

Article

LEAF-Net: A Unified Framework for Leaf Extraction and Analysis in Multi-Crop Phenotyping Using YOLOv11

Ameer Tamoor Khan *  and Signe Marie Jensen 

Department of Plant and Environmental Sciences, University of Copenhagen, 1172 Copenhagen, Denmark; smj@plen.ku.dk

* Correspondence: atk@plen.ku.dk

Abstract: Accurate leaf segmentation and counting are critical for advancing crop phenotyping and improving breeding programs in agriculture. This study evaluates YOLOv11-based models for automated leaf detection and segmentation across spring barley, spring wheat, winter wheat, winter rye, and winter triticale. The key focus is assessing whether a unified model trained on a combined multi-crop dataset can outperform crop-specific models. Results show that the unified model achieves superior performance in bounding box tasks, with mAP@50 exceeding 0.85 for spring crops and 0.7 for winter crops. Segmentation tasks, however, reveal mixed results, with individual models occasionally excelling in recall for winter crops. These findings highlight the benefits of dataset diversity in improving generalization, while emphasizing the need for larger annotated datasets to address variability in real-world conditions. While the combined dataset improves generalization, the unique characteristics of individual crops may still benefit from specialized training.

Keywords: plant phenotyping; YOLOv11; leaf segmentation; precision agriculture; deep learning



Academic Editor: Kohei Koyama

Received: 5 December 2024

Revised: 7 January 2025

Accepted: 14 January 2025

Published: 17 January 2025

Citation: Khan, A.T.; Jensen, S.M. LEAF-Net: A Unified Framework for Leaf Extraction and Analysis in Multi-Crop Phenotyping Using YOLOv11. *Agriculture* **2025**, *15*, 196. <https://doi.org/10.3390/agriculture15020196>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In modern agriculture, precise plant phenotyping is essential for advancing crop breeding programs, optimizing agricultural practices, and addressing the global challenge of food security [1]. With the growing demand for food driven by an expanding population, alongside challenges posed by climate change and limited arable land, the development of efficient and accurate methods for analyzing plant traits has become a critical priority [2]. Among various traits, accurate segmentation and identification of leaves during the early stages of crop development are particularly important for understanding crop health, growth dynamics, and development across different crop varieties. However, automated analysis of these traits remains challenging due to occlusion, environmental variability, and the complex morphology of crop plants.

Traditional plant phenotyping methods rely heavily on manual observations and measurements [3]. While these methods have served as the foundation of crop breeding for decades, they are inherently time-consuming, labor-intensive, and prone to human error. These limitations make them unsuitable for large-scale breeding programs, where evaluating numerous traits across extensive populations is required. Additionally, the increasing adoption of precision agriculture underscores the need for high-throughput and non-destructive phenotyping techniques that enhance the efficiency and sustainability of modern farming systems [4].

Early-stage leaf identification and counting play a crucial role in monitoring crop health, understanding growth patterns, and evaluating early-stage crop maturity [5]. Accurate identification of leaves provides valuable insights for breeders, enabling a better understanding of growth dynamics and supporting the selection of superior crop varieties. Moreover, segmentation is critical for estimating the Leaf Area Index (LAI), a key metric for understanding crop growth and canopy structure [6]. LAI estimation plays a vital role in predicting photosynthetic efficiency, yield potential, and resource allocation in plants. However, manual counting is impractical for large-scale studies under diverse field conditions, necessitating automated solutions enabled by advancements in artificial intelligence (AI) and computer vision [7].

Classical approaches to leaf counting have primarily relied on manual or semi-automated methods. These techniques often involve human experts counting leaves in the field [4]. Image processing methods, such as edge detection, thresholding, and region-based segmentation, have also been applied to digital images of crops [8]. While these methods show promise in controlled environments, their effectiveness diminishes under real-world conditions where variability in lighting, occlusion, and plant morphology present significant challenges. Furthermore, many classical methods are tailored to specific crop types, limiting their applicability across multiple species [9].

To the best of our knowledge, no prior work has comprehensively addressed early-stage leaf segmentation and counting across multiple crop types—including spring barley, spring wheat, winter wheat, winter rye, and winter triticale—under diverse field conditions, highlighting a gap in the literature.

Recent advancements in AI, particularly in deep learning and computer vision, have transformed plant phenotyping by enabling automated analysis [10,11]. State-of-the-art models such as convolutional neural networks (CNNs) and object detection frameworks have demonstrated remarkable success in tasks such as segmentation of crop leaves, early-stage trait analysis, and biomass estimation [7,12,13]. Additionally, high-throughput phenotyping platforms using unmanned aerial vehicles (UAVs) and robotic systems equipped with advanced imaging sensors have made large-scale data collection feasible [14].

Despite these advancements, existing AI-driven methods often face significant limitations when applied to diverse crop types and real-world field conditions. Challenges such as variable lighting, occlusion from overlapping plant structures, and the inherent diversity of crop morphologies pose barriers to robust early-stage leaf detection and segmentation [7].

The present study seeks to evaluate whether distinct models are necessary for leaf segmentation and identification across different crop types at the early growth stage or whether a unified model, trained on a dataset encompassing multiple crops, can achieve superior performance. The rationale behind the latter approach lies in the hypothesis that shared morphological features among crops may enable the model to generalize effectively, thereby enhancing its robustness and accuracy. This investigation critically examines the adaptability and transferability of AI models in agricultural applications, with a specific focus on early-stage crop monitoring and leaf analysis, aiming to advance automated phenotyping methodologies.

The remainder of this paper is organized as follows: Section 2 discusses the dataset used in this study, including data collection and annotation processes. Section 3 outlines the methodology employed for leaf segmentation and counting. Section 4 describes the training and validation of the models. Section 5 presents the analysis and results. Section 6 provides a discussion of the findings, and Section 7 concludes the paper with final remarks and future directions.

2. Dataset Overview

The dataset used in this study was collected from five agricultural fields at the University of Copenhagen. It included images from five different crops: spring barley, spring wheat, winter wheat, winter rye, and winter triticale. The primary objective of this dataset was to support leaf segmentation tasks, with leaves manually annotated in green for visibility, as shown in Figure 1. Sample images from the dataset illustrated the diversity in crop types and imaging conditions. The manual annotations focused on leaves, which were often challenging to distinguish in raw images, emphasizing the dataset's importance in segmentation tasks. These annotations provided essential ground truth data for developing and evaluating segmentation models.

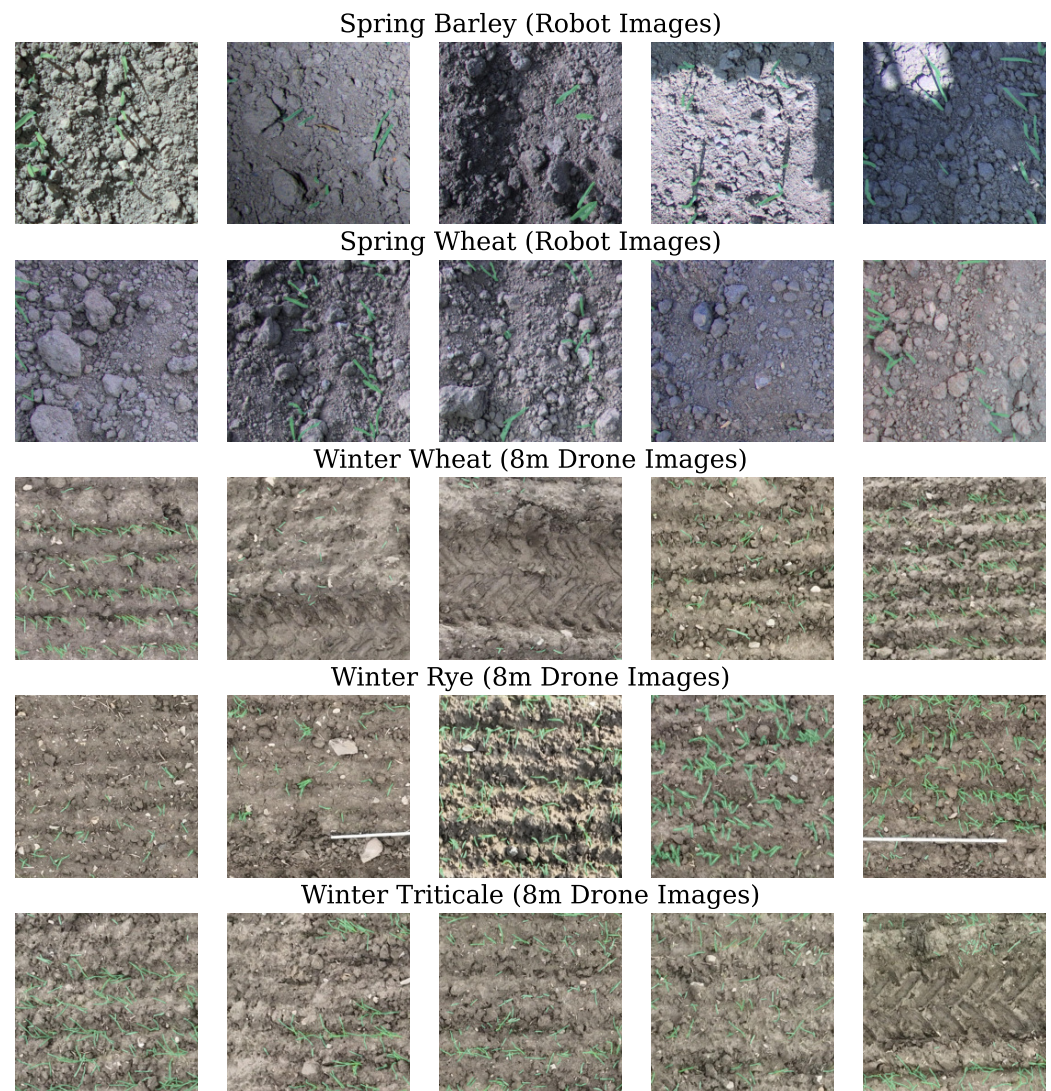


Figure 1. Sample images from the spring and winter crops dataset, displaying annotated leaves highlighted in green, which were otherwise challenging to distinguish in raw images. Spring barley and spring wheat images were captured using an agricultural robot (overall height approximately 2.15 m), while winter wheat, winter rye, and winter triticale images were collected using a drone flying at an altitude of 8 m.

The dataset was acquired using two distinct platforms: Spring barley and spring wheat images were collected using an agricultural robot, which had an overall height of approximately 2.15 m, a ground clearance of 80 cm, and an image resolution of 2448×2048 . Winter wheat, winter rye, and winter triticale images were captured using a drone flying at

an altitude of 8 m with an image resolution of 5280×395 . The image sources and heights are detailed in Figure 1.

The dataset consisted of 907 training and 112 testing images across all crop types, with a notable imbalance in distribution. Specifically, spring barley had 419 training images and 43 testing images, while spring wheat had 410 training images and 38 testing images. The winter crops—winter wheat, winter rye, and winter triticale—were evenly distributed, with each having 25 training images and 3 testing images, as shown in Figure 2.

In addition to the image distribution, the number of annotated leaves in the dataset was a critical aspect. Figure 3 shows the leaf count for each crop type in the training and testing datasets. The spring barley training set contained 4987 leaves, with 646 in the testing set. Similarly, spring wheat had 6990 leaves in the training set and 832 in the testing set. The winter crops had considerably fewer leaves, with winter wheat having 2721 leaves in training and 506 in testing, winter rye having 3309 in training and 229 in testing, and winter triticale containing 2730 in training and 564 in testing. These counts demonstrated that the training datasets were significantly more comprehensive than the testing datasets, ensuring robust model development while still providing sufficient data for evaluation.

To analyze the differences between images captured by the robot and the drone, Principal Component Analysis (PCA) was performed using features extracted from a pretrained ResNet50 model. Images were passed through the model, and the resulting features were used to perform PCA with two components. The PCA results, shown in Figure 4, demonstrated distinct clustering based on the imaging source.

This clustering reflects variations in leaf size and imaging conditions. Robot-captured images, taken closer to the ground under controlled conditions, exhibited larger and more detailed leaf representations. These images avoided environmental disturbances such as wind or inconsistent lighting. In contrast, drone-captured images, taken from higher altitudes in natural field conditions, were influenced by environmental factors, resulting in smaller and less uniform leaf representations.

Overall, the spring and winter crop dataset was a valuable resource for segmentation tasks, particularly for leaf identification in diverse crop types. The dataset's variety in imaging conditions—spanning robot-collected and drone-captured images—added to its robustness. While the imbalance between spring and winter crop samples was evident, the dataset's carefully annotated leaves and comprehensive training sets provided a solid foundation for developing high-performance segmentation models. This dataset was well suited for applications in precision agriculture, enabling researchers to build solutions for automated crop analysis and monitoring.

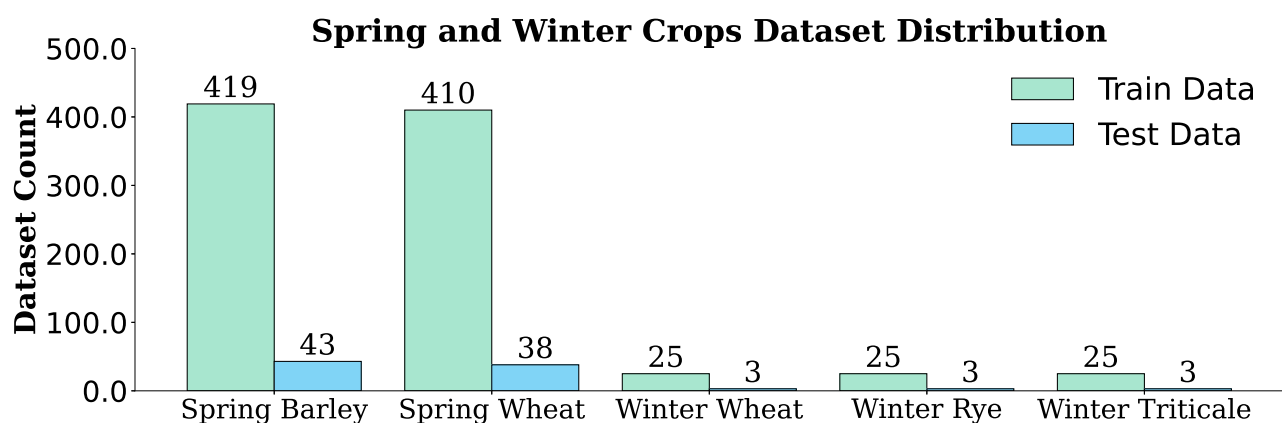


Figure 2. Spring and winter crop dataset distribution illustrating the count of images in the training and testing datasets for each crop type.

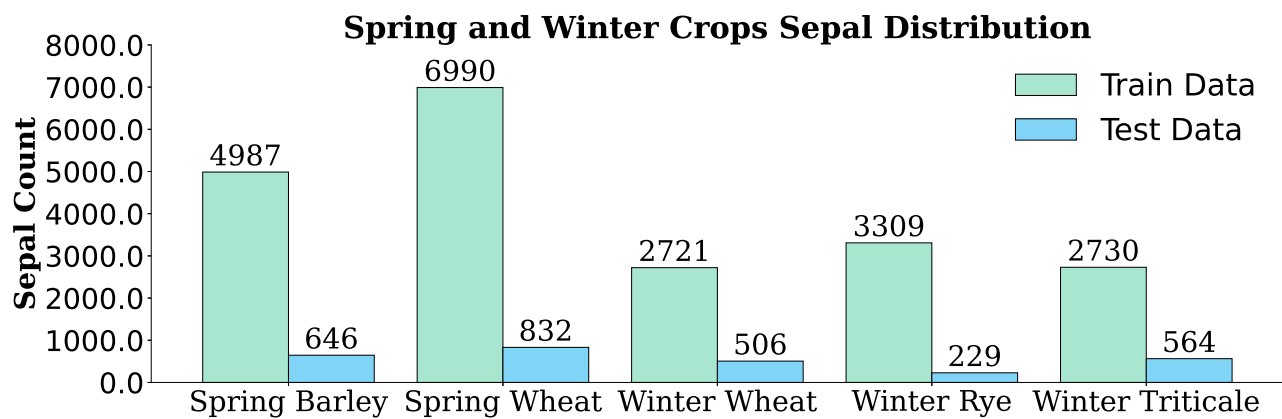


Figure 3. Spring and winter crop leaf distribution showing the count of manually annotated leaves across the training and testing datasets for each crop type.

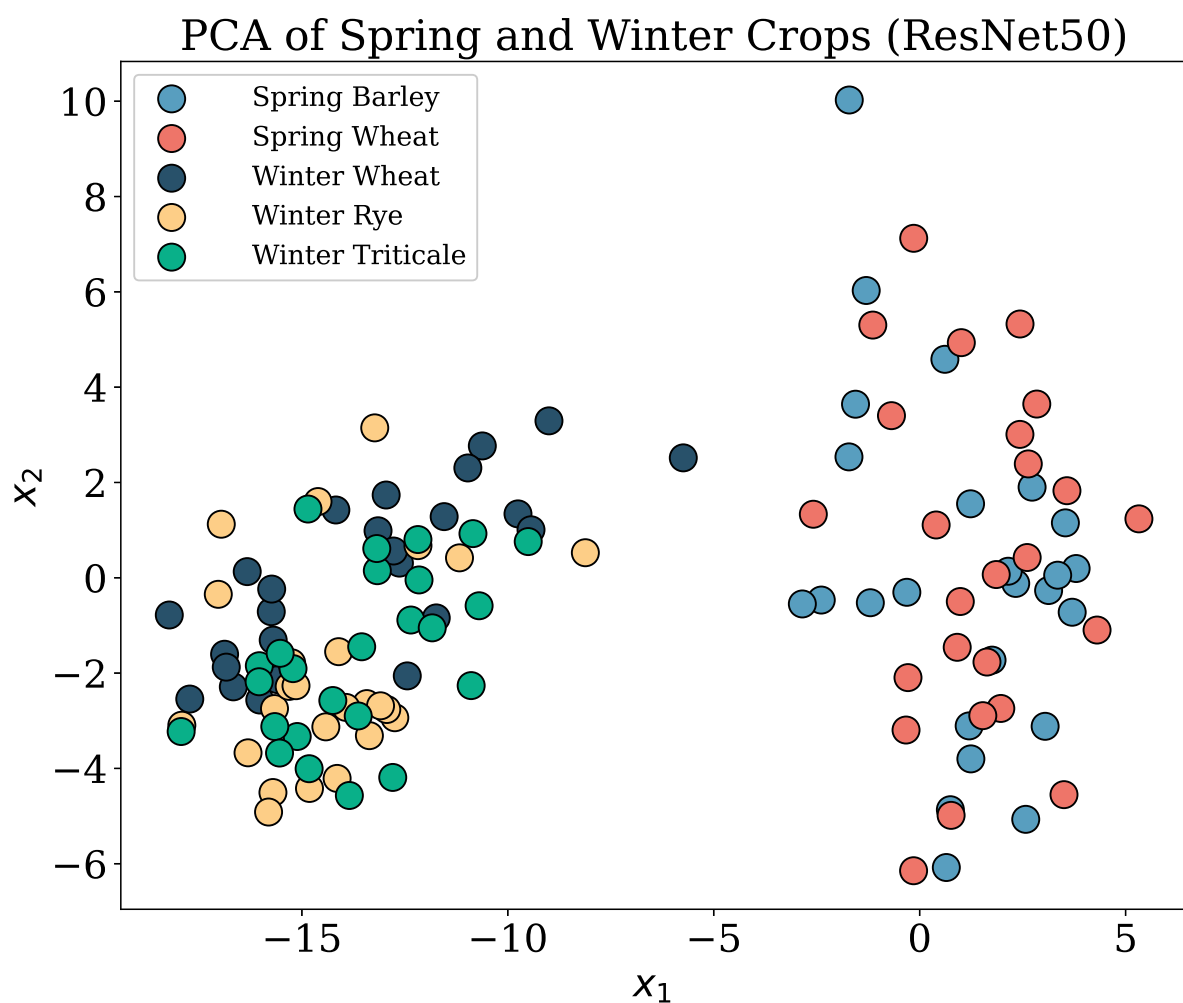


Figure 4. PCA of spring and winter crops visualized using features from a pretrained ResNet50 model. The clustering reflected height-based differences, with robot images capturing larger leaf sizes due to proximity to the ground and controlled conditions, while drone images showed smaller leaf sizes influenced by natural, uncontrolled environmental factors.

3. Architecture of YOLOv11

The YOLOv11 architecture introduces several critical enhancements to improve efficiency and detection accuracy, as illustrated in Figure 5. The design comprises three primary

components: Backbone, Neck, and Head, each incorporating innovative mechanisms to optimize performance.

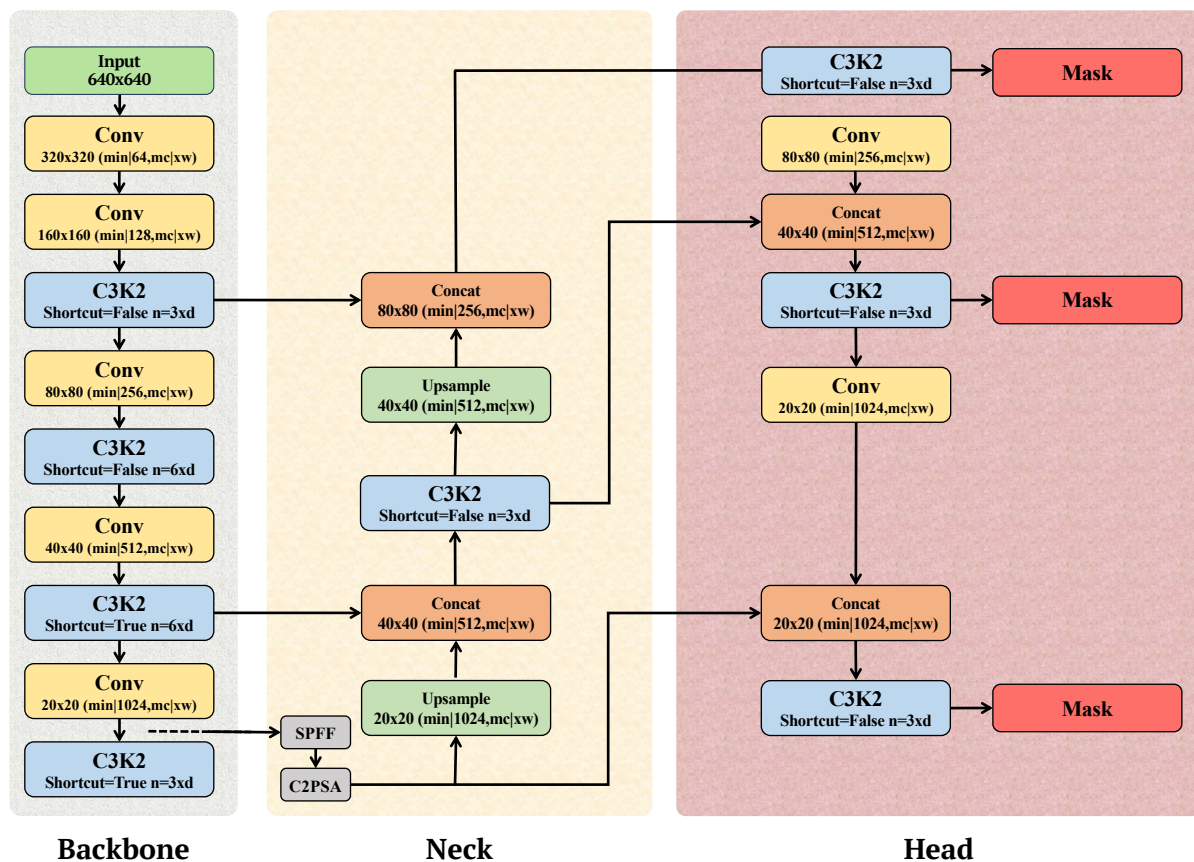


Figure 5. A schematic diagram of YOLOv11 illustrating its three core components: Backbone, Neck, and Head. The Backbone handles multi-scale feature extraction using advanced blocks such as the C3k2 and Spatial Pyramid Pooling Fast (SPPF), designed for efficient feature representation. The Neck aggregates and refines these features with additional mechanisms like Cross-Stage Partial with Spatial Attention (C2PSA). Finally, the Head predicts object bounding boxes, masks, and classifications using enhanced multi-scale processing.

3.1. Backbone

The Backbone is the foundation for multi-scale feature extraction, leveraging advancements to ensure computational efficiency:

- **C3k2 Block:** This Cross-Stage Partial (CSP) bottleneck employs a kernel size of two, replacing older, more computationally intensive blocks. The design reduces the model's overall complexity while retaining its ability to capture essential features.
- **SPPF (Spatial Pyramid Pooling Fast):** Adapted from YOLOv8, this block consolidates features across multiple receptive fields, enhancing the representation of objects at varying scales.
- **C2PSA (Cross-Stage Partial with Spatial Attention):** This novel addition integrates spatial attention, allowing the model to focus on critical regions within the input image. It is particularly effective for detecting small or occluded objects, which are common challenges in agricultural datasets.

3.2. Neck

The Neck serves as a bridge, refining features extracted by the Backbone and preparing them for prediction. Enhancements include the following:

- **C3k2 Block:** These blocks continue to process features efficiently, maintaining high throughput while reducing the computational cost.
- **C2PSA Mechanism:** By refining feature maps through spatial attention, the Neck enhances the model's ability to distinguish between relevant and irrelevant features.

3.3. Head

The Head outputs the final predictions, including object bounding boxes, masks, and class probabilities:

- **C3k2 Blocks:** These blocks enhance the processing of multi-scale features, ensuring fine-grained detection across various object sizes.
- **CBS (Convolution-BatchNorm-SiLU):** This combination improves feature normalization and training stability, enhancing the overall performance of the network.
- **Final Detection Layer:** Outputs Masks, bounding box coordinates, objectness scores, and class probabilities with a focus on efficiency and accuracy, critical for real-time applications.

3.4. Applications in Agriculture

The architectural enhancements in YOLOv11, as depicted in Figure 5, make it particularly well suited for agricultural applications. Its ability to detect small or occluded objects is advantageous in tasks such as the following:

- **Crop Monitoring [15]:** Identifying and counting crop-specific traits, such as leaves, flowers, or fruit, to estimate yield.
- **Weed Detection [16]:** Differentiating between crops and weeds for targeted herbicide application.
- **Pest and Disease Identification [17]:** Detecting early signs of pests or diseases to mitigate potential yield losses.
- **Automated Harvesting [18]:** Recognizing ripe crops or fruits for efficient harvesting.

The modular design of YOLOv11 ensures flexibility, enabling adaptation to various agricultural use cases. The integration of spatial attention mechanisms and multi-scale feature processing improves its accuracy in diverse field conditions, from densely planted crops to heterogeneous landscapes.

4. Training and Validation Settings

4.1. Training Procedure

The YOLOv11 architecture was employed to perform leaf detection and counting across five crop types: Spring barley, spring wheat, winter wheat, winter rye, and winter triticale. The dataset was manually annotated, with leaves delineated to provide ground truth data. Images were split into training and testing subsets, ensuring representative distributions for all crop types.

To evaluate the impact of training strategies, two models were trained for each crop. In the first approach, datasets from all five crops were combined to train a single model for the specific crop, leveraging the combined data to improve generalization. In the second approach, only the dataset of the specific crop under evaluation was used for training, focusing the model on crop-specific features. This dual strategy allowed for a comparative analysis of performance between generalized and crop-specific training, providing insights into the benefits and limitations of both approaches under real-world conditions.

4.1.1. Data Augmentation

To improve model generalization, several data augmentation techniques were applied. These included random rotations, horizontal and vertical flips, random cropping,

and adjustments to brightness and contrast. Additionally, advanced augmentations such as Gaussian blur, median blur, grayscale transformations, and CLAHE were utilized to simulate diverse real-world imaging conditions.

4.1.2. Training Configuration

The model was trained using the Adam optimizer with an initial learning rate of 0.0001, managed via a cosine decay schedule. A batch size of eight was used to optimize computational efficiency given the hardware constraints. Early stopping was employed with a patience value of 10 epochs, ensuring the prevention of overfitting. Training hyperparameters such as momentum, weight decay, and warm-up strategies were fine-tuned for optimal performance, as outlined in Table 1.

Table 1. YOLOv11 training hyperparameters and dataset specifications.

Hardware Specifications	
CPU	Google Colab
GPU	NVIDIA T4
RAM	40 GB
Model Specifications (YOLOv11)	
Optimizer	Adam
Batch Size	8
Learning Rate	0.0001 (Cosine LR Decay)
Epochs	50
Patience	100 (Early Stopping)
NMS	False
LR Decay Rate	0.01
Momentum	0.937
Weight Decay	0.0005
Warmup Epochs	3.0
Warmup Momentum	0.8
Warmup Bias LR	0.1
Box Loss Coeff.	7.5
Classification Loss Coeff.	0.5
DFL Loss Coeff.	1.5
Pose Loss Coeff.	12.0
K-Object Loss Coeff.	1.0
Label Smoothing	0.0
Data Specifications	
Spring Barley Train	419 Images
Spring Barley Test	43 Images
Spring Wheat Train	410 Images
Spring Wheat Test	38 Images
Winter Wheat Train	25 Images
Winter Wheat Test	3 Images
Winter Rye Train	25 Images
Winter Rye Test	3 Images
Winter Triticale Train	25 Images
Winter Triticale Test	3 Images
Image Size	640 × 640 Pixels
Data Augmentation	
Blur	$p = 0.01$, blur_limit = (3, 7)
Median Blur	$p = 0.01$, blur_limit = (3, 7)
To Gray	$p = 0.01$
CLAHE	$p = 0.01$, clip_limit = (1, 4.0), tile_grid_size = (8, 8)

4.1.3. Loss Functions

The following loss functions were employed. For bounding box tasks, the *Distribution Focal Loss* (DFL) was used to enhance sensitivity to challenging examples by emphasizing harder-to-classify bounding boxes, while the *Complete Intersection over Union* (CIoU) Loss provided a comprehensive metric for bounding box alignment by considering over-

lap, center distance, and aspect ratio. For segmentation tasks, the *Binary Cross-Entropy (BCE) Loss* ensured effective pixel-wise classification, and the *IoU Loss* was utilized to maximize the overlap between predicted and ground truth masks, which is crucial for accurate segmentation.

4.1.4. Hardware Specifications

The training experiments were performed on a Google Colab environment equipped with an NVIDIA T4 GPU and 40 GB of RAM. This setup enabled efficient processing of high-resolution 640×640 pixel images and accelerated training cycles.

4.2. Validation and Evaluation

The performance of YOLOv11 models was evaluated on the validation set of each crop type. Evaluation metrics primarily focused on the following:

- **Detection Accuracy:** The ability to correctly identify and localize leaves within images.
- **Loss Metrics:** Training and validation loss trends were monitored throughout the training process to evaluate model convergence and generalization.

5. Analysis and Results

The performance of YOLOv11-based models was evaluated using validation datasets for each crop type, focusing on bounding box (BBox) and segmentation metrics. These evaluations assessed the models' ability to detect and count leaves accurately across spring barley, spring wheat, winter wheat, winter rye, and winter triticale. The results are summarized below, with detailed analysis and visualizations provided in Figures 6–9 and Table 2.

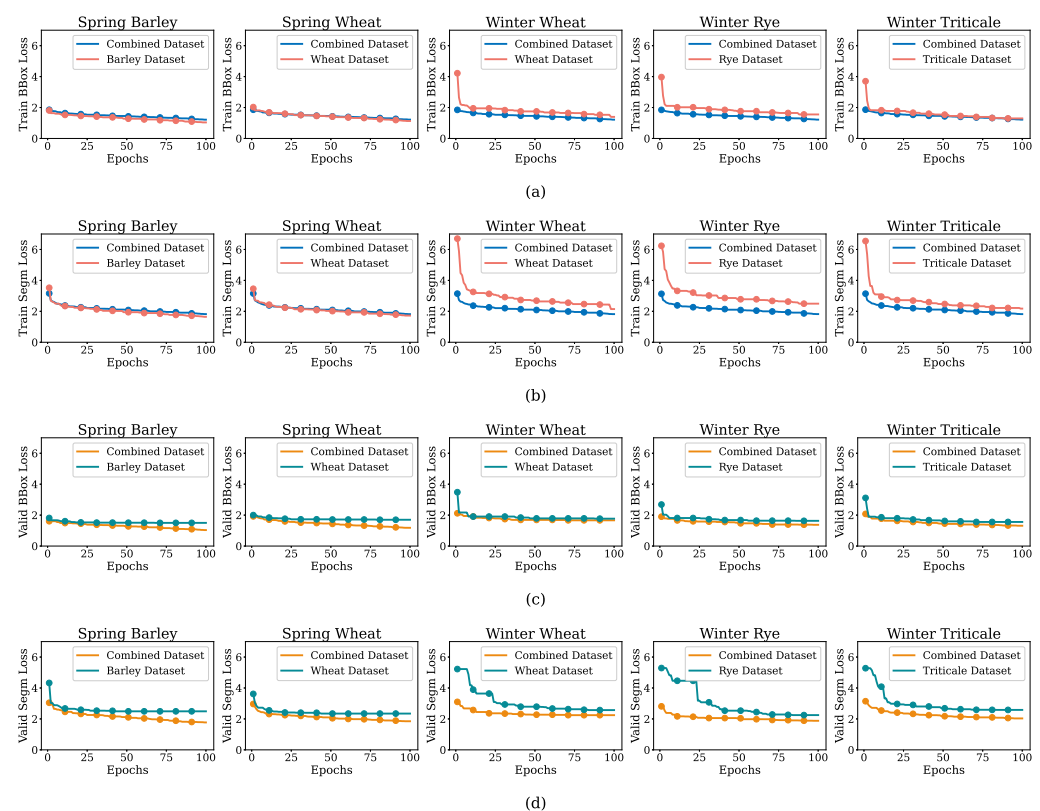


Figure 6. (a–d) Training and validation losses for spring barley, spring wheat, winter wheat, winter rye, and winter triticale across combined and individual datasets, showing box and segmentation losses.

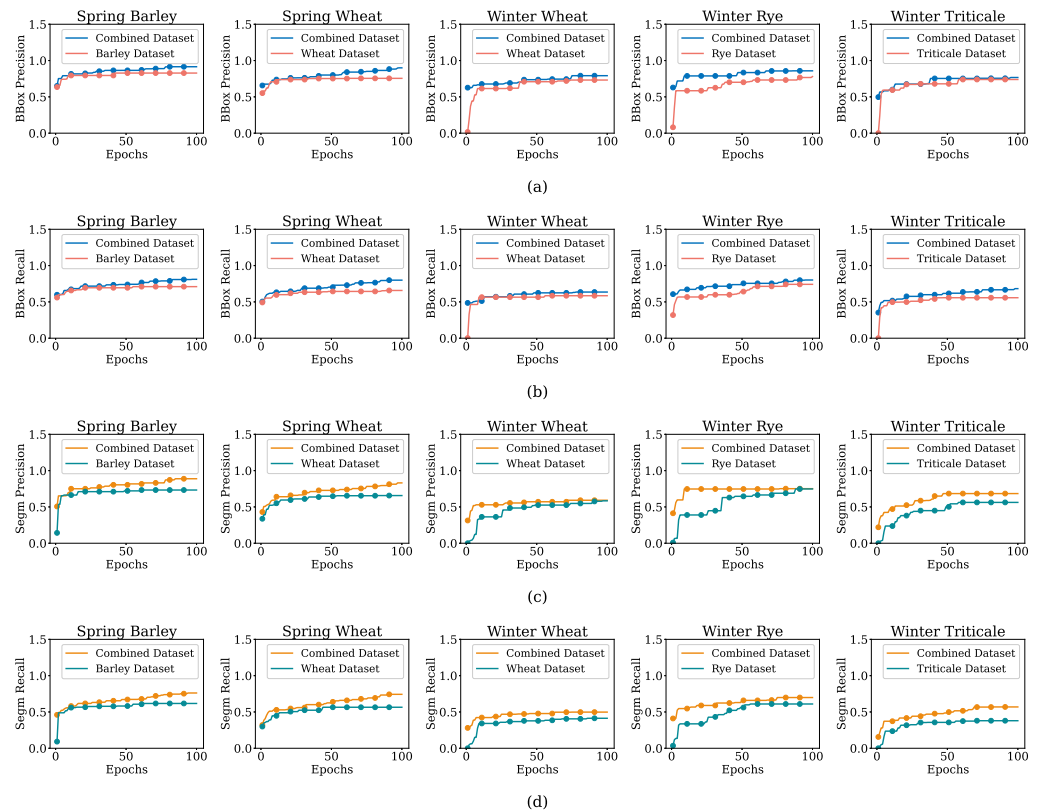


Figure 7. (a–d) Precision and recall for spring barley, spring wheat, winter wheat, winter rye, and winter triticale across combined and individual datasets, showing box and segmentation precision and recall.

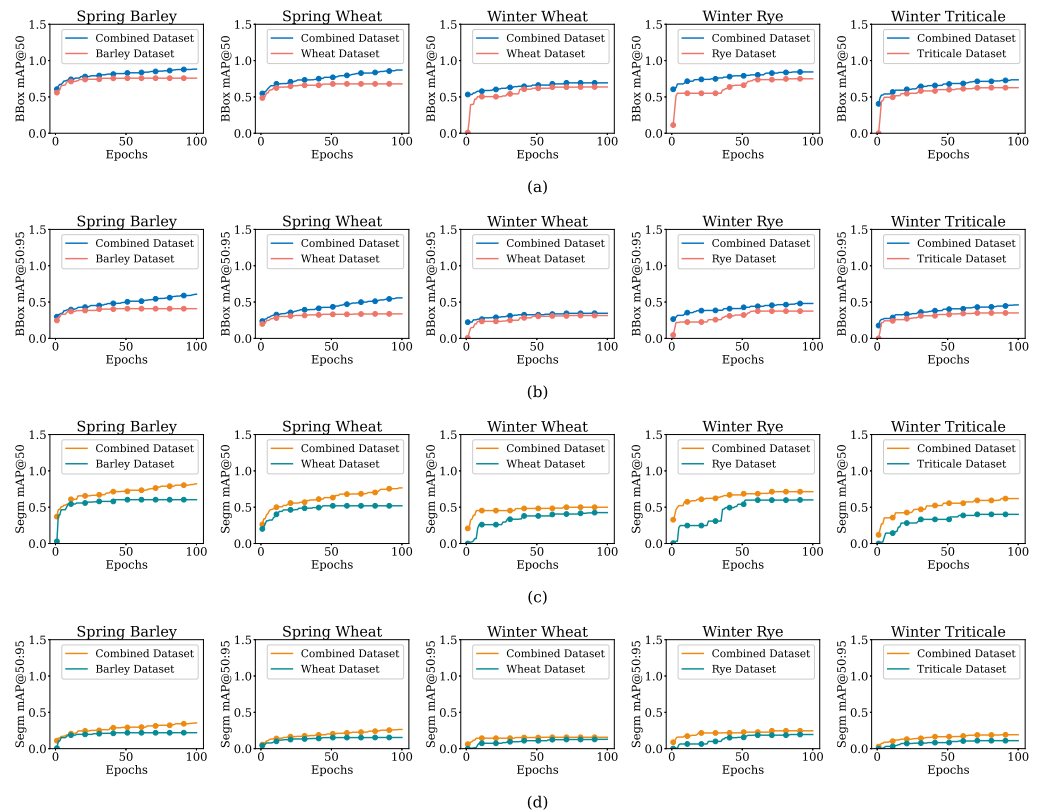


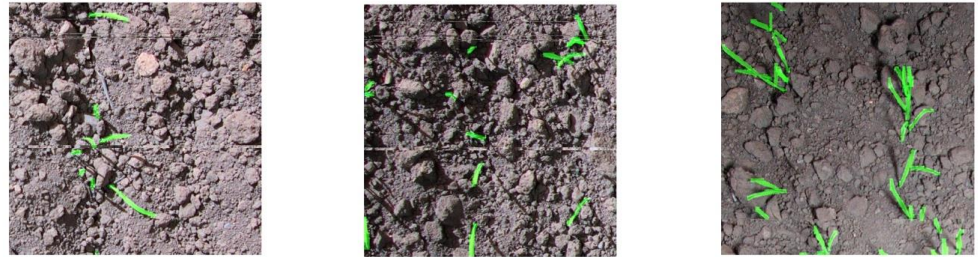
Figure 8. (a–d) Mean Average Precision (mAP) for spring barley, spring wheat, winter wheat, winter rye, and winter triticale across combined and individual datasets, showing box and segmentation @mAP50 and @mAP50:95.

Spring Barley (Robot Image)



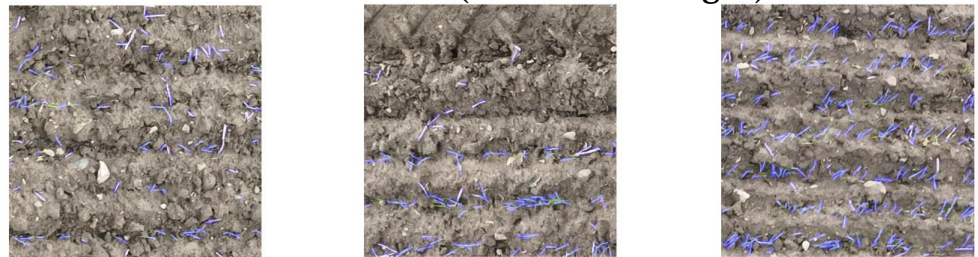
(a)

Spring Wheat (Robot Images)



(b)

Winter Wheat (8m Drone Images)



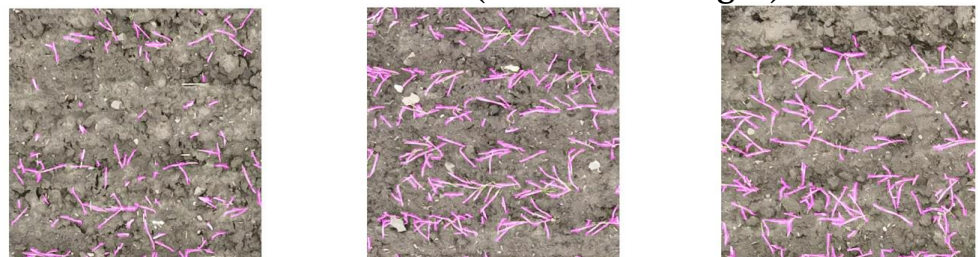
(c)

Winter Rye (8m Drone Images)



(d)

Winter Triticale (8m Drone Images)



(e)

Figure 9. (a–e) Prediction results for spring barley, spring wheat (robot images), and winter wheat, rye, and triticale (8 m drone images), with detected crops highlighted in color-coded annotations.

Table 2. Evaluation metrics for spring and winter crops across combined and individual datasets.

Crop	Dataset	Images	Box (P)	Box (R)	Box mAP50	Box mAP50:95	Mask (P)	Mask (R)	Mask mAP50	Mask mAP50:95
Spring Barley	Combined	43	0.902	0.795	0.881	0.607	0.862	0.755	0.822	0.355
	Spring Barley	43	0.828	0.633	0.755	0.410	0.699	0.573	0.604	0.219
Spring Wheat	Combined	38	0.884	0.787	0.865	0.558	0.813	0.726	0.755	0.263
	Spring Wheat	38	0.884	0.787	0.865	0.558	0.813	0.726	0.755	0.263
Winter Wheat	Combined	3	0.731	0.613	0.683	0.345	0.564	0.452	0.470	0.151
	Winter Wheat	3	0.705	0.567	0.631	0.312	0.574	0.391	0.421	0.127
Winter Rye	Combined	3	0.816	0.764	0.825	0.478	0.685	0.655	0.699	0.228
	Winter Rye	3	0.802	0.685	0.764	0.420	0.668	0.563	0.554	0.188
Winter Triticale	Combined	3	0.735	0.663	0.732	0.460	0.654	0.541	0.591	0.183
	Winter Triticale	3	0.709	0.534	0.623	0.351	0.534	0.385	0.405	0.110

5.1. Loss and Convergence Analysis

Figure 6 illustrates the training and validation loss curves for BBox and segmentation tasks across all crop types. Models trained on the combined dataset consistently exhibited lower loss values during training and validation, particularly for spring crops. This reflects the ability of the combined dataset to provide more diverse and representative features, facilitating better generalization. In contrast, models trained on individual datasets had higher loss values, especially for winter crops, underscoring the challenges posed by limited data and increased variability.

5.2. Precision and Recall Analysis

Figure 7 presents the precision and recall curves for BBox and segmentation tasks. Models trained on the combined dataset consistently achieved higher precision and recall values for both spring and winter crops, with a more noticeable advantage for bounding box tasks. For instance, spring barley and spring wheat achieved over 0.9 precision and recall for BBox tasks when using the combined dataset. However, for segmentation tasks, the difference between combined and individual datasets was smaller, particularly for winter crops, where the individual datasets occasionally outperformed the combined dataset in recall. This indicates that while the combined dataset enhances generalization, specific nuances of individual crops may still require specialized training.

5.3. Mean Average Precision (mAP) Analysis

The mAP metrics for BBox and segmentation tasks, as shown in Figure 8, further highlight the superior performance of the combined dataset. For spring crops, the mAP@50 values for both BBox and segmentation exceeded 0.85, with segmentation metrics slightly trailing BBox metrics. For winter crops, the combined dataset still outperformed individual datasets in BBox mAP, achieving values above 0.7, while segmentation mAP remained lower due to the inherent challenges of fewer samples and more complex imaging conditions. The combined dataset's higher mAP values indicate its ability to leverage cross-crop similarities for robust detection.

5.4. Performance Metrics Overview

The quantitative metrics summarized in Table 2 underscore the combined dataset's advantage across most tasks. For instance, spring barley achieved a BBox mAP@50 of 0.881 and a segmentation mAP@50 of 0.822 using the combined dataset, compared to 0.755 and 0.604, respectively, for the individual dataset. Similarly, for winter crops like winter triticale, the combined dataset achieved a segmentation mAP@50 of 0.591, outperforming

the individual dataset's 0.405. These results emphasize that a larger and more diverse dataset significantly enhanced model performance, particularly for bounding box tasks, while segmentation tasks remained more sensitive to dataset variability.

5.5. Detection Insights from Visual Outputs

Figure 9 illustrates the visual results for all crop types. Models trained on the combined dataset provided denser and more accurate detections, particularly for spring crops, with minimal false positives or missed detections. For winter crops, while detections were generally less dense, the combined dataset still produced more reliable outputs in challenging regions with occlusion or complex backgrounds. The visual outputs align with the quantitative metrics, highlighting the effectiveness of leveraging diverse datasets for robust performance.

5.6. Discussion of Combined vs. Individual Training

Overall, the combined dataset demonstrated superior performance across most metrics, particularly for bounding box tasks, due to the larger and more diverse training pool. This generalization ability is crucial for practical applications where crop-specific annotations may be limited. However, for segmentation tasks, individual datasets occasionally provided better recall, indicating that crop-specific features might still benefit from targeted training. This trade-off highlights the importance of balancing dataset diversity with crop-specific fine-tuning to optimize performance across all phenotyping tasks.

6. Discussion

The application of YOLOv11 for leaf detection and counting across diverse crop types highlights the transformative potential of advanced deep learning methods in automated plant phenotyping. By leveraging architectural innovations and real-time processing capabilities, YOLOv11 has shown promise in addressing the challenges of large-scale agricultural assessments, where manual methods are both labor-intensive and error-prone. The model's ability to detect and quantify leaves under varying environmental and imaging conditions underscores its versatility and practical utility.

A key strength of YOLOv11 lies in its architectural enhancements, such as the introduction of the C3k2 block and the C2PSA module. These components have substantially improved feature extraction and spatial attention, enabling the model to effectively handle complex leaf morphologies and varying background conditions. The integration of such mechanisms has set YOLOv11 apart from its predecessors, allowing it to adapt to different imaging scenarios, including high-resolution robot images and drone-captured data at higher altitudes. Additionally, the real-time inference capabilities of YOLOv11 make it well suited for field deployment in precision agriculture, where rapid decision-making is often critical.

However, certain challenges remain that highlight the need for further refinement. Occlusion, caused by overlapping plant structures, continues to hinder detection performance, particularly in densely populated or highly vegetative areas. The inability to consistently separate overlapping leaves limits the model's ability to achieve perfect accuracy in such scenarios. Similarly, extreme lighting variations, ranging from overexposure in bright sunlight to underexposure in shaded regions, can introduce inconsistencies in detection. These factors, inherent to real-world agricultural environments, represent areas where the model's robustness can be improved.

The PCA analysis presented in Figure 4 underscores the distinct differences between robot- and drone-captured images, which are driven by environmental factors and imaging conditions. Robot-captured images, taken closer to the ground, provide larger and more

detailed leaf representations under controlled conditions. In contrast, drone-captured images, taken from higher altitudes, reflect smaller and less uniform leaf features influenced by natural field conditions such as wind and lighting variability. Despite these differences, the combined dataset approach adopted in this study enabled the model to generalize effectively across varying environmental conditions. This is evidenced by the consistently high mAP scores for both spring crops (robot-captured) and winter crops (drone-captured). However, segmentation tasks for winter crops, which involve greater environmental variability, occasionally showed limitations, emphasizing the need for further dataset expansion and domain-specific fine-tuning. Future work will focus on enhancing dataset diversity and training strategies to ensure robust generalization across heterogeneous imaging conditions.

The variability in imaging conditions across crop types further underscores the importance of dataset quality and diversity. While YOLOv11 performs well with larger, diverse datasets such as those for spring crops, smaller datasets for winter crops present a significant challenge. Limited data availability reduces the model's ability to generalize effectively, particularly when faced with complex soil textures, sparse leaf distributions, or drone images captured at varying altitudes. Strategies such as employing advanced data augmentation techniques, including rotations, brightness adjustments, and synthetic data generation using Generative Adversarial Networks (GANs) [19], can help mitigate the impact of dataset imbalance. GANs can create realistic synthetic images for underrepresented winter crops, enabling more balanced training. Additionally, diffusion models [20], known for their ability to generate high-quality and diverse synthetic data, can serve as an effective alternative to GANs by better capturing the variability and complexity of real-world agricultural datasets. Furthermore, the integration of Large Language Models (LLMs) [21] can streamline the annotation process by generating descriptive meta-data and assisting in semi-automated pipelines, making data preparation more efficient and scalable.

One of the most valuable applications of precise leaf segmentation is in the estimation of the Leaf Area Index (LAI), a key parameter for assessing plant canopy structure and health [6]. LAI estimation relies heavily on accurate detection and segmentation of leaves, particularly under dense canopy conditions, where overlapping leaves and occlusion present challenges. By providing high-resolution segmentation outputs, YOLOv11 can facilitate the derivation of LAI metrics, which are instrumental for understanding photosynthetic efficiency, water use, and light interception in crops. This makes it an indispensable tool in optimizing crop management and modelling resource allocation under real-world agricultural conditions.

One significant finding of this study is that the combined dataset generally led to improved performance, particularly for bounding box tasks, largely due to its larger size and diversity. However, this improvement was not universal. For segmentation tasks, individual models occasionally outperformed the unified model for certain crops, such as winter wheat and winter triticale. This suggests that while a larger dataset helps generalization, specific nuances and challenges associated with particular crops, such as unique leaf morphologies or environmental conditions, may require specialized training. For instance, winter crops often face greater occlusion and variability in natural field conditions, which could limit the benefits of a generalized model [22]. Semi-automated annotation pipelines using pre-trained YOLOv11 models for pseudo-labeling, followed by human verification, could also accelerate dataset creation for underrepresented crops while maintaining annotation quality.

One of the most critical areas for future research is the expansion of annotated datasets to enhance the model's performance and generalization capabilities [23]. In-

creasing the volume and diversity of annotated data is paramount, particularly for crops with smaller datasets, such as winter wheat, winter rye, and winter triticale. Larger datasets will enable the model to better capture the inherent variability in leaf morphology, growth stages, and environmental conditions, leading to more robust detection across all crop types. Furthermore, increasing the diversity of annotated data by including samples from multiple geographical regions and agricultural practices will enhance the model's adaptability. By capturing the variability in crop appearances across different climates, soil types, and cultivation methods, the expanded dataset will provide the model with a more comprehensive understanding of leaf characteristics. The annotation process should focus on incorporating samples from diverse field conditions, including variations in lighting, soil textures, and plant densities. Expanding the range of annotated images to include more occluded and overlapping leaves will allow the model to learn from challenging scenarios, thereby improving its ability to handle real-world complexities. Additionally, integrating annotated images from different sensor modalities, such as multispectral or hyperspectral cameras, could provide richer information for both detection and segmentation tasks.

To efficiently generate large-scale annotated datasets, semi-automated or automated annotation pipelines should be explored [24]. Leveraging pre-trained models to assist with initial annotations and refining these using human verification can significantly reduce the time and effort required for data preparation. This approach would be particularly useful for crops with limited datasets, allowing researchers to scale up annotations without excessive manual effort [25].

The impact of dataset size and diversity cannot be overstated, as it directly influences the model's ability to generalize across crop types and environmental conditions. However, the results also indicate that the effectiveness of a combined dataset depends on the specific crop and task, highlighting the need for a balanced approach that combines the benefits of diverse data with crop-specific fine-tuning. As the annotated dataset grows, the YOLOv11 models will become increasingly capable of handling the nuances of leaf detection and counting, paving the way for more reliable and scalable applications in precision agriculture.

7. Conclusions

This study demonstrated the effectiveness of YOLOv11 for automated leaf detection and counting across spring and winter crops. The use of high-resolution robot images and drone-captured data enabled robust evaluations, with high accuracy observed for spring crops. However, challenges such as occlusion, lighting variability, and smaller dataset sizes impacted performance for winter crops. These findings highlight the importance of larger and more diverse annotated datasets to improve model robustness and generalization. Expanding dataset diversity and leveraging semi-automated annotation tools can address these limitations. Overall, YOLOv11 offers a scalable and efficient solution for leaf detection and segmentation, advancing automated plant phenotyping in precision agriculture.

Author Contributions: Conceptualization, A.T.K. and S.M.J.; Methodology, A.T.K.; Formal analysis, A.T.K.; Writing—original draft, A.T.K.; Writing—review & editing, S.M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Promilleafgiftsfonden for landbrug.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This work was part of the project *Halm til det hele*, which is financed by the Danish Promilleafgiftsfonden for Landbrug.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Furbank, R.T.; Tester, M. Phenomics-technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [[CrossRef](#)] [[PubMed](#)]
2. Reynolds, M.; Atkin, O.K.; Bennett, M.; Cooper, M.; Dodd, I.C.; Foulkes, M.J.; Frohberg, C.; Hammer, G.; Henderson, I.R.; Huang, B.; et al. Addressing research bottlenecks to crop productivity. *Trends Plant Sci.* **2021**, *26*, 607–630. [[CrossRef](#)]
3. Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.; Yan, J. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* **2020**, *13*, 187–214. [[CrossRef](#)] [[PubMed](#)]
4. Chawade, A.; van Ham, J.; Blomquist, H.; Bagge, O.; Alexandersson, E.; Ortiz, R. High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy* **2019**, *9*, 258. [[CrossRef](#)]
5. Ronse De Craene, L.P. Are petals sterile stamens or bracts? The origin and evolution of petals in the core eudicots. *Ann. Bot.* **2007**, *100*, 621–630. [[CrossRef](#)]
6. Das, P.; Rahimzadeh-Bajgirani, P.; Livingston, W.; McIntire, C.D.; Bergdahl, A. Modeling forest canopy structure and developing a stand health index using satellite remote sensing. *Ecol. Inform.* **2024**, *84*, 102864. [[CrossRef](#)]
7. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
8. Rose, J.C.; Paulus, S.; Kuhlmann, H. Accuracy analysis of a multi-view stereo approach for phenotyping of tomato plants at the organ level. *Sensors* **2015**, *15*, 9651–9665. [[CrossRef](#)] [[PubMed](#)]
9. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.G. Unmanned Aerial Vehicles (UAV) in precision agriculture: Applications and challenges. *Energies* **2021**, *15*, 217. [[CrossRef](#)]
10. Singh, A.; Jones, S.; Ganapathysubramanian, B.; Sarkar, S.; Mueller, D.; Sandhu, K.; Nagasubramanian, K. Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* **2021**, *26*, 53–69. [[CrossRef](#)] [[PubMed](#)]
11. Johannsen, L.C.; Khan, A.T.; Jensen, S.M.; Kruppa-Scheetz, J. Innovative Leaf Disease Mapping: Unsupervised Anomaly Detection for Precise Area Estimation. *Int. J. Mach. Learn. Cybern.* **2024**, preprint.
12. Madec, S.; Jin, X.; Lu, H.; De Solan, B.; Liu, S.; Duyme, F.; Heritier, E.; Baret, F. Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* **2019**, *264*, 225–234. [[CrossRef](#)]
13. Khan, A.T.; Jensen, S.M.; Khan, A.R.; Li, S. Plant disease detection model for edge computing devices. *Front. Plant Sci.* **2023**, *14*, 1308528. [[CrossRef](#)]
14. Maes, W.H.; Steppe, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [[CrossRef](#)]
15. d’Andrimont, R.; Yordanov, M.; Martinez-Sanchez, L.; Van der Velde, M. Monitoring crop phenology with street-level imagery using computer vision. *Comput. Electron. Agric.* **2022**, *196*, 106866. [[CrossRef](#)]
16. Junior, L.C.M.; Ulson, J.A.C. Real time weed detection using computer vision and deep learning. In Proceedings of the 2021 14th IEEE International Conference on Industry Applications (INDUSCON), Sao Paulo, Brazil, 15–18 August 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1131–1137.
17. Abbaspour-Gilandeh, Y.; Aghabara, A.; Davari, M.; Maja, J.M. Feasibility of using computer vision and artificial intelligence techniques in detection of some apple pests and diseases. *Appl. Sci.* **2022**, *12*, 906. [[CrossRef](#)]
18. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* **2020**, *7*, 1–19. [[CrossRef](#)]
19. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training gans. In *Advances in Neural Information Processing Systems, Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016*; Curran Associates Inc.: Red Hook, NY, USA, 2016; Volume 29, p. 29.
20. Trabucco, B.; Doherty, K.; Gurinas, M.; Salakhutdinov, R. Effective data augmentation with diffusion models. *arXiv* **2023**, arXiv:2302.07944.
21. Bonifacio, L.; Abonizio, H.; Fadaee, M.; Nogueira, R. Inpars: Data augmentation for information retrieval using large language models. *arXiv* **2022**, arXiv:2202.05144.
22. Hamuda, E.; Mc Ginley, B.; Glavin, M.; Jones, E. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Comput. Electron. Agric.* **2017**, *133*, 97–107. [[CrossRef](#)]
23. Amin, S.U.; Hussain, A.; Kim, B.; Seo, S. Deep learning based active learning technique for data annotation and improve the overall performance of classification models. *Expert Syst. Appl.* **2023**, *228*, 120391. [[CrossRef](#)]

24. Pangakis, N.; Wolken, S. Keeping Humans in the Loop: Human-Centered Automated Annotation with Generative AI. *arXiv* **2024**, arXiv:2409.09467.
25. Wang, Y.; Stevens, D.; Shah, P.; Jiang, W.; Liu, M.; Chen, X.; Kuo, R.; Li, N.; Gong, B.; Lee, D.; et al. Model-in-the-Loop (MILO): Accelerating Multimodal AI Data Annotation with LLMs. *arXiv* **2024**, arXiv:2409.10702.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.