

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

Ameer Tamoor Khan

atk@plen.ku.dk

University of Copenhagen <https://orcid.org/0000-0001-6838-992X>

Signe Marie Jensen

University of Copenhagen <https://orcid.org/0000-0002-7738-5231>

Research Article

Keywords: Plant Phenotyping, YOLOv11, Leaf Segmentation, Precision Agriculture, Deep Learning

Posted Date: December 6th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-5582314/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

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

Ameer Tamoor Khan, Signe Marie Jensen

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 above 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. This highlights that while the combined dataset improves generalization, the unique characteristics of individual crops may still benefit from specialized training. This work demonstrates the potential of AI-driven models to advance automated phenotyping for large-scale precision agriculture.

Index Terms—Plant Phenotyping, YOLOv11, Leaf Segmentation, Precision Agriculture, Deep Learning.

I. 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 II discusses the dataset used in this study, including data collection and annotation processes. Section III outlines the methodology employed for leaf segmentation and counting. Section IV describes the training and validation of the models. Section V presents the analysis and results. Section VI provides a discussion of the findings, and Section VII concludes the paper with final remarks and future directions.

II. 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 Fig. 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.

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 meters, 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 meters with an image resolution of 5280×395 . The image sources and heights are detailed in Fig. 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 Fig. 2.

In addition to the image distribution, the number of annotated leaves in the dataset was a critical aspect. Fig. 3 shows the leaf count for each crop type in the training and testing datasets. The spring barley training set contained 4,987 leaves, with 646 in the testing set. Similarly, spring wheat had

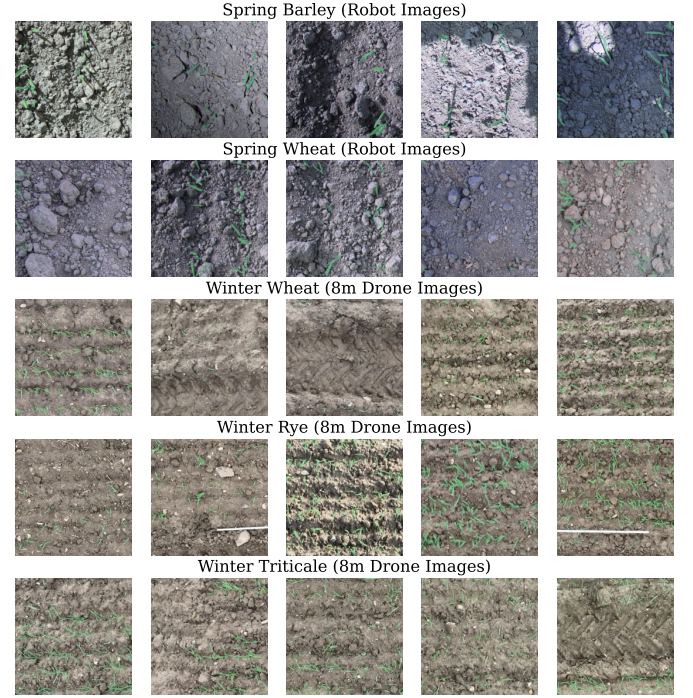


Fig. 1. Sample images from the spring and winter crops dataset, displaying annotated leaves highlighted in green, which were otherwise challenging to distinguish in raw images. spring barley and spring wheat images were captured using an agricultural robot (overall height approximately 2.15 meters),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

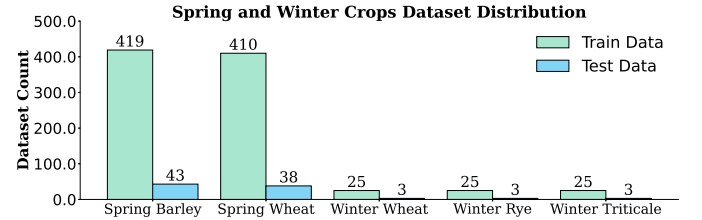


Fig. 2. Spring and winter Crops Dataset Distribution illustrating the count of images in the training and testing datasets for each crop type.

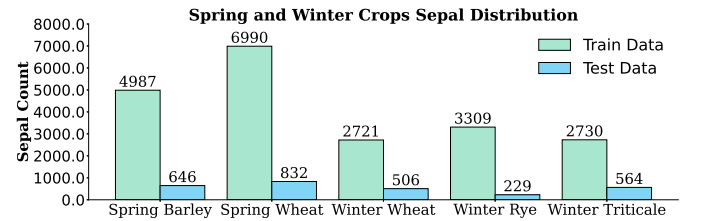


Fig. 3. Spring and winter Crops Leaf Distribution showing the count of manually annotated leaves across the training and testing datasets for each crop type.

6,990 leaves in the training set and 832 in the testing set. The winter crops had considerably fewer leaves, with winter wheat having 2,721 leaves in training and 506 in testing, winter rye having 3,309 in training and 229 in testing, and winter triticale containing 2,730 in training and 564 in testing. These counts demonstrated that the training datasets were significantly more comprehensive than the testing datasets, ensuring robust model

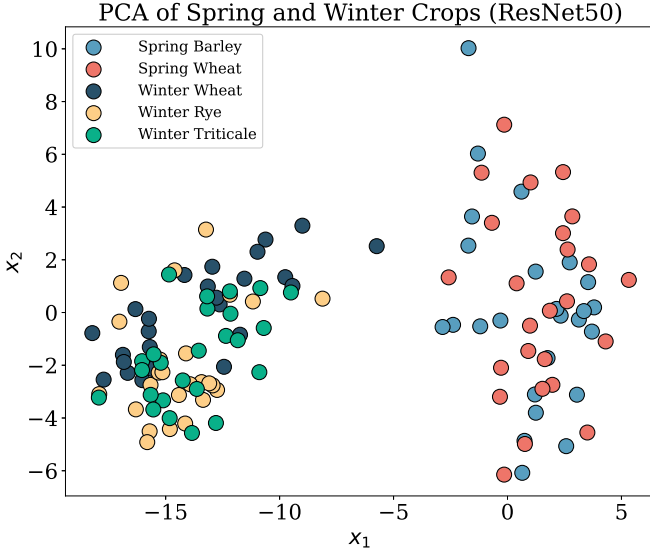


Fig. 4. PCA of spring and winter crops visualized using features from a pretrained ResNet50 model. The clustering reflected height-based differences, with robot images capturing larger leaf sizes due to proximity to the ground and controlled conditions, while drone images showed smaller leaf sizes influenced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

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. The PCA results, shown in Fig. 4, demonstrated distinct clustering based on the imaging source. This difference is attributed to variations in leaf size and imaging conditions. Robot-captured images, taken closer to the ground, provide larger and more detailed leaf representations under controlled conditions, avoiding real-world disturbances such as wind or lighting variations. In contrast, drone-captured images, taken from a higher altitude under natural field conditions, are affected by environmental factors, leading to smaller and less uniform leaf representations.

Overall, the spring and winter crops 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.

III. ARCHITECTURE OF YOLOV11

The YOLOv11 architecture introduces several critical enhancements to improve efficiency and detection accuracy, as illustrated in Fig. 5. The design comprises three primary components: Backbone, Neck, and Head, each incorporating innovative mechanisms to optimize performance.

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 2, 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.

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

- **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) *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.

A. Applications in Agriculture

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

- **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.

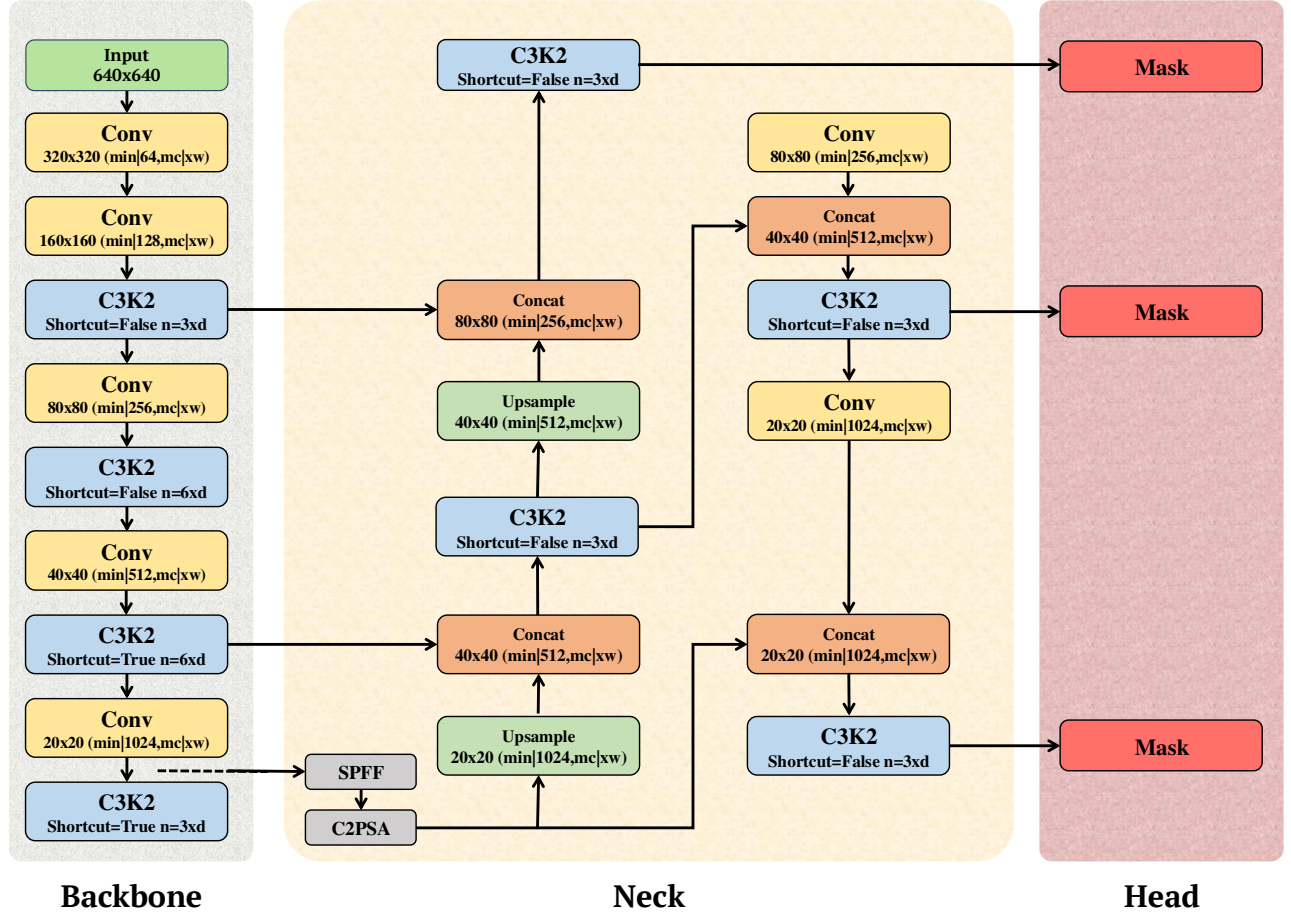


Fig. 5. The schematic diagram of YOLOv11 illustrating its three core components: Backbone, Neck, and Head. The Backbone handles multi-scale feature extraction using advanced blocks such as the C3k2 and Spatial Pyramid Pooling Fast (SPPF), designed for efficient feature representation. The Neck aggregates 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.

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.

IV. TRAINING AND VALIDATION SETTINGS

A. 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 sepals 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.

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.

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 8 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 I.

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	640x640 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)

TABLE I
YOLOV11 TRAINING HYPERPARAMETERS AND DATASET
SPECIFICATIONS

3) *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 640x640 pixel images and accelerated training cycles.

B. Validation and Evaluation

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

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

V. 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 Fig. 6–9 and TABLE II.

A. Loss and Convergence Analysis

Fig. 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.

B. Precision and Recall Analysis

Fig. 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.

C. Mean Average Precision (mAP) Analysis

The mAP metrics for BBox and segmentation tasks, as shown in Fig. 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.

D. Performance Metrics Overview

The quantitative metrics summarized in TABLE II 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,

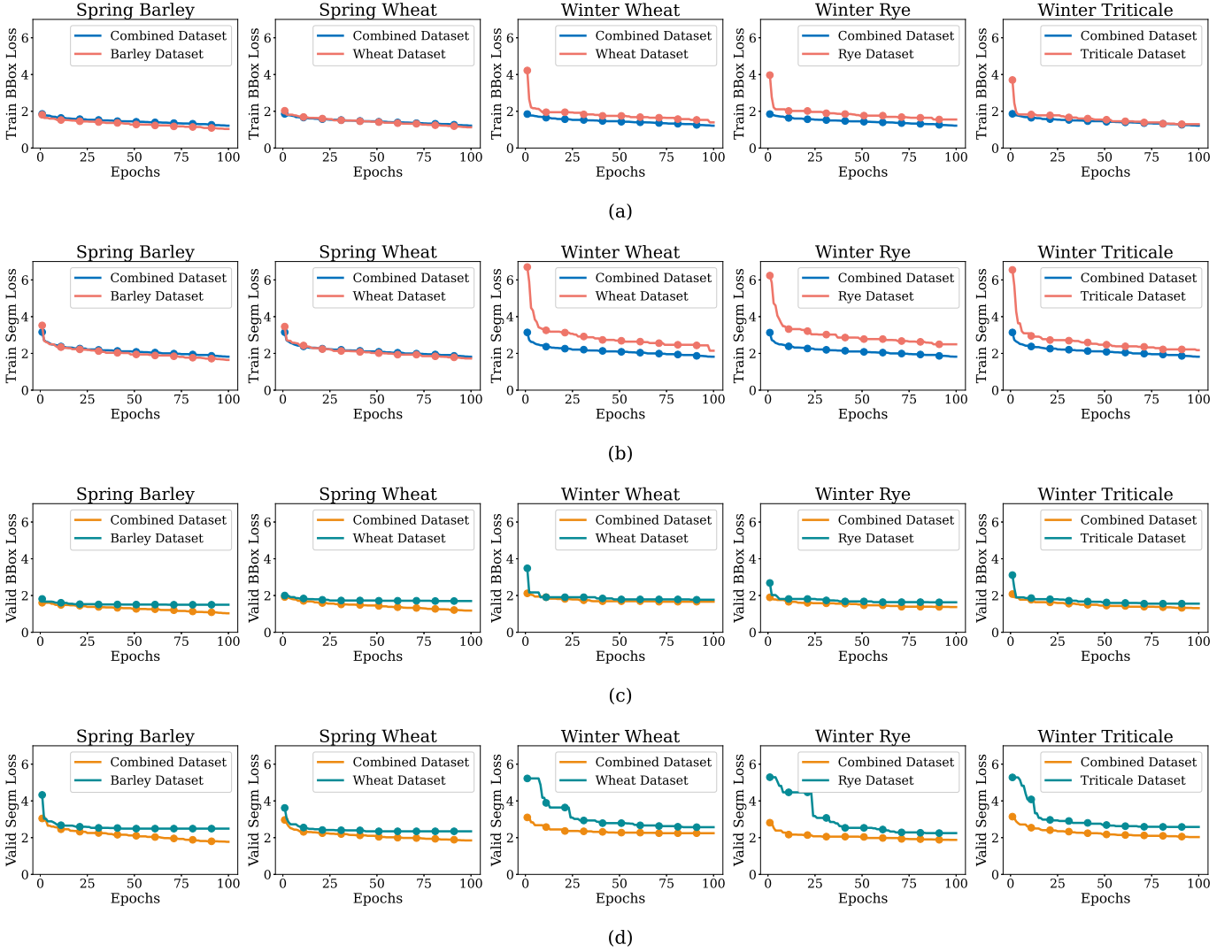


Fig. 6. Training and validation losses for spring barley, spring wheat, winter wheat, winter rye, and winter triticale across combined and individual datasets, showing box and segmentation losses.

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.

E. Detection Insights from Visual Outputs

Fig. 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.

F. 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.

VI. 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

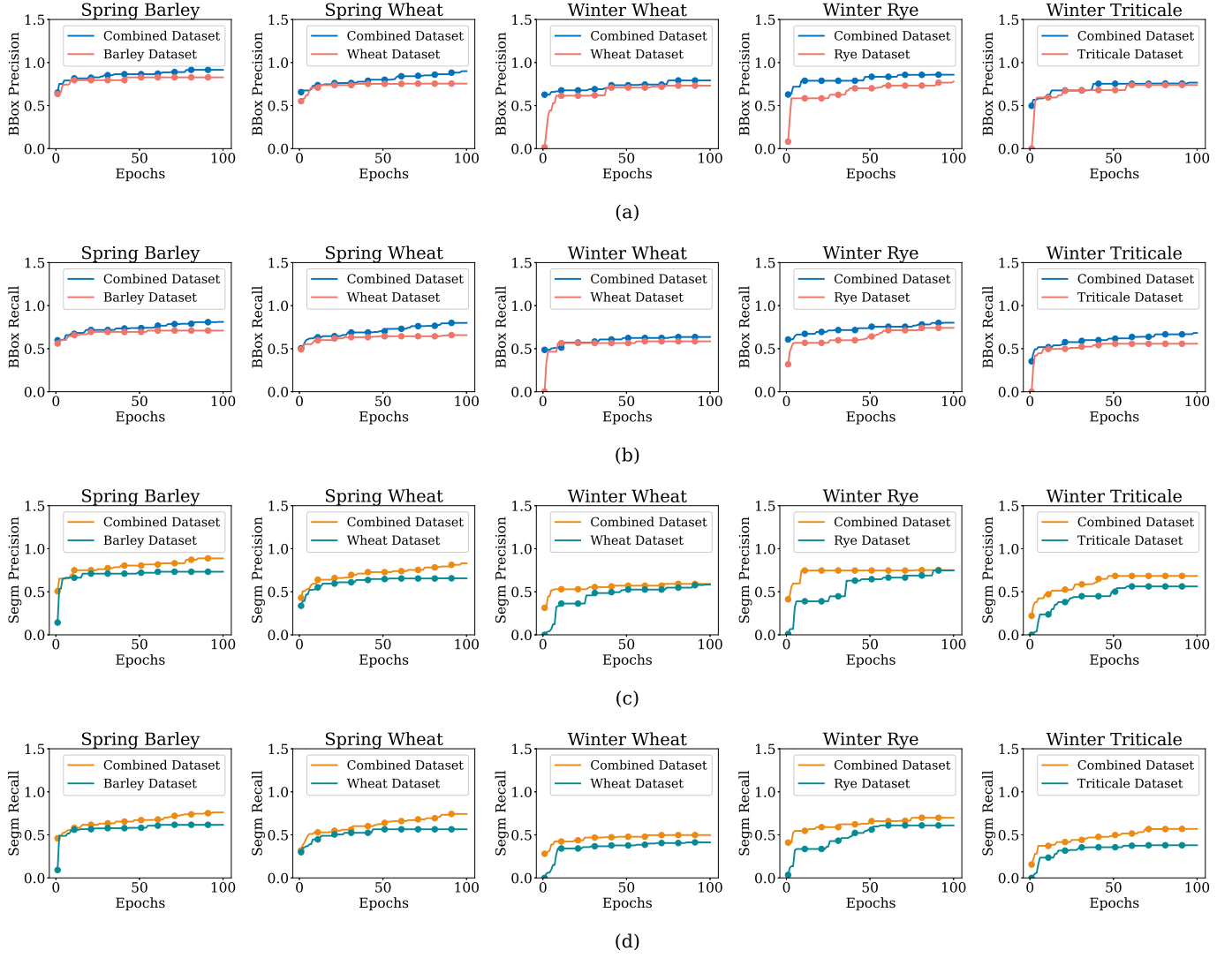


Fig. 7. Precision and Recall for spring barley, spring wheat, winter wheat, winter rye, and winter triticale across combined and individual datasets, showing box and segmentation precision and recall.

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 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. Addressing this imbalance is crucial for ensuring consistent performance across all crop types and conditions.

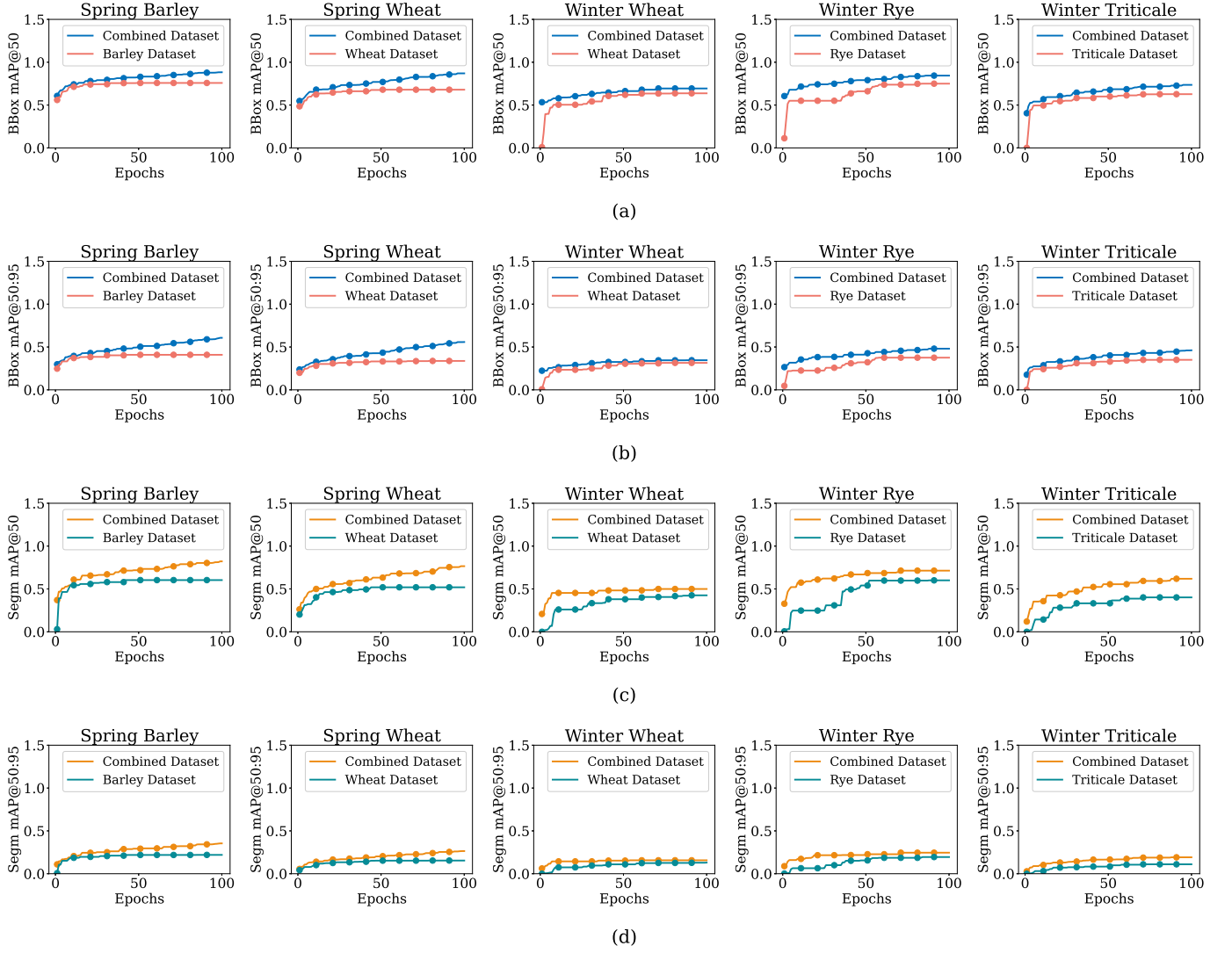


Fig. 8. Mean Average Precision (mAP) for spring barley, spring wheat, winter wheat, winter rye, and winter triticale across combined and individual datasets, showing box and segmentation @mAP50 and @mAP50:95.

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 modeling 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 [19].

One of the most critical areas for future research is the expansion of annotated datasets to enhance the model's performance and generalization capabilities [20]. Increasing 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

Crop	Dataset	Images	Instances	Box P	Box R	Box mAP50	Box mAP50:95	Mask P	Mask R	Mask mAP50	Mask mAP50:95
Spring Barley	Combined Dataset	43	645	0.902	0.795	0.881	0.607	0.862	0.755	0.822	0.355
	Spring Barley Dataset	43	645	0.828	0.633	0.755	0.410	0.699	0.573	0.604	0.219
Spring Wheat	Combined Dataset	38	832	0.884	0.787	0.865	0.558	0.813	0.726	0.755	0.263
	Spring Wheat Dataset	38	832	0.884	0.787	0.865	0.558	0.813	0.726	0.755	0.263
Winter Wheat	Combined Dataset	3	506	0.731	0.613	0.683	0.345	0.564	0.452	0.470	0.151
	Winter Wheat Dataset	3	506	0.705	0.567	0.631	0.312	0.574	0.391	0.421	0.127
Winter Rye	Combined Dataset	3	229	0.816	0.764	0.825	0.478	0.685	0.655	0.699	0.228
	Winter Rye Dataset	3	645	0.802	0.685	0.764	0.420	0.668	0.563	0.554	0.188
Winter Triticale	Combined Dataset	3	564	0.735	0.663	0.732	0.460	0.654	0.541	0.591	0.183
	Winter Triticale Dataset	3	564	0.709	0.534	0.623	0.351	0.534	0.385	0.405	0.110

TABLE II
EVALUATION METRICS FOR SPRING AND WINTER CROPS ACROSS COMBINED AND INDIVIDUAL DATASETS

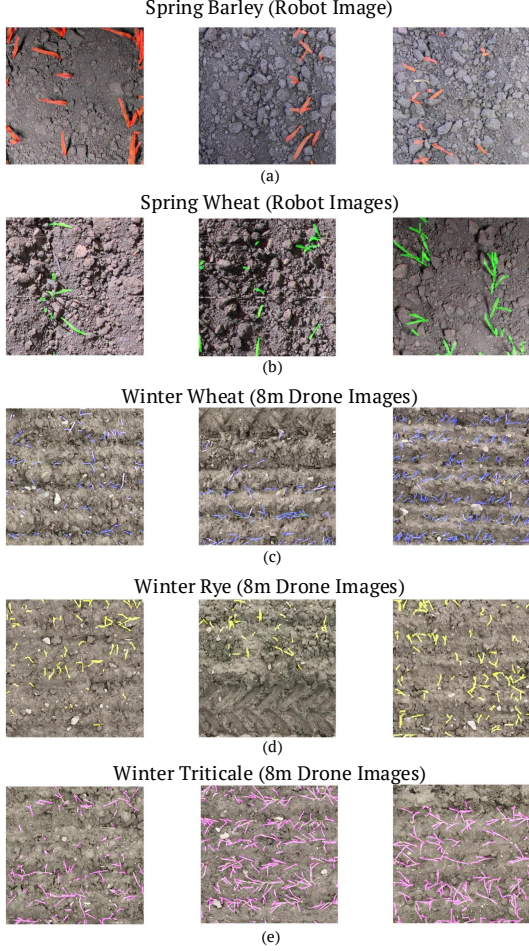


Fig. 9. Prediction results for spring barley, spring wheat (Robot images), and winter wheat, rye, and triticale (8m drone images), with detected crops highlighted in color-coded annotations.

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 [21]. 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 [22].

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.

VII. CONCLUSION

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.

Conflict of Interest

The authors declare that they have no conflicts of interest related to this project.

Acknowledgment

This work was part of the project, *Halm til det hele*, financed by the Danish Promilleaflgiftsfonden for Landbrug.

REFERENCES

- [1] R. T. Furbank and M. Tester, "Phenomics—technologies to relieve the phenotyping bottleneck," *Trends in plant science*, vol. 16, no. 12, pp. 635–644, 2011.
- [2] M. Reynolds, O. K. Atkin, M. Bennett, M. Cooper, I. C. Dodd, M. J. Foulkes, C. Frohberg, G. Hammer, I. R. Henderson, B. Huang, *et al.*, "Addressing research bottlenecks to crop productivity," *Trends in Plant Science*, vol. 26, no. 6, pp. 607–630, 2021.
- [3] W. Yang, H. Feng, X. Zhang, J. Zhang, J. H. Doonan, W. D. Batchelor, L. Xiong, and J. Yan, "Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives," *Molecular plant*, vol. 13, no. 2, pp. 187–214, 2020.
- [4] A. Chawade, J. van Ham, H. Blomquist, O. Bagge, E. Alexandersson, and R. Ortiz, "High-throughput field-phenotyping tools for plant breeding and precision agriculture," *Agronomy*, vol. 9, no. 5, p. 258, 2019.
- [5] L. P. Ronse De Craene, "Are petals sterile stamens or bracts? the origin and evolution of petals in the core eudicots," *Annals of Botany*, vol. 100, no. 3, pp. 621–630, 2007.
- [6] P. Das, P. Rahimzadeh-Bajgiran, W. Livingston, C. D. McIntire, and A. Bergdahl, "Modeling forest canopy structure and developing a stand health index using satellite remote sensing," *Ecological Informatics*, vol. 84, p. 102864, 2024.
- [7] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.
- [8] J. C. Rose, S. Paulus, and H. Kuhlmann, "Accuracy analysis of a multi-view stereo approach for phenotyping of tomato plants at the organ level," *Sensors*, vol. 15, no. 5, pp. 9651–9665, 2015.
- [9] P. Velusamy, S. Rajendran, R. K. Mahendran, S. Naseer, M. Shafiq, and J.-G. Choi, "Unmanned aerial vehicles (uav) in precision agriculture: Applications and challenges," *Energies*, vol. 15, no. 1, p. 217, 2021.
- [10] A. Singh, S. Jones, B. Ganapathysubramanian, S. Sarkar, D. Mueller, K. Sandhu, and K. Nagasubramanian, "Challenges and opportunities in machine-augmented plant stress phenotyping," *Trends in Plant Science*, vol. 26, no. 1, pp. 53–69, 2021.
- [11] L. C. Johannsen, A. T. Khan, S. M. Jensen, and J. Kruppa-Scheetz, "Innovative leaf disease mapping: Unsupervised anomaly detection for precise area estimation," 2024.
- [12] S. Madec, X. Jin, H. Lu, B. De Solan, S. Liu, F. Duyme, E. Heritier, and F. Baret, "Ear density estimation from high resolution rgb imagery using deep learning technique," *Agricultural and forest meteorology*, vol. 264, pp. 225–234, 2019.
- [13] A. T. Khan, S. M. Jensen, A. R. Khan, and S. Li, "Plant disease detection model for edge computing devices," *Frontiers in Plant Science*, vol. 14, p. 1308528, 2023.
- [14] W. H. Maes and K. Steppe, "Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture," *Trends in plant science*, vol. 24, no. 2, pp. 152–164, 2019.
- [15] R. d'Andrimont, M. Yordanov, L. Martinez-Sanchez, and M. Van der Velde, "Monitoring crop phenology with street-level imagery using computer vision," *Computers and Electronics in Agriculture*, vol. 196, p. 106866, 2022.
- [16] L. C. M. Junior and J. A. C. Ulson, "Real time weed detection using computer vision and deep learning," in *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*, pp. 1131–1137, IEEE, 2021.
- [17] Y. Abbaspour-Gilandeh, A. Aghabara, M. Davari, and J. M. Maja, "Feasibility of using computer vision and artificial intelligence techniques in detection of some apple pests and diseases," *Applied Sciences*, vol. 12, no. 2, p. 906, 2022.
- [18] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—a review," *Information Processing in Agriculture*, vol. 7, no. 1, pp. 1–19, 2020.
- [19] E. Hamuda, B. Mc Ginley, M. Glavin, and E. Jones, "Automatic crop detection under field conditions using the hsv colour space and morphological operations," *Computers and electronics in agriculture*, vol. 133, pp. 97–107, 2017.
- [20] S. U. Amin, A. Hussain, B. Kim, and S. Seo, "Deep learning based active learning technique for data annotation and improve the overall performance of classification models," *Expert Systems with Applications*, vol. 228, p. 120391, 2023.
- [21] N. Pangakis and S. Wolken, "Keeping humans in the loop: Human-centered automated annotation with generative ai," *arXiv preprint arXiv:2409.09467*, 2024.
- [22] Y. Wang, D. Stevens, P. Shah, W. Jiang, M. Liu, X. Chen, R. Kuo, N. Li, B. Gong, D. Lee, *et al.*, "Model-in-the-loop (milo): Accelerating multimodal ai data annotation with llms," *arXiv preprint arXiv:2409.10702*, 2024.