9 - Comparison of Weighted and Absolute Losses under PBOS and Conventional Training

In (a), the green arrows indicate the rapid loss reduction at the beginning of each stage, and the pink arrows indicate the stable convergence at the end of each stage, suggesting sufficient optimization within each stage. In (b), the yellow stars mark the beginning of each stage, and the red arrows indicate the performance gains of each branch under PBOS relative to conventional training. The red boxes and yellow circles highlight the collaboration between the confidence and classification branches, which rely on consistent features, and their competition with the localization branch, which depends on different features. The yellow box indicates the jump in the dilation coefficient in SDConv, which causes training fluctuation and alters the convergence direction.

As shown in Fig. 9, PBOS exhibits clear stage-wise optimization behavior. After each branch-weight adjustment, the total weighted loss decreases sharply and then converges stably within the stage, while the absolute loss curves show clear improvement of the target branch. More importantly, the final combined loss remains consistently lower than that of conventional training, indicating that PBOS improves both stage-wise convergence and global optimization, and ultimately yields better prediction performance.

Detection Performance of Soy-MOPNet

Soy-MOPNet achieved strong point-level detection performance for soybean seeds and stem nodes, with an *MAE* of 12.966, an *RMSE* of 19.182, an *Acc* of 0.943, and a *Pre* of 0.965 and 0.876 for seeds and stem nodes, respectively. Meanwhile, the model size was reduced to 18.8% of the initial design, indicating a favorable balance between accuracy and efficiency.

As shown in Fig. 10, Soy-MOPNet maintained stable detection across soybean plants with different shapes, orientations, and densities, and remained effective under dense distributions and complex backgrounds. The local results further indicate that Soy-MOPNet captures both fine-grained pod features and global plant posture, allowing it to detect seeds within elongated pods and distinguish interference from neighboring plants near image boundaries, while avoiding false counting of empty or burst pods. However, some false detections still occurred under rare interference patterns, such as field markers, curled dead leaves, and boundary disturbances, suggesting that the current square receptive field is still limited in capturing the longitudinal structure of soybean plants.

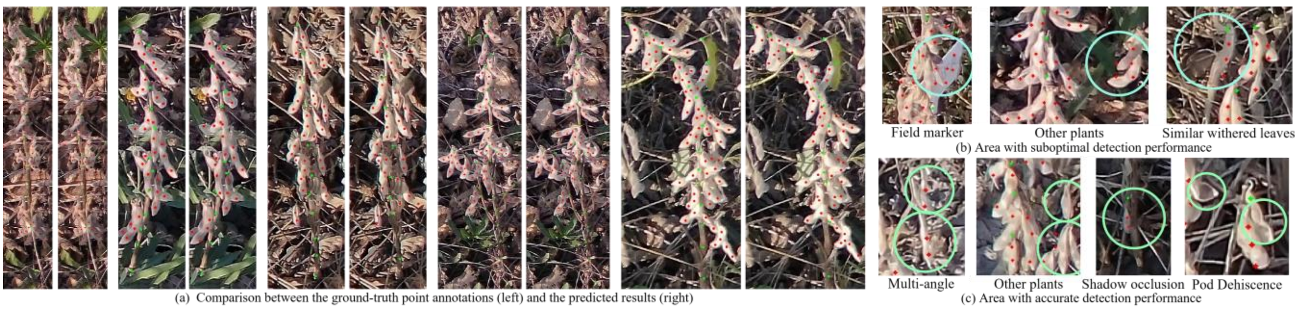


Fig. 10 - Visualization of detection results and local performance analysis

Comparative and Ablation Experiments

Table 1 presents the comparative and ablation results of Soy-MOPNet. The results show that Soy-MOPNet achieves the best overall trade-off among localization, classification, and efficiency, making it the most suitable framework for field soybean multi-organ detection.

Table 1

Performance Metrics of Comparative and Ablation Experiments

Experiment	Models	Counting and Localization Indexes			Lightweight Indexes		Classification Indexes		
		MAE	*RMSE*	R^2	*FPS*	*Param*	*Acc*	Pre_{seed}	Pre_{node}
Comparative Experiment	YOLOv8n	126.936	164.283	0.435	588	3.01M	0.572	0.599	0.485
	YOLOv10n	131.746	158.074	0.442	552	2.71M	0.554	0.574	0.449
	YOLOv10m	92.643	117.324	0.538	253	16.49M	0.599	0.695	0.525
	TasselNetV2+	25.023	33.251	0.672	239	31.35M	-	-	-
	SoybeanNet-S	10.327	12.453	0.841	152	52.04M	-	-	-
	P2PNet	11.237	14.925	0.738	249	21.58M	-	-	-
Ablation Experiment	VGG16+P2P	44.975	61.398	0.741	224	23.98M	0.808	0.923	0.451
	StarNet+P2P	50.292	74.942	0.737	472	4.50M	0.789	0.934	0.335
	ODConv+StarNet+P2P	37.834	48.707	0.824	476	4.51M	0.828	0.841	0.786
	SD_StarNet+P2P	25.726	31.224	0.809	452	4.51M	0.891	0.910	0.833
	Soy-MOPNet+PBOS(ours)	12.966	19.182	0.868	457	4.51M	0.943	0.965	0.876

Among the compared methods, YOLO supports classification but suffers from severe missed detections in dense and occluded scenes because overlapping anchor-based predictions are suppressed, particularly for stem nodes. TasselNetV2+ improves dense target modeling, but its density-based representation lacks explicit structural and category information, which limits robustness in complex field conditions. P2PNet and SoybeanNet-S further confirm the advantage of point-based detection for this task, whereas SoybeanNet-S relies on a Transformer backbone and therefore introduces substantially higher model complexity.

The ablation results verify the effectiveness of the proposed modules and strategy. The multiplicative lightweight backbone improves efficiency, dynamic convolution and adaptive dilation enhance the representation of dense, small, and structurally heterogeneous targets while improving the joint extraction and fusion of fine-grained local details and large-scale global plant posture features within the same layer, and PBOS further strengthens multi-branch optimization. Their combination ultimately yields the best overall performance in accuracy, speed, and robustness.

Multi-trait Regression Analysis for Mature Soybeans

Regression analysis was conducted for six phenotypic traits, including plant height, stem node count, seed count, yield per plant, main stem length, and lodging resistance. As shown in Fig. 11, except for lodging resistance, the other five traits showed strong linear correlations with the ground truth, indicating that Soy-MOPNet can effectively support multi-trait prediction in mature soybeans.

Among them, plant height achieved the highest consistency, further confirming the accuracy of UAV-ZSAR in geometric correction and scale recovery. Traits derived from seed detection generally showed higher accuracy than those relying on stem node localization, mainly because stem nodes were more easily affected by occlusion, structural complexity, and background interference. Otherwise, although larger deviations were observed in plants with more complex architecture, denser pod distribution, and stronger occlusion, the predicted values still followed clear regression trends rather than becoming highly scattered, indicating that the model preserved stable quantitative characterization ability under natural field conditions.

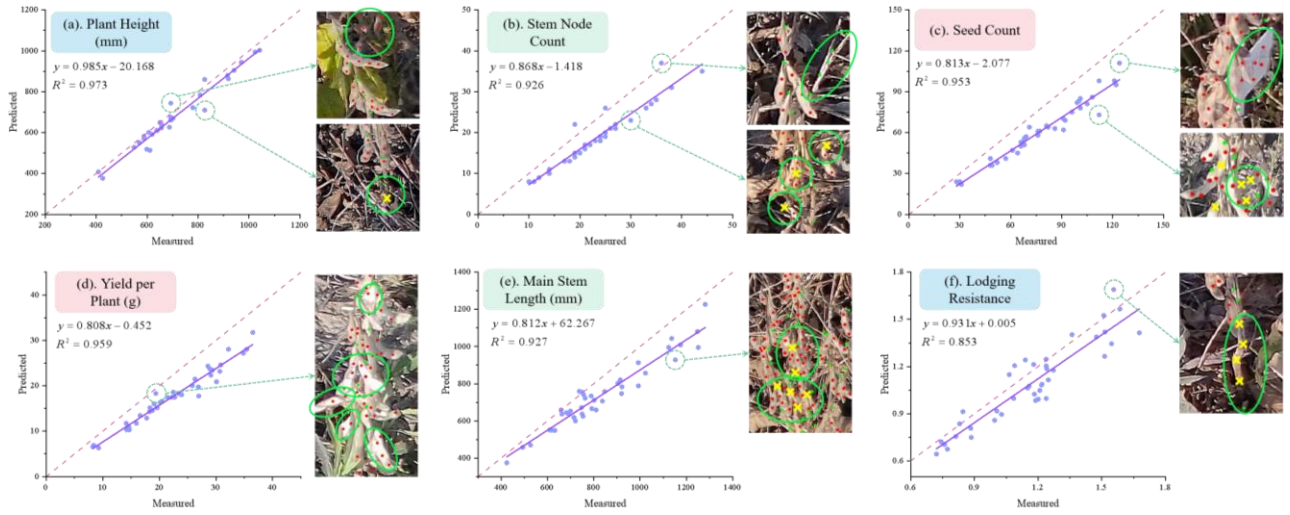
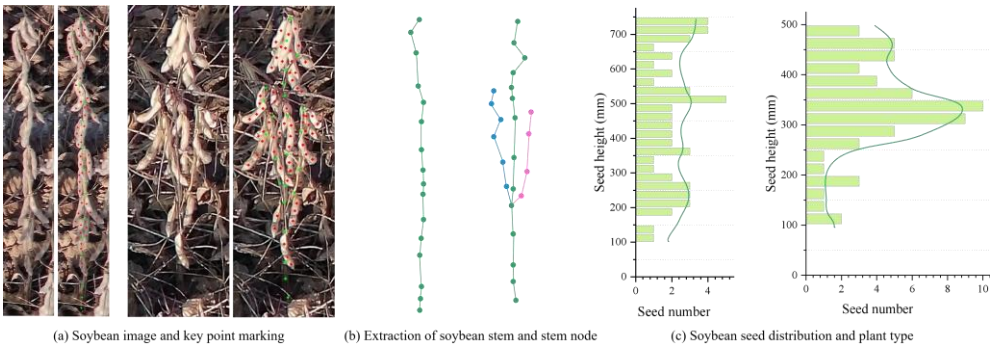


Fig. 11 - Visualization of detection results and local performance analysis



(a) Soybean image and key point marking (b) Extraction of soybean stem and stem node (c) Soybean seed distribution and plant type

Fig. 12 - Visualization of detection results and local performance analysis

Based on the detected stem nodes, the stem architecture was further reconstructed using a directional search algorithm, as shown in Fig. 12, enabling the extraction of additional structural traits such as branch number, stem curvature, vertical seed distribution, bottom pod height, and effective stem length. These results indicate that the proposed framework can provide not only key maturity traits but also richer structural phenotypes, offering practical support for breeding selection and mechanized harvesting.

CONCLUSIONS

To address parallel multi-phenotype extraction of mature soybeans in complex and dense field environments, this study established a lightweight point-based multi-organ phenotyping framework for UAV imagery. By combining oblique-image rectification, dynamic feature modeling, and progressive multi-branch optimization, the proposed method enables effective extraction of soybean organ information and stable prediction of multiple phenotypic traits under field conditions. Its effectiveness has been verified through systematic experiments, demonstrating its potential for high-throughput phenotyping, intelligent breeding, and related agricultural applications.

ACKNOWLEDGEMENT

This research was funded by the Central Guided Local Science and Technology Development Project (YDZX2024060), the Key Project of Shandong Marine Service Industry (First Batch, 2025), and supported by the Taishan Industrial Experts Program (tscy20251042).

REFERENCES

[1] Ariza-Sentís, M., Baja, H., Vélez, S., Valente, J., (2023). Object detection and tracking on UAV RGB videos for early extraction of grape phenotypic traits. *Comput. Electron. Agric.* 211, 108051. https://doi.org/10.1016/j.compag.2023.108051

[2] Awais, M., Wang, X., Hussain, S., Aziz, F., Mahmood, M.Q., (2025). Advancing Precision Agriculture Through Digital Twins and Smart Farming Technologies: A Review. *AgriEngineering* 7, 137.

[3] Chen, Q., Liao, S., Kui, X., Hu, Z., Zhou, J., (2025). Paying attention to the minute details: Supervised keypoint detection on dense, complex point clouds. *Eng. Appl. Artif. Intell.* 151, 110668. https://doi.org/10.1016/j.engappai.2025.110668

[4] De Pretto, C., Giordano, R.D.C., Tardioli, P.W., Costa, C.B.B., (2018). Possibilities for Producing Energy, Fuels, and Chemicals from Soybean: A Biorefinery Concept. *Waste Biomass Valorization* 9, 1703-1730. https://doi.org/10.1007/s12649-017-9956-3

[5] Feng, L., Chen, S.S., Zhang, C., Zhang, Y.C., He, Y., (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Comput. Electron. Agric.* 182, 106003. https://doi.org/10.1016/j.compag.2021.106033

[6] Guo, Y.X., Gao, Z.Q., Zhang, Z.G., Li, Y., Hu, Z.B., Xin, D.W., Chen, Q.S., Zhu, R.S., (2022). Automatic and Accurate Acquisition of Stem-Related Phenotypes of Mature Soybean Based on Deep Learning and Directed Search Algorithms. *Front. Plant Sci.* 13, 906751. https://doi.org/10.3389/fpls.2022.906751

[7] He, H.T., Ma, X.D., Guan, H.O., (2022). A calculation method of phenotypic traits of soybean pods based on image processing technology. *Ecological Informatics* 69, 101676. https://doi.org/10.1016/j.ecoinf.2022.101676

[8] He, H.T., Ma, X.D., Guan, H.O., Wang, F.Y., Shen, P.P., (2023). Recognition of soybean pods and yield prediction based on improved deep learning model. *Front. Plant Sci.* 13, 1096619. https://doi.org/10.3389/fpls.2022.1096619

[9] He, J., Weng, L., Xu, X., Chen, R., Peng, B., Li, N., Xie, Z., Sun, L., Han, Q., He, P., Wang, F., Yu, H., Bhat, J.A., Feng, X., (2024). DEKR-SPrior: An Efficient Bottom-Up Keypoint Detection Model for Accurate Pod Phenotyping in Soybean. *Plant Phenomics* 6, 0198. https://doi.org/10.34133/plantphenomics.0198

[10] Hernandez, C.M., Correndo, A., Kyveryga, P., Prestholt, A., Ciampitti, I.A., (2023). On-farm soybean seed protein and oil prediction using satellite data. *Comput. Electron. Agric.* 212, 108096.

[11] Hu, Y., Du, H., Lu, J., Zhang, H., Li, S., Du, X.,(2024). Interface synergistic stabilization of zinc anodes via polyacrylic acid doped polyvinyl alcohol ultra-thin coating. *Journal of Energy Storage* 87, 111444.

[12] Kanglei, W., Tan, W., Yuan, R., Xiu, J., Xiaobo, W., Jiajia, L., Zhe, Z., Zhaohui, J., Xing, S., Wu, Z.,(2025). Practical framework for generative on-branch soybean pod detection in occlusion and class imbalance scenes. *Eng. Appl. Artif. Intell.* 139, 109613.

[13] Kerec, J., Machidon, A.L., Machidon, O.M., (2026). Deployment-Aware NAS for Lightweight UAV Object Detectors in Precision Agriculture Crop Monitoring. *AgriEngineering* 8, 43.

[14] Li, J.J., Magar, R.T., Chen, D., Lin, F., Wang, D.C., Yin, X., Zhuang, W.C., Li, Z.J., (2024). SoybeanNet: Transformer-based convolutional neural network for soybean pod counting from Unmanned Aerial Vehicle (UAV) images. *Comput. Electron. Agric.* 220, 108861. https://doi.org/10.1016/j.compag.2024.108861

[15] Li, S., Yan, Z.Z., Guo, Y.X., Su, X.Y., Cao, Y.Y., Jiang, B.F., Yang, F., Zhang, Z.G., Xin, D.W., Chen, Q.S., Zhu, R.S., (2022). SPM-IS: An auto-algorithm to acquire a mature soybean phenotype based on instance segmentation. *Crop Journal* 10, 1412-1423. https://doi.org/10.1016/j.cj.2021.05.014

[16] Li, Y.L., Zhan, X.H., Liu, S.Y., Lu, H., Jiang, R.B., Guo, W., Chapman, S., Ge, Y.F., de Solan, B., Ding, Y.F., Baret, F., (2023). Self-Supervised Plant Phenotyping by Combining Domain Adaptation with 3D Plant Model Simulations: Application to Wheat Leaf Counting at Seedling Stage. *Plant Phenomics* 5, 0041. https://doi.org/10.34133/plantphenomics.0041

[17] Liu, F., Liu, H., Wu, Q., Han, Z., Pang, S., Wang, S., Zhao, L., (2025). Pod-pose : an efficient top-down keypoint detection model for fine-grained pod phenotyping in mature soybean. *Plant Methods* 21, 82. https://doi.org/10.1186/s13007-025-01399-0

[18] Lu, W., Du, R.T., Niu, P.S., Xing, G.N., Luo, H., Deng, Y.M., Shu, L., (2022). Soybean Yield Preharvest Prediction Based on Bean Pods and Leaves Image Recognition Using Deep Learning Neural Network Combined With GRNN. *Front. Plant Sci.* 12, 791256. https://doi.org/10.3389/fpls.2021.791256

[19] Maraveas, C.,(2024). Image Analysis Artificial Intelligence Technologies for Plant Phenotyping: Current State of the Art. *AgriEngineering* 6, 3375-3407.

[20] Oliveira, B.R.d., Sobrinho, R.L., Ferreira, F.R.T., Putti, F.F., Bodini, M., Saporetti, C.M., Goliatt, L., (2025). Machine Learning-Based Prediction of Soybean Plant Height from Agronomic Traits Across Sequential Harvests. *AgriEngineering* 7, 408.

[21] Shen, Y.T., Lu, X.Q., Lyu, M.Q., Zhou, H.Y., Guan, W.X., Jiang, L.X., He, Y.H., Cen, H.Y., (2024). High-throughput phenotyping of individual plant height in an oilseed rape population based on Mask-RCNN and UAV images. *Precision Agriculture* 25, 811-833. https://doi.org/10.1007/s11119-023-10095-9

[22] Song, Q., Wang, C., Jiang, Z., Wang, Y., Tai, Y., Wang, C., Li, J., Huang, F., Wu, Y., (2021). Rethinking Counting and Localization in Crowds:A Purely Point-Based Framework, p. arXiv:2107.12746.

[23] Wang, G., Jiang, B., Liu, Y., Wang, L., Zhang, Y., Yan, J., Wang, K., (2025). Source-Load Coordinated Optimization Framework for Distributed Energy Systems Using Quasi-Potential Game Method. *Protection and Control of Modern Power Systems* 10, 103-122.

[24] Xie, C.Q., Yang, C., (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Comput. Electron. Agric.* 178, 105731. https://doi.org/10.1016/j.compag.2020.105731

[25] Yang, S., Zheng, L., Wu, T., Sun, S., Zhang, M., Li, M., Wang, M., (2024). High-throughput soybean pods high-quality segmentation and seed-per-pod estimation for soybean plant breeding. *Eng. Appl. Artif. Intell.* 129, 107580. https://doi.org/10.1016/j.engappai.2023.107580

[26] Yao, M.W., Li, W., Chen, L., Zou, H.J., Zhang, R., Qiu, Z.J., Yang, S., Shen, Y., (2024a). Rice Counting and Localization in Unmanned Aerial Vehicle Imagery Using Enhanced Feature Fusion. *Agronomy-Basel* 14, 868. https://doi.org/10.3390/agronomy14040868

[27] Yao, Z.S., Zhang, D.S., Tian, T., Zain, M., Zhang, W.J., Yang, T.L., Song, X.X., Zhu, S.L., Liu, T., Ma, H.J., Sun, C.M.,(2024b). APW: An ensemble model for efficient wheat spike counting in unmanned aerial vehicle images. *Comput. Electron. Agric.* 224, 109204. https://doi.org/10.1016/j.compag.2024.109204

[28] Yu, Z.H., Wang, Y.X., Ye, J.X., Liufu, S.J., Lu, D.L., Zhu, X.L., Yang, Z.M., Tan, Q.J., (2024). Accurate and fast implementation of soybean pod counting and localization from high-resolution image. *Front. Plant Sci.* 15, 1320109. https://doi.org/10.3389/fpls.2024.1320109

[29] Zhang, C.X., Lu, X., Ma, H.M., Hu, Y.H., Zhang, S.A., Ning, X.M., Hu, J.W., Jiao, J., (2023). High-Throughput Classification and Counting of Vegetable Soybean Pods Based on Deep Learning. *Agronomy-Basel* 13, 1154. https://doi.org/10.3390/agronomy13041154

[30] Zhang, Y.H., Liu, M.F., He, J.B., Wang, Y.F., Xing, G.N., Li, Y., Yang, S.P., Zhao, T.J., Gai, J.Y., (2015). Marker-assisted breeding for transgressive seed protein content in soybean Glycine max (L.) Merr. *Theor. Appl. Genet.* 128, 1061-1072. https://doi.org/10.1007/s00122-015-2490-4

[31] Zhang, Z., Jin, X., Rao, Y., Wan, T.Y., Wang, X.B., Li, J.J., Chen, H.R., Wu, K.L., Kong, F.C., Tian, Z., Shao, X., (2024). DSBEAN: An innovative framework for intelligent soybean breeding phenotype analysis based on various main stem structures and deep learning methods. *Comput. Electron. Agric.* 224, 109135. https://doi.org/10.1016/j.compag.2024.109135

[32] Zhao, J.S., Kaga, A., Yamada, T., Komatsu, K., Hirata, K., Kikuchi, A., Hirafuji, M., Ninomiya, S., Guo, W., (2023). Improved Field-Based Soybean Seed Counting and Localization with Feature Level Considered. *Plant Phenomics* 5, 0026. https://doi.org/10.34133/plantphenomics.0026