



OPEN Transfer learning models for wheat ear detection on multi-source dataset

Željana Grbović^{1✉}, Marko Panić¹, Dimitrije Stefanović¹, Sanja Brdar¹, Vladimir Crnojević¹ & Tatjana Lončar-Turukalo²

Monitoring wheat growth, as one of the most important food grain sources for human nutrition, and forecasting yields are done through different phenological phases. Reliable estimates on yields play a crucial role in securing sufficient food supplies for the world's growing population. Currently, farmers estimate a wheat yield during the later stages of growth and are often biased in this process. Plant breeding scientists use a more accurate approach that collects data on the number of wheat ears manually counted at various locations throughout the field. A sufficiently precise count of wheat ears is one of the most important parameters for reliable early-stage prediction of wheat yield. To support the development of an affordable and trustworthy automated wheat ear detection approach, this work introduces a novel high-quality RGB smartphone image dataset, BioS-Wheat, comprising 5,696 annotated images across four wheat varieties. Additionally, it evaluates six deep learning models for wheat ear detection. Among the F-RCNN-based models, RetinaNet, YOLOv8, and a Vision Transformer-based detector, RT-DETR, achieved the highest mean Average Precision (mAP@50) of 91%, with significantly higher computational complexity. BioS-Wheat complements Global Wheat Head Detection datasets, introducing a meaningful shift in data complexity with high sowing density and minimal row spacing, emphasizing the impact of agronomic diversity on model performance by an increased object occlusion and dense spatial arrangements. Enriched and agronomically diverse datasets support model robustness at different varieties, growth stages, and locations. This work offers a good baseline for establishing the procedure for image crowdsourcing, further dataset expansions, and model improvements.

Keywords Bios-Wheat dataset, Wheat ear detection, Mask-RCNN based models, Yolov8, Vision transformer based model

The exponential growth of the human population invokes the necessity for more precise crop monitoring and sets higher crop yields as one of the most demanding challenges. As part of the global plan, 17 goals have been set by 2030 to end hunger, achieve food security, and improve nutrition, according to the Food and Agriculture Organization of the United Nations¹. The key to accomplishing the entire set of goals is the investment in agriculture—crops, livestock, forestry, fisheries, and aquaculture². Monitoring wheat yield, as the primary grain source for human consumption, is a treasured part of the global population nutrition plan³.

Although there is a worldwide trend that more and more agricultural fields are devoted to wheat cultivation compared to other crops, as evidenced by the wheat yield of more than 750 million tons per year, further expansion of arable land is limited⁴. Therefore, to enhance crop production in the future, producers are willing to incorporate new precision agriculture technologies during the decision-making process to achieve high-quality and robust crop types with higher yields.

There are only a few wheat ear counting datasets, in most cases collected by single entities, which directly introduces bias and limitations, such as certain genotypes and acquisition environments. Moreover, the labeling process is tedious and usually differs among institutions, which requires additional adjustment for further usage on shared and expanded open-source and public datasets⁵. Additionally, those datasets often represent moderate-density fields and lack sufficient examples of severe occlusion, limiting their applicability in real-world high-density scenarios. The BioS-Wheat dataset addresses these gaps by providing a standardized and low-cost RGB dataset that captures dense wheat canopies with extensive wheat ear overlap, multiple genotypes, and varying fertilization levels. This design enables comprehensive evaluation of model robustness under realistic field

¹BioSense Institute, University of Novi Sad, Novi Sad, Serbia. ²Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia. ✉email: zeljanagrbovic@biosense.rs

conditions, bridging the gap between controlled imaging studies and operational crop monitoring environments. Field crop producers and breeding programs within breeding research institutions can significantly benefit from the information obtained by rapid yield prediction based on automatically counted wheat ears^{6,7}.

In this work, we facilitate an early wheat yield prediction approach by proposing a methodology for an automated, low-cost, and easy-to-adapt user-affordable wheat ear detection based on acquired smartphone RGB wheat images in outdoor conditions.

The main contributions of this work are a new dataset with 4 wheat varieties from a location in Europe, Serbia, and a transfer-learning methodology based on deep learning for wheat ear detection in collected images.

The paper is organized as follows: The next two sections cover related work on traditional and deep learning-based computer vision methods, followed by materials and methods, including an overview of open-source datasets, our dataset creation and harmonization process, image quality assessment, image acquisition recommendations, and the proposed wheat ear detection method with evaluation metrics. The Results section presents the findings of the image quality assessment and compares the performance of the proposed deep learning models. The discussion section examines the outcomes, advantages, and limitations of the proposed approach. The final section provides concluding remarks and outlines potential directions for future research.

Related work

According to⁸, the number of wheat ears and the weight of one thousand grains within the square meter are the most significant parameters that contribute to the wheat yield estimation. Specifically, the number of wheat ears per unit area is one of the most important agronomic yield traits, since it has the highest correlation with the grain yield⁹. The conventional methods usually infer the wheat yield¹⁰ based on comparing crop performances across years or from manually counted wheat ears per region of interest (ROI) and therefore are insufficiently precise and prone to subjective judgment^{11,12}. Manual wheat ear counting involves constraining the ROI using a square meter (or less) wooden frame placed over the wheat^{9,11}. For fields of one hectare in size or less, the total number of ears is obtained by extrapolating the wheat number obtained per ROI. If the field size is significantly larger than one hectare, estimating the total number of ears requires repeating ROI counting procedures at multiple locations in the field⁹. The models for wheat ear detection trained based on manually counted data usually lack generalization quality, as they do not achieve matching accuracy and precision on new datasets and acquisitions⁵. The approach proposed in this work facilitates an automated wheat ear detection and count, drastically reducing the manual effort that is time-consuming, labor-intensive, and error-prone, aggravating widespread field implementation^{6,13}.

Following the rapid growth and improvements in sensor technologies and advanced image processing, computer vision, and artificial intelligence algorithms, data-driven approaches are gradually being adopted and implemented in agriculture. Some application areas target various problems in agriculture^{14–16}, such as weed detection^{17,18}, plant identification, plant phenotyping^{19,20} and disease detection^{21–26}. For example, in²⁷, the automatic plant species identification has been described; in²⁸, the extraction of *Arabidopsis Thaliana* phenotyping features is explored; and in^{29,30}, the detection of wheat ears in controlled environments is described in^{19,31,32}. However, wheat ear identification is a challenging computer vision problem due to variability in observational conditions and wheat ear orientation, appearance variability in different phenological phases, and inherent genetic differences. Moreover, image acquisition might be hampered by possible blurring or movement or wind, which introduces overlapping wheat ears in dense population areas, thus making the automatic wheat ear counting problem even more challenging⁵.

Traditional computer vision methods

Starting in 2008, the first idea to facilitate automatic ear counting from the image taken in the field was proposed using color-texture image analysis in³³, while in³¹ wheat ear phenotypic traits such as length, width, surface area, and perimeter are extracted from images based on morphology. Authors in³² presented a three-layer neural network classifier³⁴ to extract RGB color features such as ear length. In⁹, the Laplacian frequency filter³⁵ and the median filter³⁶ are used for suppressing the presence of noise around wheat ears by removing high-frequency components in the image. The final segmentation of ears is done using the *Find Maxima* detection algorithm³⁷. Filtration using Gabor filter banks³⁸ followed by principal component analysis³⁹ and the K-means clustering algorithm⁴⁰ is proposed in⁴¹ for segmentation of regions belonging to the wheat ears. It reached an average accuracy of 90.7% compared to human manual counting and was tested with a real-world dataset containing images of wheat from flowering to ripening stages. In⁴², the number of wheat ears from images is obtained by a Simple Linear Iterative Clustering Method (SLIC)⁴³, transforming the unit of the wheat image from pixels to superpixel blocks. This study ranged from 80.8% to 94.2% depending on different nitrogen fertilizers. The dataset is a part of the supplementary material. A twin-support-vector-machine segmentation (TWSVM-Seg) model⁴⁴ is utilized in⁸ for the segmentation of wheat ears on the pixel level. The results of the proposed approach showed a strong correlation in the range of 0.79 to 0.82 with the data obtained by manual counting. The majority of data collected in these studies is adjusted for specific purposes and objectives, making it harder to use them for more general applications.

Deep learning-based computer vision methods

Relying on the increasing performance of graphics processing units (GPUs) and the wide availability of data-collecting sensors facilitating the emergence of large datasets^{45,46}, deep learning has become the cutting edge of computer vision for object recognition⁴⁷, instance segmentation⁴⁸, semantic segmentation⁴⁹ and image regression^{50,51}, providing high performance, thus outperforming other image processing techniques and classical machine learning approaches^{14,52}. The application of convolutional neural networks (CNNs) plays a crucial role in image-based phenotyping²⁸. As in other similar applications, in the context of wheat ear counting, a common

approach is to train CNN models using data from one part of a field and evaluate their performance on data from another part^{52–56}. The relative wheat counting error was estimated at 10% in⁵⁷ for such a method if the image resolution is controlled. In⁵⁸, the authors developed an encoder/decoder-based CNN model for the semantic segmentation of wheat ears that outperforms traditional manual methods of computer vision algorithms. In⁵⁹, the authors developed a wheat ear detection and probabilistic tracking model for wheat growth in the field. Considering these problem-specific challenges that occur in the images, existing developed and pre-trained models for the semantic segmentation of wheat ears so far have not achieved encouraging results. The problem of wheat ear occlusion is difficult for traditional image processing methods with high-level semantic features, while the existing deep learning-based counting methods have not solved the occlusion efficiently. In parallel with neural networks designed for semantic segmentation, many instance segmentation architectures are developed: Mask R-CNN⁶⁰, Mask Lab⁶¹, PANet⁶², Hybrid task cascade⁶³, GCNET⁶⁴, Yolact⁶⁵, Mask2Former⁶⁶ and object detection architectures: RCNN⁶⁷, Fast RCNN⁶⁸, MultiPath Network⁶⁹, Faster RCNN (F-RCNN)⁴⁷, Non-local Neural Networks⁷⁰, Tensor Mask⁷¹, Mask Scoring RCNN⁷², the Yolo versions, etc. Recently, several authors proposed deep learning image-based plant phenotyping methods^{19,30,73}. In⁶, the usage of CNNs for solving the wheat ear instance segmentation task in controlled indoor conditions is presented. The latest developed approaches for wheat ear detection rely on CNN-based object detection baselines: F-RCNN^{57,74,75}, YOLO⁷⁶, YOLOv3⁷⁷, YOLOv4⁷⁸, YOLOv5 model^{79,80}, YOLOv7⁸¹, YOLOv8⁸² and RIA-SpikeNet from⁸³. These models are utilized on a publicly available dataset generated through a challenge named the Global Wheat Head Challenge⁵, from seven different institutions. The training data consists of more than 3,000 images from France, the UK, Switzerland, and Canada, and the test data includes about 1,000 images from Australia, Japan, and China. A brief review of the state-of-the-art deep learning approaches for wheat ear detection has been presented in⁷. In⁸⁴ authors developed WheatNet, which uses a MobileNetV2⁸⁵ as a backbone and two additional sub-networks developed for each module: for wheat detection and density-based counting. In⁸⁶, a YOLOv5⁸⁷ is applied and used as a baseline for wheat ear detection, achieving performance with an average F1 score of 0.93. The performance of the cutting-edge DeepMAC segmentation model⁸⁸ is also reported in⁸⁶ with the average F1 score of ear bounding box detection of 0.93 and segmentation F1 score of 0.86. The method proposed in⁸⁹ is based on RetinaNet⁹⁰ and achieves the Intersection over Union (IoU) with the applied threshold of 0.5 (mAP@50) of 0.92. In⁹¹, the approach based on an improved EfficientDet-D0 object detection model⁹² has been proposed using transfer learning methods, i.e., the pre-training of the model backbone network to extract the high-level semantic features of wheat ears. In the same work, the random-cutout image augmentation method is proposed based on the removal of some of the identified bounding boxes guided by the wheat ear number and its size in the images to simulate occlusion in real-environment wheat images. In the final architecture, after the backbone suppresses irrelevant information, an additional convolutional block attention module (CBAM) is implemented, reaching an accuracy of 94%, which is about 2% higher than the original model. The CTWheatNet model from⁹³ uses diverse attention-based modules and achieves an mAP@50 of 93%, while one of the novel transformer-based approaches, WH-DETR, reaches 95.7%. The RIA-SpikeNet from⁸³ is a model that is designed to detect and count wheat spikes in complex field environments. The obtained results showed that it outperforms the state-of-the-art YOLOv8 detection model, achieving a competitive 81.54% mAP and 90.29% R². All referenced studies primarily use the Global Wheat Head Detection (GWHD) Datasets⁵ for wheat ear detection.

In this paper, we build upon the foundational GWHD framework, using it as the basis for creating our dataset. Our contribution involves expanding this dataset by increasing the number of images, varieties, and locations, while also improving the quality of the data.

We contribute an additional dataset, BioS-Wheat, containing wheat ears with four different genotypes from one additional European location (see Fig. 1). The BioS-Wheat dataset extends the existing GWHD dataset in terms of genotypes, type of sensor, and location. The aim is to address the specific problem of wheat ear counting, contributing a new dataset that can support future research in solving this issue.

The deep learning architectures used in this work are modifications of Mask RCNN with three different backbones and RetinaNet, which are trained both on the reference GWHD dataset and on the extended dataset. Additionally, YOLOv8 and RT-DETR architectures are used to demonstrate the contribution of the data enhancement to the performance of the learned models.

Materials and methods

Although deep learning models are widely used in many applications, it is still challenging to achieve their robustness in plant phenotyping studies because there aren't many real-world datasets produced in outdoor settings⁵. Furthermore, since these deep learning-based techniques need a lot of data, data variation is also a necessary condition for achieving the highest level of accuracy and objectivity. Due to the limited time frame and weather conditions for image capture, the inevitable processes—such as planning the experimental field and gathering, labeling, and cleaning datasets produced outdoors—face additional challenges. Weather, imaging timing, which is dependent on phenological phases, and labor-intensive, time-consuming fieldwork are the primary obstacles to data collection. The following paragraph provides an explanation of the BioS-Wheat dataset that was gathered for this study. It also reviews previous datasets in the fields of agriculture and phenotyping, emphasizing the GWHD as a reference dataset that is used in this paper. With variations in genotypes, location, and treatments used, the BioS-Wheat dataset expands the current GWHD dataset.

Open-source dataset overview

The Open Images V4⁹⁴, MS COCO⁹⁵, and ImageNet⁹⁶ are a few well-known large open-source datasets that are used for deep neural network training and can meet a variety of application demands in object detection tasks. Some of the segregated datasets for detecting objects with a focus on dense contexts are PUCPR⁹⁷, CARPK⁹⁸, which contains Unmanned Aerial Vehicle (UAV) images, and SKU-110K, comprising images of objects on

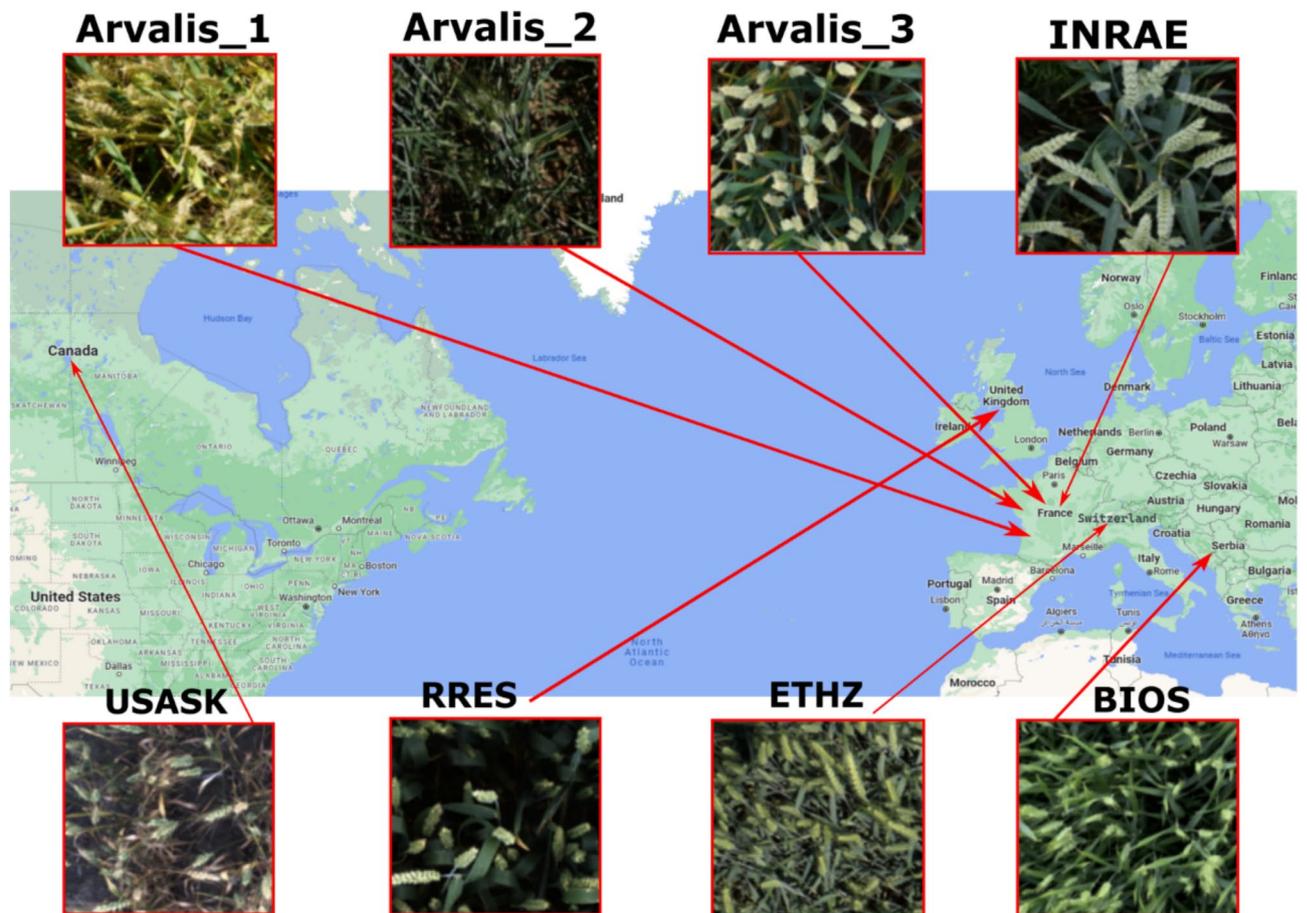


Figure 1. Example images per sources.

supermarket shelves⁹⁹. Nevertheless, images from those datasets usually contain fewer objects belonging to generic classes (cars, balloons, etc.), which are slightly overlapped or not overlapped at all. This underscores the necessity for a specialized dataset specifically designed for our task. The models we are utilizing have been pre-trained on datasets that do not fully capture the complexities of our problem. For instance, many datasets consist of images with just two objects occupying 80% of the image pixels, while our dataset contains images with over 50 wheat ears. The scale relationships between objects in these images differ significantly, which hampers the network's ability to effectively learn from the data. Additionally, our images often feature more complex object overlap, which further complicates the learning process compared to simpler scenarios. Among the publicly available datasets created in plant phenotyping studies, the vast majority are made in controlled indoor conditions, while only a few of them contain data generated in outdoor (in-field) conditions, such as datasets for weed classification^{100,101}, counting of sorghum^{102,103} and wheat ears¹⁰⁴. Overall, existing datasets and pre-trained models can serve as a starting point in agricultural applications, but when the outdoor scene is specific, such as the plot of wheat, the problem of detecting the objects of interest is grossly aggravated. The GWHD dataset is currently the largest open dataset containing high-resolution RGB images of wheat acquired in field conditions, together with instance-level annotations and geolocations of wheat ears within them. Within our work, for the reference dataset, we use a GWHD dataset and extend it with the newly created BioS-Wheat dataset.

GWHD dataset

The GWHD dataset consists of RGB images acquired during flowering, post-flowering, and ripening wheat phenological phases from nine institutions. For purposes of the GWHD challenge^{105,106} and this study, seven datasets are involved (see Table 1).

The wheat is sown with three densities (low, normal, and high) and row spacing from 12.5 cm to 30.5 cm. During image acquisition, different cameras mounted on phenotyping platforms are used with varying heights in the range from 1.8 m to 3 m above the ground. The focal length of cameras used goes from 7.7 to 60 mm, and the Ground Sample Distance (GSD) is in the range from 0.1 to 0.6 mm. Knowing from the literature that the average wheat ear length is around 1.5 cm in diameter¹⁰⁷, the obtained GSDs are sufficient for a detailed representation of wheat ear texture. Even though the high diversity of used camera sensors can help deep learning models to generalize better, that also imposes some other challenges due to data harmonization⁵. For the GWHD dataset, the authors did a manual examination of images and removed those that were acquired in an earlier phenological growth stage prior to flowering. Since the precise wheat ear detection⁵⁷ depends on image

Institution	Country	Lat	Long	Stage	Row spacing (cm)	Sowing density (seeds per m ²)	Genotypes
Arvalis1	FR	43.7N	5.8E	Post-flowering-ripening	17.5	300	20
Arvalis2	FR	43.7N	5.8E	Post-flowering	17.5	300	20
Arvalis3	FR	49.7N	3.0E	Post-flowering-ripening	17.5	300	4
ETHZ	CH	47.4N	8.6E	n.a	12.5	400	354
INRAE	FR	43.5N	1.5E	Post-flowering	16	300	7
RRES	UK	51.8N	0.36W	n.a	n.a	350	6
USASK	CA	52.1N	106E	n.a	30.5	250	16
BioS-Wheat	RS	45.45N	19.62E	Post-flowering	12.5	500-550	4
BioS-Wheat	RS	45.45N	19.62E	Medium-milk	12.5	500-550	4

Table 1. The extended GWHD table⁵ with BioS-Wheat data.

resolution and wheat ear size, which is conditioned by selected genotype and growth stage during the time of acquisition, the authors additionally rescaled images. Finally, the GWHD dataset consists of 3442 RGB images with a resolution of 1024x1024, of which 3373 images have wheat ears in them with an average number of 40 ears per image. The rest of the images do not contain wheat ears but ground, grass, or weeds. The GWHD dataset was labeled by several institutions using the COCO Annotator, the existing web-based labeling platform^{108,109}. For the unlabeled datasets, authors from⁵ used a “weakly supervised deep learning framework”¹⁰³, that relies on the YoloV3 model¹¹⁰ trained on datasets from two sources (Arvalis1 and UTokio1) to obtain labels for the remaining unlabeled images. Given that labels in the GWHD dataset are obtained through a semi-automatic approach, several issues related to human error can arise. These include occlusion of wheat ears, shadows obscuring certain ears, lighting inconsistencies, and perspective-related distortions, which can lead to variations in ear sizes across images. Such factors contribute to errors in labeling, necessitating meticulous human verification to ensure accuracy in labeling processes.

BioS-Wheat dataset

For the aim of this research study, an experimental outdoor trial with the design presented in Fig. 2 is created at the field near Ravno Selo (45.45N 19.62E), in Serbia. Four wheat varieties: Ingenio (In) by Syngenta, NS40S (NS), Pobeda (Po), and Futura (Fu) by NS Seme are sown within 24 plots each, with four fertilization treatments (see the amounts in Fig. 2), leading to variability in the wheat yield caused by different genotypes and fertilizing conditions. A summary of the BioS-Wheat dataset, with the number of images and wheat ears, is presented in Table 2. This dataset is notable for its exceptionally high sowing density (500–550 seeds per m²) and minimal row spacing (12.5 cm), conditions that are markedly more intensive than those observed in other GWHD subsets. This results in increased inter-plant competition and a dense spatial arrangement of wheat ears, leading to frequent occlusions, overlaps, and irregular ear orientations. These factors introduce considerable complexity in object detection, particularly for models that rely on axis-aligned bounding boxes or are trained on more sparsely populated datasets, such as GWHD. Incorporating the BioS-Wheat subset into the training data enhances the diversity and difficulty of the learning scenarios, enabling models to better adapt to high-density field conditions. This, in turn, improves their robustness and generalization, particularly for applications in regions where intensive cultivation practices are common.

Table 3 presents a multi-dimensional comparison of the GWHD and BioS-Wheat datasets, highlighting differences in dataset size, annotation granularity, domain coverage, and imaging conditions. While GWHD provides diversity in sowing densities and imaging platforms, the inclusion of BioS-Wheat contributes unique benefits: it represents extremely dense plots with frequent occlusions, overlapping wheat ears, and irregular ear orientations, reflecting more challenging real-world scenarios. Incorporating BioS-Wheat into the study, therefore, enhances the robustness and generalizability of phenotyping models, enabling them to handle high-density, complex field conditions that are not fully captured by existing datasets. This combination of datasets provides a more comprehensive foundation for model training and evaluation, supporting both methodological innovation and practical applicability in wheat ear detection.

The image acquisition procedure was conducted twice in two wheat growth stages: watery ripe and medium milk, according to the BBCH scale¹¹¹, in the morning hours in the period from 9 AM to 12 PM on cloud-free days to mitigate or altogether avoid possible variations in illumination conditions. A smartphone with a triple-camera system consisting of a 13 mm ultra-wide lens (*f*/2.4), a 26 mm wide lens (*f*/1.6), and a 52 mm telephoto lens (*f*/2.0)¹¹² with autofocus and dual optical image stabilization is used for imaging. The smartphone was positioned 1 meter above the crop canopy such that the whole ROI, bounded by a 0.5 m x 0.5 m wooden frame, was in the center of the image and occupied the middle part of the one wheat plot (see Fig. 2b). Although a standardized protocol for image acquisition is not established in the literature, a GSD of 0.3 mm is recommended by⁵⁷, and this recommendation was followed in our study. Acquired raw images were further appropriately rotated such that the wooden frame was parallel to the image axis, and the ROI was segmented by extracting all pixels within the wooden frame (see Fig. 2). First, the frame was segmented by converting the RGB into the HSI (hue, saturation, intensity) color model¹¹³ and by applying thresholding within the hue channel, followed by a morphological closing operation for noise reduction. The final region denoting the frame is obtained using the connected component algorithm, implemented within the OpenCV library^{114,115}, and by selecting the component with the largest area. The horizontal lines of the frame were estimated using the Canny edge

**Basic N treatment
(Autumn)**

In	0	60	180	120	60	180	0	120	180	120	60	0
Ns	120	0	60	180	120	0	60	180	0	60	120	180
Po	60	180	120	0	180	0	120	60	60	120	180	0
Fu	180	120	0	60	60	120	180	0	120	0	180	60

**Top-dress N treatment
(Spring)**

In	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0
Ns	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0
Po	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0
Fu	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0

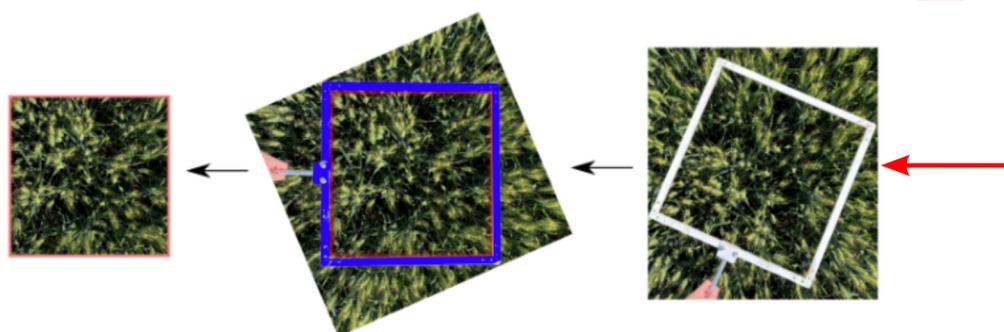


Figure 2. The experiment design with the plots numbered according to different amounts of N treatment in autumn and spring, followed by preprocessing steps for the extraction of the frame from raw images. After the segmentation of the wooden frame (blue) and image rotation, the final ROI (red) is selected for further analysis.

Varieties	images	Wheat ears
Ingenio	1472	14,528
NS40S	1344	14,080
Pobeda	1472	13,141
Futura	1408	12,736
Summary	5696	54,485

Table 2. A composition of Bios-Wheat dataset.

detector^{116,117} followed by Hough transformation¹¹⁸. The angle for rotation was estimated as the angle between the detected horizontal lines of the frame and the horizontal image axis. After the rotation is performed (see Fig. 2 b), the ROI within the frame is extracted. Finally, the obtained images of ROIs were divided following the same procedure as in⁵, resulting in 5696 square images of size 1024 x 1024. Moreover, studies are highlighting the inherent difficulty of accurately labeling densely overlapping wheat ears or similar complex cases in the GWHD dataset. As shown in the second figure from⁵, these challenging scenarios include overlapping wheat ears, heads partially truncated at the image boundaries, and images with low illumination or other suboptimal conditions. To address these issues that happen by applying the object detection semi-supervised labeling approach, our proposed method for labeling mitigates the impact of high overlap and other annotation challenges by effectively separating clustered instances, as illustrated in Fig. 3.

The ground truth labels from extracted ROI images were initially created for semantic segmentation labels that were created manually and were further adjusted for instance segmentation tasks. Imprecision within some of the ground truth labels, induced by crossing or overlapping among wheat ears, was overcome by utilizing distance transformation¹¹⁹ followed by appropriate thresholding, where half of the maximum distance value per connected component was selected as the threshold. The final bounding box labels for each denoted wheat ear were made using the connected component algorithm^{114,115}. This approach ensures more reliable instance-level

Feature	GWHD Dataset	BioS-Wheat Dataset
Number of images	3442	5696
Number of annotated objects	197,606	54,485
Classes	1 (wheat head)	1 (wheat head)
Annotation type	Axis-aligned bounding boxes	Axis-aligned bounding boxes
Annotation precision	High; semi-automatic labeling refined manually; minor errors from occlusion, lighting, and perspective	High; manual verification; very dense plots with frequent occlusions and overlaps
Field conditions	Outdoor wheat plots; sowing densities: low, normal, high; row spacing 12.5–30.5 cm	Outdoor wheat plots; extremely high sowing density (500–550 seeds per m ²); minimal row spacing 12.5 cm
Growth stage	Post-flowering (fully emerged ears)	Watery, ripe, and medium milk stages (BBCH scale)
Viewpoint/perspective	Nadir or near-nadir; phenotyping platforms or hand-held cameras	Nadir; smartphone positioned 1 m above canopy with 0.5 m × 0.5 m ROI frame
Lighting conditions	Natural daylight, mostly morning hours; moderate variation in brightness	Morning hours, cloud-free days; minimized variation
Background/clutter	Soil, grass, weeds; moderate occlusion and visual complexity	Dense wheat plots; frequent occlusion, overlapping heads, irregular ear orientations
Image resolution / GSD	1024 × 1024 px; GSD ≈ 0.1 – 0.6 mm depending on platform	4032 × 3024 px; GSD 0.3 mm per pixel
Camera/sensor types	Various RGB cameras on phenotyping platforms	Smartphone with triple-camera system (ultra-wide 13 mm, wide 26 mm, telephoto 52 mm)

Table 3. Comparative overview of the GWHD and Bios-Wheat datasets used in the study.

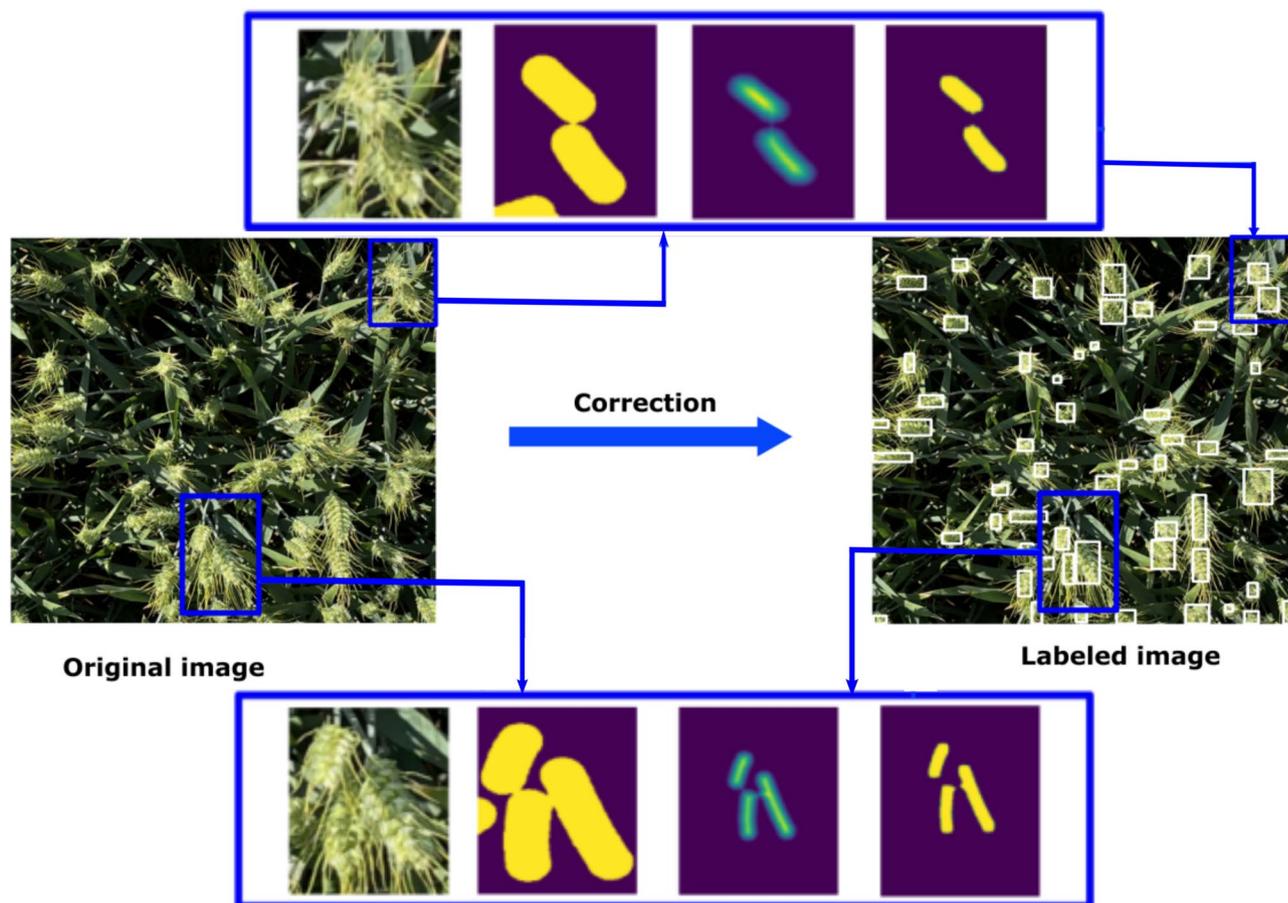


Figure 3. Visualization of the proposed labeling method handling dense and overlapping instances. The approach mitigates annotation challenges by effectively separating clustered wheat ears.

detection and quantification, even under dense and complex visual conditions, thereby improving robustness and annotation consistency in the dataset.

The BioS-Wheat dataset uses the following naming convention: the first two letters of the variety's name are used as markers to distinguish varieties, while the following number from 0 to 23 indicates the number of plots

that belong to a specified variety but also different fertilization treatments. Even though this dataset consists of additional information, in this study, all varieties with different fertilizer treatments belong to the same source, using the same practice as GWHD. The required form consists of a dictionary with the ID of the image, the width and height of the image, the bounding box for each wheat ear, and the image source.

Image quality assessment in datasets

Millions of digital images are shared daily across different platforms, making accurate Image Quality Assessments (IQA) very important for enhancing visual experiences.¹²⁰ Besides its usage in daily social media, IQA plays a vital role in academic and technical purposes^{121,122}, particularly when comparing images from different sensors. Reliable image quality predictions ensure that images are processed, stored, and displayed correctly, influencing their overall effectiveness and interpretation¹²⁰. To justify the usage of a smartphone next to scientific and professional cameras, in this study, we investigate the quality of acquired images within the BioS-Wheat and GWHD datasets using objective and commonly used image quality metrics. We will focus on no-reference methods for IQA, as no reference image could be singled out, and we aim at an overall assessment of image quality in both datasets¹²³. These methods calculate features that describe the image's structure relevant to the human opinion of the image quality. The following three methods will be used for the assessment: Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)¹²⁴, Natural Image Quality Evaluator (NIQE)¹²⁵, and Perception-based Image Quality Evaluator (PIQE)¹²⁶. However, using multiple metrics is essential to confidently validate that smartphone images are suitable for the study, ensuring robust quality assessment across various distortions. BRISQUE has the advantage of being opinion-aware, as it is trained on subjective human quality scores. PIQE provides local and global quality scores, but could be less computationally efficient compared to the other two methods. BRISQUE performs well with specific distortions but is complemented by NIQE and PIQE, which handle arbitrary distortions¹²⁷. All three methods score image quality on a scale from 0 to 100, where 100 indicates the poorest quality. In the “Results” are presented results of the metrics that show the images in the BioS-Wheat dataset are of high quality. In the following paragraph, we provide a recommendation for the high-quality smartphone imaging of wheat ears in outdoor conditions to facilitate data collection and its further enrichment.

Image acquisition recommendations

A standard protocol for acquiring wheat ear images in the field, under different outdoor conditions, is still lacking because related studies are not aligned¹²⁸. Different factors, such as the time of day, which causes varying lighting conditions, wheat phenological stages, camera perspective, and field of view, need to be considered during image acquisition, as they affect the proper performance of wheat ear recognition within acquired images. In⁹ and⁶, the nadir perspective is presented as a relevant choice during the procedure of image acquisition, especially if varieties are sown with high density. Since the height of the wheat can vary about 20%, setting the height of the camera during image acquisition is still an open question, which directly affects the visibility of wheat ears in images⁵. RGB cameras are usually used in studies, while the potential of thermal camera usage is mentioned in⁹ with a cautious assessment due to its high cost and lower resolution. For similar reasons, the use of cameras for professional photography in the field is discouraged, which further supports the idea of investigating the possibility of wheat ear detection using standard, widely available cameras on smartphone devices. Our study supports the conclusions of⁵ that there is no single phenological stage ideal for imaging all genotypes. According to the achieved recognition performance in this study, the best period for image acquisition is after flowering, when the wheat ears emerge fully, being still light, thus maintaining an upright position and avoiding or minimizing the potential overlap between the ears and leaves. Based on these recommendations and gathered experience in the field, the BioS-Wheat images were acquired under the following protocol: (i) a nadir perspective; (ii) the camera height was set to 1 m above the canopy to ensure the wooden frame was captured, aligning with the requirements from⁵, and (iii) the camera used was with a resolution of 4032 × 3024 pixels, resulting in a ground sampling distance (GSD) of 0.3 mm per pixel.

The proposed method for wheat ear detection

In this study, pre-trained models from Detectron2¹²⁹ and Ultralytics¹³⁰ are used. Detectron2 is a modular PyTorch library developed for the problem of object detection, containing trained models on different publicly available datasets and providing baseline results for most of them. Ultralytics publicly shared the YOLO (You Only Look Once) evolution series for real-time object detection and image segmentation. The latest YOLO versions aim to enhance previous versions with new features for improved performance, flexibility, and efficiency. Since this study aims to detect wheat ears in RGB images, it can be referred to as an object detection problem, as in the initial report with the GWHD dataset⁵ where the localization and counting of wheat ears in RGB images rely on the faster region-based convolutional neural network (F-RCNN) architectures, consisting of a Region Proposal Network (RPN) and Object Detection Network. This architecture achieved the best result of 0.77 mAP@50 with a mean RMSE of 12.63 wheat ears per image and R² 0.57. Moreover, the same architecture is used as one of the most prominent by the first three ranked teams applied to the GWHD challenge^{105,106}. Besides this baseline, RetinaNet reported the highest precision of around 90% for wheat ear detection on the different classification problems, such as wheat ear maturity stage¹³¹ or focusing on the wheat ears from the certain area size range on the same dataset (GWHD)⁸⁹. Within the studies^{81,132,133}, very high mAP@50 results are obtained, with YOLOv8 achieving 93%, while transformer-based architectures^{93,134} and RT-DETR¹³⁵ report even higher results, around 95%. Although these studies demonstrate exceptional precision in addressing the challenges of wheat ear detection, they utilized only publicly available GWHD datasets and improved their accuracies through data augmentation and synthetic data. In this work, the baseline models are three models that have the same backbone as a model in⁵, and RetinaNet, YOLOv8, and RT-DETR, for their higher performance as

the state-of-the-art: F-RCNN R50 FPN 3x, F-RCNN R101 FPN 3x, F-RCNN X101 32x8d FPN 3x, RetinaNet R101 FPN 3x, YOLOv8, and RT-DETR. The GWHD challenge did not result in novel architectures that could potentially improve the detection of wheat ear overlap or smaller wheat ears. The majority of the participants used existing open-source available architectures: EfficientDet⁹², F-RCNN⁴⁷, and Yolo-v3¹¹⁰. Nevertheless, that might indicate that the network architecture selection is not the prime factor in the final model performance, as none of the models used generalize significantly better.¹⁰⁶ Additionally, all winners used pseudo-labeling¹³⁶, using the predictions over the test set and fine-tuning the model with mixed data training with pseudo-labeled data¹⁰⁶. The overall top result is 0.69¹⁰⁵, by measuring mAP, a model's performance according to the weighted accuracy depending on the weight of each class¹⁰⁶.

Evaluation metrics

The most used metrics in evaluating the performance of deep learning methods for image segmentation tasks are the Intersection over Union (IoU), with its variants, along with Average Precision (AP). For evaluation of the instance segmentation models, IoU was used as an intermediate step for calculating mAP, which was calculated as the mean of AP values calculated on a per-class basis and defined as

$$mAP = \frac{1}{n} \sum_i^n AP_i \quad (1)$$

where n represents the number of classes in the considered object detection problem. In this study, mAP is denoted as mAP@50 since it was used as an IoU threshold, as in⁵. A wheat ear counting is evaluated with Relative Counting Error (RCE) as in⁴¹:

$$RCE(i) = \frac{|y_p(i) - y_t(i)|}{y_t(i)} \quad (2)$$

where y_p and y_t are predicted and ground-truth (targeted) numbers of wheat ears per observed image, i , and with accuracy (AC) calculated by the following equation:

$$AC(i) = \begin{cases} \frac{y_p(i)}{y_t(i)}, & \text{if } y_p(i) \leq y_t(i) \\ 2 - \frac{y_p(i)}{y_t(i)}, & \text{otherwise} \end{cases} \quad (3)$$

and ensuring that AC lies between 0 and 1⁴¹.

Hardware

In this study, we used 4 high-performance graphics processing units (GPUs). The GPU employed for the deep learning task and complex computational operations was the NVIDIA GeForce RTX 2080 Ti, containing 4352 CUDA cores and 11 gigabytes of high-bandwidth memory (HBM2). Their single-precision (FP32) computational performance reaches up to 13.4 teraflops, yielding a total of 53.6 teraflops across all four GPUs. This substantial computational power enabled the efficient training of the selected deep learning models within practical time frames.

Results

In this section, we will evaluate the image quality assessment and performance of models for wheat ear detection listed in “[The proposed method for wheat ear detection](#)” on GWHD separately as well as on the merged dataset, which includes GWHD and BioS-Wheat, in the following denoted as BioS-Wheat&GWHD.

Image quality assessment

In Table 4 we summarized all obtained image quality scores (their mean values with standard deviation on datasets), while in Fig. 5 distributions of quality scores obtained from the NIQE method applied on both datasets are presented. Although the NIQE score, which performs better for GWHD, indicates more favorable results, the distribution demonstrates greater stability for the Bios-Wheat dataset compared to GWHD. The overall obtained results indicate that the BioS-Wheat dataset has very high-quality images according to these three scores, and the usage of a smartphone can be involved in practices of image acquisition in situ in outdoor conditions (Fig. 4).

Wheat ear detection

In this study, we employed 5-fold cross-validation to evaluate the generalization capability of the considered models mentioned in “[The proposed method for wheat ear detection](#)”. Although the dataset comprises eight

Dataset	BRISQUE	NIQE	PIQE
GWHD	37.08 ± 13.24	14.87 ± 3.2	23.83 ± 12.48
BioS-Wheat	10.82 ± 5.18	15.55 ± 1.07	8.05 ± 3.17

Table 4. The image quality assessment—GWHD and BioS-Wheat datasets.

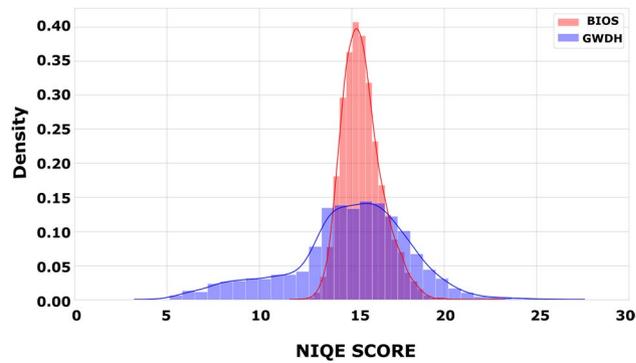


Figure 4. The image quality assessment: NIQE.

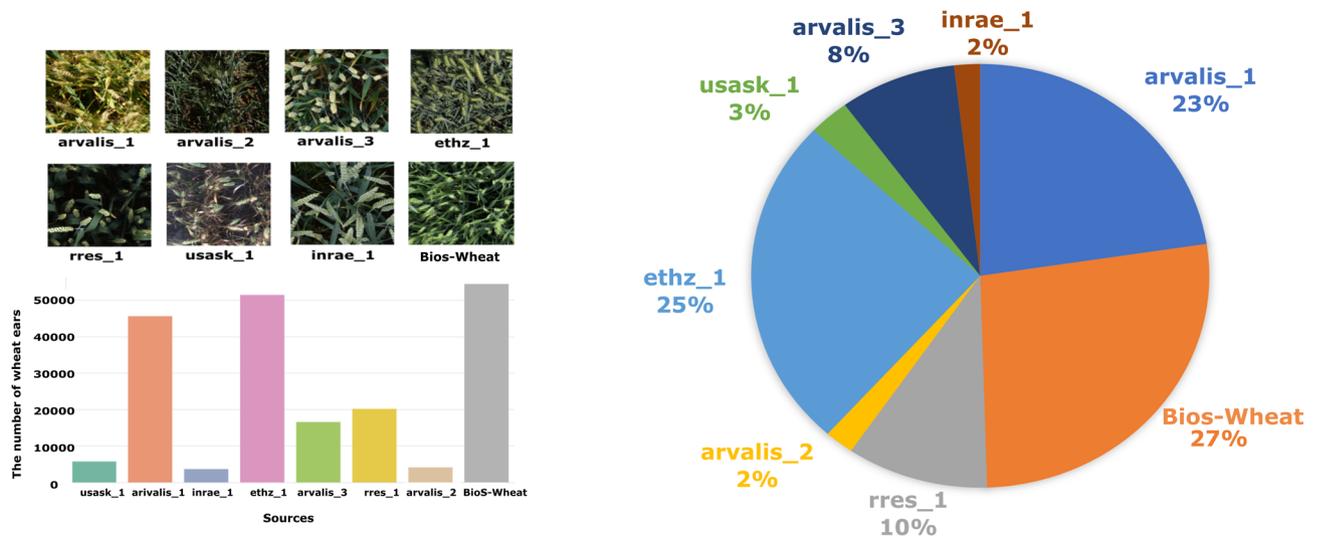


Figure 5. Left: Example images from each data source, annotated with the corresponding number of wheat ears. Right: Average representation of each source in the dataset across 5-fold cross-validation experiments.

institutions/sources, the subsets are not equally balanced (see Fig. 5); therefore, we chose a 5-fold scheme instead of a leave-one-source-out approach to ensure each fold contains sufficient representation across classes.

Leave-one-source-out (LOSO) cross-validation is not suitable for this dataset due to the extreme imbalance among sources, which range from 2 to 27% of the total dataset. Such a split would result in highly variable and misleading test metrics: sources representing only 2–3% of the dataset would yield noisy mAP@50 estimates, whereas sources representing 23–27% would disproportionately dominate the evaluation and give biased results. Consequently, training could be unstable when leaving out large sources, and overall estimates of model generalization would be misleading. In contrast, a stratified random split that respects both the overall class distribution and source proportions provides a more robust, reproducible, and interpretable evaluation of model performance. The folds were generated heuristically, and across repetitions, random splits consistently produced the most robust and reproducible results. For each fold, the dataset was split randomly into training, validation, and testing sets, while carefully maintaining the overall class distribution to ensure representative subsets, as illustrated in Fig. 5. This methodology aligns with principles from information distribution theory: random sampling maximizes the entropy of the training and test sets while minimizing potential bias introduced by structured or source-based splits, thereby providing a more objective and generalizable estimate of model performance across the entire data distribution^{137,138}. Our choice is empirically justified and ensures that reported results reliably reflect model capabilities. Ideally, a model trained on multiple datasets, as in our case, can be fine-tuned on a reasonable number of examples from a new, unseen dataset and still achieve solid performance. However, expecting the model to maintain the same level of performance under a strict leave-one-source-out regime is unrealistic due to the large number of factors contributing to variability across sources, such as illumination, crop growth stage, imaging conditions, and field management practices. This approach provides a practical trade-off between robust model training and generalization capability.

Performance metrics are reported for both the validation set and the held-out test set to ensure robust assessment. In the first experimental setting, models were fine-tuned only on the GWHD dataset, which contains 3373 images and 147,793 annotated wheat ears. Of these, 2699 images were used for training, while 337

Model	mAP@50	Training time	Inference time
F-RCNN R 50	85.37 ± 0.87	1h 40min 56s	65 ms
F-RCNN R 101	85.07 ± 0.38	2h 10min 52s	97 ms
F-RCNN X 101 32x8d	84.84 ± 0.37	4h 6min 17s	116 ms
RetinaNet R 101	84.79 ± 0.46	2h 7min 43s	85 ms
YOLOv8	85.91 ± 0.46	5h 12min 18s	18 ms
RT-DETR	87.47 ± 0.21	8h 48min 2s	137 ms

Table 5. GWHD dataset—results.

Model	mAP@50	Training duration	Inference time
F-RCNN R 50	79.58 ± 1.24	1h 45min 12s	65 ms
F-RCNN R 101	71.21 ± 1.75	2h 15min 07s	27 ms
F-RCNN X 101 32x8d	83.58 ± 2.34	4h 25min 11s	116 ms
RetinaNet R 101	86.4 ± 1.96	2h 10min 22s	85 ms
YOLOv8	87.2 ± 1.41	5h 20min 10s	18 ms
RT-DETR	90.1 ± 0.12	9h 01min 36s	137 ms

Table 6. BioS-Wheat—results.

images were allocated for validation and 337 for testing. We report both training and inference times to provide a comprehensive view of model efficiency. Training times indicate resource requirements, while inference times reflect deployment speed. Table 5 presents a comparison of six chosen detection models evaluated on the GWHD dataset, reporting average mean Average Precision at 50% IoU (mAP@50), training time, and inference time per image. All results are in the range of SOTA results from recent studies^{139–141}.

The results on the GWHD dataset demonstrate a trade-off between accuracy, training time, and inference speed across the evaluated models. RT-DETR achieves the highest detection accuracy with an mAP@50 of 87.47%, benefiting from its transformer-based architecture, but this comes with the longest training time of over 8 hours and the slowest inference speed at 137 ms per image. Among the CNN-based detectors, YOLOv8 offers a compelling balance by achieving competitive accuracy (85.91%) while maintaining the fastest inference time of 18 ms, making it well-suited for real-time applications, particularly in the context of our extended dataset, which includes smartphone data as an additional source. F-RCNN variants exhibit robust performance with mAP scores around 85%, with the ResNet-50 backbone offering the shortest training time (approximately 1 hour 40 minutes) and moderate inference speed (65 ms). Larger backbones, such as ResNeXt-101, increase training and inference times substantially, with inference latency reaching 116 ms. RetinaNet with a ResNet-101 backbone performs slightly below the best models in accuracy and inference time, positioning itself as a middle ground in this comparison. Overall, these findings highlight the importance of considering the trade-offs between model complexity, accuracy, and computational efficiency.

The second experiment focused on the performance of the same models trained on the BioS-Wheat dataset, and the results are shown in the following Table 6. The results in both Table 5 (GWHD) and Table 6 (BioS-Wheat) highlight key differences in model behavior between our proposed dataset and the existing public benchmark. On BioS-Wheat, RT-DETR achieves the highest accuracy with an mAP@50 of 90.1, surpassing all other models, though at the cost of the longest training duration and highest inference time. YOLOv8 offers a strong balance, with the second-best accuracy (87.2 mAP@50) and the fastest inference speed (18 ms), making it particularly attractive for real-time applications, running on smartphones. RetinaNet and F-RCNN variants perform competitively but generally underperform compared to YOLOv8 and RT-DETR in both accuracy and speed trade-offs. In contrast, results on GWHD show overall lower performance variation among models, with accuracies clustered in the 84–87 mAP@50 range. RT-DETR again leads with the highest accuracy (87.47), while YOLOv8 demonstrates consistently fast inference with competitive accuracy (85.91). Interestingly, models such as F-RCNN R50 achieve relatively higher accuracy on GWHD than on BioS-Wheat, suggesting that the BioS-Wheat dataset poses additional challenges due to its higher density of objects, more pronounced variability, and realistic field-specific conditions. Overall, these results confirm that while RT-DETR offers superior accuracy across both datasets, YOLOv8 provides the best compromise between speed and accuracy, especially for operational deployment. The differences observed between BioS-Wheat and GWHD underscore the importance of dataset characteristics in model evaluation, reinforcing the relevance of BioS-Wheat for testing robustness under real-world agricultural variability.

The third experiment focused on the performance of the same models trained on the standard dataset and fine-tuned on the merged dataset (BioS-Wheat + GWHD). The whole dataset consists of 9059 wheat ear images with 202278 wheat ears, where 7247 images are used for training and 453 images for validation and testing. This choice is consistent with best practices in transfer learning, where maximizing training data is prioritized, and a smaller validation set is sufficient for performance tracking^{142,143}.

Model	mAP@50	Training duration	Inference time
F-RCNN R 50	81.48 ± 2.31	1h 49min 34s	65 ms
F-RCNN R 101	72.36 ± 0.69	2h 22min 12s	27 ms
F-RCNN X 101 32x8d	84.33 ± 0.37	4h 36min 21s	116 ms
RetinaNet R 101	87.79 ± 0.76	2h 17min 41s	85 ms
YOLOv8	88.00 ± 0.56	5h 34min 11s	18 ms
RT-DETR	91.00 ± 0.41	9h 14min 42s	137 ms

Table 7. BioS-Wheat + GWHD dataset—Results.

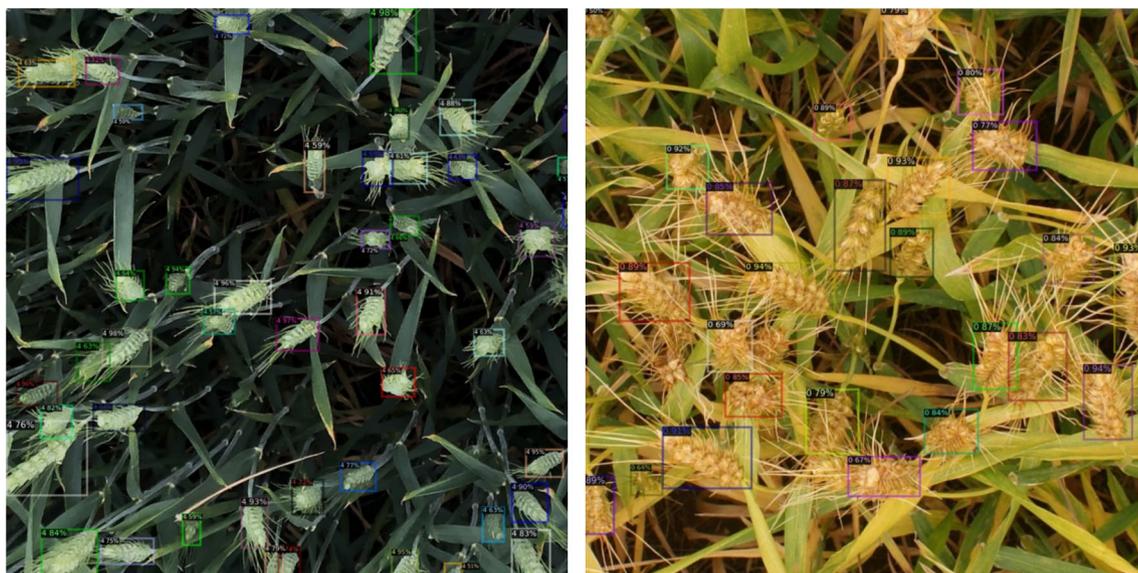


Figure 6. Detected wheat ears on test images by RetinaNet R 101 from the merged dataset; left the result on the example from the BioS-Wheat dataset, right from the GWHD dataset.

All experiments were executed under a consistent training procedure, using the same set of hyperparameters across all datasets. The learning rate was set to 0.0001, the batch size was fixed at 4 images per iteration, and the learning rate scheduler was configured as WarmupCosineLR to provide a gradual warmup followed by cosine decay, promoting stable convergence. The optimizer was stochastic gradient descent (SGD) with momentum, as in the default configuration. Empirically, learning rates in the range 0.0001 to 0.01 and batch sizes from 2 to 8 were tested, with the selected values yielding the most stable training and highest validation performance, while higher learning rates or larger batch sizes often caused oscillations in the loss or slower convergence.

Models that have demonstrated the most stable and SOTA results across various studies: YOLOv8¹⁴⁴ and RT-DETR, a Vision Transformer-Based Real-Time Object Detector¹⁴⁵ showed substantial improvements in accuracy compared to F-RCNN models, while the training time and resources were considerable, with YOLOv8 requiring 5.921 hours and RT-DETR taking 9.2 hours to complete 100 epochs (see Table 7).

When evaluated on the merged dataset including BioS-Wheat, a few models exhibited reduced detection precision compared to training solely on less dense subsets. The increased visual complexity introduced by the BioS-Wheat data, characterized by tighter spacing and higher plant density, proved particularly challenging for models like F-RCNN R101 and RetinaNet, which showed slight performance degradation. However, models with stronger generalization capabilities, such as YOLOv8 and RT-DETR, maintained relatively stable mAP scores, indicating improved resilience to occlusion and object overlap. The examples of images with detected wheat ears from the merged dataset are shown in Fig. 6.

The integration of the BioS-Wheat dataset revealed divergent behaviors across the evaluated architectures (see Table 8). The F-RCNN models, particularly the R101 backbone, exhibited a decline in performance, with a sharp reduction of 12.71 mAP50, suggesting limited robustness to data heterogeneity and possible overfitting to the original distribution. In contrast, RetinaNet, YOLOv8, and RT-DETR benefited from the extended dataset, indicating stronger generalization capabilities; among them, RT-DETR demonstrated the highest gain (3.53 mAP50), underscoring its capacity to use additional variability in the training data. Training times increased consistently across all models, which is expected given the expanded input, while inference times remained generally stable because that is most related to the architecture, not the data. On the other side, the integration of the BioS-Wheat dataset into the GWHD dataset results in a statistically significant improvement in overall detection performance. Notably, the BioS-Wheat subset introduces agronomically dense wheat varieties

Model	Δ mAP@50	Δ Training duration
F-RCNN R 50	- 3.89	+08min 38s
F-RCNN R 101	- 12.71	+11min 20s
F-RCNN X 101 32x8d	- 0.51	+30min 04s
RetinaNet R 101	+3.00	+09min 58s
YOLOv8	+2.09	+21min 53s
RT-DETR	+3.53	+26min 40s

Table 8. A relative performance comparison of models trained on the merged dataset against the GWHD baseline.

Arvalis1	BioS-Wheat	RRES1	Arvalis2	ETHZ1	USASK1	Arvalis3	INRAE1
35.33 \pm 0.76	53.17 \pm 0.40	53.83 \pm 1.93	32.31 \pm 3.81	41.68 \pm 1.52	42.43 \pm 3.31	41.30 \pm 8.26	52.82 \pm 2.37

Table 9. Average mAP@50 obtained by RT-DETR across sources.

and complex field conditions—characterized by high sowing density, pronounced occlusion, overlapping wheat ears, and increased visual clutter—which are largely underrepresented in the original GWHD distribution. This enrichment enhances the model's ability to generalize across more diverse and challenging scenarios.

The modest quantitative gains observed in model metrics are, therefore, not a reflection of redundancy or noise, but rather of the dataset pushing the limits of existing architectures. Indeed, such data exposes blind spots in conventional detectors, making it invaluable for stress-testing robustness and generalization. Moreover, from an applied perspective, BioS-Wheat represents realistic field conditions in high-intensity agricultural practices where visual complexity is inevitable. While more complex detection architectures (e.g., incorporating oriented bounding boxes, transformer-based attention, or instance segmentation) may better handle such scenarios, they often come at the cost of inference speed and computational overhead. Given our long-term objective of achieving real-time wheat ear detection in practical field deployments, this trade-off is non-trivial. Therefore, incorporating BioS-Wheat in model development is a strategic step: it injects complexity that enhances robustness without overfitting to artificially idealized cases, and it does so while preserving the possibility of lightweight inference.

Overall, the problem of multi-detection and over-counting of wheat ears is manifest in this challenging topic. Many authors, such as⁸⁹, seek simplification and reach higher precision by removing certain critical bounding box areas using an adjusted RetinaNet model. This approach boosts the precision score but reduces the ability to detect very small or very large wheat ears in images (the applied threshold on a calculated area of the bounding box excludes those that are smaller than 20 and larger than 2×10^5 pixels). One additional way to address this issue is to align all images in the dataset to a horizontal orientation, as suggested by David et al.¹⁰⁶, or by employing oriented bounding boxes that can better capture non-axis-aligned objects, as proposed by Yu et al.¹⁴⁶. We further analyzed the RT-DETR model, which achieved the highest mAP@50 on the merged dataset (see Table 7), in detection performances per source (see Table 9). High variability in the standard deviation of the mAP@50 metric suggests sensitivity to intra-source differences. Incorporating such diverse sources, though challenging, enhances the overall merged dataset's robustness.

In summary, the intention of the BioS-Wheat dataset inclusion is not primarily for improving detection performances but to ensure that detection models are exposed to the full spectrum of real-world variability, thereby aligning development with agronomic relevance and operational feasibility. In addition, robustness experiments were conducted to assess model performance under realistic image perturbations, such as variations in brightness, Gaussian and motion blur, and JPEG compression. These tests demonstrate that the model maintains a consistently high mAP_{50} , the main evaluation metric emphasized in this study, across a wide range of conditions. For example, brightness adjustments of $\pm 30\%$ or moderate Gaussian blur (kernel sizes 3×3 to 7×7) result in minimal changes to mAP_{50} , while moderate JPEG compression (quality $q = 30-50$) also preserves high performance ($mAP_{50} \geq 0.86$). Only extreme compression ($q = 10$) leads to a notable decrease ($mAP_{50} = 0.81$), illustrating expected sensitivity. Overall, these results indicate that the model is robust to common real-world perturbations, further supporting its reliability and suitability for practical agricultural deployment (Table 10).

Wheat ear counting

The precision of wheat ear counting is evaluated based on the mentioned metrics: RCE and AC. The best model achieved an RCE of 0.11 and an AC of 0.885 for the BioS-Wheat dataset. In addition to those metrics, for precision evaluation, we used the R2 score¹⁴⁷ and standard correlation coefficients: Pearson¹⁴⁸, Spearman¹⁴⁹, and Kendall¹⁵⁰. Moreover, to get insight into the number of ears detected multiple times or undetected, we calculate a Mean Squared Error (MSE) and Root Mean Square Error (RMSE) between the detected number and the true number of wheat ears. The results are presented in Table 11.

The overall metrics indicate that the model has a good but not perfect fit across the BioS-Wheat dataset. The Pearson, Spearman, and Kendall correlations demonstrate a strong alignment between true number of

Perturbation	Precision	Recall	F1	mAP_{50}
Brightness -30	0.91	0.87	0.89	0.90
Brightness 0	0.92	0.89	0.90	0.91
Brightness +30	0.91	0.87	0.88	0.90
Gaussian 3	0.90	0.87	0.88	0.89
Gaussian 5	0.89	0.85	0.87	0.88
Gaussian 7	0.88	0.834	0.86	0.87
JPEG 10	0.82	0.77	0.80	0.81
JPEG 30	0.87	0.84	0.85	0.86
JPEG 50	0.90	0.87	0.88	0.89
Motion 3	0.91	0.88	0.89	0.90
Motion 5	0.90	0.87	0.88	0.89
Motion 7	0.88	0.85	0.90	0.87

Table 10. Robustness evaluation of the model under various image perturbations. mAP_{50} is emphasized as the main metric.

Pearson	Spearman	Kendall	R2 score	MSE	RMSE
0.849	0.724	0.846	0.693	2.67	1.634

Table 11. Correlations between labeled and detected wheat ears in the BioS-Wheat dataset.

Variety	Pearson	Spearman	Kendall	R ² score
Ingenio	0.809	0.684	0.812	0.616
NS40S	0.846	0.707	0.839	0.694
Pobeda	0.889	0.778	0.886	0.767
Futura	0.905	0.896	0.795	0.805

Table 12. Correlations between the numbers of detected and the true number of wheat ears in the BioS-Wheat dataset per variety.

wheat ears and detected counts, indicating that the model effectively captures the general trend and ranking. However, the R² score, MSE, and RMSE indicate that the model's predictions are not perfectly aligned with the true number of wheat ears, revealing areas where accuracy could be enhanced. To better understand the model's performance and identify specific areas for improvement, we conducted a detailed analysis of each variety within our dataset (see results in Table 12).

The analysis reveals that for the Futura wheat ears, the model achieved the strongest overall correlations and R² score, suggesting the detection model performs best on this variety, providing both high linear and monotonic agreement with the true number of wheat ears. It is a variety known as a high-yielding bread wheat variety characterized by dense, well-defined, and uniformly shaped ears—traits that likely enhance its visibility in aerial imagery. For Pobeda wheat ears, the model also shows strong correlations and a high R² score, making it the second most accurately detected variety. It is a robust variety widely grown in the Balkans for its adaptability and moderate resistance to lodging, often featuring large, erect ears with consistent spacing, which aids detection precision. Wheat ears of NS40S and Ingenio varieties are the most imprecisely detected, and thus lower correlations and R² scores are achieved, especially for the Ingenio variety.

Additionally, the KDE (Kernel Density Estimation) plot in Fig. 7 illustrates how closely the number of detected wheat ears matches the number counted in different varieties. For Pobeda and Futura, the data points are densely clustered along the diagonal, indicating strong agreement between the numbers of the detected and the true number of wheat ears. Futura exhibits a slightly broader distribution, suggesting minor deviations but overall consistency. Ingenio shows more variability compared to Pobeda and Futura, though its core density still follows the diagonal, implying that, despite greater variation, the model generally performs well. NS40S displays the widest spread, indicating the greatest inconsistency and lowest detection accuracy among the four varieties.

Discussion

The published studies in the field of wheat ear detection are dedicated to limited datasets that depend on geographical location, imposing challenges during the migration of developed and trained models to other regions. To the best of our knowledge, no other wheat ear-labeled dataset has been systematically collected by a smartphone. The inclusion of the BioS-Wheat subset to GWHD introduces a meaningful shift in data complexity, emphasizing the impact of agronomic diversity on model performance. High sowing density and

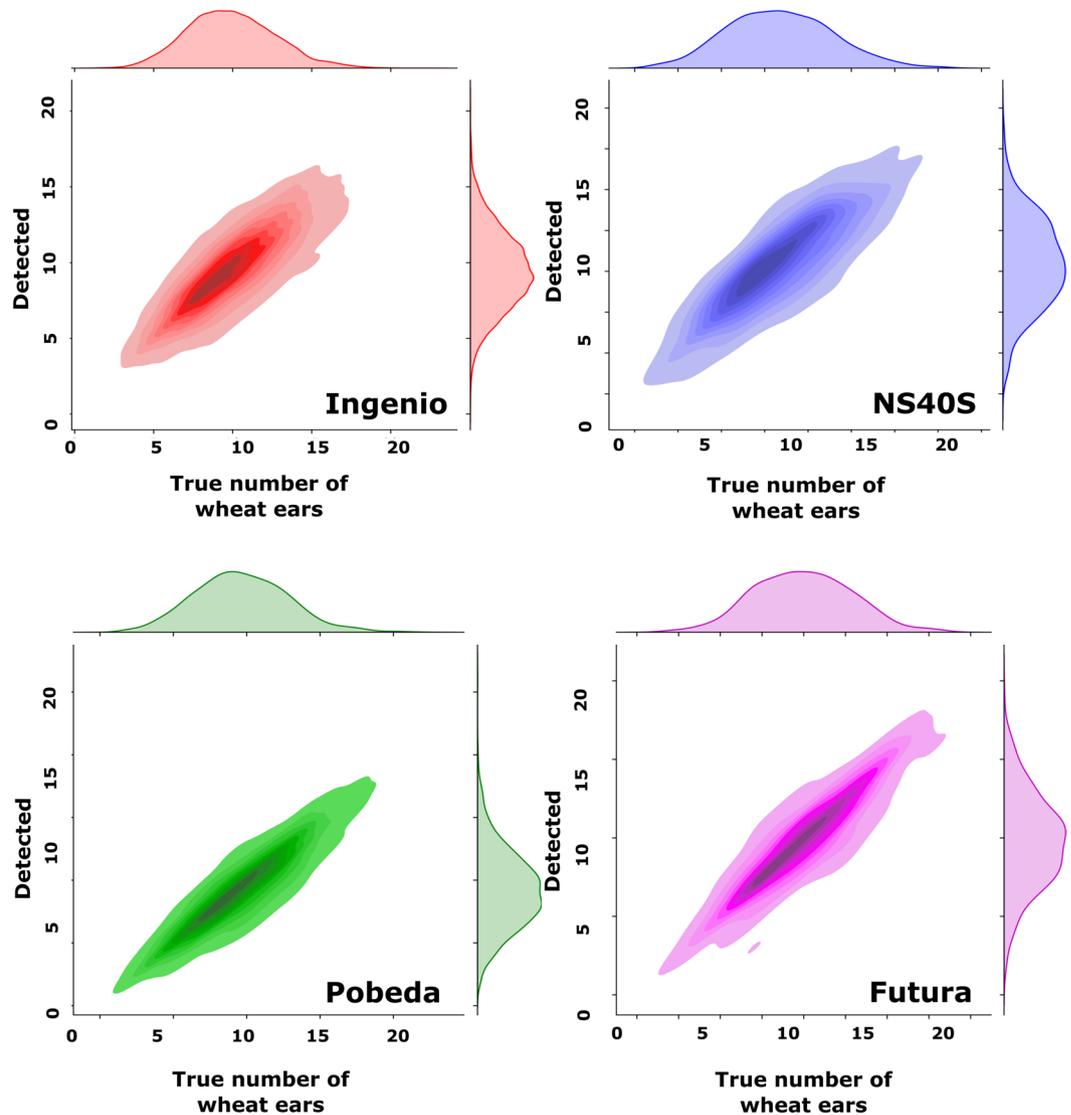


Figure 7. KDE plots showing the density of detection results for each wheat variety in the BioS-Wheat dataset, highlighting differences in model performance and alignment with true number of wheat ears.

minimal row spacing lead to increased object occlusion and dense spatial arrangements, which challenge standard object detectors, particularly those with fixed anchor sizes or limited capacity to model irregular shapes. Nonetheless, training with such data is beneficial, as it exposes models to real-world variability. This fosters robustness, especially in deployment scenarios where field conditions are less controlled and denser crop structures are prevalent. This study includes data collection and harmonization steps that are validated by the image quality methods. Moreover, applying deep learning methods gives insights into how adding different geographical locations affects distinguishing wheat ears across the spatial distribution.

The deep learning methods that are compared in this study are F-RCNN and RetinaNet, recognized as those achieving the highest precision in^{57,131}, YOLOv8, and RT-DETR. The superior performance of RT-DETR and YOLOv8 in our study can be additionally explained by their ability to effectively model densely distributed and highly occluded objects, as exemplified in the BioS-Wheat dataset. In the case of RT-DETR, the transformer-based architecture uses global self-attention, enabling the model to capture long-range dependencies and contextual relationships across the image. This facilitates better differentiation between individual wheat ears, even when they are heavily overlapping or partially occluded, by learning relational cues beyond local receptive fields. Such global reasoning allows RT-DETR to map individual instances more precisely and reduces the misclassification or suppression of adjacent objects. Additionally, we include confidence–performance curves illustrating precision, recall, and F1 across confidence thresholds (see Fig. 8), which demonstrate stable calibration and consistent performance trends across folds, reinforcing the systematic advantages of this architecture.

YOLOv8, in contrast, achieves robustness through an anchor-free detection paradigm and an optimized multi-scale backbone. By directly predicting object centers instead of relying on fixed anchor boxes, YOLOv8 circumvents the limitations of anchor matching in crowded regions. This design is particularly advantageous in wheat plots, where wheat ears appear at varying sizes and often overlap. The improved multi-scale feature

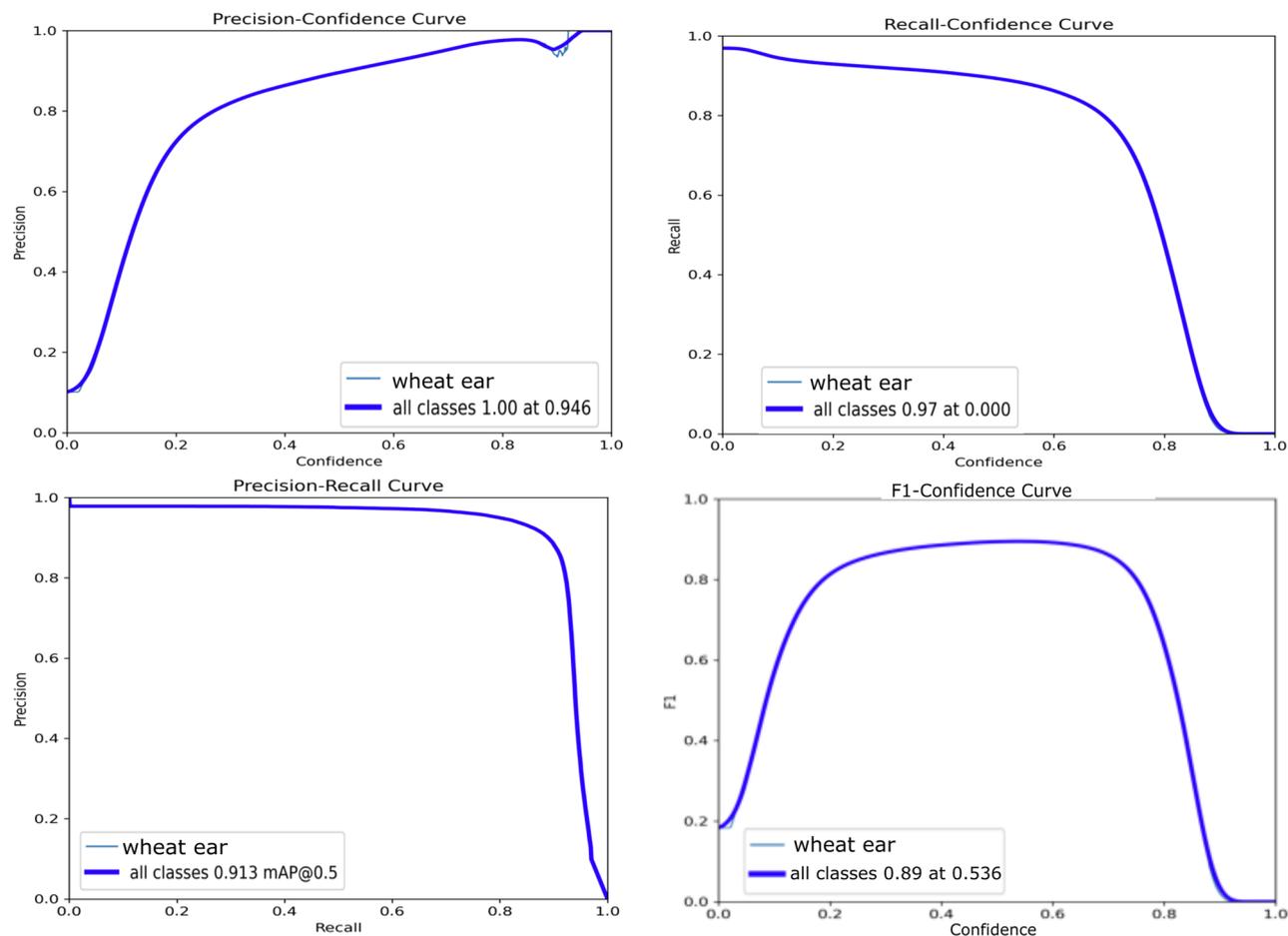


Figure 8. Confidence-based performance evaluation. Top panels: Precision–Confidence (left) and Recall–Confidence (right) curves showing precision increasing and recall decreasing smoothly with confidence, indicative of good calibration. Bottom panels: Precision–Recall (left) and F1–Confidence (right) curves, demonstrating stable high precision across recall values and identifying an optimal F1 operating point (confidence around 0.54, F1 is 0.89). The consistent monotonic behavior across all curves confirms the model’s reliable probability calibration and supports its suitability for field deployment.

extraction further enhances the detection of smaller and partially visible ears, resulting in more stable localization and reduced overlap errors. By comparison, traditional anchor-based detectors such as F-RCNN and RetinaNet exhibit performance degradation under these conditions, as dense object distributions lead to anchor redundancy, misaligned proposals, and excessive non-maximum suppression. Collectively, these findings highlight how the architectural choices of RT-DETR and YOLOv8 enable superior adaptation to complex, high-density agricultural imagery, underscoring their robustness relative to older detection frameworks.

In studies, such as^{57,131}, models that usually achieved high performances are applied with some simplifications that include the division of the image dataset into two classes based on wheat ear maturity, which introduced another layer of complexity, and by excluding some potentially erroneous samples to avoid learning from inaccurate data. The results from this study support the previous findings that deep learning models and transfer learning approaches for object detection can be successfully learned and applied to datasets containing images across many countries with different wheat ear densities and growing environments. Even though the shape and grains of wheat ears tend to be more stable during growing stages, the contrast in images between the wheat ears, background, and leaves is more pronounced, which further aggravates wheat ear detection^{57,74,151}.

To face a common challenge for all automatic and semi-automatic labeling methods used for wheat ears, which can even occur during manual annotation, in future work the alternative strategies for instance separation, such as watershed-based methods, learnable instance separation, or direct object detection approaches using models like the latest versions of YOLO, will be considered, highlighting potential avenues for improving robustness in dense and overlapping scenarios.

The incorporation of the BioS-Wheat dataset effectively acts as a “stress test” for model generalization. While classical anchor-based models, such as F-RCNN with a ResNet-101 backbone, exhibited a notable performance decline under these conditions, modern architectures like RT-DETR and YOLOv8 remained stable or improved. This difference arises because BioS-Wheat introduces higher object density and increased occlusion, which challenge the adaptability of models reliant on region proposal networks and predefined anchors. In contrast,

architectures with global attention mechanisms or anchor-free, multi-scale detection capabilities are better equipped to generalize under such complex scenarios, highlighting the enhanced robustness of contemporary object detectors in dense and occluded environments.

By analyzing varieties within the Bios-Wheat dataset, Pobeda and Futura demonstrate stronger alignment and more consistency between the number of wheat ears detected and true value, indicating a higher detection accuracy for these varieties. Ingenio and NS40S show a slight increase in variability, which could mean lower detection precision for these varieties. This analysis suggests that the detection model performs best with Pobeda and Futura, while it may require further refinement to handle other varieties with more precision.

In this study, several models were evaluated to identify the models that achieve sufficient robustness in performance for data from different geographic locations and sources, and different growth stages, in order to avoid the restrictive model usage on images acquired during a specific growth stage. This means that potential users will not be restricted by any established image timing protocols since there aren't any currently in place. Achieving sufficient robustness at different varieties, growth stages, and locations can provide a good baseline for establishing the procedure for image crowdsourcing, which would further facilitate dataset expansion and model improvements. Even though the RetinaNet showed lower time complexity when struggling with a few classes, the F-RCNN showed better overall results in basic DNN models, while RT-DETR showed the best results in average precision but the highest complexity, resource demand, and training duration.

Conclusions

The BioS-Wheat dataset advances the standardization of agricultural computer vision by providing a reproducible and high-density benchmark for wheat ear detection under realistic and complex field conditions. Developed under the principles of open science, BioS-Wheat, together with its accompanying framework, enhances transparency, comparability, and methodological rigor within the research community. Its scalable design enables the development of robust, field-adaptable models that extend beyond wheat to diverse genotypes, cultivation systems, and phenotyping tasks, thereby facilitating large-scale crop monitoring and precision agriculture applications. Incorporating the BioS-Wheat subset into the GWHHD dataset significantly enhances its representativeness by introducing high-density, occlusion-heavy imagery. Although this complexity can lead to slight model performance drops, it ultimately supports the development of more robust, adaptable wheat ear detectors. The results underscore the importance of including agronomically diverse and visually challenging data in model training to ensure performance reliability across a broader range of field conditions. Moreover, this study shows that the usage of affordable sensors and 'good enough' technology can provide many benefits. It facilitates the data collection, which directly results in the improvement of data-driven learning algorithms and their performance, but also extends databases among countries, different genotypes, and species. The agricultural industry has been transformed by the emergence of affordable technologies, particularly smartphones, which have provided farmers with new ways to manage their fields more effectively. Smartphones can be used to monitor soil and crop health, access real-time weather information, remotely manage farm operations, practice precision agriculture, and collect data efficiently. These technological advancements have boosted productivity and efficiency in agriculture, empowering farmers to make better-informed decisions and increase their yields.

Data availability

The code used in this study is available from the corresponding author upon request. The dataset will be made publicly available on the Zenodo platform following the publication of the paper.

Received: 24 July 2025; Accepted: 10 November 2025

Published online: 29 December 2025

References

1. FAO, IFAD, UNICEF, WFP, WHO. The state of food security and nutrition in the world 2024 – building climate resilience for food security and nutrition, <https://doi.org/10.4060/cd1254en> (2018).
2. Food and Agriculture Organization of the United Nations. Sdgs helpdesk – sustainable development goals helpdesk. <https://www.fao.org/sustainable-development-goals-helpdesk/en/> (2025). Accessed: 2025-09-23.
3. Food and Agriculture Organization of the United Nations. Faostat – commodities by country. https://www.fao.org/faostat/en/#rankings/commodities_by_country (2025). Accessed: 2025-09-22.
4. Curtis, B. C., Rajaram, S. & Gómez Macpherson, H. *Bread Wheat: Improvement and Production*, vol. 30 of *FAO Plant Production and Protection Series* (Food and Agriculture Organization of the United Nations (FAO), Rome, 2002). Accessed: 2025-09-23, <https://www.fao.org/4/y4011e/y4011e00.htm>.
5. David, E. 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* <https://doi.org/10.34133/2020/3521852> (2020).
6. Hasan, M., Chopin, J., Laga, H. & Miklavcic, S. Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* <https://doi.org/10.1186/s13007-018-0366-8> (2018).
7. Thakur, A. K., Singh, S., Goyal, N. & Gupta, K. A comparative analysis on the existing techniques of wheat spike detection. In *2021 2nd International Conference for Emerging Technology (INCET)*, 1–6, <https://doi.org/10.1109/INCET51464.2021.9456284> (IEEE, 2021).
8. Zhou, C. et al. Wheat ears counting in field conditions based on multi-feature optimization and twsvm. *Front. Plant Sci.* **9**, 1024. <https://doi.org/10.3389/fpls.2018.01024> (2018).
9. Fernandez-Gallego, J. A., Kefauver, S. C., Gutiérrez, N. A., Nieto-Taladriz, M. T. & Araus, J. L. Wheat ear counting in-field conditions: high throughput and low-cost approach using rgb images. *Plant Methods* **14**, 1–12. <https://doi.org/10.1186/s13007-018-0289-4> (2018).
10. Key, J. M. & Qualset, C. O. Conventional methods of wheat breeding. In Smith, E. L. (ed.) *Genetic Improvement in Yield of Wheat*, 7–24, <https://doi.org/10.2135/cssaspecpub13.c2> (Crop Science Society of America, Madison, WI, 1986).
11. Crop, N. D. & Report, P. 5 steps for estimating wheat yields. <https://www.farmprogress.com/wheat/5-steps-estimating-wheat-yields> (2018).

12. Cai, Y. et al. Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches. *Agric. For. Meteorol.* **274**, 144–159. <https://doi.org/10.1016/j.agrformet.2019.03.010> (2019).
13. Grbovic, Z., Panic, M., Marko, O., Brdar, S. & Crnojevic, V. Wheat ear detection in rgb and thermal images using deep neural networks. *Environments* **11**, 13 (2019).
14. Kamilaris, A. & Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **147**, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016> (2018).
15. Zhu, N. et al. Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. *Int. J. Agric. Biol. Eng.* **11**, 32–44. <https://doi.org/10.25165/j.ijabe.20181104.4475> (2018).
16. Kondaveeti, H. K., Brahma, G. P. & Sahithi, D. V. Deep learning applications in agriculture: The role of deep learning in smart agriculture. In *Artificial Intelligence and IoT-Based Technologies for Sustainable Farming and Smart Agriculture*, 325–345, <https://doi.org/10.4018/978-1-7998-1722-2.ch020> (IGI Global, 2021).
17. Dyrmann, M., Jørgensen, R. N. & Midtby, H. S. Roboweedsupport: Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Adv. Anim. Biosci.* **8**, 842–847. <https://doi.org/10.1017/S2040470017000206> (2017).
18. McCool, C., Perez, T. & Upcroft, B. Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics. *IEEE Robot. Autom. Lett.* **2**, 1344–1351. <https://doi.org/10.1109/LRA.2017.2675698> (2017).
19. Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P. & French, A. P. Deep learning for multi-task plant phenotyping. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2055–2063, <https://doi.org/10.1101/204552> (2017).
20. Taghavi Namin, S., Esmailzadeh, M., Najafi, M., Brown, T. B. & Borevitz, J. O. Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods* **14**, 66. <https://doi.org/10.1186/s13007-018-0333-4> (2018).
21. Lu, J., Hu, J., Zhao, G., Mei, F. & Zhang, C. An in-field automatic wheat disease diagnosis system. *Comput. Electron. Agric.* **142**, 369–379. <https://doi.org/10.1101/204552> (2017).
22. Saleem, M. H., Potgieter, J. & Arif, K. M. Plant disease detection and classification by deep learning. *Plants* <https://doi.org/10.3390/plants8110468> (2019).
23. Genaev, M. A. et al. Image-based wheat fungi diseases identification by deep learning. *Plants* <https://doi.org/10.3390/plants10081500> (2021).
24. Goyal, L., Sharma, C. M., Singh, A. & Singh, P. K. Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. *Inf. Med. Unlock.* **25**, 100642. <https://doi.org/10.1016/j.imu.2021.100642> (2021).
25. Li, L., Zhang, S. & Wang, B. Plant disease detection and classification by deep learning—a review. *IEEE Access* **9**, 56683–56698. <https://doi.org/10.1109/ACCESS.2021.3069646> (2021).
26. Kumar, D. & Kukreja, V. Deep learning in wheat diseases classification: A systematic review. *Multimedia Tools Appl.* **81**, 10143–10187. <https://doi.org/10.1007/s11042-022-12160-3> (2022).
27. Lad, A. & Raval, M. S. Improving wheat head detection: A data-centric approach by domain variance reduction. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, 1140–1143, <https://doi.org/10.1145/3477314.3507190> (2022).
28. Barré, P., Stöver, B. C., Müller, K. F. & Steinhage, V. Leafnet: A computer vision system for automatic plant species identification. *Eco. Inform.* **40**, 50–56. <https://doi.org/10.1016/j.ecoinf.2017.05.005> (2017).
29. Namin, S. T., Esmailzadeh, M., Najafi, M., Brown, T. B. & Borevitz, J. O. Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods* **14**, 66. <https://doi.org/10.1186/s13007-018-0333-4> (2018).
30. Ubbens, J. R. & Stavness, I. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* **8**, 1190. <https://doi.org/10.3389/fpls.2017.01190> (2017).
31. Bi, K., Jiang, P., Li, L., Shi, B. & Wang, C. Non-destructive measurement of wheat spike characteristics based on morphological image processing. *Trans. Chin. Soc. Agric. Eng.* **26**, 212–216. <https://doi.org/10.3969/j.issn.1002-6819.2010.12.036> (2010).
32. Kun, B., Pan, J., Chongwei, T., Feifei, H. & Cheng, W. The design of wheat variety bp classifier based on wheat ear feature. *Chin. Agric. Sci. Bull.* **6**, 465–467. <https://doi.org/10.11924/j.issn.1000-6850.2010-2527> (2011).
33. Cointault, F., Guerin, D., Guillemin, J.-P. & Chopinet, B. In-field triticum aestivum ear counting using colour-texture image analysis. *N. Z. J. Crop. Hortic. Sci.* **36**, 117–130. <https://doi.org/10.1080/01140670809510227> (2008).
34. Hecht-Nielsen, R. Theory of the backpropagation neural network. In Wechsler, H. (ed.) *Neural Networks for Perception*, 65–93, <https://doi.org/10.1016/B978-0-12-741252-8.50010-8> (Elsevier, 1992).
35. Woods, J. W. Chapter 7 - image enhancement and analysis. In Woods, J. W. (ed.) *Multidimensional Signal, Image, and Video Processing and Coding (Second Edition)*, 223–256, <https://doi.org/10.1016/B978-0-12-381420-3.00007-2> (Academic Press, Boston, 2012), second edition edn.
36. Brownrigg, D. R. K. The weighted median filter. *Commun. ACM* **27**, 807–818. <https://doi.org/10.1145/358198.358222> (1984).
37. Thorsten Wagner, J.-Y. T. Find maxima detection algorithm, image j. [https://imagej.net/imagej-wiki-static/Find_maxima_\(Track_mate_module\)](https://imagej.net/imagej-wiki-static/Find_maxima_(Track_mate_module)) (2016).
38. Matz, G. & Hlawatsch, F. Linear time–frequency filters: On-line algorithms and applications. In *Applications in Time-Frequency Signal Processing*, 205–272, <https://doi.org/10.1201/9781315220017-6> (CRC Press, 2018).
39. Shlens, J. A tutorial on principal component analysis. arXiv preprint [arXiv:1404.1100](https://arxiv.org/abs/1404.1100) <https://doi.org/10.13140/2.1.1593.1684> (2014).
40. Fix, E. & Hodges, J. L. Discriminatory analysis. nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **57**, 238–247. <https://doi.org/10.2307/1403797> (1989).
41. Alharbi, N., Zhou, J. & Wang, W. Automatic counting of wheat spikes from wheat growth images. *University of East Anglia, Digital repository* <https://doi.org/10.5220/0006580403460355> (2018).
42. Du, Y. et al. Field wheat ears counting based on superpixel segmentation method. *Sci. Agric. Sin.* **52**, 21–33. <https://doi.org/10.3864/j.issn.0578-1752.2024.12.005> (2019).
43. Kim, K.-S., Zhang, D., Kang, M.-C. & Ko, S.-J. Improved simple linear iterative clustering superpixels. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, 259–260, <https://doi.org/10.1109/ISCE.2013.6570216> (2013).
44. Huang, H., Wei, X. & Zhou, Y. Twin support vector machines: A survey. *Neurocomputing* **300**, 34–43. <https://doi.org/10.1016/j.neucom.2018.01.093> (2018).
45. Alom, M. Z. et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics* **8**, 292. <https://doi.org/10.3390/electronics8030292> (2019).
46. Russakovsky, O. et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252. <https://doi.org/10.1007/s11263-015-0816-y> (2015).
47. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural. Inf. Process. Syst.* <https://doi.org/10.5555/2969239.2969250> (2015).
48. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969, <https://doi.org/10.1109/ICCV.2017.322> (2017).
49. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28 (Springer, 2015).
50. Aich, S. & Stavness, I. Global sum pooling: A generalization trick for object counting with small datasets of large images. arXiv preprint [arXiv:1805.11123](https://arxiv.org/abs/1805.11123) <https://doi.org/10.1109/CVPRW.2019.00104> (2018).

51. Xiong, H. et al. From open set to closed set: Counting objects by spatial divide-and-conquer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8362–8371, <https://doi.org/10.1109/ICCV.2019.00845> (2019).
52. Minaee, S. et al. Image segmentation using deep learning: A survey. arXiv preprint [arXiv:2001.05566](https://arxiv.org/abs/2001.05566)<https://doi.org/10.48550/arXiv.2001.05566> (2020).
53. Fukushima, K., Miyake, S. & Ito, T. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Trans. Syst. Man Cybern.* <https://doi.org/10.1109/29.21701> (1983).
54. Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. & Lang, K. J. Phoneme recognition using time-delay neural networks. *IEEE Trans. Acoust. Speech Signal Process.* **37**, 328–339. <https://doi.org/10.1109/29.21701> (1989).
55. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324. <https://doi.org/10.1109/5.726791> (1998).
56. Jin, X. et al. High-throughput measurements of stem characteristics to estimate ear density and above-ground biomass. *Plant Phenomics* <https://doi.org/10.34133/2019/7532758> (2019).
57. Mader, S. et al. Ear density estimation from high resolution rgb imagery using deep learning technique. *Agric. For. Meteorol.* **264**, 225–234. <https://doi.org/10.1016/j.agrformet.2018.10.013> (2019).
58. Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P. & Hawkesford, M. J. Deepcount: In-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Front. Plant Sci.* **10**, 1176. <https://doi.org/10.3389/fpls.2019.01176> (2019).
59. Gibbs, J. A., Burgess, A. J., Pound, M. P., Pridmore, T. P. & Murchie, E. H. Recovering wind-induced plant motion in dense field environments via deep learning and multiple object tracking. *Plant Physiol.* **181**, 28–42. <https://doi.org/10.1104/pp.18.01083> (2019).
60. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175> (2020).
61. Chen, L.-C. et al. Masklab: Instance segmentation by refining object detection with semantic and direction features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4013–4022, <https://doi.org/10.1109/CVPR.2018.00421> (2018).
62. Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. Path aggregation network for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8759–8768, <https://doi.org/10.1109/CVPR.2018.00913> (2018).
63. Chen, K. et al. Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4974–4983, <https://doi.org/10.1109/CVPR.2019.00509> (2019).
64. Cao, Y., Xu, J., Lin, S., Wei, F. & Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0, <https://doi.org/10.1109/ICCVW.2019.00014> (2019).
65. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J. Yolact: Real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9157–9166, <https://doi.org/10.1109/ICCV.2019.00926> (2019).
66. Cheng, B., Misra, I., Schwing, A. G., Kirillov, A. & Girdhar, R. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1290–1299, <https://doi.org/10.1109/CVPR52688.2022.00135> (2022).
67. Liu, L. et al. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **127**, 1659–1679. <https://doi.org/10.1007/s11263-019-01247-4> (2019).
68. Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448, <https://doi.org/10.1109/ICCV.2015.169> (2015).
69. Zagoruyko, S. et al. A multipath network for object detection. arXiv preprint [arXiv:1604.02135](https://arxiv.org/abs/1604.02135)<https://doi.org/10.48550/arXiv.1604.02135> (2016).
70. Wang, X., Girshick, R., Gupta, A. & He, K. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7794–7803, <https://doi.org/10.1109/CVPR.2018.00813> (2018).
71. Chen, X., Girshick, R., He, K. & Dollár, P. TensorMask: A foundation for dense object segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2061–2069, <https://doi.org/10.1109/ICCV.2019.00215> (2019).
72. Huang, Z., Huang, L., Gong, Y., Huang, C. & Wang, X. Mask scoring r-cnn. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6409–6418, <https://doi.org/10.1109/CVPR.2019.00655> (2019).
73. Singh, A. K., Ganapathysubramanian, B., Sarkar, S. & Singh, A. Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* **23**, 883–898. <https://doi.org/10.1016/j.tplants.2018.07.004> (2018).
74. Hasan, M. M., Chopin, J. P., Laga, H. & Miklavcic, S. J. Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* **14**, 1–13. <https://doi.org/10.1186/s13007-018-0366-8> (2018).
75. Narisetti, N., Neumann, K., Röder, M. S. & Gladilin, E. Automated spike detection in diverse european wheat plants using textural features and the frangi filter in 2d greenhouse images. *Front. Plant Sci.* **11**, 666. <https://doi.org/10.3389/fpls.2020.00666> (2020).
76. Gong, B., Ergu, D., Cai, Y. & Ma, B. A method for wheat head detection based on yolov4. *Res. Square* <https://doi.org/10.21203/rs.3.rs-100445/v1> (2020).
77. Zhan, J. Wheat ear detection based on one-stage object detector, <https://doi.org/10.1117/12.2614597> (2023).
78. Gong, B., Ergu, D., Cai, Y. & Ma, B. Real-time detection for wheat head applying deep neural network. *Sensors* **21**, 191. <https://doi.org/10.3390/s21010191> (2020).
79. Xu, X. et al. An automatic wheat ear counting model based on the minimum area intersection ratio algorithm and transfer learning. *Measurement* **216**, 112849. <https://doi.org/10.1016/j.measurement.2023.112849> (2023).
80. Qian, G. Wheatnet-cs: A wheat ear detection algorithm for complex background, https://doi.org/10.1007/978-3-031-26662-3_32 (2023).
81. Li, Z. et al. Real-time detection and counting of wheat ears based on improved yolov7. *Comput. Electron. Agric.* **218**, 108670. <https://doi.org/10.1016/j.compag.2024.108670> (2024).
82. Li, R., Meng, J., Wu, Y., Zhang, D. & He, Y. Wheat ear detection based on faster-rcnn-yolov8s algorithm. *N. Z. J. Crop. Hortic. Sci.* <https://doi.org/10.1080/01140671.2024.2134567> (2024).
83. Wen, C. et al. A generalized model for accurate wheat spike detection and counting in complex scenarios. *Sci. Rep.* **14**, 24189. <https://doi.org/10.1038/s41598-024-75523-w> (2024).
84. Khaki, S., Safaei, N., Pham, H. & Wang, L. Wheatnet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing* **474**, 78–89. <https://doi.org/10.1016/j.neucom.2022.03.017> (2022).
85. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520, <https://doi.org/10.1109/CVPR.2018.00474> (2018).
86. Dandrifosse, S. et al. Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity. *Comput. Electron. Agric.* **199**, 107161. <https://doi.org/10.1016/j.compag.2022.107161> (2022).
87. Solovyev, R., Wang, W. & Gabruseva, T. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image Vis. Comput.* **107**, 104117. <https://doi.org/10.1016/j.imavis.2021.104117> (2021).
88. Birodkar, V., Lu, Z., Li, S., Rathod, V. & Huang, J. The surprising impact of mask-head architecture on novel class segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7015–7025, <https://doi.org/10.1109/ICCV48922.2021.00701> (2021).

89. Wen, C. et al. Wheat spike detection and counting in the field based on spikeretinanet. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2022.821717> (2022).
90. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988, <https://doi.org/10.1109/ICCV.2017.00332> (2017).
91. Wang, Y., Qin, Y. & Cui, J. Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* **12**, 1139. <https://doi.org/10.3389/fpls.2021.645899> (2021).
92. Tan, M., Pang, R. & Le, Q. V. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10781–10790, <https://doi.org/10.1109/CVPR42600.2020.01078> (2020).
93. Guan, Y. et al. Ctwheatnet: Accurate detection model of wheat ears in field. *Comput. Electron. Agric.* **225**, 109272. <https://doi.org/10.1016/j.compag.2024.109272> (2024).
94. Kuznetsova, A. et al. The open images dataset v4. *Int. J. Comput. Vision* **128**, 1956–1981. <https://doi.org/10.1007/s11263-020-01316-z> (2020).
95. Lin, T.-Y. et al. Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755, https://doi.org/10.1007/978-3-319-10602-1_48 (2014).
96. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255, <https://doi.org/10.1109/CVPR.2009.5206848> (IEEE, 2009).
97. De Almeida, P. R., Oliveira, L. S., Britto, A. S. Jr., Silva, E. J. Jr. & Koerich, A. L. Pklot-a robust dataset for parking lot classification. *Expert Syst. Appl.* **42**, 4937–4949. <https://doi.org/10.1016/j.eswa.2015.02.009> (2015).
98. Hsieh, M.-R., Lin, Y.-L. & Hsu, W. H. Drone-based object counting by spatially regularized regional proposal network. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 4145–4153, <https://doi.org/10.1109/ICCV.2017.446> (2017).
99. Goldman, E., Herzig, R., Eisenschat, A., Goldberger, J. & Hassner, T. Precise detection in densely packed scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5227–5236, <https://doi.org/10.1109/CVPR.2019.00536> (2019).
100. Sa, I. et al. Weedmap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sens.* **10**, 1423. <https://doi.org/10.3390/rs10091423> (2018).
101. Teimouri, N. et al. Weed growth stage estimator using deep convolutional neural networks. *Sensors* **18**, 1580. <https://doi.org/10.3390/s18051580> (2018).
102. Guo, W. et al. Aerial imagery analysis-quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Front. Plant Sci.* **9**, 1544. <https://doi.org/10.3389/fpls.2018.01544> (2018).
103. Ghosal, S. et al. A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* <https://doi.org/10.34133/2019/1525874> (2019).
104. Xiong, H. et al. Tasselnetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* **15**, 1–14. <https://doi.org/10.1186/s13007-019-0160-2> (2019).
105. Global wheat detection competition leaderboard. <https://www.kaggle.com/competitions/global-wheat-detection/leaderboard> (2022). Accessed: December 14, 2022.
106. David, E., Ogidi, F., Guo, W., Baret, F. & Stavness, I. Global wheat challenge 2020: Analysis of the competition design and winning models. arXiv preprint [arXiv:2105.06182](https://arxiv.org/abs/2105.06182)<https://doi.org/10.48550/arXiv.2105.06182> (2021).
107. Genaev, M. A. et al. Morphometry of the wheat spike by analyzing 2d images. *Agronomy* **9**, 390. <https://doi.org/10.3390/agronomy9070390> (2019).
108. Brooks, J. Coco annotator. *GitHub Repository* <https://doi.org/10.1145/3578495.3578502> (2019). <https://github.com/jsbroks/coco-annotator>.
109. Brooks, J. COCO Annotator (2022). Accessed: October 28, 2022, <https://github.com/jsbroks/coco-annotator>.
110. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)<https://doi.org/10.48550/arXiv.1804.02767> (2018).
111. Monograph, B. Growth stages of mono- and dicotyledonous plants. *Federal Biol. Res. Centre Agric. For.* <https://doi.org/10.1002/9781119402442.biblio> (2001).
112. Apple Inc. Apple Support: Enabling Accessibility Features (2020). Accessed: February 27, 2025, <https://support.apple.com/en-us/111875>.
113. Joblove, G. H. & Greenberg, D. Color spaces for computer graphics. In *Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, 20–25, <https://doi.org/10.1145/965139.807362> (1978).
114. Di Stefano, L. & Bulgarelli, A. A simple and efficient connected components labeling algorithm. In *Proceedings 10th international conference on image analysis and processing*, 322–327, <https://doi.org/10.5555/839281.840794> (IEEE, 1999).
115. OpenCV. Connected Components (2022). Accessed: October 28, 2022, https://docs.opencv.org/3.4/d3/dc0/group__imgproc__shape.html.
116. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698. <https://doi.org/10.1109/TPAMI.1986.4767851> (1986).
117. Xu, Z., Baojie, X. & Guoxin, W. Canny edge detection based on open cv. In *2017 13th IEEE international conference on electronic measurement & instruments (ICEMI)*, 53–56, <https://doi.org/10.1109/ICEMI.2017.00029> (IEEE, 2017).
118. Hart, P. E. & Duda, R. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM* **15**, 11–15. <https://doi.org/10.1145/361237.361242> (1972).
119. Felzenszwalb, P. & Huttenlocher, D. Distance transforms of sampled functions. technical report tr2004–1963. *Cornell Comput. Inf. Sci.* <https://doi.org/10.4086/toc.2012.v008a019> (2004). Accessed: September 23, 2025.
120. Saha, A., Mishra, S. & Bovik, A. C. Re-iqu: Unsupervised learning for image quality assessment in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5846–5855, <https://doi.org/10.1109/CVPR45688.2023.00574> (2023).
121. Chen, H., Wang, Z., Yang, Y., Sun, Q. & Ma, K. Learning a deep color difference metric for photographic images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22242–22251, <https://doi.org/10.1109/CVPR45688.2023.02217> (2023).
122. Erlenbusch, F. et al. Thermal infrared single image dehazing and blind image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 459–469, <https://doi.org/10.1109/CVPRW59228.2023.00052> (IEEE, 2023).
123. Mittal, A., Moorthy, A. K. & Bovik, A. C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **21**, 4695–4708. <https://doi.org/10.1109/TIP.2012.2214050> (2012).
124. Mittal, A., Moorthy, A. K. & Bovik, A. C. Blind/referenceless image spatial quality evaluator. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 723–727, <https://doi.org/10.1109/ACSSC.2011.6190099> (IEEE, 2011).
125. Mittal, A., Soundararajan, R. & Bovik, A. C. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **20**, 209–212. <https://doi.org/10.1109/LSP.2012.2227726> (2012).
126. Venkatanath, N., Praneeth, D., Bh, M. C., Channappayya, S. S. & Medasani, S. S. Blind image quality evaluation using perception based features. In *2015 Twenty First National Conference on Communications (NCC)*, 1–6, <https://doi.org/10.1109/NCC.2015.7084843> (IEEE, 2015).

127. MathWorks. Image Quality Metrics (2022). Accessed: November 10, 2022, <https://www.mathworks.com/help/images/image-quality-metrics.html>.
128. Fernandez-Gallego, J. A. et al. Automatic wheat ear counting using thermal imagery. *Remote Sens.* **11**, 751. <https://doi.org/10.3390/rs11070751> (2019).
129. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick, R. Detectron2. <https://github.com/facebookresearch/detectron2> (2019).
130. Ultralytics. Yolo documentation (2023). Accessed: 2024-10-31, <https://docs.ultralytics.com/>.
131. Li, J. et al. Wheat ear recognition based on retinanet and transfer learning. *Sensors* **21**, 4845. <https://doi.org/10.3390/s21144845> (2021).
132. Ban, X., Liu, P., Xu, L. & Zhao, J. A lightweight model based on yolov8n in wheat spike detection. In *2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–6, <https://doi.org/10.1109/Agro-Geoinformatics55657.2023.1000012> (IEEE, 2023).
133. Mas, M. S., Saidah, S. & Ibrahim, N. Detection and counting of wheat ear using yolov8. *International Journal of Electrical & Computer Engineering (2088-8708)* **14**, <https://doi.org/10.11591/ijece.v14i5.pp5813-5823> (2024).
134. Zhao, Z., Feng, W., Zou, X. & Li, Q. Method for wheat ear detection based on csr-detr. In *2024 6th International Conference on Electronic Engineering and Informatics (EEI)*, 464–467, <https://doi.org/10.1109/EEI52411.2024.1000014> (IEEE, 2024).
135. Yang, Z., Yang, W., Yi, J. & Liu, R. Wh-detr: An efficient network architecture for wheat spike detection in complex backgrounds. *Agriculture* **14**, 961. <https://doi.org/10.3390/agriculture14060961> (2024).
136. Lee, D.-H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning, ICML* **3**, 896. <https://doi.org/10.1145/3042573.3042576> (2013).
137. Reza, F. M. *An introduction to information theory* (Courier Corporation, 1994).
138. Hajar, M., El Badaoui, M., Raad, A. & Bonnardot, F. Discrete random sampling: Theory and practice in machine monitoring. *Mech. Syst. Signal Process.* **123**, 386–402. <https://doi.org/10.1016/j.ymssp.2018.12.035> (2019).
139. Ye, J. et al. Wheatlfanet: in-field detection and counting of wheat heads with high-real-time global regression network. *Plant Methods* **19**, 103. <https://doi.org/10.1186/s13007-023-01079-x> (2023).
140. Shi, L. et al. Yolov5s-t: A lightweight small object detection method for wheat spikelet counting. *Agriculture* **13**, 872. <https://doi.org/10.3390/agriculture13040872> (2023).
141. Shen, X. et al. A lightweight network for improving wheat ears detection and counting based on yolov5s. *Front. Plant Sci.* **14**, 1289726. <https://doi.org/10.3389/fpls.2023.1289726> (2023).
142. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks?. *Adv. Neural. Inf. Process. Syst.* <https://doi.org/10.5555/2969033.2969197> (2014).
143. Shen, M. Rethinking data selection for supervised fine-tuning. arXiv preprint [arXiv:2402.06094](https://arxiv.org/abs/2402.06094)<https://doi.org/10.48550/arXiv.2402.06094> (2024).
144. Jocher, G., Chaurasia, A. & Qiu, J. Ultralytics YOLOv8 (2023). Version 8.0.0, License: AGPL-3.0.
145. Lv, W. et al. Detsr beat yolos on real-time object detection (2023). [arXiv:2304.08069](https://arxiv.org/abs/2304.08069).
146. Yu, J., Chen, W., Liu, N. & Fan, C. Oriented feature pyramid network for small and dense wheat heads detection and counting. *Sci. Rep.* **14**, 8106. <https://doi.org/10.1038/s41598-024-35829-1> (2024).
147. Gogtay, N. J. & Thattai, U. M. Principles of correlation analysis. *J. Assoc. Phys. India* **65**, 78–81. https://doi.org/10.1007/978-3-642-00296-0_5 (2017).
148. Cohen, I., Huang, Y., Chen, J. & Benesty, J. Pearson correlation coefficient. *Noise Reduction in Speech Processing* 1–4, https://doi.org/10.1007/978-3-642-00296-0_5 (2009).
149. Sedgewick, P. Spearman's rank correlation coefficient. *BMJ* <https://doi.org/10.1136/bmj.g7327> (2014).
150. Schaeffer, M. S. & Levitt, E. E. Concerning kendall's tau, a nonparametric correlation coefficient. *Psychol. Bull.* **53**, 338. <https://doi.org/10.1037/h0045013> (1956).
151. Zhang, J. et al. *Wheat-Net: An Automatic Dense Wheat Spike Segmentation Method Based on an Optimized Hybrid Task Cascade Model*, 87–106 (Springer, 2022). <https://doi.org/10.3389/fpls.2022.834938>.

Author contributions

Z.G. collected data and, with D.S. and M.P., conceived the experiment(s). Z.G. and D.S. did data processing and analysis. M.P., V.C. and T.L.T. validated the results. All authors analyzed the results. Z.G., M.P., and T.L.T. wrote the draft of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

This research was supported by the European Union's Horizon 2020 research and innovation programmes AN-TARES project (SGA-CSA. No. 739570 under FPA No. 664387, <https://doi.org/10.3030/739570>), the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-136/2025-03/200358) and the Provincial Secretariat for Higher Education and Scientific Research of the Autonomous Province of Vojvodina through the project 'Development of a decision support system for agricultural production using data fusion and artificial intelligence' (Grant No. 003074633 2024 09418 003 000 000 001 04 002).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ž.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025