Data-Efficient and Accurate Rapeseed Leaf Area Estimation by Self-supervised Vision Transformer for Germplasms Early Evaluation

Pengfei Hao
Zhejiang Academy of Agricultural Sciences

Jianpeng An
Tianjin University

Qing Cai
Tiangong University

Junqin Cao
Zhejiang Academy of Agricultural Sciences

Chaochao He
Zhejiang Academy of Agricultural Sciences

Zhiqi Ma
Zhejiang Academy of Agricultural Sciences

Shuijin Hua
Zhejiang Academy of Agricultural Sciences

Baogang Lin
bglin@126.com

Zhejiang Academy of Agricultural Sciences

Research Article

Keywords: Deep Learning, Leaf Area Estimation, Self-Supervised Learning, Plant Phenotyping, Rapeseed, Breeding

Posted Date: August 20th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7278813/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Methods on December 5th, 2025.
See the published version at https://doi.org/10.1186/s13007-025-01478-2.

Data-Efficient and Accurate Rapeseed Leaf Area Estimation by Self-supervised Vision Transformer for Germplasms Early Evaluation

Pengfei Hao[a,1], Jianpeng An[b,1], Qing Cai[c], Junqin Cao[a], Chaochao He[a], Zhiqi Ma[a], Shuijin Hua[a] and Baogang Lin[a,*]

[a]*Institute of Crops and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, 310021, Zhejiang, China*
[b]*School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, Tianjin, China*
[c]*School of Artificial Intelligence, Tiangong University, Tianjin, 300387, Tianjin, China*

ARTICLE INFO

Keywords:
Deep Learning
Leaf Area Estimation
Self-Supervised Learning
Plant Phenotyping
Rapeseed
Breeding

ABSTRACT

Early-stage, accurate and high-throughput phenotyping through leaf area estimation is critical for future rapeseed breeding, but faces two key constraints: expensive data annotation and persistent challenge of leaf occlusion. To address these issues, we present a data-efficient deep learning framework using smartphone-captured top-down RGB images for rapeseed leaf area quantification. Our approach utilizes a two-stage strategy where a Vision Transformer (ViT) backbone is first pre-trained on a large, aggregated corpus of diverse, non-rapeseed public plant datasets using the DINOv2 self-supervised learning method. This pre-trained model is then fine-tuned on a custom rapeseed dataset using a novel Canopy-Mix data augmentation technique to handle fragmented views analogous to occlusion, and a hybrid loss function combining Smooth L1 and Log-Cosh for robust convergence. Through rigorous 5-fold cross-validation, our proposed model achieved state-of-the-art predictive performance (Coefficient of Determination, R^2=0.805). What's more, the predicted leaf area demonstrated a remarkably strong correlation with both fresh weight (r=0.900) and dry weight (r=0.885). The model significantly outperformed a range of baselines, including models trained from scratch, those pre-trained on ImageNet, and a heuristic method based on manually annotated bounding boxes. Ablation studies confirmed the essential contribution of each component, while qualitative analysis of attention maps demonstrated the model's ability to precisely localize the leaf canopy and ignore background distractors. This study demonstrates that domain-specific self-supervised pre-training offers a powerful solution to overcome data limitations in agricultural vision, providing a robust and scalable tool for non-destructive phenotyping that can potentially accelerate the rapeseed breeding cycle.

1. Introduction

As the world's largest oilseed consumer, China has long faced a chronic shortage in oilseed supply (Liu et al., 2019; Mehmood et al., 2021). While rapeseed breeders have dedicated significant efforts to enhancing yield, quality, and stress tolerance in rapeseed varieties, three critical constraints persist: (1) the excessively long total growth period of winter rapeseed, (2) time-consuming field characterization processes, and (3) labor shortages compounded by an aging agricultural workforce (Sun et al., 2024). These challenges underscore the urgent need for accurate and efficient rapeseed growth monitoring systems to support breeder decision-making and accelerate breeding cycles (Li et al., 2022b).

Leaf Area Index (LAI) and Above Ground Biomass (AGB) serve as pivotal physiological indicators for crop growth monitoring, providing direct insights into growth rates, water/nutrient status, and stress resistance (Araus and Cairns, 2014; Li et al., 2020). Early-stage estimation of these parameters is particularly crucial, enabling breeders to make informed decisions regarding: (1) plant nutrition management, (2) growth status evaluation, (3) preliminary material selection, (4) yield forecasting, and (5) field cultivation strategies (Murchie and Burgess, 2022; Li et al., 2022a).

While the importance are well-established, their accurate and efficient measurement remains a significant challenge. Traditional methods of measuring LAI and AGB were punching-weighting method for LAI and destructive sampling weighting for AGB. However, although such methods were easy to implement and had high precision, it relied on the

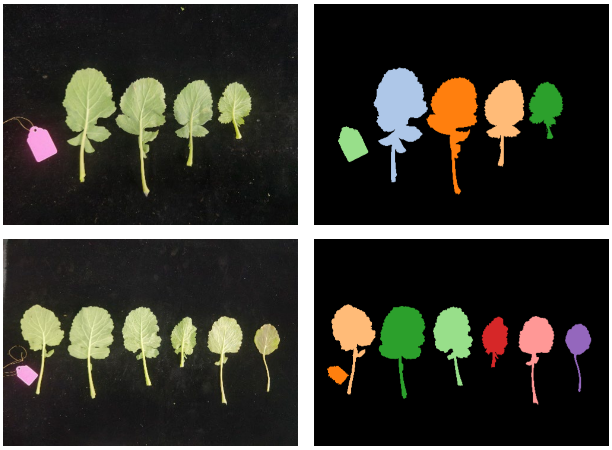
*Corresponding author
bglin@126.com (B. Lin)
[1]These authors contributed equally to this work.

professional level of inspectors, which had the disadvantages of strong subjectivity, weak real-time performance, and high labor costs. As a result, it cannot meet the development needs of efficient and intelligent agriculture at field scale (Chang et al., 2017; Zeng et al., 2018).

In the past decades, as the appearance and development of computer techniques, especially with the combination of image information acquisition, such as smart phone or Unmanned Aerial Vehicle (UAV), contributes a lot to achieve this goal (A et al., 2021; Shao et al., 2022; Yue et al., 2023). Recently, researches have conducted lots of studies on AI+agrculture, for example estimation of cotton leaf area using UAV point cloud data (Zhang et al., 2024), assessment of bergamot essential oil content using smartphone images (Anello et al., 2024), estimation of tea yield using UAV images and deep learning (Wang et al., 2024), AI-lot based plant diseases detection and treatment (Ibrahim et al., 2025), and so on. These results demonstrate the capacity of machine and deep learning models to interrogate large datasets, generate new and testable outputs and predict crop behaviour, highlighting the powerful role of data in the future of agriculture (Newman and Furbank, 2021). However, early computational methods often relied on classical computer vision techniques, such as color space transformation (e.g., to HSV or Lab*), followed by thresholding or clustering to segment plant pixels from the background (Li et al., 2024; Zhou et al., 2020). What's more, the performance of these methods is notoriously fragile, showing high sensitivity to fluctuations in ambient light, shadows, and the complexity of the background, such as varying soil color or the presence of debris (Guo et al., 2023).

The advent of deep learning, especially Convolutional Neural Networks (CNNs) and more recently Vision Transformers (ViTs), marked a significant leap forward, achieving state-of-the-art results in plant segmentation and phenotyping tasks (Murphy et al., 2024; Lin et al., 2025). However, an approach based purely on segmentation is inherently constrained by a fundamental challenge in plant analysis: leaf occlusion, where overlapping leaves—a natural aspect of plant growth—hide significant portions of the total leaf area from the camera's view (Chen et al., 2023). Consequently, methods that only quantify visible pixels will systematically and increasingly underestimate the true leaf area as the plant canopy develops (Miao et al., 2021). Beyond this conceptual limitation, the effectiveness of this supervised learning paradigm is contingent upon overcoming a critical bottleneck: the immense data requirement. Training powerful models like CNNs or ViTs demands vast, manually annotated datasets, for which creating pixel-perfect masks is an extremely tedious and costly process (Singh et al., 2018; Nagasubramanian et al., 2022). This challenge is exacerbated by the problem of domain shift; a model meticulously trained on one crop or in a controlled environment often fails to generalize to a new species—such as rapeseed in our case—or to more variable field conditions (Ilyas et al., 2023; Zhu et al., 2025). Addressing the occlusion problem therefore requires moving beyond simple segmentation to a method that can infer total area, which necessitates a model capable of learning implicit knowledge about plant morphology (Fan et al., 2022). Thus, a clear research gap exists for a methodology that is not only high-throughput and robust but can also overcome the reliance on extensive labeled datasets and effectively estimate total leaf area in the presence of occlusion.

To address these challenges, we pivot from the traditional supervised learning paradigm and instead leverage the power of self-supervised learning (SSL), a methodology that learns rich visual representations from unlabeled data, thereby mitigating the need for extensive manual annotation (Jing and Tian, 2020). Specifically, we employ DINOv2, a state-of-the-art SSL framework that utilizes a ViT backbone (Oquab et al., 2024). This approach has been shown to excel at learning highly generalizable and semantically meaningful features directly from images, making it an ideal candidate for diverse visual recognition tasks (Naseer et al., 2021). Our core strategy is built on a two-stage transfer learning process designed to maximize data efficiency and model performance (Yosinski et al., 2014). In the first stage, we pre-train the DINOv2 model on a large, aggregated corpus of varied public plant datasets. Crucially, this pre-training data contains numerous crop species and leaf types but deliberately excludes our target, rapeseed. This process forces the model to learn a robust and fundamental representation of plant morphology, a technique that has proven effective for domain adaptation in other data-scarce fields like medical imaging (Azizi et al., 2023). Subsequently, this powerful, pre-trained feature extractor is fine-tuned using our small, specific dataset of 833 rapeseed images to perform leaf area regression. By directly predicting a continuous value, our model learns to infer the total leaf area, implicitly accounting for overlapping regions, rather than simply segmenting visible pixels (Zdrazil et al., 2025). To further enhance model robustness and prevent overfitting on the limited fine-tuning data, we also introduce and evaluate a novel, improved Mixup data augmentation technique. Unlike standard Mixup (Zhang et al., 2018), our proposed method is designed to create more challenging and realistic interpolated samples that better preserve the structural integrity of plant canopies (Wang et al., 2025). Through this combined approach, we hypothesize that we can develop an accurate and high-throughput model for rapeseed leaf area estimation that effectively overcomes the key limitations of data dependency and occlusion inherent in previous methods.



(a) (b)

Figure 1: Example of leaf masks. (a) The original input image. (b) The corresponding segmentation mask, where each color represents a distinct object instance.

The primary objective of this study is to develop and validate a data-efficient deep learning framework for the accurate and high-throughput estimation of rapeseed leaf area from top-down RGB images. This research directly addresses the practical challenges faced by rapeseed breeders, aiming to automate the laborious process of manual phenotyping. To achieve this, we pursue three specific goals: (i) to investigate the efficacy of self-supervised pre-training on diverse, non-target plant datasets as a strategy to create robust and generalizable features; (ii) to propose and systematically evaluate an improved Mixup data augmentation technique for enhanced model robustness; and (iii) to integrate these components into a final regression model and rigorously validate its predictive accuracy. By achieving these goals, this work aims to significantly increase the efficiency of breeding pipelines and ultimately contribute to shortening the rapeseed breeding cycle.

Our study represents a preliminary attempt to integrate deep learning with rapeseed breeding. Designed to address practical challenges encountered by breeders, this research aims to replace manual labor with AI technology, thereby enhancing efficiency and significantly shortening the breeding cycle.

2. Materials and Methods

2.1. Field trials and plant materials

This study was conducted at Huzhou academy of agricultural sciences, Zhejiang province, China (30°48'35.8"N, 120°11'32.4"E) during 2024-2025. The rapeseed (Brassica napus L.) were sowing at October 20th, and the seedlings were transplanted to the field at November 15th, 2024, with planting density at 225,000 plants·hm^2. The soil type was paddy soil, and the fertilization rates was 750 kg·hm^2 rapeseed specific slow-release fertilizer (N;P_2O_5:K_2O=25:7:8). The field management of all the plants were the same during all the time.

Total of 833 different rapeseeds materials which differs mainly on leaf area, leaf type and growth were select and top-down photoed at January 1st 2025. All the materials were growing at seedling to 5-leaf stage.

2.2. Acquisition of rapeseed leaf images and measurement of ground-truth leaf area

The acquisition of the images were carried out in the field, and a stick marked at 60 cm was used to keep the shooting height the same. The parameters of the mobile camera (Xiaomi 14 Pro, 16 GB RAM+512 GB ROM, Qualcomm snapdragon 8 Gen 3) was f/1.42, 1/954s, ISO 50. Each image was matched to a rapeseed seedling and a 4 cm * 3 cm label paper.

After photoed, all the leaves of each plant were picked off and tiled to a black background to ensure high contrast. A transparent glass plate were pressed against the leaves to keep them unfolded. The images were then processed using a computer vision pipeline implemented with OpenCV (Bradski, 2000). This process first identified a pink-colored calibration object of a known size in each image to establish a pixel-to-cm^2 conversion ratio. Subsequently, leaf regions

Table 1
Public Datasets Used for Self-Supervised Pre-training.

Dataset Name	No. of Images	Description	Key Features
CVPPP (Scharr et al., 2014; Minervini et al., 2014)	810	Top-down images of *Arabidopsis* and tobacco rosettes in controlled settings.	Rosette structure, leaf overlap patterns.
Flavia (Wu et al., 2007)	1,907	Scanned images of single, detached leaves from 32 different species.	High-fidelity leaf shape, edge, and vein details from a wide variety of species.
Plant Pathology 2021	18,635	Field images of apple leaves exhibiting various health statuses.	Disease symptoms, and variations in leaf color, texture, and health.
Plant Seedlings (Giselsson et al., 2017)	408	Top-down images of seedlings of 12 different crop and weed species.	Early growth stages, inter-species variation among young plants.
VegAnn (Madec et al., 2023)	3,775	Field images of 26+ vegetable crops and weeds in cluttered agricultural scenes.	Complex backgrounds, real-world shadows, and diverse lighting conditions.

were segmented based on a predefined color range in the HSV color space, with the resulting binary mask used to calculate the final area, as illustrated in Figure 1.

2.3. Measurement of rapeseed fresh weight and dry weight

After finishing the leaf area laboratory photoing, all the above ground organs were weighted using a high-precision balance (accuracy at 0.01 g) for fresh weight. Then all the samples were placed to an oven at 105°C for 1h, and followed by 75°C until the weight keep the same, and the dry weight were measured the same as the fresh weight.

2.4. Public Datasets for Self-Supervised Pre-training

To build a robust and generalized feature extractor, we leveraged a large, aggregated corpus of images from five publicly available plant-related datasets. The deliberate selection of these datasets was designed to provide the model with a broad and diverse understanding of plant morphology, leaf texture, and appearance under various conditions, none of which include rapeseed. A summary of these datasets is provided in Table 1.

Our pre-training corpus was constructed to capture a wide array of visual information. It includes the CVPPP dataset (Scharr et al., 2014; Minervini et al., 2014), which provides clear, top-down views of Arabidopsis and tobacco rosettes to teach the model fundamental overlap patterns. The Flavia dataset (Wu et al., 2007) contributes high-resolution scans of single leaves from 32 species, allowing the model to learn intricate shape and venation details on clean backgrounds. To ensure real-world applicability, we incorporated the Plant Pathology 2021 dataset[2] for its diversity in leaf health status and the VegAnn dataset (Madec et al., 2023) for its complex field backgrounds and varied lighting conditions across 26+ crop species. Finally, the Plant Seedlings dataset (Giselsson et al., 2017) was included to teach the model to recognize 12 different species at various early growth stages.

2.5. Model Architecture and Training Strategy

Our methodology for leaf area estimation is based on a two-stage transfer learning framework, designed to maximize data efficiency and predictive accuracy. The overall pipeline is illustrated in Figure 2. The first stage consists of self-supervised pre-training, where a ViT backbone learns robust and generalizable visual representations from a large, unlabeled corpus of diverse public plant imagery. In the second stage, this pre-trained backbone is adapted for our specific task through supervised fine-tuning. During this stage, the model, appended with a regression head, is trained on our labeled rapeseed dataset to predict leaf area. To enhance robustness and mitigate overfitting on our

[2] https://www.kaggle.com/competitions/plant-pathology-2021-fgvc8

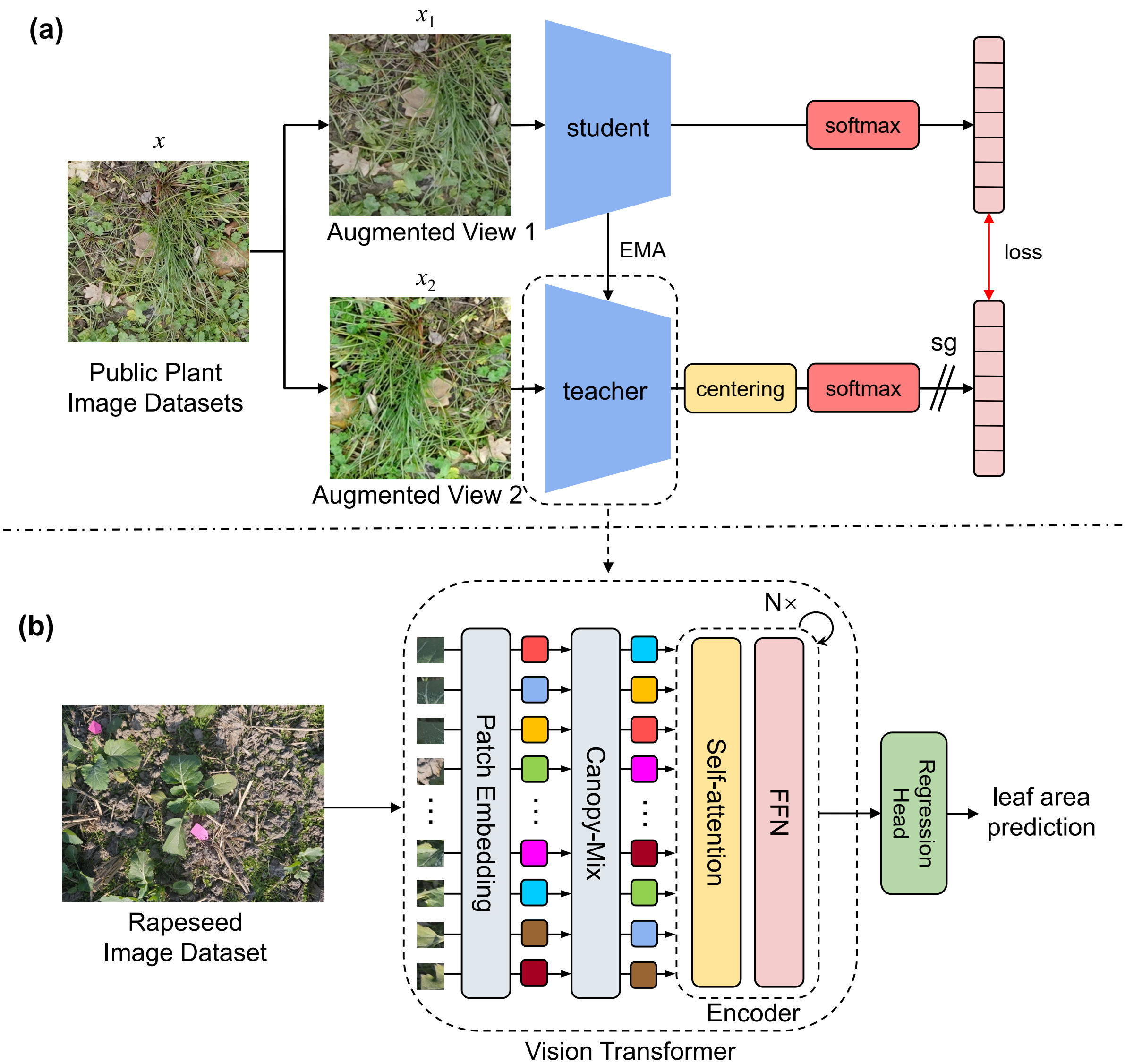


Figure 2: Overview of our proposed two-stage framework. (a) The self-supervised pre-training stage. A ViT based student network is trained to match the output of a teacher network on a large corpus of unlabeled public plant datasets. The teacher's weights are an EMA of the student's weights, and a stop-gradient (sg) operation is applied to the teacher's output to prevent model collapse. (b) The supervised fine-tuning stage for leaf area regression. An input rapeseed image is converted into patch embeddings, processed by our proposed Canopy-Mix augmentation at the token level, and then passed through the ViT Encoder. A final regression head maps the model's output features to a continuous leaf area prediction.

limited dataset, we incorporate a novel data augmentation strategy during this fine-tuning process. The subsequent subsections will detail each component of this framework.

2.5.1. Self-Supervised Pre-training with DINOv2

To build a powerful feature extractor without relying on labeled data, we employed a self-supervised pre-training strategy based on the DINOv2 framework (Oquab et al., 2024). DINOv2 is a state-of-the-art method that learns robust visual representations by distilling knowledge from a teacher network to a student network, using only a large corpus of unlabeled images. This approach is particularly well-suited for our study, as it allows the model to learn fundamental patterns of plant morphology from a wide range of crop imagery before being exposed to our specific rapeseed dataset.

The core mechanism of DINOv2 is a form of self-distillation, as illustrated in Figure 2 (a). Both the student and teacher networks share the same Vision Transformer (ViT-Base) architecture. During training, a single input image from our aggregated public plant datasets is processed to create multiple augmented views, specifically a set of high-resolution "global" crops and lower-resolution "local" crops. All crops are passed through the student network, while only the global crops are passed through the teacher network. The objective is to train the student network to match the output probability distribution of the teacher network for the same global crops. This probability distribution, P, is generated from the network's output logits, z, using a temperature-scaled softmax function:

$$P(z)_i = \frac{\exp(z_i/\tau)}{\sum_{k=1}^{K} \exp(z_k/\tau)} \tag{1}$$

where τ is a temperature parameter that controls the sharpness of the distribution, and K is the dimension of the output logits. Different temperatures are used for the student (τ_s) and the teacher (τ_t). The final training objective is achieved by minimizing the cross-entropy loss between these two distributions.

A critical component of this framework is how the teacher network is updated. Instead of being trained via backpropagation, the teacher's weights (θ_t) are an exponential moving average (EMA) of the student's weights (θ_s). This update is performed at each training step according to the following rule:

$$\theta_t \leftarrow \lambda\theta_t + (1-\lambda)\theta_s \tag{2}$$

where λ is a momentum coefficient that smoothly increases from 0.996 to 1 during training. This process, which incorporates a stop-gradient operation to prevent trivial solutions, creates a more stable and effective "teacher," guiding the student to learn discriminative and semantically rich features. For our study, this pre-training phase was conducted on the combined dataset of over 25,535 images from the CVPPP, Flavia, Plant Pathology 2021, Plant Seedlings, and VegAnn datasets for 300 epochs, leveraging the official DINOv2 implementation and its recommended hyperparameters. The result of this stage is a ViT backbone with a deep understanding of general plant features, poised for effective fine-tuning on our specific task.

2.5.2. Fine-tuning for Leaf Area Regression

Following the self-supervised pre-training stage, the ViT backbone, now enriched with general-purpose visual features, was adapted for the specific downstream task of leaf area estimation through supervised fine-tuning. This stage leverages our labeled rapeseed dataset to teach the model to map learned features to a continuous scalar output.

To achieve this, the pre-trained ViT backbone was utilized as the primary feature extractor. We appended a lightweight regression head to the model, which consists of a Multi-Layer Perceptron (MLP) with 2 fully-connected layers and a ReLU activation function. This head is responsible for taking the high-dimensional class token embedding from the ViT output and mapping it to a single numerical value, which corresponds to the predicted leaf area in square centimeters (cm^2).

The fine-tuning process was conducted on our custom dataset of 833 labeled rapeseed images. During this stage, the weights of the entire model were unfrozen to allow for end-to-end training. A differential learning rate was employed, where the newly added regression head was trained with a higher learning rate, while the pre-trained ViT backbone was updated with a much smaller learning rate. This strategy allows the model to preserve the powerful features learned during pre-training while subtly adapting them to the specific visual characteristics of our rapeseed data. To prevent overfitting and enhance model generalization on this limited dataset, our improved Mixup data augmentation technique was applied to the training set. The model was trained to minimize the Mean Squared Error (MSE) between the predicted and ground-truth leaf areas, using the AdamW optimizer for 200 epochs. This fine-tuning process yields the final specialized model, optimized for accurate rapeseed leaf area prediction.

2.5.3. Canopy-Mix Augmentation

A primary difficulty in accurately estimating leaf area from top-down imagery stems from two related challenges: leaf occlusion, which renders parts of the canopy invisible, and the presence of background noise, such as soil, shadows, or other non-target plants. Standard data augmentations like linear interpolation (i.e., standard Mixup) are ill-suited for these structural challenges, as they create physically implausible, semi-transparent images rather than realistic examples of occlusion. To address this, we propose Canopy-Mix, a new data augmentation strategy designed to simultaneously enhance the model's ability to handle occlusion and improve its robustness to contextual noise.

The core objective of Canopy-Mix is to explicitly train the model to infer total area from fragmented and incomplete visual evidence, a situation directly analogous to real-world occlusion. By shuffling and recombining token patches from two different plant images, we create complex, chimeric canopy structures. This process forces the model to learn how to aggregate leaf area from non-contiguous patches, a critical skill for evaluating an occluded plant where the canopy is perceived as a collection of disconnected parts. In doing so, the model is compelled to develop a more fundamental understanding of leaf morphology itself, rather than memorizing the holistic shape of a "typical" plant.

Simultaneously, this token-shuffling mechanism provides the secondary benefit of reducing the impact of background noise. By breaking the fixed spatial relationships between the target plant and any adjacent distracting elements, Canopy-Mix prevents the model from learning spurious correlations based on location. The model must learn to identify target leaf tokens in a context-independent manner, thereby improving its focus and robustness.

Specifically, the Canopy-Mix operation is performed on the token representations of two image-label pairs, (X_1, Y_1) and (X_2, Y_2). After passing the images through the ViT embedding layer, we obtain their corresponding token sequences, $T_1 = [x_1^1, ..., x_1^n]$ and $T_2 = [x_2^1, ..., x_2^n]$. A mixing ratio λ is first sampled from a symmetric Beta distribution, $\lambda \sim \mathrm{Beta}(\alpha, \alpha)$. This λ then determines the proportion of tokens to be drawn from each image. A binary mask $B \in \{0, 1\}^n$ is generated with a proportion of λ ones, and this mask is randomly shuffled. The new, mixed token sequence $\tilde{T}$ is generated by shuffling the tokens within each original sequence and then combining them using the binary mask B:

$$\tilde{T} = s(T_1) \odot B + s(T_2) \odot (1 - B) \tag{3}$$

where $s(\cdot)$ represents the random shuffling operation at the token-level and $\odot$ denotes element-wise multiplication. This operation effectively creates a new set of feature tokens where patches from two different plants are interwoven in a random, patch-wise manner, guided by the mask B.

Unlike classification tasks where token-level labels can be mixed, our ground truth is a single continuous value for the entire image. Therefore, the corresponding leaf area label $\tilde{Y}$ for the new token sequence $\tilde{T}$ is generated by applying the same mixing ratio λ to the original image-level labels:

$$\tilde{Y} = \lambda Y_1 + (1 - \lambda) Y_2 \tag{4}$$

By training on these structurally complex, chimeric plant examples, the network learns to decouple leaf area estimation from holistic plant shape. It is forced to identify and sum the contributions of leaf-like features regardless of their spatial arrangement, thereby enhancing its ability to infer total leaf area from the fragmented and overlapping evidence present in real-world images.

2.5.4. Hybrid Loss Function

The choice of loss function is critical for the performance and stability of regression models. While standard losses like Mean Squared Error (L2 loss) are common, they can be overly sensitive to outliers. To address this and ensure robust convergence, we employ a hybrid loss function that combines the strengths of Smooth L1 loss and Log-Cosh loss.

The first component is the Smooth L1 Loss, which is less sensitive to large errors than the L2 loss. It behaves like the L2 loss for small errors, providing smooth gradients near the target, while behaving like the L1 loss for larger errors, which prevents exploding gradients caused by outliers.

The second component is the Log-Cosh Loss, which is a twice-differentiable and smoother function that also acts similarly to the L2 loss but is more robust to large, occasional errors. The final combined loss, L_{combined}, is a weighted sum of these two functions:

$$L_{\mathrm{combined}} = L_{\mathrm{SmoothL1}} + \alpha \cdot L_{\mathrm{LogCosh}} \tag{5}$$

where L_{SmoothL1} is the Smooth L1 loss, and L_{LogCosh} is the mean of $\log(\cosh(y_{\mathrm{pred}} - y_{\mathrm{true}}))$. By combining these two loss functions, our model benefits from both the outlier robustness of Smooth L1 and the smooth gradient properties of

Table 2
Comparative performance of the proposed model against baselines across 5-fold cross-validation. The main values represent the mean, and the subscripts are the corresponding standard derivations.

Methods	R^2 ↑	MAE ↓	RMSE ↓	RRMSE ↓
ResNet-50	$0.325_{0.273}$	$42.486_{8.050}$	$56.243_{12.345}$	$0.355_{0.075}$
ResNet-50 (Pre-trained on ImageNet)	$0.604_{0.051}$	$32.273_{1.827}$	$43.788_{1.958}$	$0.277_{0.010}$
ViT	$0.257_{0.292}$	$44.440_{10.177}$	$58.929_{13.040}$	$0.371_{0.074}$
ViT (Pre-trained on ImageNet)	$0.694_{0.022}$	$28.010_{1.811}$	$38.465_{3.315}$	$0.243_{0.011}$
Ours	$\mathbf{0.805_{0.022}}$	$\mathbf{22.207_{0.791}}$	$\mathbf{30.299_{1.233}}$	$\mathbf{0.191_{0.008}}$

Log-Cosh, leading to more stable training and improved predictive accuracy. Based on empirical validation through grid search, the weighting hyperparameter α was set to 0.2.

2.6. Experimental Setup and Evaluation Metrics

2.6.1. Experimental Setup

All experiments were conducted using the PyTorch deep learning framework on NVIDIA A100 GPUs. Our model employs a Vision Transformer Base (ViT-B/16) architecture, with the backbone initialized from the DINOv2 weights pre-trained on the public plant datasets. To ensure a robust and unbiased evaluation of our model on the limited dataset of 833 images, we implemented a 5-fold cross-validation procedure. The entire dataset was randomly partitioned into five equally sized, non-overlapping folds. The training and evaluation process was iterated five times; in each iteration, one distinct fold was held out as the test set, while the remaining four folds were used for training. For hyperparameter tuning and model selection within each iteration, one of these four training folds was designated as a validation set. The final performance of the model is reported as the mean and standard deviation across all five folds.

During the fine-tuning within each fold, the model was trained for 200 epochs with a batch size of 64. We utilized the AdamW optimizer with a starting learning rate of 5e-5 and a weight decay of 0.05. A cosine annealing scheduler was employed to gradually decrease the learning rate. For our proposed Canopy-Mix data augmentation, the hyperparameter α for the Beta distribution was set to 1.0. The model checkpoint that achieved the best performance on the validation set within each fold was selected for final evaluation on the respective hold-out test set.

2.6.2. Evaluation Metrics

To quantitatively assess the performance of our leaf area regression model, we employed a set of four standard statistical metrics. For all formulas, let y_i be the ground-truth leaf area, $\hat{y}_i$ be the model's predicted leaf area for the i-th sample, $\bar{y}$ be the mean of the ground-truth values, and n be the total number of samples in the test set.

The primary metric is the Coefficient of Determination (R^2), which measures the proportion of variance in the dependent variable that can be predicted from the independent variable(s). An R^2 value closer to 1.0 indicates a stronger correlation and a better model fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

The Mean Absolute Error (MAE) measures the average of the absolute differences between predictions and actual values, treating all errors with equal weight and providing a direct interpretation of the average prediction error.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

In addition, we report two error-based metrics in the original units of measurement (cm^2). The Root Mean Square Error (RMSE) quantifies the standard deviation of the prediction errors and is particularly sensitive to large errors due to the squaring term.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

Finally, we compute the Relative Root Mean Square Error (RRMSE) by normalizing the RMSE with the mean of the ground-truth observations. This dimensionless metric facilitates comparisons across datasets of different scales and is often expressed as a percentage.

$$RRMSE = \frac{RMSE}{\bar{y}} \times 100\% \tag{9}$$

3. Results and Discussion

3.1. Overall Model Performance

The primary objective of this study was to develop a robust model for rapeseed leaf area estimation. By leveraging a two-stage framework combining self-supervised pre-training with a specialized fine-tuning strategy, our proposed model demonstrated high accuracy and stability. The performance, evaluated through a rigorous 5-fold cross-validation on our test dataset, is summarized in Table 2. Our framework achieved a mean R^2 of 0.805 ± 0.022, indicating a strong linear relationship between the model's predictions and the ground-truth measurements. Furthermore, the model yielded a low RMSE of 30.299 ± 1.233 cm^2 and a MAE of 22.207 ± 0.791 cm^2, signifying a high degree of predictive accuracy with small average prediction errors. The low standard deviation across all metrics highlights the model's stability and its ability to generalize consistently across different subsets of the data.

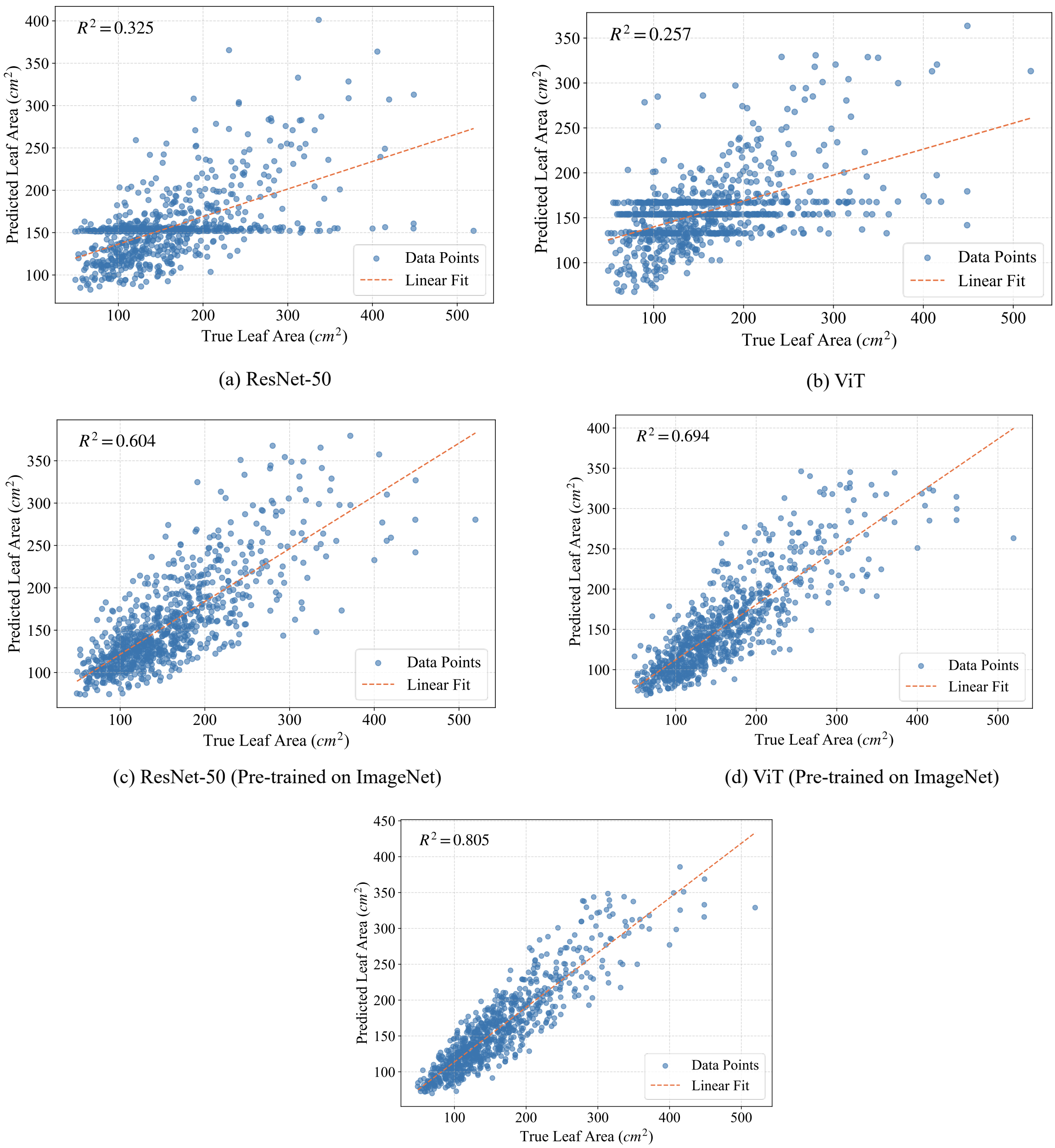
To provide a robust horizontal comparison, our framework was evaluated against several baseline and alternative approaches. As shown in Table 2, our proposed method significantly outperforms models trained from scratch, such as a standard ResNet-50 or a ViT-Base, which struggled to converge effectively on the limited training data. More importantly, our model also shows a marked improvement over a ViT backbone pre-trained on the generic ImageNet dataset, underscoring the benefits of domain-specific self-supervised pre-training on plant-related imagery. The features learned from various plant morphologies appear to be more transferable to our specific task than those learned from general objects.

The strong quantitative performance of our proposed method, summarized in Table 2, is visually corroborated by the comparative scatter plots in Figure 3 and the error distribution shown in Figure 4. The models trained from scratch, ResNet-50 (Figure 3(a)) and ViT (Figure 3(b)), exhibit extremely poor performance. Their scatter plots show significant dispersion and contain many points collapsed into horizontal lines, indicating the models failed to learn discriminative features and instead predicted a constant value for large subsets of the input data. In contrast, leveraging pre-training on ImageNet (Figure 3(c) and Figure 3(d)) yields a dramatic improvement, with the points forming a much clearer linear trend. Finally, our proposed model (Figure 3(e)) clearly demonstrates superior performance; the data points are tightly clustered around the identity line with minimal dispersion. The box plot of RMSE values in Figure 4 further summarizes these findings, visually highlighting the large error distribution and instability of the models trained from scratch. In stark contrast, it shows our proposed model not only achieves the lowest median error but also the most compact distribution, indicating superior stability across all cross-validation folds. This comprehensive visual and quantitative comparison affirms that our framework, which integrates domain-specific self-supervised learning, provides a more robust and accurate solution than training from scratch or using generic ImageNet pre-training for this challenging agricultural regression task.

3.2. Ablation Studies

To systematically dissect our framework and quantify the contribution of each proposed component, we conducted a series of ablation studies. The goal was to validate the effectiveness of our three core components: 1) domain-specific SSL pre-training, 2) the Canopy-Mix data augmentation, and 3) the hybrid loss function. By evaluating different combinations of these components, we isolated their individual and synergistic impacts on the final performance. The comprehensive results of this analysis are presented in Table 3, with corresponding visualizations in Figure 5.

The most significant performance gain was unequivocally attributed to the self-supervised pre-training strategy. Any model configuration without SSL initialization performed dramatically worse than its SSL-enabled counterpart.



(a) ResNet-50

(b) ViT

(c) ResNet-50 (Pre-trained on ImageNet)

(d) ViT (Pre-trained on ImageNet)

(e) Ours

Figure 3: Visual comparison of predicted vs. ground-truth leaf area for the proposed model and various baselines. (a) ResNet-50 from scratch, (b) ViT from scratch, (c) ResNet-50 with ImageNet pre-training, (d) ViT with ImageNet pre-training, and (e) our proposed model. The red dashed line represents the linear fit line. (a) and (b) show significant dispersion and contain many points collapsed into horizontal lines, indicating the models failed to learn discriminative features and instead predicted a constant value for large subsets of the input data.

For instance, the model using only Canopy-Mix and our custom loss function achieved a modest R^2 of just 0.447. However, by simply adding SSL pre-training, the R^2 of the full model jumped to 0.805. This substantial improvement, also visible when comparing other pairs in Table 3, confirms our hypothesis that leveraging diverse, unlabeled plant data to learn foundational visual features is the most critical step for success on our limited dataset.

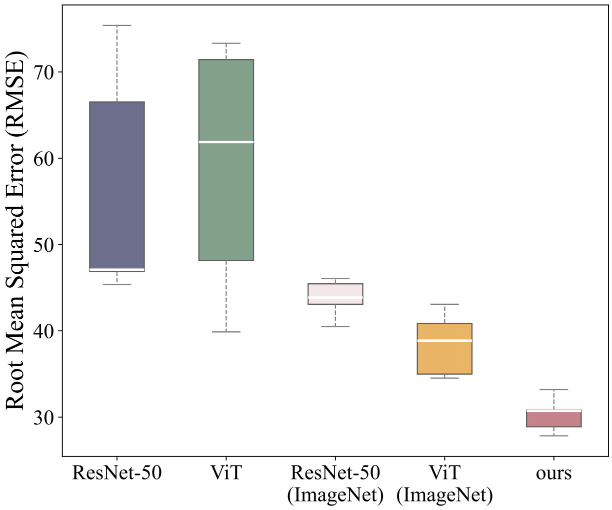


Figure 4: Box plot of RMSE distribution for each model across the 5 cross-validation folds.

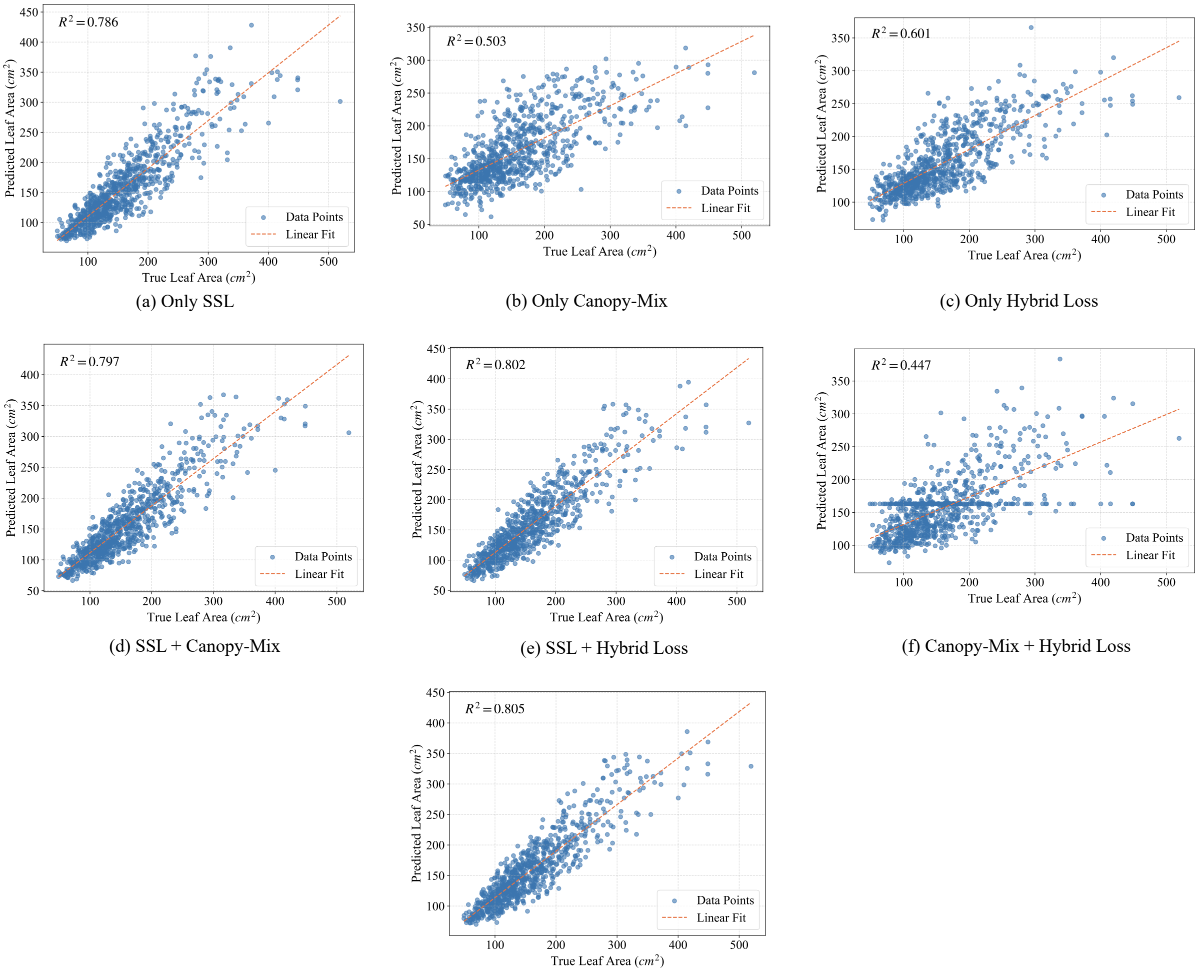
Table 3
Ablation study results for the key components of our framework.

SSL	Canopy-Mix	Hybrid Loss	R^2 ↑	MAE ↓	RMSE ↓	RRMSE ↓
✓			$0.786_{0.024}$	$23.957_{1.743}$	$32.275_{2.180}$	$0.204_{0.013}$
	✓		$0.503_{0.061}$	$37.649_{2.921}$	$49.316_{4.802}$	$0.311_{0.026}$
		✓	$0.601_{0.075}$	$33.128_{4.249}$	$44.167_{6.354}$	$0.278_{0.035}$
✓	✓		$0.797_{0.027}$	$22.979_{1.119}$	$31.411_{2.081}$	$0.198_{0.012}$
✓		✓	$0.802_{0.024}$	$23.096_{1.123}$	$31.034_{1.608}$	$0.196_{0.009}$
	✓	✓	$0.447_{0.231}$	$38.275_{9.231}$	$51.424_{12.449}$	$0.323_{0.070}$
✓	✓	✓	$\mathbf{0.805_{0.022}}$	$\mathbf{22.207_{0.791}}$	$\mathbf{30.299_{1.233}}$	$\mathbf{0.191_{0.008}}$

Next, we evaluated the impact of our hybrid loss function. Comparing the model with "SSL + Canopy-Mix" (R^2 = 0.797) to the final model which also includes the hybrid loss (R^2 = 0.805), we observe a clear improvement in performance and stability. The hybrid loss, by combining the outlier-robustness of Smooth L1 with the smooth gradients of Log-Cosh, guides the model to a more precise and stable convergence, as evidenced by the consistently lower RMSE and smaller standard deviations across the folds.

Finally, we assessed the contribution of our Canopy-Mix data augmentation. When adding Canopy-Mix to the "SSL + Loss" configuration, the R^2 increased from 0.802 to 0.805. While a modest gain in R^2, Canopy-Mix consistently helped in lowering the prediction errors (MAE and RMSE). This suggests that Canopy-Mix acts as an effective regularizer. By creating structurally challenging training samples, it forces the model to learn a more generalized representation of leaf area, preventing overfitting and improving its ability to handle the natural variability in unseen test images.

The visual results in Figure 5 corroborate these quantitative findings. The scatter plot for the model without SSL (e.g., Canopy-Mix + Loss) shows significant dispersion, while the plot for our full, final model shows the tightest clustering of points along the identity line. In summary, these studies provide clear evidence that while SSL provides the foundational leap in performance, both the hybrid loss function and our Canopy-Mix augmentation make distinct and valuable contributions, working synergistically to achieve the final state-of-the-art result.



(a) Only SSL (b) Only Canopy-Mix (c) Only Hybrid Loss

(d) SSL + Canopy-Mix (e) SSL + Hybrid Loss (f) Canopy-Mix + Hybrid Loss

(g) Ours

Figure 5: Visual results of the ablation study.

3.3. Qualitative Analysis and Model Interpretability

Beyond quantitative metrics, it is crucial to understand how our model arrives at its predictions. To gain a qualitative insight into the model's decision-making process, we visualized its internal attention mechanism. This analysis allows us to interpret what regions of the input image the model deems most important. We contrast our model's automated predictions with a baseline approach that relies on the area of manually annotated bounding boxes. This comparison highlights the sophistication of our feature-learning model against a method that, despite requiring human labor for localization, is fundamentally unsuited for precise area quantification. The significant performance gap between these two approaches is detailed in Table 4.

We first investigated the attention maps generated by the final layer of the ViT, visualized as heatmaps in Figure 6 (c). These maps reveal that the model has successfully learned to focus with surgical precision on the target plant's canopy. A key observation is the model's robustness to visual distractors; despite the presence of other, smaller rapeseed plants, colored markers, and tools within the scene, the model's attention remains tightly concentrated on the primary plant of interest. This demonstrates that our model has learned not only to identify leaf tissue but also to distinguish the target subject from other visually similar or distracting elements, a critical capability for reliable deployment.

The manually annotated bounding box, while accurately localizing the plant, is fundamentally flawed as a proxy for leaf area for two primary reasons. First, the rectangular area inevitably includes significant non-leaf background

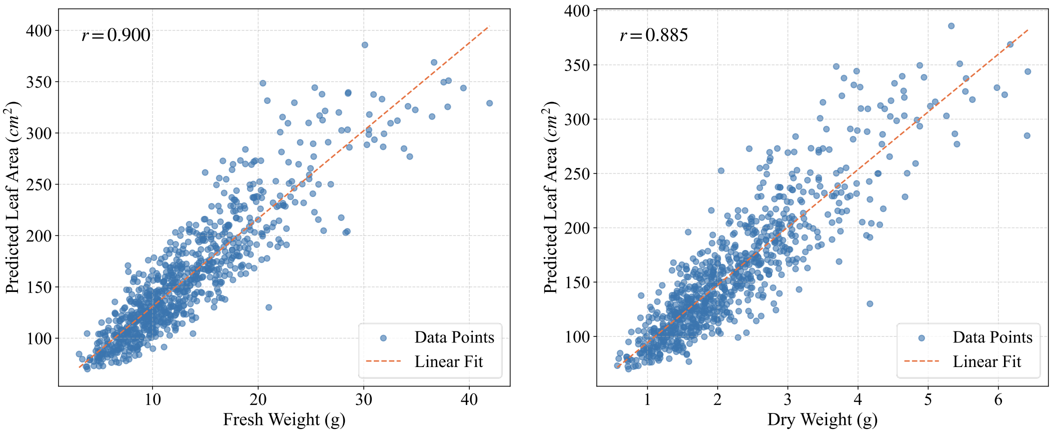
(a) (b) (c)

Figure 6: Qualitative visualization of the model's interpretability. (a) Original input images. (b) Manually annotated bounding boxes accurately localizing the target plant, but including significant background soil. (c) Our model's attention map.

soil. Second, and more critically, the 2D box cannot account for the three-dimensional structure of the canopy and therefore fails to capture the area of any occluded leaves hidden from view. This limitation becomes a significant source of systematic error as plants grow and self-occlusion increases. This inherent inadequacy is reflected in the quantitative results in Table 4. This confirms that a feature-learning approach like ours is essential, as it can learn the morphological cues needed to infer the area of occluded leaves, a capability a simple geometric approximation entirely lacks.

Table 4
Performance comparison between our proposed model and manually annotated bounding box areas.

	R^2 ↑	MAE ↓	RMSE ↓	RRMSE ↓
Box Area	$0.636_{0.042}$	$31.923_{1.551}$	$42.080_{2.443}$	$0.266_{0.014}$
Ours	$\mathbf{0.805_{0.022}}$	$\mathbf{22.207_{0.791}}$	$\mathbf{30.299_{1.233}}$	$\mathbf{0.191_{0.008}}$



(a) Predicted Leaf Area vs. Fresh Weight (b) Predicted Leaf Area vs. Dry Weight

Figure 7: Scatter plots showing the strong positive linear correlation between the model's predicted leaf area and (a) fresh weight, and (b) dry weight. The Pearson correlation coefficients (r) are displayed in the top-left corner of each plot, indicating that the predicted leaf area is a reliable proxy for plant biomass.

3.4. Correlation with Above Ground Biomass

To further validate the practical utility of our model, we investigated the relationship between the model's non-destructive leaf area predictions and key destructive biomass measurements, namely fresh weight and dry weight. A strong correlation between predicted leaf area and these biomass indicators would confirm that our model's output serves as a reliable proxy for monitoring plant growth and vigor. This analysis is critical for applications in breeding and agronomy, where biomass is a primary determinant of final yield.

Our analysis revealed a strong positive linear relationship between the predicted leaf area and both biomass metrics. As visually presented in the scatter plots in Figure 7, the data points form a clear and tight linear trend. We calculated the Pearson correlation coefficient (r) to quantify this relationship. The Pearson coefficient is used here as our goal is to measure the strength of the linear association between these two variables, rather than to evaluate a predictive regression model's performance with R^2. The correlation between predicted leaf area and fresh weight was found to be exceptionally high, with an r-value of 0.900. A similarly strong correlation was observed with the dry weight, yielding a slightly lower but still very high r-value of 0.885. These high correlation coefficients provide quantitative evidence that the leaf area estimated by our model is a powerful and accurate indicator of the plant's overall biomass accumulation.

The implications of this finding are significant. It validates that our non-destructive, image-based estimation method can be confidently used to approximate critical growth parameters that would otherwise require destructive harvesting. This provides a powerful, high-throughput tool for breeders to monitor plant development, assess responses to different treatments, and make early selections in the breeding cycle, thereby enhancing the efficiency and effectiveness of agricultural research.

4. Conclusion

In this study, we successfully developed and validated a novel deep learning framework to address the critical challenge of accurately and efficiently estimating rapeseed leaf area from top-down imagery. By integrating self-supervised pre-training on diverse plant datasets with a specialized fine-tuning strategy that included an a new Canopy-Mix data augmentation method and a hybrid Loss Function, our model achieved strong predictive performance with $R^2 = 0.805$ and robustness. Our analyses confirmed that domain-specific pre-training is superior to generic pre-training and that each component of our methodology made a significant contribution to handling key challenges like leaf occlusion and data scarcity. This work provides a powerful and data-efficient pipeline for non-destructive plant phenotyping, demonstrating the potential of modern self-supervised learning to accelerate agricultural research, the future close directions can be:

1. Evaluating the rapeseed plant growth rates through twice non-destructive photoing;
2. Assessing biotic/abiotic stress response by measuring damage severity and recovery rates in rapeseed plants.
3. Development of objective phenotypic characterization tools for rapeseed seedlings to facilitate early-stage and high-throughput screening of breeding germplasms.

Notwithstanding these advancements, certain limitations warrant consideration: The accuracy of the model needs to be improved due to the limited sample size. Additionally, while the methodology demonstrates efficacy during seedling stages, its application to later growth phases necessitates supplementary investigations to overcome challenges posed by complex shadow artifacts and extreme canopy density conditions.

Author contributions: CRediT

Baogang Lin: Supervision, Resources, Investigation, Conceptualization. **Qing Cai**: Methodology. **Junqin Cao**: Investigation. **Chaochao He**: Investigation. **Zhiqi Ma**: Investigation. **Shuijin Hua**: Supervision. **Jianpeng An**: Writing – original draft, Formal analysis, Methodology, Software. **Pengfei Hao**: Writing – original draft, Investigation, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Funding

This work was supported by National Natural Science Foundation of China (32201693) and Zhejiang Key Laboratory of Digital Dry Land Crops (2022E10012).

Acknowledgements

Great appreciations to the editor and reviewers for their enthusiastic assistance, guidance, and suggestions during the submission process.

References

A, F.S., B, M.A.K., A, M.S., C, M.M., D, L.M.G., E, S.R., 2021. Deep neural network features fusion and selection based on pls regression with an application for crops diseases classification. Applied Soft Computing .

Anello, M., Mateo, F., Bernardi, B., Giuffrè, A.M., Blasco, J., Gómez-Sanchis, J., 2024. Convolutional neural networks to assess bergamot essential oil content in the field from smartphone images. Industrial Crops and Products 220, 10.

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science 19, 52–61.

Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., Chen, T., Tomasev, N., Mitrović, J., Strachan, P., et al., 2023. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. Nature Biomedical Engineering 7, 756–779.

Bradski, G., 2000. The opencv library. Dr. Dobb's Journal: Software Tools for the Professional Programmer 25, 120–123.

Chang, A., Jung, J., Maeda, M.M., Landivar, J., 2017. Crop height monitoring with digital imagery from unmanned aerial system (uas). Computers and Electronics in Agriculture 141, 232–237.

Chen, H., Liu, S., Wang, C., Wang, C., Gong, K., Li, Y., Lan, Y., 2023. Point cloud completion of plant leaves under occlusion conditions based on deep learning. Plant Phenomics 5, 0117.

Fan, X., Zhou, R., Tjahjadi, T., Das Choudhury, S., Ye, Q., 2022. A segmentation-guided deep learning framework for leaf counting. Frontiers in Plant Science 13, 844522.

Giselsson, T.M., Dyrmann, M., Jørgensen, R.N., Jensen, P.K., Midtiby, H.S., 2017. A Public Image Database for Benchmark of Plant Seedling Classification Algorithms. arXiv preprint .

Guo, X., Qiu, Y., Nettleton, D., Schnable, P.S., 2023. High-throughput field plant phenotyping: a self-supervised sequential cnn method to segment overlapping plants. Plant Phenomics 5, 0052.

Ibrahim, A.S., Mohsen, S., Selim, I.M., Alroobaea, R., Alsafyani, M., Baqasah, A.M., Eassa, M., 2025. Ai-iot based smart agriculture pivot for plant diseases detection and treatment. Scientific Reports 15, 16576.

Ilyas, T., Lee, J., Won, O., Jeong, Y., Kim, H., 2023. Overcoming field variability: unsupervised domain adaptation for enhanced crop-weed recognition in diverse farmlands. Frontiers in Plant Science 14, 1234616.

Jing, L., Tian, Y., 2020. Self-supervised visual feature learning with deep neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence 43, 4037–4058.

Li, J., Chen, W., Ying, T., Yang, L., 2024. Double-exposure algorithm: A powerful approach to address the accuracy issues of fractional vegetation extraction under shadow conditions. Applied Sciences 14, 7719.

Li, L., Qiao, J., Yao, J., Li, J., Li, L., 2022a. Automatic freezing-tolerant rapeseed material recognition using uav images and deep learning. Plant methods 18, 5.

Li, P., Zhang, X., Wang, W., Zheng, H., Cheng, T., 2020. Estimating aboveground and organ biomass of plant canopies across the entire season of rice growth with terrestrial laser scanning. International Journal of Applied Earth Observation and Geoinformation 91, 102132.

Li, R., Wang, D., Zhu, B., Liu, T., Sun, C., Zhang, Z., 2022b. Estimation of nitrogen content in wheat using indices derived from rgb and thermal infrared imaging. Field Crops Research .

Lin, J., Zhang, X., Qin, Y., Yang, S., Wen, X., Cernava, T., Chen, X., 2025. Fg-unet: fine-grained feature-guided unet for segmentation of weeds and crops in uav images. Pest management science 81, 856–866.

Liu, Y., Liu, S., Li, J., Guo, X., Lu, J., 2019. Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from uav multispectral images. Computers and Electronics in Agriculture 166, 105026.

Madec, S., Irfan, K., Velumani, K., Baret, F., David, E., Daubige, G., Samatan, L.B., Serouart, M., Smith, D., James, C., et al., 2023. Vegann, vegetation annotation of multi-crop rgb images acquired under diverse conditions for segmentation. Scientific Data 10, 302.

Mehmood, S.S., Lu, G., Luo, D., Hussain, M.A., Lv, Y., 2021. Integrated analysis of transcriptomics and proteomics provides insights into the molecular regulation of cold response in brassica napus. Environmental and Experimental Botany 187.

Miao, C., Guo, A., Thompson, A.M., Yang, J., Ge, Y., Schnable, J.C., 2021. Automation of leaf counting in maize and sorghum using deep learning. The Plant Phenome Journal 4, e20022.

Minervini, M., Abdelsamea, M.M., Tsaftaris, S.A., 2014. Image-based plant phenotyping with incremental learning and active contours. Ecological Informatics 23, 35–48.

Murchie, E., Burgess, A., 2022. Casting light on the architecture of crop yield. Crop and Environment .

Murphy, K.M., Ludwig, E., Gutierrez, J., Gehan, M.A., 2024. Deep learning in image-based plant phenotyping. Annual Review of Plant Biology 75.

Nagasubramanian, K., Singh, A., Singh, A., Sarkar, S., Ganapathysubramanian, B., 2022. Plant phenotyping with limited annotation: Doing more with less. The Plant Phenome Journal 5, e20051.

Naseer, M.M., Ranasinghe, K., Khan, S.H., Hayat, M., Shahbaz Khan, F., Yang, M.H., 2021. Intriguing properties of vision transformers. Advances in Neural Information Processing Systems 34, 23296–23308.

Newman, S.J., Furbank, R.T., 2021. Explainable machine learning models of major crop traits from satellite-monitored continent-wide field trial data. Nature Plants , 7.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P., Li, S., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jégou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2024. Dinov2: Learning robust visual features without supervision. Trans. Mach. Learn. Res. 2024. URL: https://openreview.net/forum?id=a68SUt6zFt.

Scharr, H., Minervini, M., Fischbach, A., Tsaftaris, S.A., 2014. Annotated image datasets of rosette plants, in: European conference on computer vision, Suisse Zürich. pp. 6–12.

Shao, M., Nie, C., Cheng, M., Yu, X., Bai, Y., Ming, B., Song, H., Jin, X., 2022. Quantifying effect of tassels on near-ground maize canopy rgb images using deep learning segmentation algorithm. Precision Agriculture 23, 400–418.

Singh, A.K., Ganapathysubramanian, B., Sarkar, S., Singh, A., 2018. Deep learning for plant stress phenotyping: trends and future perspectives. Trends in plant science 23, 883–898.

Sun, C., Zhang, W., Zhao, G., Wu, Q., Liang, W., Ren, N., Cao, H., Zou, L., 2024. Mapping rapeseed (brassica napus l.) aboveground biomass in different periods using optical and phenotypic metrics derived from uav hyperspectral and rgb imagery. FRONTIERS IN PLANT SCIENCE 15.

Wang, C., Xia, Y., Xia, L., Wang, Q., Gu, L., 2025. Dual discriminator gan-based synthetic crop disease image generation for precise crop disease identification. Plant Methods 21, 46.

Wang, S.M., Yu, C.P., Ma, J.H., Ouyang, J.X., Zhao, Z.M., Xuan, Y.M., Fan, D.M., Yu, J.F., Wang, X.C., Zheng, X.Q., 2024. Tea yield estimation using uav images and deep learning. Industrial Crops and Products 212, 118358.

Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., Xiang, Q.L., 2007. A leaf recognition algorithm for plant classification using probabilistic neural network, in: 2007 IEEE international symposium on signal processing and information technology, IEEE. pp. 11–16.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Advances in neural information processing systems 27.

Yue, J., Yang, H., Yang, G., Fu, Y., Wang, H., Zhou, C., 2023. Estimating vertically growing crop above-ground biomass based on uav remote sensing. Computers and electronics in agriculture 205, 14.

Zdrazil, J., Kong, L., Klimeš, P., Jasso-Robles, F.I., Saiz-Fernández, I., Güder, F., Spíchal, L., Snášel, V., De Diego, N., 2025. Next-generation high-throughput phenotyping with trait prediction through adaptable multi-task computational intelligence. Computers and Electronics in Agriculture 235, 110390.

Zeng, Y., Xu, B., Yin, G., Wu, S., Hu, G., Yan, K., Yang, B., Song, W., Li, J., 2018. Spectral invariant provides a practical modeling approach for future biophysical variable estimations. Remote Sensing 10.

Zhang, H., Cissé, M., Dauphin, Y.N., Lopez-Paz, D., 2018. mixup: Beyond empirical risk minimization, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, OpenReview.net. URL: https://openreview.net/forum?id=r1Ddp1-Rb.

Zhang, L., Sun, B., Zhao, D., Shan, C., Wang, B., Wang, G., Song, C., Chen, P., Lan, Y., 2024. Improved estimation of cotton (gossypium hirsutum l.) lai from multispectral data using uav point cloud data. Industrial Crops and Products 217, 17.

Zhou, C., Hu, J., Xu, Z., Yue, J., Ye, H., Yang, G., 2020. A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks. Frontiers in plant science 11, 402.

Zhu, L., Tang, L., Ren, S., 2025. Deep learning-based text generation for plant phenotyping and precision agriculture. Frontiers in Plant Science 16, 1564394.