End-to-End Deep Learning Models for Gap Identification in Maize Fields

Rana Waqar[1], Željana Grbović[1], Maryam Khan[2], Nina Pajević[1],
Dimitrije Stefanović[1], Vladan Filipović[1], Marko Panić[1], Nemanja Djuric[1]
[1]BioSense Institute, Novi Sad, Serbia [2]Farmevo Technologies, New York, USA
{rana.waqar, zeljana.grbovic}@biosense.rs, maryamkhan@student.ku.edu.pk, {nina.pajevic,
vladan.filipovic, dimitrije.stefanovic, panic}@biosense.rs, nemanja@temple.edu

Abstract

We propose an approach to jointly count plants and detect gaps in maize fields using end-to-end deep-learning models. Unlike previous efforts that focused solely on plant detection, our methodology also integrates the task of gap identification, offering a holistic view of the state of the agricultural field. Moreover, we consider different data sources in our experiments and explore the benefits of using multispectral over RGB images, which are commonly used in the industry. The findings suggest that multi-task learning on multispectral images significantly outperforms other model configurations, demonstrating the potential of the proposed approach for precision agriculture.

1. Introduction

Smart agriculture has emerged as a cutting-edge approach in modern agricultural practices, leveraging technological innovations to optimize crop production and resource management [21]. While traditional methods of manual field assessment are labor-intensive and prone to errors, recent advancements in computer vision and remote sensing technologies offer promising avenues for automating these tasks with unprecedented accuracy and efficiency. The usability and efficiency of such digital agricultural solutions rely on the use of unmanned aerial vehicles (UAVs) that are equipped with various imaging sensors, such as red–green–blue (RGB), thermal, hyperspectral, and multispectral cameras, as well as light detection and ranging (LIDAR) sensors. These sensors generate detailed, high-res data that provide precise insight into can be used for several important tasks, such as crop health monitoring, soil moisture estimation, or crop yield prediction [16]. Moreover, UAVs exhibit minimal turnaround time and field operational delays. For instance, a UAV can cover 50-100 acres per day, approximately 30 times more than a traditional knapsack sprayer, resulting in exceptional field efficiency.

While a common high-resolution RGB camera offers a useful visual depiction of crops, advanced image sensors like multispectral cameras provide deeper insights into biochemical properties of the plant's surface, related to chlorophyll content and plant nitrogen adoption [33]. Due to their capabilities, these sensors and associated technologies can be invaluable assets in precision agriculture, enabling more efficient and effective crop management strategies [17]. However, despite their potential, they are often not utilized in practice by farmers or the industry due to their high prices and knowledge gap in data interpretation, leading to a lack of trust among farmers in adopting such technologies. We address this issue and investigate the benefits of different data sources to increase confidence in this valuable technology, thus helping to facilitate its wider adoption in agricultural settings.

According to the Economic Research Service of the United States Department of Agriculture (USDA) [2], maize stands as one of the principal feed grains, alongside sorghum, oats, and barley, constituting over 95% of the total feed grain production and utilization. On average, an expansive 90 million acres of maize are planted annually in the United States. However, being a cornerstone of global food security, maize cultivation faces the challenge of achieving optimal plant density to maximize yield potential. The optimal density of plants per acre, known as plant population, stands as a critical determinant for maize yield within a designated area [43]. Achieving ideal density includes ensuring that each plant has sufficient area for growth, but also detecting and filling in excessive gaps between plants that can occur due to errors during the seeding process. Then, attaining the ideal plant count optimizes resource utilization and minimizes intra-plant competition, in addition to enhancing overall crop productivity [25]. Traditionally, evaluating plant density relies on manual tallies on a field section, a very labor-intensive process in extensive farming setups. Additionally, plant detection and gap identification are usually assessed separately, which is susceptible to human error during the sampling and counting stages. This is then further compounded by inaccuracies through

assuming uniformity across the entire field when aggregating information from the sampled segments [23, 25, 29, 43].

In this study we propose a novel multi-task deep architecture for plant and gap detection, utilizing and exploring the use of RGB and multispectral images collected by UAVs. Unlike existing methods that operate in two steps, first detecting plants and then computing the gaps given the plant detections [14], we propose a single-step, end-to-end approach where plants and gaps are jointly detected. Additionally, the data from this study will be made publicly available, providing a valuable resource for further research and development in the field of precision agriculture.

Here we give a summary of our paper. In Section 2 is given literature overview, while in Section 3 we provide a concise overview of the data utilized in this study, covering data acquisition, image registration, and labeling. Section 4 details our proposed methodology for the multi-task architecture, while Section 5 presents explanations and discussion of the experiments, along with the obtained results and qualitative analysis. Finally, in Section 6 we provide conclusions and future work.

2. Related Work

One of the areas where machine learning methods have yet to reach sufficient integration is agriculture, an application area that directly affects billions of people across the globe. In a literature review on machine learning algorithms used in agriculture [6], the authors presented how artificial intelligence deals with various crop management applications: weed and pest detection, plant disease detection, stress detection in plants, smart farms or automation in farms, and crop yield estimation and prediction [6]. Recent studies [4, 5] give a brief review of the latest deep learning techniques used in agriculture. Additionally, the study [30] provided an analysis of enhancing crop protection precision in the context of Agriculture 5.0 through the utilization of machine learning and emerging technologies. This advancement heralds a transformative phase in intelligent crop management, characterized by automated decision-making, unmanned operations, and reduced human involvement.

In these agricultural systems, the integration of UAV data is indispensable, owing to its efficiency and capacity to deliver high-resolution data with diverse applications. In the recent study [7] the usage of UAVs in phenotyping applications was explored, including the key parameters of UAV image acquisition. Furthermore, the study discussed the strengths and limitations of various UAV imaging sensors across different applications. For example, the authors argued that multispectral UAV data has a valuable role in modern agriculture, offering valuable insights into crop health and environmental conditions. In comparison, RGB data provides visual information limited to the red, green, and blue channels, offering a common representation of the crop's appearance without the additional insights provided by multispectral data. While RGB data is suitable for basic visual assessments, such as plant morphology or position, multispectral data enables more precise analysis, allowing farmers to detect subtle variations in plant health, identify stress factors, and optimize management practices for improved crop productivity. In [44] authors focused on comparing the benefits of RGB versus multispectral UAV data for mapping plant communities, while in [9] authors explore the utilization of UAV-based RGB and multispectral imagery for the evaluation of maize crop damage. In most of these studies, one-stage deep learning models take the lead for crop assessment or predictive tasks, such as for maturity estimation [15] or ground biomass estimation [48]. While previous studies predominantly rely on single-task deep learning models, often utilizing RGB imagery, in our work we explore both RGB and multispectral data. Furthermore, we introduce an end-to-end approach that simultaneously detects plants and gaps between them, thereby offering a comprehensive solution for agricultural monitoring tasks shown to be both efficient and effective.

In this study, we focus on RGB and multispectral data gathered by UAVs of the maize field, since maize is one of the most important food sources on a global scale, having an immense significance in global agricultural production [11]. It ranks among the most widely cultivated cereal crops worldwide and plays a critical role in food security and economic development by serving as a vital feedstock for livestock, contributing to the production of meat, dairy, and other animal products. Its versatility extends beyond food and feed, with maize also being utilized in industrial processes, such as the production of biofuels, starches, and other bio-based products [2]. Given its widespread cultivation and diverse uses, maize stands as a cornerstone of global agriculture, impacting economies, livelihoods, and food systems worldwide [2, 12]. Our approach contributes to the understanding and management of maize cultivation by offering a novel solution for plant count and gap detection in maize fields. By providing an efficient and accurate method for maize monitoring, the proposed approach can help enhance agricultural productivity and contribute to food security and economic development.

In the context of deep learning predictive models, techniques for plant counting explored in this study these methods have garnered significant attention in recent literature [22, 27, 32, 40–42, 46]. While previous studies have predominantly relied on RGB images and advanced object detection techniques, such as YoloV5 [19] used in [13, 28] or YoloV8 [47] used in [18], our proposed approach considers both RGB and multispectral UAV images. The limitations identified in the single-task deep learning models served as a basis for establishing a baseline for conducting a novel end-to-end model, which also focuses on gap identification

Figure 1. Bird's eye view of the experimental field

Table 1. Drone specifications used during data collect

Specification	Description
Model	Mavic 3 Multispectral
Flight height	20 meters
Imaging	RGB/ Multispectral
GSD for RGB images	0.54 cm
GSD for MS images	0.92 cm

in a multi-task setting in addition to plant detection. This approach lays down the groundwork for optimizing yield through gap-filling during the germination stage, reducing the dependence on the plant detection task. To the best of our knowledge, our study is the first to tackle these dual objectives using a single-stage multi-task learning approach.

3. Data

The study area encompasses a 5.95-acre maize field near Khanewal, Punjab, Pakistan, serving as the location for data acquisition; see Figure 1 for the bird's-eye view illustration of the field. Aerial imagery was collected in July 2023 using the Mavic 3 Multispectral UAV by DJI equipped with a multispectral sensor, adhering to a 60% side and 70% frontal overlap, as shown in Figure 2. The key flight and drone specifications are detailed in Table 1. To enable a unified multi-source and multi-task approach, our input data required both a standard RGB aerial image and a multispectral image consisting of near infrared (NIR) at 860 nm ± 26 nm, red at 650 nm ± 16 nm, red edge at 730 nm ± 16 nm, and green channel at 560 nm ± 16 nm.

Figure 2. Mavic 3 Multispectral drone in flight on the left, and flight parameters on the right

Table 2. Comparison of sensor specifications

Specification	RGB	Multispectral
True focal length	12.29 mm	4.34 mm
Image size (megapixels)	20 MP	5 MP
Sensor type	4/3" CMOS	1/2.8" CMOS
Sensor width	17.4 mm	5.2 mm
Sensor height	13 mm	3.9 mm
Equivalent focal length	24 mm	25 mm

3.1. Data Acquisition and Geometric Correction

For our data acquisition process we carefully selected an operational altitude of 20 meters. This strategic choice was made to achieve the optimal Ground Sample Distance (GSD), allowing us to detect gaps and count plants during the early stages of crop development. During the drone's operation, it captured both multispectral and RGB channels using separate imaging sensors. These sensors were specifically designed for their respective bands and had varying focal lengths, as detailed in Table 2. This setup ensured that we acquired comprehensive data covering a wide range of spectral information, as necessary for our analyses.

Following the data collection we utilized Agisoft Metashape [3] to process the acquired data and generate high-resolution orthomosaics for both RGB and multispectral images. The resulting GSD for RGB images was 0.54 cm, providing us with the capability for a thorough visual examination of the entire crop field. Meanwhile, the multispectral images achieved a GSD of 0.92 cm, offering us valuable insights into plant health and vitality that go beyond what is visible to the naked eye. As a result, the orthomosaics processed from RGB and multispectral (MS) images exhibit different GSDs. Furthermore, the Real-Time Kinematic Positioning (RTK) module captures the individual location of each camera's imaging center. Different mounting locations of the sensors, combined with the positional inaccuracies, lead to image offsets during post-processing of the RGB and MS data, as depicted in Figure 3.

The data set is categorized into two main groups: UAV-acquired 3-band RGB and multispectral imagery comprising near-infrared, green, and red bands. It includes a total of

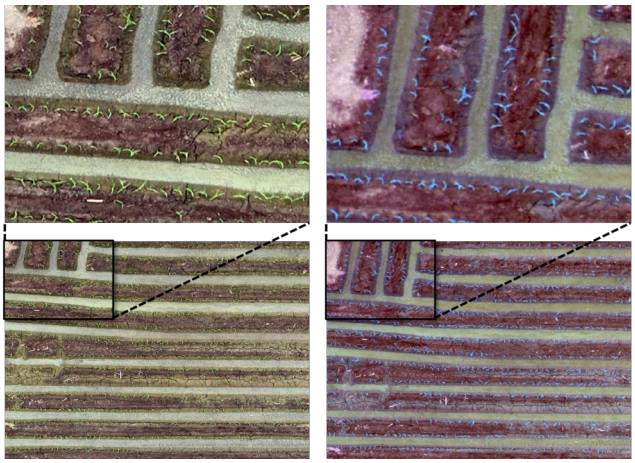


Figure 3. Image offsets between RGB (left) and MS (right)

Table 3. Data distribution for training and testing

Data	RGB	MS	Gap labels	Plant labels
Train	2,789	2,789	48,007	42,228
Test	678	678	13,566	12,641
Total	3,467	3,467	61,573	54,869

3,467 high-resolution RGB and multispectral images, each scene of size 430×371 pixels. The train/test split was set at 80/20, with 2,789 images allocated for training and 678 for testing across each data set as shown in Table 3. Notably, the number of gap labels is 61,573, of which 48,007 training and 13,566 test labels, while there are 54,869 plant labels, comprising 42,228 training and 12,641 test labels.

To ensure accurate comparative assessment, achieving pixel-to-pixel alignment between both imagery data was crucial. This alignment process involved resampling using the nearest neighbor method [1], where the multispectral data was adjusted to conform to the optimal image dimensions of the higher-resolution RGB data, specifically at 430×371 pixels. We interpolate the original image using the nearest neighbor resampling method. Additionally, in our research we share label annotation between RGB and multispectral data, thus using identical labels for various inputs. This is possible due to the alignment between the two data sources, which we discuss in more detail below.

As we aim to achieve the necessary alignment for both data sets, this approach necessitates optimized image registration to ensure valid comparative assessments and to provide a foundation for the future utilization of identical training data across multiple inputs. In this study, for image registration we use Oriented Features from Accelerated Segment Test (FAST) and Rotated Binary Robust Independent Elementary Features (BRIEF), introduced as ORB in [37]. It combines efficient keypoint detection and compact descriptor, tailored to respect keypoint orientation for

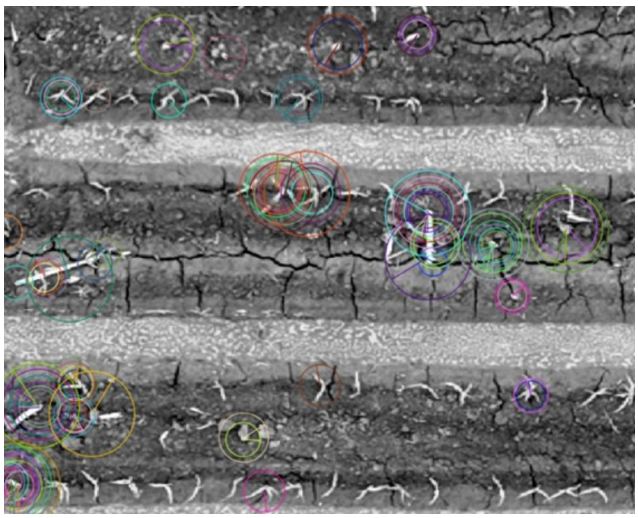


Figure 4. Keypoint identification

improved matching accuracy across various image orientations. The process begins with FAST detecting key points by evaluating a circular pixel arrangement around candidate pixels, ensuring feature significance through intensity contrast [36], as illustrated in Figure 4. Subsequently, ORB calculates each keypoint's orientation using intensity centroid, ensuring rotation invariance, and utilizes a rotated version of the BRIEF descriptor that adapts to the keypoint's orientation, capturing the local intensity pattern [8]. Keypoints across images are matched by performing a brute-force comparison of their binary descriptors. Random Sample Consensus (RANSAC) is then employed iteratively to refine the homography matrix, helping to accurately align images despite potential distortions or discrepancies caused by factors such as perspective changes, sensor noise, or environmental conditions. It selects match subsets and evaluates transformation accuracy through inlier assessment, which is crucial for precise image stitching [10].

3.2. Data Labeling

We used a semi-supervised approach to label the data, which involved the creation of bounding box annotations marked on multispectral imagery using AnyLabeling software [31]. In particular, 200 images were manually labeled and used to train a YoloV8 model, which was utilized to expedite the process by automatically labeling the remaining unlabeled data. Every label was then manually verified, and all labels found to be incorrect or misdirected were manually corrected by human annotators.

The distribution of bounding box labels for each class (i.e., maize plants and gaps) used in this semi-supervised approach is summarized in Table 3, with detailed data labels statistics further illustrated in Figure 5. As can be seen in the figure, maize plant bounding boxes tend to be very

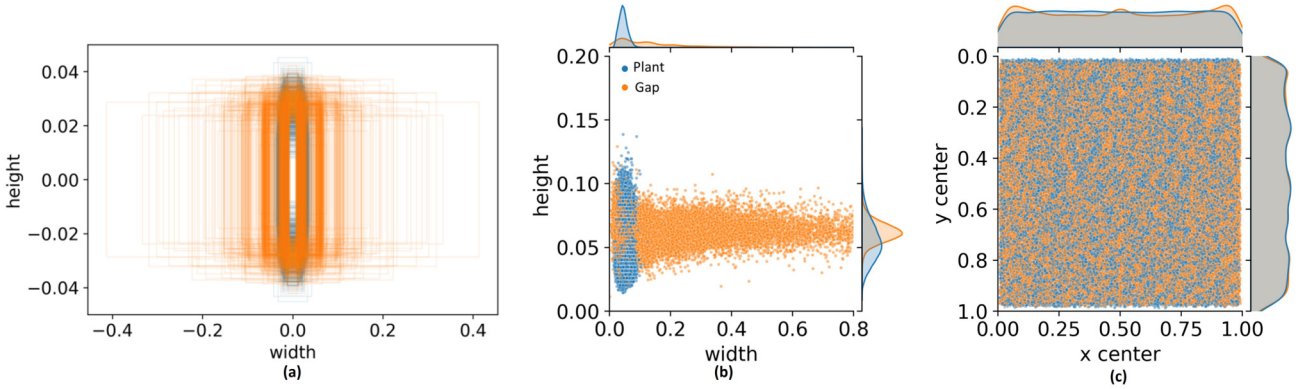


Figure 5. Label statistics showing the detailed distribution of positions and bounding box dimensions of training data: (a) shape of bounding boxes; (b) size of bounding boxes; (c) coordinates of bounding boxes

compact, while gaps exhibit larger and wider bounding box shapes (as observed in Figures 5a and 5b). Moreover, in Figure 5c we can see that neither of the classes is localized and both can instead be found in all parts of the field, with near-uniform distribution along both x- and y-axes.

4. Methodology

In this section we describe a novel approach of multi-task end-to-end architecture for joint detection of maize plants and gaps between them. In particular, we address the objectives of plant and gap detection concurrently, while optimizing a unified loss function. This approach ensures a holistic optimization of both tasks, as well as simplified training and inference processes.

In our multitask end-to-end architecture, each plant and gap instance is represented by a set of coordinates denoted as $\{x, y, w, h\}$, where x and y denote the x- and y- coordinates of the bounding box's top-left corner, respectively, w denotes the width and h denotes the height of the bounding box. These coordinates are used to precisely localize the position and describe the size of each plant and the gap within the input images.

When it comes to loss computation, the model distinguishes between maize plants and gaps by calculating separate losses for each task using their bounding box parameters. This tailored approach enables effective model training by optimizing parameters specifically for each detection task, thus improving performance and accuracy. We utilize the smooth L1 loss function for both tasks to quantify the disparity between predicted and ground truth bounding box parameters. In particular, we optimize the bounding box regression loss L_{reg}, computed as

$$L_{reg} = \sum_{r \in \{x,y,h,w\}} \text{smooth-L1}(r - r^*), \qquad (1)$$

where smooth-L1 is computed as

$$\text{smooth-L1}(r) = \begin{cases} 0.5r^2 & \text{if } |r| < 1, \\ |r| - 0.5 & \text{otherwise.} \end{cases} \qquad (2)$$

Here, r represents the predicted bounding box regression targets, while r^* denotes the corresponding regression ground truth. The regression loss is coupled with class-based cross-entropy loss with soft-max activation, and the combined per-pixel losses result in a unified framework that ensures precise localization and accurate classification of plants and gaps in input images.

As another important aspect of our methodology, we want to emphasize that we used both RGB and multispectral images as inputs to the model for which the above losses are minimized. The description of utilized data sources and image type is elaborated more extensively in Section 3.

5. Experiments

5.1. Experimental Setup

For the task encompassing plant count and gap identification, we evaluated the performance of two popular object detectors, namely Faster R-CNN [35] and YoloV8 [20]. Faster R-CNN has widespread adoption having demonstrated robust performance in agriculture applications [24, 26, 34, 38, 39], and we use its Detectron2 implementation [45]. Conversely, YoloV8 gained popularity due to good detection performance and efficient real-time inference, and we use its Ultralytics implementation [20]. We compare the detection frameworks using RGB and multispectral imagery data sets, each comprising three spectral bands.

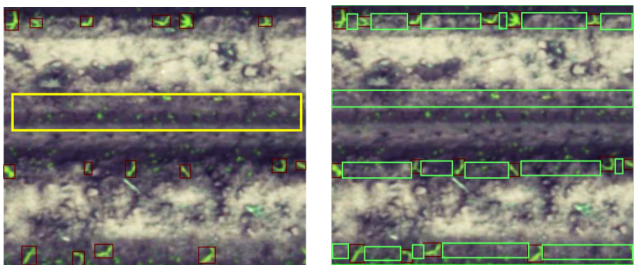
For the single tasks of plant count and gap identification, Faster-RCNN and YoloV8 were trained with batch sizes of 4 and 8 for 500 and 200 epochs, with learning rates of 0.05 and 0.01, respectively, with a weight decay of 0.0005 for

both data sets. For multi-task, Faster R-CNN underwent training for 1000 epochs with a batch size of 4, a learning rate of 0.0005, and a weight decay of 0.0005. This model exhibited divergence during training when higher learning rates were applied, leading us to opt for a lower learning rate. However, YoloV8 was trained for 310 epochs as there was no significant improvement observed during the final 50 epochs, prompting us to stop training at 310 epochs. The training procedure employed a batch size of 8, a learning rate of 0.01, and a weight decay of 0.0005. Despite employing a larger batch size, YoloV8 demonstrated faster training, which led us to use a batch size of 8.

Both models were applied to both RGB and multispectral data sets to perform the tasks of plant count and gap detection. This comprehensive evaluation enabled us to assess the efficacy of Faster R-CNN and YoloV8 across varied spectral representations, providing insights into their applicability in real-world scenarios. The model performance was evaluated using a range of commonly utilized metrics, including precision, recall, and F1-score. To assess the model performance utilizing inputs of RGB and MS for the single and multi-task detection of plants and gaps in maize fields, the precision and recall scores were calculated for each model across every class using a specified threshold value, obtained based on the highest F1-score value over a complete threshold range. An optimized threshold range from 0.3 to 0.55 was obtained, varying depending on the task, data source, and model. Additionally, cumulative metrics of F1-score, precision, and recall were computed based on average PR curves of both classes (denoted respectively as Cum.F1, Cum.P, and Cum.R).

5.2. Two-Stage Baseline for Gap Identification

The baseline approach utilizes an object detection model to detect plants in images by outlining them with bounding boxes. Once the plants are detected, their spatial arrangement is analyzed to identify gaps between them. This is achieved by examining the distances between adjacent plants and applying criteria such as minimum distance thresholds. In an earlier study [14] the plant rows were defined by connecting centroids of identified plants, ensuring connections within a typical row gap distance of around $75cm$. These connections formed polylines, which were extended to intersect neighboring polylines, marking the inter-row spacing. These row lines also facilitated grouping plants within each row and identifying gaps between them. Gaps exceeding the expected spacing between plants (approximately $45cm$) were targeted for reseeding during germination to optimize yield. In this procedure, if a row lacks any plants it becomes impossible to detect a gap since the gap identification relies on the presence of plants within the row, as illustrated by the yellow box in Figure 6. In contrast, our novel approach proposes to simultaneously identify both the maize plants and the gaps between them. This concurrent detection streamlines the gap identification process, enhancing overall efficiency. It reduces the risk of overlooking gaps, resulting in higher accuracy in gap detection and overall field analysis. Additionally, this approach is adaptable to various agricultural settings and crop types, offering flexibility in its application across different scenarios and environments.



(a) Maize plant detection (b) Jointly detecting maize and gaps

Figure 6. Variations in detection results; yellow box in (a) indicates missed gap due to absence of any plant, while green boxes in (b) indicate gap detection through multi-task deep-learned model

Gap detection in maize fields was conducted using the Faster R-CNN model, aiming to improve upon previous methods that relied solely on precise plant counting. Initially, the Faster R-CNN model was employed for maize plant detection [14]. The study involved classifying maize and non-maize plant objects using a data set of a total of 3467 images. The evaluation showed promising performance metrics, with a plant detection precision of 96%. However, instances of mis-detected gaps were observed, mainly due to inaccuracies in plant detection, highlighting the model's reliance on accurate plant identification for effective gap delineation.

5.3. Results

This section presents the findings and performance evaluation of the proposed methodology in plant and gap detection. The results are organized to highlight key findings, trends, and comparisons with different approaches, providing insights into the efficacy and potential applications. In Table 4 we provide the results of Faster RCNN and YoloV8 models related to plant density and gap identification using RGB and MS data sets. In our experiments, we evaluated the performance of both models for individual tasks, as well as conducted a comparison between single-task and multi-task approaches on both data sets.

In our first experiment, detecting maize plants using the RGB data set (see Table 4, named Plants Only), Faster-RCNN achieved a precision of 95% and a similar recall of 95.3%, while YoloV8 attained a precision of 95% and a recall of 94%. Slightly better, in the MS data set Faster-RCNN demonstrated a precision of 96% and 96.7% re-

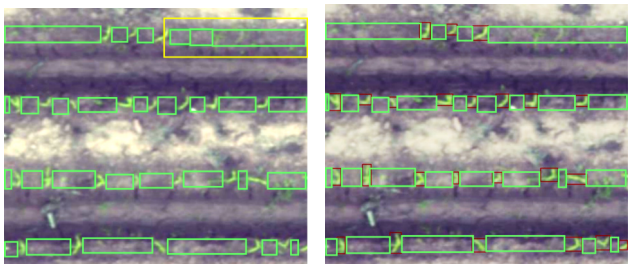
Table 4. Evaluation summary of single- and multi-task performance on plant counting and gap detection

Task	Data set	Models	Class	Prec.	Recall	Cum.F1	Cum.P	Cum.R
Plant Only	RGB	Faster-RCNN	Maize Plant	0.950	0.953	0.951	0.950	0.953
		YoloV8		0.950	0.940	0.946	0.951	0.941
	MS	Faster-RCNN		0.960	0.967	0.963	0.960	0.967
		YoloV8		0.961	0.951	0.956	0.961	0.951
Gaps Only	RGB	Faster-RCNN	Gaps	0.943	0.941	0.942	0.943	0.941
		YoloV8		0.933	0.948	0.940	0.933	0.949
	MS	Faster-RCNN		0.963	0.954	0.958	0.963	0.954
		YoloV8		0.960	0.947	0.954	0.96	0.947
Plant + Gaps	RGB	Faster-RCNN	Maize Plant	0.963	0.951	0.948	0.949	0.946
			Gaps	0.937	0.942			
		YoloV8	Maize Plant	0.948	0.950	0.941	0.945	0.937
			Gaps	0.942	0.925			
	MS	Faster-RCNN	Maize Plant	0.975	0.966	0.965	0.970	0.960
			Gaps	0.965	0.955			
		YoloV8	Maize Plant	0.955	0.969	0.956	0.956	0.955
			Gaps	0.957	0.942			

call, whereas YoloV8 with 95.1% recall reached a precision of 96.1%. Notably, both Faster R-CNN and YoloV8 performed slightly better across all evaluation metrics using the MS data set.

However, when focusing on gap detection, our second experiment (see Table 4, named as Gaps Only) evaluated model performance for detecting gaps irrespective of plant presence, which was previously considered crucial for achieving satisfactory precision [14]. In the RGB data set, Faster-RCNN achieved a precision of 94.3% and a recall of 94.1%, while YoloV8 reached a precision of 93.3% and a recall of 94.8%. In the MS data set, Faster-RCNN excelled with a precision of 96.3% and a recall of 95.4% while YoloV8 attained a precision of 96% and a recall of 94.7%. Significantly, both models demonstrated considerable improvements, notably surpassing the performance of the multispectral data set when compared to the RGB data set by approximately 3% across all metrics. Considering the richer information offered by multispectral data compared to RGB, it showed superior performance in both plant count and gap detection tasks.

In the third experiment (see Table 4, named Plants + Gaps), we analyzed the multi-task approach to plant and gap detection, and the performance of Faster-RCNN and YoloV8 for RGB and MS data. Using a multi-task approach to plant and gap detection with RGB data, the F1-score of Faster-RCNN and YoloV8 was 0.948 and 0.941, comparatively lower than their performance using the MS data that yielded an F1-score of 0.965 and 0.956, respectively. The assessment revealed promising results, particularly with Faster RCNN achieving a precision of 97.5% for plant and 96.5% for gap detection on multispectral data, challenging previous assumptions regarding the dependency of gap identification on accurate plant detections.



(a) Gaps Only (b) Plant & Gaps

Figure 7. Comparison of single-task baseline versus multi-task approach for gap detection; the yellow rectangle indicates the area of false positive gap detections

5.4. Qualitative Analysis

Across all tasks, the RGB data consistently displayed a higher number of false positives (FPs) compared to the multispectral images, indicating a tendency to misclassify non-plant or non-gap features as plants or gaps. For instance, in RGB-based multiclass identification 860 FPs were recorded compared to 459 in the multispectral data for the same task. Furthermore, the evaluation metrics displayed better results in multispectral data compared to RGB (see Table 4, Plant

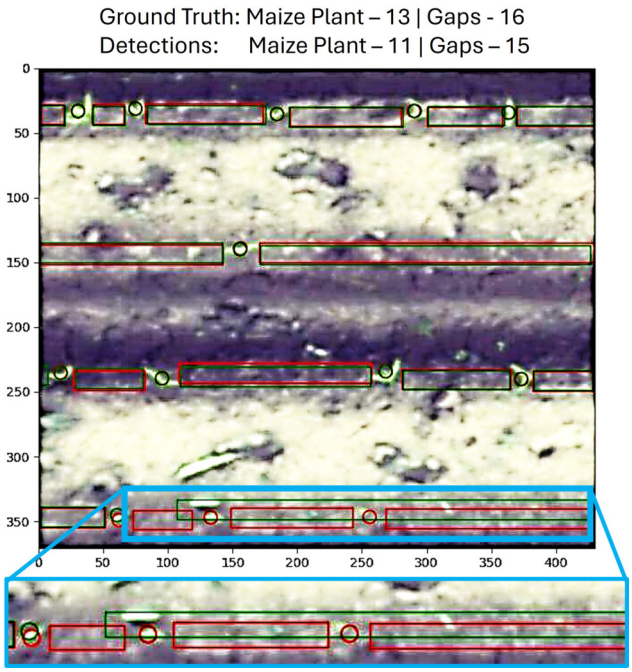


Figure 8. FP detections of plants and gaps; red box indicates ground truth data and green corresponds to detected objects

+ Gaps) with a cumulative precision of 0.970 for multispectral, notably higher than 0.949 for RGB, when utilizing Faster-RCNN for multi-task detection. These findings underscore the benefit of multispectral over the RGB data.

In addition, the tendency of mis-detection or false positives was observed in single task approach of gap detection. Highlighted in Figure 7, the yellow area indicates the false positive gaps, which represent the instances where the single-task approach for gap detection incorrectly marked overlapping gaps. This failure is however not observed with multi-task approach, where the model could better outline the gaps along with plant identification. The multi-task identification of plants and gaps not only decreased the incidence of false positives but also yielded enhanced cumulative precision from 96% to 97% on multispectral data (as shown in Table 4).

The concurrent application of end-to-end deep learning models for gap identification on multispectral data through a multi-task approach revealed certain limitations, besides promising results. As shown in Figure 8, the discrepancies between the detected objects and the ground truth data are evident. The highlighted instances, where the model overlooked the presence of plants and gaps, resulting in false positive detections, incorrectly elongating gaps across multiple instances. The occurrence of false positive gaps and plants in the multi-task detection underscores the shortcomings of the present approach, emphasizing the necessity for further refinement of the models to rectify overlooked detections. These observed limitations indicate the scope for potential improvement in methodology.

6. Conclusion

In this study we presented a detailed evaluation of the performance of different models across various tasks and data sets, related to plant and gap detection in maize fields. This analysis offers valuable insights into the models' effectiveness and suitability for early problem detection and mitigation in agricultural applications. Notably, our findings highlight the potential of multi-task learning, demonstrating the models' ability to identify multiple classes within single instances with high precision and recall. Moreover, our study emphasizes the promising results of using the multispectral data set over RGB, yielding better performances despite its lower resolution. This underscores the importance of considering spectral information in agricultural applications.

7. Acknowledgements

This research was funded by the ANTARES project under the European Union's Horizon 2020 research and innovation program (SGA-CSA no. 739570 under FPA no. 664387).

References

[1] Gdalgridnearestneighboroptions struct reference. https://gdal.org/doxygen/gdal__alg_8h.html. Accessed: 2024-03-19. 4

[2] Feed grains sector at a glance. https://www.ers.usda.gov/topics/crops/corn-and-other-feed-grains/feed-grains-sector-at-a-glance/. Online; Last updated: 21-December-2023, accessed 21-March-2023. 1, 2

[3] AgiSoft LLC. Agisoft metashape, 2024. Version 2.1.0. 3

[4] Yaganteeswarudu Akkem, Saroj Kumar Biswas, and Aruna Varanasi. Smart farming using artificial intelligence: A review. *Engineering Applications of Artificial Intelligence*, 120:105899, 2023. 2

[5] Ishana Attri, Lalit Kumar Awasthi, Teek Parval Sharma, and Priyanka Rathee. A review of deep learning techniques used in agriculture. *Ecological Informatics*, page 102217, 2023. 2

[6] Ishana Attri, Lalit Kumar Awasthi, and Teek Parval Sharma. Machine learning in agriculture: a review of crop management applications. *Multimedia Tools and Applications*, 83 (5):12875–12915, 2024. 2

[7] Ocident Bongomin, Jimmy Lamo, Joshua Mugeziaubwa Guina, Collins Okello, Gilbert Gilibrays Ocen, Morish Obura, Simon Alibu, Cynthia Awuor Owino, Agnes Akwero, and Samson Ojok. Uav image acquisition and processing for high-throughput phenotyping in agricultural research and breeding programs. *The Plant Phenome Journal*, 7(1): e20096, 2024. 2

[8] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *2010 European Conference on Computer Vision*, pages 778–792. Springer, 2010. 4

[9] Barbara Dobosz, Dariusz Gozdowski, Jerzy Koronczok, Jan Žukovskis, and Elżbieta Wójcik-Gront. Evaluation of maize crop damage using uav-based rgb and multispectral imagery. *Agriculture*, 13(8):1627, 2023. 2

[10] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 4

[11] FAO Food and Agriculture Organization of United States. Agricultural production statistics. 2000–2021. faostat analytical brief series no. 60. https://doi.org/10.4060/cc3751en, 2022. 2

[12] FAO Food and Agriculture Organization of United States. Statistical yearbook, world food and security. https://www.fao.org/3/cb4477en/cb4477en.pdf, 2022. 2

[13] Hailiang Gong, Xi Wang, and Weidong Zhuang. Research on real-time detection of maize seedling navigation line based on improved yolov5s lightweighting technology. *Agriculture*, 14(1):124, 2024. 2

[14] Ž. Grbović, R. Waqar, N. Pajević, D. Stefanović, V. Filipović, and M. Panić. Closing the gaps in crop management: Uav-guided approach for detecting and estimating targeted gaps among corn plants. 8th workshop on Computer Vision in Plant Phenotyping and Agriculture at IEEE/ICCV International Conference on Computer Vision, 2023. 2, 6, 7

[15] Jingyu Hu, Hao Feng, Qilei Wang, Jianing Shen, Jian Wang, Yang Liu, Haikuan Feng, Hao Yang, Wei Guo, Hongbo Qiao, et al. Pretrained deep learning networks and multispectral imagery enhance maize lcc, fvc, and maturity estimation. *Remote Sensing*, 16(5):784, 2024. 2

[16] Bojana Ivošević, Marko Kostić, Nataša Ljubičić, and Grbović. Application of unmanned aerial systems to address real-world issues in precision agriculture. In *Unmanned Aerial Systems in Agriculture*, pages 51–69. Elsevier, 2023. 1

[17] Bojana Ivošević, Marko Kostić, Nataša Ljubičić, Željana Grbović, and Marko Panić. A drone view for agriculture. In *Unmanned Aerial Systems in Agriculture*, pages 25–47. Elsevier, 2023. 1

[18] Zhijie Jia, Xinglong Zhang, Hongye Yang, Yuan Lu, Jiale Liu, Xun Yu, Dayun Feng, Kexing Gao, Jianfu Xue, Bo Ming, et al. Comparison of drone-assisted maize seedling detection models based on deep learning. 2024. 2

[19] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, Yonghye Kwon, Jiacong Fang, Kalen Michael, Diego Montes, Jebastin Nadar, Piotr Skalski, et al. ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference. *Zenodo*, 2022. 2

[20] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. 5

[21] EMBM Karunathilake, Anh Tuan Le, Seong Heo, Yong Suk Chung, and Sheikh Mansoor. The path to smart farming: Innovations and opportunities in precision agriculture. *Agriculture*, 13(8):1593, 2023. 1

[22] Bruno T. Kitano, Caio C. T. Mendes, André R. Geus, Henrique C. Oliveira, and Jefferson R. Souza. Corn plant counting using deep learning and uav images. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5, 2019. 2

[23] Adrian A Koller, Jorge A Rascon, Eric Lam, Max Metcalf, Austin Childers, Bill Raun, Randy Taylor, and Paul Weckler. Design and development of a stick planter for the developing world. In *2012 Dallas, Texas, July 29-August 1, 2012*, page 1. American Society of Agricultural and Biological Engineers, 2012. 2

[24] Xiaohong Kong, Xinjian Li, Xinxin Zhu, Ziman Guo, and Linpeng Zeng. Detection model based on improved faster-rcnn in apple orchard environment. *Intelligent Systems with Applications*, 21:200325, 2024. 5

[25] Marko Kostić, Dušan Rakić, Dragi Radomirović, Lazar Savin, Nebojša Dedović, Vladimir Crnojević, and Nataša Ljubičić. Corn seeding process fault cause analysis based on a theoretical and experimental approach. *Computers and Electronics in Agriculture*, 151:207–218, 2018. 1, 2

[26] Yuanhong Li, Chaofeng Wang, Congyue Wang, Xiaoling Deng, Zuoxi Zhao, Shengde Chen, and Yubin Lan. Detection of the foreign object positions in agricultural soils using mask-rcnn. *International Journal of Agricultural and Biological Engineering*, 16(1):220–231, 2023. 5

[27] Minguo Liu, Wen-Hao Su, and Xi-Qing Wang. Quantitative evaluation of maize emergence using uav imagery and deep learning. *Remote Sensing*, 15(8), 2023. 2

[28] Chenghao Lu, Emmanuel Nnadozie, Moritz Paul Camenzind, Yuncai Hu, and Kang Yu. Maize plant detection using uav-based rgb imaging and yolov5. *Frontiers in Plant Science*, 14:1274813, 2024. 2

[29] Angie Rieck-Hinz Meaghan Anderson, Rebecca Vittetoe and Mark Licht. Stand assessments - corn. https://crops.extension.iastate.edu/encyclopedia/stand-assessments-corn. Online; Published: 05-February-2023, accessed: 07-July-2023. 2

[30] Gustavo A Mesías-Ruiz, María Pérez-Ortiz, José Dorado, Ana I De Castro, and José M Peña. Boosting precision crop protection towards agriculture 5.0 via machine learning and emerging technologies: A contextual review. *Frontiers in Plant Science*, 14:1143326, 2023. 2

[31] Viet Anh Nguyen. AnyLabeling - Effortless data labeling with AI support. 4

[32] Yan Pang, Yeyin Shi, Shancheng Gao, Feng Jiang, Arun-Narenthiran Veeranampalayam-Sivakumar, Laura Thompson, Joe Luck, and Chao Liu. Improved crop row detection with deep neural network for early-season maize stand count in uav imagery. *Computers and Electronics in Agriculture*, 178:105766, 2020. 2

[33] Prajakta Patane and Anup Vibhute. Chlorophyll and nitrogen estimation techniques: A review. *International Journal of Engineering Research and Reviews*, 2(4):33–41, 2014. 1

[34] Nitin Rane. Yolo and faster r-cnn object detection for smart industry 4.0 and industry 5.0: applications, challenges, and opportunities. *Available at SSRN 4624206*, 2023. 5

[35] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 5

[36] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*, pages 430–443. Springer, 2006. 4

[37] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International Conference on Computer Vision*, pages 2564–2571. IEEE, 2011. 4

[38] Toqi Tahamid Sarker, Taminul Islam, and Khaled R Ahmed. Cannabis seed variant detection using faster r-cnn. *arXiv preprint arXiv:2403.10722*, 2024. 5

[39] Yi-Shiang Shiu, Re-Yang Lee, and Yen-Ching Chang. Pineapples' detection and segmentation based on faster and mask r-cnn in uav imagery. *Remote Sensing*, 15(3):814, 2023. 5

[40] Wilbur N. Chiuyari Veramendi and Paulo E. Cruvinel. Method for maize plants counting and crop evaluation based on multispectral images analysis. *Computers and Electronics in Agriculture*, 216:108470, 2024. 2

[41] Chin Nee Vong, Lance S. Conway, Jianfeng Zhou, Newell R. Kitchen, and Kenneth A. Sudduth. Early corn stand count of different cropping systems using uav-imagery and deep learning. *Computers and Electronics in Agriculture*, 186: 106214, 2021.

[42] Biwen Wang, Jing Zhou, Martin Costa, Shawn M. Kaeppler, and Zhou Zhang. Plot-level maize early stage stand counting and spacing detection using advanced deep learning algorithms based on uav imagery. *Agronomy*, 13(7), 2023. 2

[43] RW Willey and DSO Osiru. Studies on mixtures of maize and beans (phaseolus vulgaris) with particular reference to plant population. *The Journal of Agricultural Science*, 79 (3):517–529, 1972. 1, 2

[44] Franziska Wolff, Tiina HM Kolari, Miguel Villoslada, Teemu Tahvanainen, Pasi Korpelainen, Pedro AP Zamboni, and Timo Kumpula. Rgb vs. multispectral imagery: Mapping aapa mire plant communities with uavs. *Ecological Indicators*, 148:110140, 2023. 2

[45] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019. 5

[46] Juan Xiao, Stanley Anak Suab, Xinyu Chen, Chander Kumar Singh, Dharmendra Singh, Ashwani Kumar Aggarwal, Alexius Korom, Wirastuti Widyatmanti, Tanjinul Hoque Mollah, Huynh Vuong Thu Minh, Khaled Mohamed Khedher, and Ram Avtar. Enhancing assessment of corn growth performance using unmanned aerial vehicles (uavs) and deep learning. *Measurement*, 214:112764, 2023. 2

[47] Guoliang Yang, Jixiang Wang, Ziling Nie, Hao Yang, and Shuaiying Yu. A lightweight yolov8 tomato detection algorithm combining feature enhancement and attention. *Agronomy*, 13(7):1824, 2023. 2

[48] Danyang Yu, Yuanyuan Zha, Zhigang Sun, Jing Li, Xiuliang Jin, Wanxue Zhu, Jiang Bian, Li Ma, Yijian Zeng, and Zhongbo Su. Deep convolutional neural networks for estimating maize above-ground biomass using multi-source uav images: A comparison with traditional machine learning algorithms. *Precision Agriculture*, 24(1):92–113, 2023. 2