Barley Head Detection Using UAV Imagery and YOLOv10

Ameer Tamoor Khan
atk@plen.ku.dk

University of Copenhagen

Saiful Azim
University of Copenhagen

Signe Marie Jensen
University of Copenhagen

Research Article

Keywords: Barley, YOLOv10, UAV Imagery, Precision Agriculture, Agricultural Computer Vision

Posted Date: May 21st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6709033/v1

Additional Declarations: The authors declare no competing interests.

Barley Head Detection Using UAV Imagery and YOLOv10

Ameer Tamoor Khan, Saiful Azim, Signe Marie Jensen

***Abstract*—Barley head detection is a crucial task for agricultural applications such as yield estimation and crop monitoring. Unlike wheat, automated barley head detection has not been extensively studied due to challenges posed by its complex head structures and the lack of annotated datasets. In this paper, we leverage YOLOv10, a state-of-the-art object detection framework, to detect barley heads from high-resolution images captured using UAVs. Our dataset, consisting of UAV-captured images and supplemented with the Global Wheat Head Dataset, provides a robust foundation for model training. The proposed approach achieves a mean Average Precision of 0.83 at Intersection of Union 0.5, setting a new benchmark for barley head detection. This work contributes to advancing automated crop monitoring systems in precision agriculture.**

***Index Terms*—Barley, YOLOv10, UAV Imagery, Precision Agriculture, Agricultural Computer Vision**

I. Introduction

Yield components in cereal crops, including barley, are critical factors determining the overall productivity of different varieties. These components are influenced by genetic traits, agronomic practices, and environmental conditions [1, 2].

The primary yield components include plant density (the number of plants per unit area), the number of panicles or heads per plant, the number of grains per panicle, and grain weight [3]. Each component significantly contributes to the final yield, and understanding their interactions is crucial for optimizing breeding programs and cereal production.

The number of plants per unit area significantly affects yield, as higher densities can lead to increased competition for resources but can also enhance total yield potential if managed correctly. Tiller formation, which refers to the development of additional stems from the base of the plant, is particularly important in cereals like wheat and barley, as it directly correlates with the potential number of heads and grains produced [4]. Furthermore, the number of grains per head and the grain weight are also critical yield components [5]. Finally, agronomic practices such as seed rate and fertilization, as well as environmental factors including temperature and precipitation, play crucial roles in determining the effect of these yield components [2].

Head number per plant in cereal varieties is a trait under strong genetic control [6, 7]. Estimating the number of heads is therefore an essential phenotypic measure that plays a pivotal role in identifying traits for breeding goals [2, 6].

Integrating advanced phenotyping technologies and genetic insights into head development can enhance the efficiency of breeding programs, ultimately leading to high-yielding and resilient cereal varieties. Visual counting methods are labor-intensive and prone to inaccuracies due to individual variability and spatial differences within plots or fields. Additionally, assessing the entire area manually is often impractical. Consequently, while grain yield is typically measured through harvested plots and post-harvest seed weight analysis, accurately assessing other yield components, such as head number per plant, remains a challenge due to the limitations of current measurement methods.

To improve efficiency, methods such as counting heads using ground and aerial imagery provide rapid and precise estimates of the number of heads per unit area. These estimates are essential for assessing traits related to head development and overall yield potential in cereal crops.

Despite its importance, automated detection of heads in barley remains an underexplored area in the field of computer vision. While much research has focused on detecting wheat heads using computer vision techniques, the same level of attention has not been given to barley due to the unique challenges posed by this crop.

Several challenges contribute to the difficulty of barley head detection:

- **Complex and Overlapping Head Structures**: Barley heads often overlap and form dense clusters in field environments, making them difficult to distinguish.
- **Lack of Public Datasets**: Unlike wheat, which benefits from datasets like the Global Wheat Head Dataset (GWHD) [8], there is no publicly available annotated dataset for barley head detection, limiting the ability of researchers to develop and train machine learning models.
- **Morphological**: Barley heads share some visual similarities with wheat, but their differences in size,

structure, and texture make it difficult to directly apply wheat detection models to barley.

In prior research, efforts have been made to detect wheat heads, primarily through models such as YOLO [9–11] and Faster R-CNN [12, 13], using large-scale datasets like the Global Wheat Head Dataset (GWHD). Wheat head detection has proven successful due to relatively well-formed structures and a more straightforward visual appearance in comparison to barley. Studies leveraging the GWHD have demonstrated the effectiveness of computer vision techniques in precision agriculture, but these approaches have not yet been extended to barley.

To address this challenge, we have collected and annotated a novel dataset consisting of barley heads captured in field conditions. Given the lack of publicly available data for barley, we have augmented our dataset with GWHD data to create a robust training set. While barley and wheat heads are morphologically distinct, their similarities in structure allow us to use a combination of both datasets to improve detection accuracy.

For this task, we utilize YOLOv10, a cutting-edge model in object detection. YOLOv10 introduces a dual approach to matching predicted objects with ground truth, capturing both broad and focused matches to improve detection accuracy. It also refines how predicted bounding boxes align with actual barley heads through a consistent matching process. Combining our newly collected barley dataset with GWHD, we have significantly enhanced the model's training, allowing it to better generalize to barley heads.

The rest of this paper is structured as follows: Section II presents the details of the **Dataset**, highlighting both the GWHD and our curated barley dataset. Section III introduces the **Detection Model**, focusing on the YOLOv10 architecture used for barley head detection. In Section IV, we describe the **Training and Testing** process, including data augmentation techniques and model parameters. Section V covers the **Analysis and Results**, where we evaluate the performance of the YOLOv10 model. The **Discussion** in Section VI explores the insights, limitations, and future directions of this work. Finally, Section VII concludes the paper with a summary of findings.

II. Dataset

For training our barley head detection model, we utilized two datasets: a newly collected Barley Head Dataset and the GWHD [8]. Due to the unique complexities of barley head structures and the scarcity of available data, we combined these two datasets to create a robust training set that would facilitate accurate detection of barley heads.

A. Barley Head Dataset

The Barley Head Dataset consists of images acquired using an Unmanned Aerial Vehicle (UAV) equipped with an RGB camera, specifically the DJI Mavic 3 Multispectral RTK. The images were collected from a field trial at Højebakkegaard research farm, owned by The University of Copenhagen. This research farm, located in Høje Taastrup, sits on sandy loam soil (55°40'09.4"N, 12°18'18.8"E). The trial followed a randomized complete block design, and image acquisition was carried out on June 14, 2024.

The UAV was flown at an altitude of 8m with an 80% × 80% overlap in a north–south flight pattern. The dataset includes high-resolution images (5280 × 3956 pixels) with a ground sampling distance (GSD) of 0.21 cm/pixel. The UAV camera settings were configured as follows:

- Camera Focus: Auto Focus Continuous (AFC)
- Capture Mode: Auto (A)
- Fixed Aperture: 6.3
- Fixed Exposure Time: 1/2000 s
- Fixed ISO: 100
- Fixed EV: -0.3
- Image Format: JPG

The dataset was annotated using Hasty.ai [14], where trained personnel manually labeled approximately 5,000 barley heads across 80 images. The sample images (Fig. 1(a)) illustrate the density of barley heads and the challenges in distinguishing them from one another. The bounding box area distribution (Fig. 2b) highlights the relatively smaller size of barley heads compared to wheat heads.

B. Global Wheat Head Dataset

The Global Wheat Head Dataset (GWHD) [8] is a publicly available dataset widely used for wheat head detection tasks. It consists of 3,421 images with a resolution of 1024×1024 pixels, capturing approximately 140,000 wheat heads under various field conditions. The dataset includes images acquired from ground-based phenotyping platforms at heights ranging from 1.8 meters to 3 meters, providing diverse perspectives for wheat head detection.

While wheat and barley heads exhibit morphological differences, they share similar elongated shapes. Incorporating the GWHD dataset enhances model generalization by introducing variations in crop density, head visibility, and background complexity. Studies utilizing GWHD (Fig. 1(b)) have demonstrated the effectiveness of computer vision in precision agriculture, yet such approaches have not yet been extended to barley. The bounding box area distribution (Fig. 2a) shows that wheat heads tend to occupy larger bounding boxes compared to barley heads.

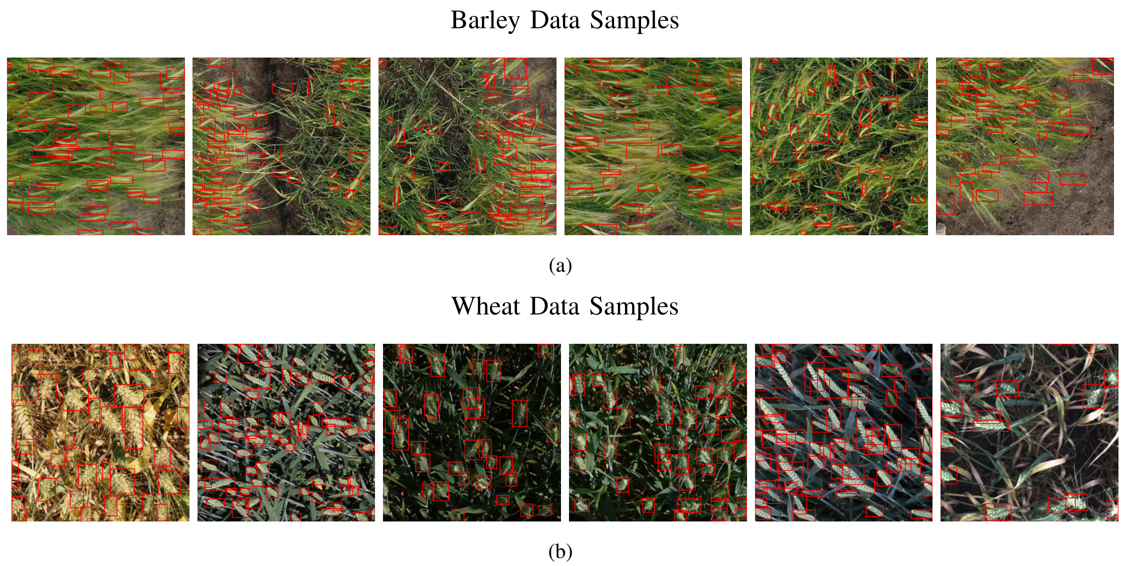


Fig. 1: (a) Barley dataset, with annotated bounding boxes. (b) Sample from the global wheat dataset with annotated bounding boxes.

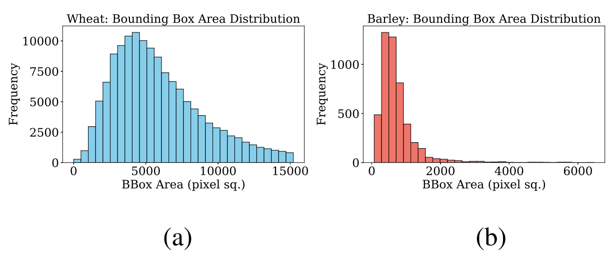


Fig. 2: Bounding box area distribution for wheat and barley datasets. (a) Distribution of bounding box areas (in pixel square units) for the wheat dataset. (b) Distribution of bounding box areas for the barley dataset.

III. Detection Model

A. *Yolov10 Architecture*

YOLOv10 [15] brings substantial improvements to object detection tasks. This section details the architecture of YOLOv10 (Fig. 3) and explains how it enhances the performance and efficiency of object detection, making it a superior choice for barley head detection over previous YOLO versions.

B. *Backbone: CSPNet*

The backbone of YOLOv10 is based on an enhanced version of CSPNet (Cross-Stage Partial Network [15]). CSPNet is designed to efficiently extract key features from the input image while balancing computational resources and accuracy. This is particularly important for plant phenotyping using high-resolution image processing. The backbone of YOLOv10 is based on an enhanced version of CSPNet (Cross-Stage Partial Network [15]). CSPNet has an ability to handle large feature maps without introducing significant computational overhead making it well-suited for tasks that require detailed spatial information while maintaining real-time performance.

C. *Neck: Path Aggregation Network (PAN)*

The Path Aggregation Network (PAN) [15] serves as the neck of YOLOv10, which is responsible for multi-scale feature fusion. PAN allows the model to combine features from different layers and scales of the network, improving its ability to detect objects of varying sizes. This multi-scale feature aggregation is particularly useful for detecting barley heads, which can vary significantly in size depending on the growth stage and environmental conditions. PAN ensures that the model can detect both small and large objects with high accuracy.

D. *Dual-Label Assignment*

A major innovation in YOLOv10 is the introduction of dual-label assignment [15], which combines both one-to-many and one-to-one strategies. During training, the one-to-many strategy assigns multiple potential bounding boxes to each object, improving the recall of the model by capturing more potential detections. In contrast, the one-to-one strategy is used during inference to ensure that only the most accurate bounding box is selected for each object. This dual approach improves both the precision and recall of the model, resulting in better overall detection performance.

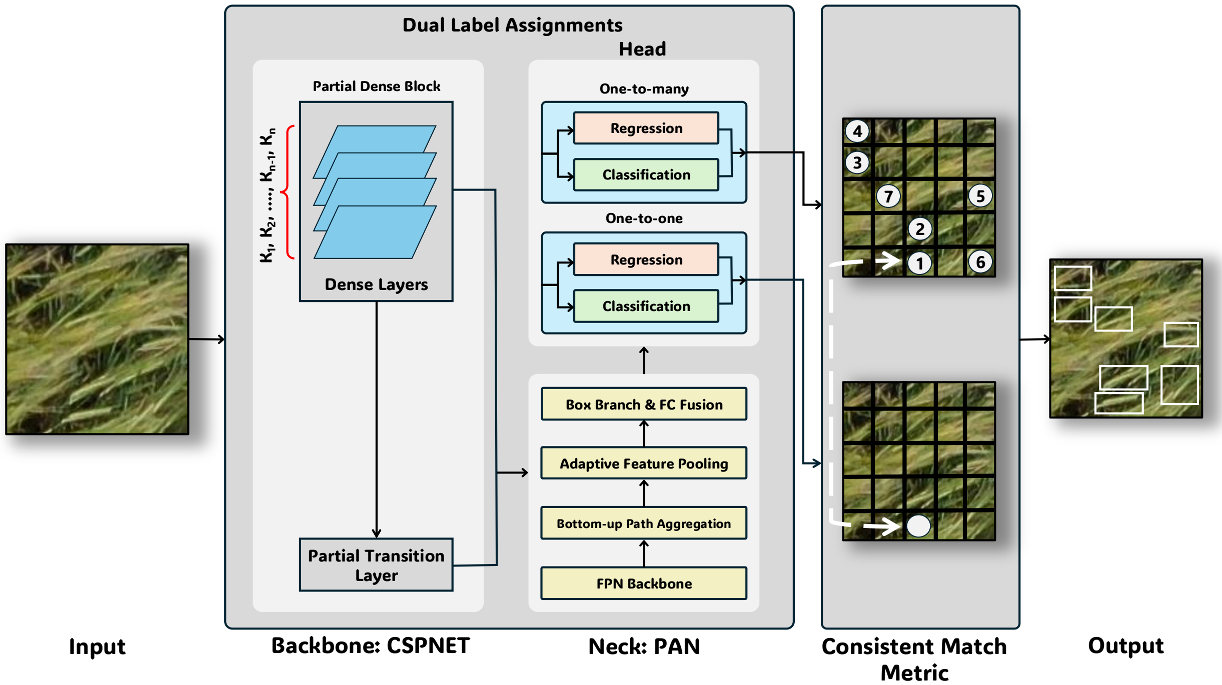


Fig. 3: The architecture of YOLOv10, an object detection model with three main components: Backbone (CSPNet) for feature extraction, Neck (PAN) for refining features, and a Head with dual-label assignments. The head uses one-to-many and one-to-one branches for regression and classification, enhancing bounding box predictions. A consistent match metric optimizes the alignment of detected objects for improved accuracy and efficiency.

Additionally, this method reduces the need for Non-Maximum Suppression (NMS), which is typically used to remove redundant bounding boxes. By minimizing reliance on NMS, YOLOv10 reduces latency during inference, making it faster.

E. Consistent Matching Metric

To further enhance detection accuracy, YOLOv10 incorporates a consistent matching metric that aligns the predictions made by both the one-to-many and one-to-one strategies [15]. This metric refines the match between predicted bounding boxes and the actual ground truth locations, ensuring that the model's detections are precise and reliable. The consistent matching metric plays a critical role in improving the model's performance during both training and inference, leading to higher detection accuracy.

F. Lightweight Classification Head

YOLOv10 uses a lightweight classification head [15], which reduces the computational cost of the model without sacrificing accuracy. By incorporating depth-wise separable convolutions, the classification head is able to process the final feature maps more efficiently, reducing the overall computational burden. This makes the model faster, particularly in resource-constrained environments, without compromising its ability to accurately classify objects.

G. Spatial-Channel Decoupled Downsampling

YOLOv10 introduces spatial-channel decoupled downsampling [15], a technique that separates the spatial and channel operations during downsampling. This method allows the model to retain more information from the input image while still reducing the computational load. By preserving important spatial and channel information, YOLOv10 can improve detection performance, especially when working with detailed objects like barley heads in complex agricultural environments.

H. Large-Kernel Convolutions and Partial Self-Attention (PSA)

To improve the detection of smaller objects, such as individual barley heads, YOLOv10 incorporates large-kernel convolutions in the deeper stages of the network [15]. These large kernels help the model capture fine-grained details in the input image, improving its ability to detect small objects that might be missed by smaller kernels.

TABLE I: Hardware, Model, Data, And Data Augmentation Specifications

Hardware Specifications	
CPU	Google Colab
GPU	NVIDIA T4
RAM	40 GB
Model Specifications	
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	
Wheat Train	3,421 Images
Barley Train	75 Images
Barley Test	5 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)

In addition, YOLOv10 integrates a Partial Self-Attention (PSA) module [15], which enhances the model's ability to learn global representations of the input image. PSA applies self-attention to a subset of the feature map, allowing the model to focus on important regions of the image without incurring significant computational overhead. This selective application of self-attention helps YOLOv10 improve its detection accuracy while maintaining efficiency.

IV. Training and Testing

Our model for barley head detection was trained using a combination of the GWHD and newly annotated barley dataset. This section outlines the hardware setup, model parameters, data specifics, and augmentation techniques applied during the training process.

The model was trained using Google Colab's high-memory environment, which provided an NVIDIA T4 GPU (16GB VRAM) and 40GB RAM, essential for handling the large datasets and intensive training operations. With this setup, we were able to efficiently train the YOLOv10 model using a batch size of 8. Data loading was parallelized with 8 workers to ensure fast processing.

We employed the YOLOv10 model initialized with pre-trained weights. The model was fine-tuned for our barley head detection task over 50 epochs (TABLE I). We used the Adam optimizer, which is known for its ability to handle sparse gradients and noisy datasets. The initial learning rate was set to 0.0001, with a cosine learning rate decay to adjust the learning rate dynamically. Early stopping was implemented with a patience parameter of 100 epochs to avoid overfitting.

The training dataset included 3,421 images from the GWHD and 75 annotated barley images. To test the model's performance, we used a small set of five barley images. The images were all resized to 640×640 pixels for consistency during both training and testing. The wheat dataset includes a significantly larger number of images (3,421) and annotated heads (approximately 140,000) compared to the barley dataset, which has only 80 images and 5,000 annotated heads in total (Fig. 4). This highlights the importance of leveraging GWHD for improved generalization.

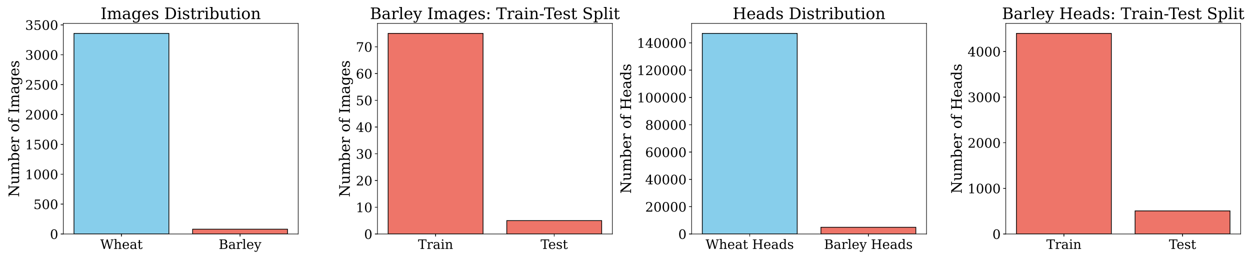
To increase the robustness of the model and to simulate a variety of field conditions, we applied a range of data augmentation techniques using the Albumentations library [16]. These included random blurring, grayscaling, and contrast adjustments (TABLE I). Such augmentations improved the model's ability to detect barley heads in diverse environmental conditions, enhancing its generalization capacity.

V. Results and Analysis

This section evaluates the performance of our YOLOv10-based barley head detection model through training and validation trends, precision-recall analysis, and confusion matrix insights.

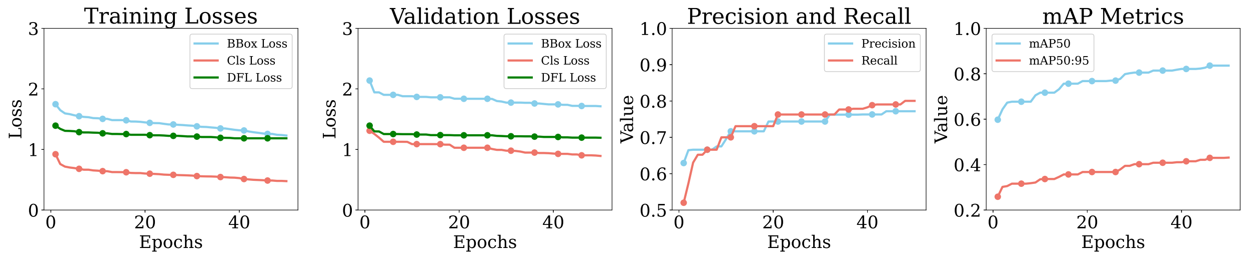
The steady reduction in both training and validation losses highlights the model's effective learning and generalization capabilities (Fig. 5). Precision demonstrates a consistent upward trend, indicating the model's ability to reduce false positives as training progresses. Similarly, recall improves, reflecting the model's increased capacity to identify barley heads accurately. The mAP metrics (mAP50 and mAP50:95) illustrate robust detection performance, with mAP50 achieving consistently higher values, making it the preferred metric for practical applications.

The training setup and use of augmentations enabled the YOLOv10 model to generalize effectively from the wheat dataset to the barley dataset, despite the limited number of barley images. The combination of robust augmentation techniques and advanced model architecture contributed significantly to the performance of the model, as reflected in the results.



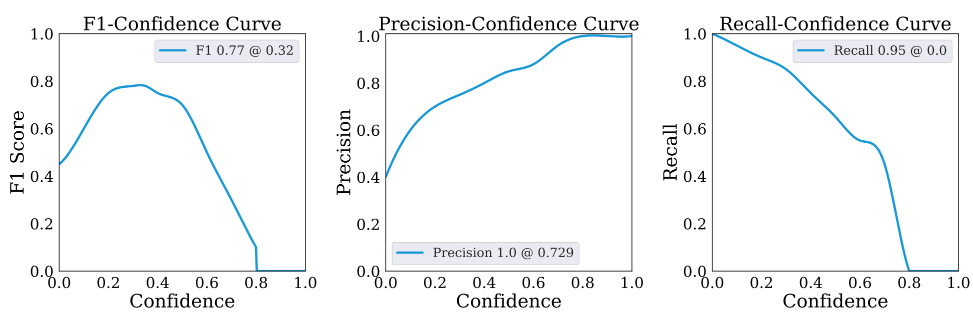
(a) (b) (c) (d)

Fig. 4: Distribution of images and crop heads for wheat and barley datasets. (a) Total number of images for wheat and barley. (b) Train-test split for the barley images. (c) The number of heads detected in wheat and barley images. (d) Train-test split for barley heads.



(a) (b) (c) (d)

Fig. 5: Comparison of training and validation losses, along with precision, recall, and mAP metrics over 50 epochs. (a) The training losses for bounding box regression (BBox), classification (Cls), and distribution focal loss (DFL). (b) The validation losses. (c) The precision and recall metrics. (d) mAP (mean average precision) metrics with mAP50 and mAP50:95 thresholds.

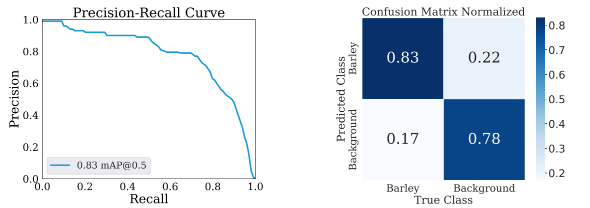


(a) (b) (c)

Fig. 6: Relationship between model confidence and performance metrics such as F1 score, precision, and recall. (a) The F1-Confidence Curve, which indicates the model's F1 score across varying confidence levels. (b) Precision-Confidence Curve. (c) The Recall-Confidence Curve.

TABLE II: Model Performance For Barley Head Detection

Model	**Precision**	**Recall**	**mAP50**	**mAP50:95**	**Fitness**	**Preprocess (ms)**	**Inference (ms)**	**Postprocess (ms)**
YOLOv10	0.7495	0.7806	0.8341	0.4311	0.4714	0.3259	70.0264	2.9637



(a) Precision-Recall Curve (b) Normalized Confusion Matrix

Fig. 7: (a) The Precision-Recall Curve gives model performance across different confidence thresholds, balancing precision and recall. (b) The Normalized Confusion Matrix is the model's accuracy in distinguishing between Barley and Background, highlighting correct and incorrect predictions.

The F1-Confidence curve suggests that the model's optimal performance is achieved at a confidence threshold of 0.32, where precision and recall are balanced (Fig. 6). Beyond this threshold, the model prioritizes precision, reducing recall, which may be desirable in applications where false positives have higher consequences.

With a precision of 74.95%, recall of 78.06%, and mAP50 of 83.41%, the YOLOv10 model for barley detection demonstrates robust detection capabilities (TABLE II). Furthermore, the fitness score of 0.4714 and inference time of 70.03 ms indicate that the model is efficient for practical applications, balancing accuracy and computational performance.

The Precision-Recall curve showcases the inherent trade-off between these metrics (Fig. 7(a)). As recall increases, precision declines, which is expected in object detection tasks.

The confusion matrix (Fig. 7(b)) reveals the model's ability to differentiate barley heads from the background. While the model performs well overall, some misclassifications persist, with a small percentage of barley instances misclassified as background and vice versa. This indicates opportunities for further fine-tuning, particularly in reducing false positives. The model is also applied to test images (Fig. 8).

VI. Discussion

The development and application of a robust model for automated barley head detection address a significant gap in agricultural research, where barley, despite its economic and agronomic importance, has remained underexplored in the domain of computer vision. Our work has demonstrated the feasibility of using state-of-the-art object detection models, such as YOLOv10, in detecting barley heads, a task that presents numerous challenges not observed to the same degree in similar crops like wheat.

YOLOv10's dual-label assignment strategy has proven effective in handling the complex and overlapping structures that are characteristic of barley heads. Traditional detection models often struggle with such intricacies [17–20], where overlapping heads result in missed detections or incorrect classifications. YOLOv10's ability to refine bounding box alignment and match predicted objects to ground truth data more effectively mitigates this issue, ensuring better precision and recall in the presence of densely packed barley heads. This demonstrates the model's capability to adapt to the unique visual challenges posed by barley crops, which are often more difficult to delineate than their wheat counterparts.

Despite these advancements, there are significant challenges associated with counting barley heads using imaging. The model can only detect what can be seen from above, meaning heads in the underlying layers of the canopy, as is often the case in barley, are missed. Additionally, the model relies entirely on the annotations provided during training. Thus, any difficulties in distinguishing or identifying barley heads within the dataset will be transferred to the model, impacting its detection capability. These challenges underscore the importance of enhancing annotation quality and incorporating new methods for capturing hidden heads to improve the model's overall accuracy.

Comparison with the literature on wheat head counting reveals that similar object detection frameworks, such as earlier versions of YOLO [9–11] and Faster R-CNN [12, 13], have achieved strong results due to the relatively simpler head structures in wheat [21, 22]. The well-organized spatial distribution of wheat heads makes detection and counting more straightforward compared to barley, which is characterized by dense clusters and significant overlapping.

However, certain limitations in our approach remain. The reliance on the GWHD, while beneficial, introduces challenges due to the morphological differences between barley and wheat heads. While combining these datasets improves generalization, targeted dataset expansion with diverse barley-specific data would likely enhance the model's performance further. Incorporating spectral imaging or other sensor modalities could also provide additional context for distinguishing barley heads from similar background elements [23–25].

Looking ahead, future research should focus on expanding the barley head dataset with images collected across different environments, seasons, and growth stages to improve robustness. Additionally, integrating multi-spectral or hyperspectral imaging could aid in better distinguishing barley heads from the background by leveraging unique spectral signatures.

(a)

(b)

Fig. 8: (a) Barley test samples with ground truth bounding boxes, reflecting the manually annotated locations of barley heads. (b) YOLOv10 model predictions on the same test samples, demonstrating the model's detection accuracy and alignment with the ground truth annotations.

Generative models, such as Generative Adversarial Networks (GANs) [26] and diffusion models [27], offer a promising solution to address the lack of diverse barley-specific datasets. These models can synthesize realistic, high-quality images of barley fields under various environmental conditions, seasons, and lighting scenarios, augmenting the training data significantly. By incorporating synthetic images generated by GANs or diffusion models, the model's robustness and generalization capabilities can be enhanced. Additionally, generative models can simulate challenging scenarios, such as overlapping barley heads or low-visibility conditions, to improve the detection model's ability to handle complex real-world environments. The use of such models has already shown success in agricultural applications, such as creating synthetic crop images to augment datasets [28].

VII. Conclusion

In this paper, we addressed the underexplored challenge of automated barley head detection using the YOLOv10 model, bridging the gap between wheat and barley head detection by leveraging the Global Wheat Head Dataset (GWHD) in combination with a newly curated barley dataset. Our experiments demonstrate that YOLOv10 achieves strong performance with a mean Average Precision (mAP) of 0.83 at an IoU threshold of 0.5, effectively managing the complexities of overlapping and densely clustered barley heads. The model's performance is attributed to the integration of advanced data augmentation techniques and a robust training dataset. However, limitations persist, particularly the relatively small size of the barley-specific dataset, which may restrict the model's ability to generalize across diverse field conditions. Future work should aim to expand the barley dataset and further improve detection accuracy by integrating additional sensor modalities—such as spectral imaging—and fine-tuning the model architecture for varied agricultural settings.

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

Data Availability Statement: The dataset used in this study is publicly available, and the link to the dataset is attached in the manuscript.

Funding Statement: No funding details available.

References

[1] V. Sadras and D. Calderini, *Crop physiology: applications for genetic improvement and agronomy*. Academic Press, 2009.

[2] N. M. Tshikunde, J. Mashilo, H. Shimelis, and A. Odindo, "Agronomic and physiological traits, and associated quantitative trait loci (qtl) affecting yield response in wheat (triticum aestivum l.): a review," *Frontiers in plant science*, vol. 10, p. 1428, 2019.

[3] V. O. Sadras and G. A. Slafer, "Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities," *Field crops research*, vol. 127, pp. 215–224, 2012.

[4] L. Ma, T. Li, C. Hao, Y. Wang, X. Chen, and X. Zhang, "Ta gs 5-3a, a grain size gene selected during wheat improvement for larger kernel and

yield," *Plant Biotechnology Journal*, vol. 14, no. 5, pp. 1269–1280, 2016.

[5] W. Grzebisz, W. Szczepaniak, J. Potarzycki, and M. Biber, "Prediction of grain yield and gluten content in winter bread wheat based on nutrient content in plant parts during the critical cereal window," *Agronomy*, vol. 13, no. 10, p. 2649, 2023.

[6] C. Moeller, J. B. Evers, and G. Rebetzke, "Canopy architectural and physiological characterization of near-isogenic wheat lines differing in the t iller in hibition gene tin," *Frontiers in plant science*, vol. 5, p. 617, 2014.

[7] Q. Xie, S. Mayes, and D. L. Sparkes, "Optimizing tiller production and survival for grain yield improvement in a bread wheat× spelt mapping population," *Annals of Botany*, vol. 117, no. 1, pp. 51–66, 2016.

[8] E. David, S. Madec, P. Sadeghi-Tehran, H. Aasen, B. Zheng, S. Liu, N. Kirchgessner, G. Ishikawa, K. Nagasawa, M. A. Badhon, *et al.*, "Global wheat head detection (gwhd) dataset: A large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods," *Plant Phenomics*, 2020.

[9] S. Qing, Z. Qiu, W. Wang, F. Wang, X. Jin, J. Ji, L. Zhao, and Y. Shi, "Improved yolo-fastestv2 wheat spike detection model based on a multi-stage attention mechanism with a lightfpn detection head," *Frontiers in Plant Science*, vol. 15, p. 1411510, 2024.

[10] J. Ye, Z. Yu, Y. Wang, D. Lu, and H. Zhou, "Wheatlfanet: in-field detection and counting of wheat heads with high-real-time global regression network," *Plant Methods*, vol. 19, no. 1, p. 103, 2023.

[11] S. Guan, Y. Lin, G. Lin, P. Su, S. Huang, X. Meng, P. Liu, and J. Yan, "Real-time detection and counting of wheat spikes based on improved yolov10," *Agronomy*, vol. 14, no. 9, p. 1936, 2024.

[12] J. Li, C. Li, S. Fei, C. Ma, W. Chen, F. Ding, Y. Wang, Y. Li, J. Shi, and Z. Xiao, "Wheat ear recognition based on retinanet and transfer learning," *Sensors*, vol. 21, no. 14, p. 4845, 2021.

[13] C. Wen, J. Wu, H. Chen, H. Su, X. Chen, Z. Li, and C. Yang, "Wheat spike detection and counting in the field based on spikeretinanet," *Frontiers in Plant Science*, vol. 13, p. 821717, 2022.

[14] Hasty.ai, "Hasty: The fastest tool for image annotation and object detection." https://app.hasty.ai/sign-in, 2024. Accessed: 2024-09-24.

[15] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458*, 2024.

[16] Ultralytics, "Basetransform.apply_semantic," 2025. Accessed: Jan. 28, 2025.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision*, pp. 21–37, 2016.

[19] S. Madec, X. Jin, H. Lu, B. De Solan, S. Liu, F. Duyme, B. Prelot, and A. Comar, "Ear density estimation from high-resolution rgb imagery using deep learning technique," *Plant Phenomics*, vol. 2019, 2019.

[20] E. David, S. Madec, P. Sadeghi-Tehran, H. Aasen, B. Zheng, S. Liu, N. Kirchgessner, G. Ishikawa, K. Nagasawa, M. A. H. Badhon, *et al.*, "Global wheat head detection (gwhd) dataset: a large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods," *Plant Phenomics*, vol. 2021, 2021.

[21] 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.

[22] J. Ma, Y. Li, H. Liu, Y. Wu, and L. Zhang, "Towards improved accuracy of uav-based wheat ears counting: A transfer learning method of the ground-based fully convolutional network," *Expert Systems with Applications*, vol. 191, p. 116226, 2022.

[23] A. M. Mutka, N. Fahlgren, K. Kuehn, K. Chen, A. L. Heuberger, S. Zaidi, R. VanBuren, and C. Buell, "Hyperspectral imaging for early detection of barley powdery mildew," *Frontiers in Plant Science*, vol. 9, p. 1962, 2018.

[24] N. Gorretta, E. Vahtmäe, Z. G. Cerovic, J. Fabre, M. Faravel, E. Anthonis, and E. Belin, "Seed quality assessment using hyperspectral imaging and chemometrics: A case study on barley seeds," *Plant Methods*, vol. 15, no. 1, pp. 1–11, 2019.

[25] W. Sun, J. Guo, X. Li, Y. Liu, and X. Wang, "Classification of barley seed varieties using hyperspectral imaging and machine learning algorithms," *Journal of Food Process Engineering*, vol. 44, no. 4, p. e13769, 2021.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[27] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion

probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[28] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2019.