Smart Agricultural Technology 3 (2023) 100129

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Estimation of spinach (*Spinacia oleracea*) seed yield with 2D UAV data and deep learning

Mar Ariza-Sentís [a,*], João Valente [b], Lammert Kooistra [a], Henk Kramer [c], Sander Mücher [c]

[a] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, the Netherlands
[b] Information Technology Group, Wageningen University & Research, the Netherlands
[c] Wageningen Environmental Research, Wageningen University & Research, the Netherlands

ARTICLE INFO

Keywords:
Precision agriculture
Seed yield estimation
Deep learning
Mask R-CNN

ABSTRACT

Precision agriculture has drawn much attention in the last few years because of the benefits it has on reducing farming costs while maximizing the harvest obtained. Yield prediction is of importance for farmers to fertilize accordingly to reach the potential yield. However, this task is still relying on manual work, which is expensive and time-consuming. Instance segmentation has been implemented in the last years for fruit detection and yield estimation, obtaining state-of-the-art metrics, and reducing the labor required. This research presents a novel approach for spinach seed yield estimation for seed production purposes, that consists of correlating the number of plants and two phenotyping variables (plant area and canopy cover percentage) with the number of harvested seeds and the thousand seed weight. Mask R-CNN is applied to count the number of detections of spinach plants and obtain the object mask from which the plant area is derived. The results show that there is a high linear correlation between a multivariate linear mixed model of the three variables and the number of seeds, with an R^2_{adj} of 0.80. Furthermore, 77.42% of the variation in the weight of thousand seeds can be explained by the number of plants. For future studies, the algorithm should be trained with more spinach images from different locations and under varying weather conditions to allow it to generalize for the crop worldwide. It can be concluded, until further research, that Mask R-CNN can be applied for spinach counting and the computation of its individual plant area, with promising results.

1. Introduction

Human iron deficiency has been and still is, one of the leading contributors to disability and death worldwide [5], causing extreme fatigue and lightheadedness [20]. It is of importance to investigate the nourishments which contain iron and provide them to the population facing iron deficiency anemia (IDA). Spinach is a good source of iron for humans, even if they are not considered “high in” iron, since they do not reach the required amount of 4.2 mg/100 g [38]. Their content of vitamin C boosts iron absorption [27], turning spinach into a true source of iron [19]. Spinach cultivation represents a global gross production value of 18 billion USD, placing it in the 33rd world position for cultivated crops, being mainland China the country with the highest gross production [39].

Predicting the yield of any crop cultivation is relevant, but its usefulness increases with vegetables and fruits since they are perishable, degrade quickly, and cannot be stored for a long period. The importance of predicting spinach yield can be seen from two different sides, the farmer and the freezing industry. Observed by the farmer side, if the yield in the middle of the campaign is predicted, fertilization can be applied more accurately to reach the potential yield and adjust the fertilization to the crop stage and nitrogen status, as the guidelines of Precision Agriculture settle [6]. Regarding the freezing industry’s point of view, the industry knows the maximum amount of tons they can handle per day. Hence, the farmers should only provide that amount. In the case of overproduction and that the spinach needs to be frozen that same day (because of its perishable nature), the farmer should find a second industry source for processing [26]. If the final purpose of spinach cultivation is seed production, as in this paper, it is important to predict the seed yield (seen for instance as the number of seeds produced per plant) to understand the behavior of each spinach variety and continue breeding the varieties until the highest seed yield is obtained.

Nowadays, agriculture has to face many food production challenges to be able to feed the increasing world population. Among them, the

* Corresponding author.
E-mail address: mar.arizasentis@wur.nl (M. Ariza-Sentís).

https://doi.org/10.1016/j.atech.2022.100129
Received 14 September 2022; Received in revised form 7 October 2022; Accepted 8 October 2022
Available online 9 October 2022

most important concerns sustainable food production, while reducing greenhouse emissions, conserving groundwater sources, and meeting the Millennium Development Goal of eradicating hunger [10]. With the boost of Machine Learning (ML) and especially Deep Learning (DL), the most time-consuming field tasks (such as manually counting plants) have been able to be easily carried out, which has drawn the attention of many researchers. The developed approaches range from basic Deep Learning algorithms to models with state-of-the-art metrics, such as Mask R-CNN, developed for instance segmentation. State-of-the-art accuracy is defined as the highest accuracy obtained in other studies. Mask R-CNN is chosen for this project as it has shown high accuracy -higher than Faster R-CNN [34]- with multiple datasets and it provides an object mask, which refers to the silhouette and the inside of the object tinted on a color. Once the object mask is obtained, as it includes the mask coordinates, the plant area (m^2) is computed straightforwardly and is correlated with seed yield.

Mask R-CNN has already been adopted for agricultural purposes, such as pathogen detection [36,37] and weed detection [28], to precisely apply localized pesticides and herbicides. Several studies accounting for Mask R-CNN in 2D have been conducted in different crops to automatize harvesting. These crops include pomegranate [47], apple [18], orange [8], cucumber [23], and strawberry [46], with a minimum F1-score of 89.47% and a maximum of 96.49%, among them all. The F1-score is the measure of the accuracy of a test, computed as the ratio between the number of true positives results (precision) and the number of all positive results (recall). Most of the articles claim that missing detections are still an important issue to be solved. Machefer et al. [24] introduced a refitted strategy for potato and lettuce counting and sizing and concluded that transfer learning can be used to reduce the required labeled data and used a large coarsely-annotated dataset. Nonetheless, they accepted that their model is crop-specific even if both potato and lettuce are low-density crops. In addition, counting limitations occur when plants reach a later growing stage since they have a high overlap [21]. Nevertheless, it also happens with human annotators because they get biased with the overlap. Lastly, their model was only trained on 256 $\times$ 256 pixel frames and hence should be generalized to other pixel frames.

Most of the studies conducted to predict yield using some phenotyping variables are carried out with UAVs [7,43–45]. Among them, height is the most mentioned variable to predict yield [7,44,45]. Bendig et al. [2] derived the crop surface model from RGB images based on several observations over the growing season and extracted the height to correlate it with barley's biomass, with an R^2 of 0.81 for fresh biomass and 0.82 for dry biomass. Nevertheless, an earlier time window for seed yield forecasting is required for the farmer to deal with under/overproduction.

Experiments with rice have correlated its yield with several variables, such as Leaf Area Index (LAI) and different vegetation indices (VI) [48]. However, including data from multiple growth stages is necessary to increase the linear relationship between the mentioned variables and yield. Yang et al. [43] used a CNN with RGB and multispectral images as input data to predict rice yield and they concluded that DL leads to more accurate results than using common variables as VI since the latest have limited capacities at advanced stages such as ripening. Furthermore, the closer to the harvest moment the more spatial features the crop has, and hence, the better the accuracy of the predicted yield.

Regarding spinach, most of the studies check for correlation between yield components (for instance number of leaves and plant height) [13,33], but none of them offers the correlation of a yield component with seed yield itself. Valente et al. [40] proposed a machine vision and transfer learning algorithm to count the number of spinach plants, with an accuracy of 95%. However, their method assumes that all plants are spinach, and hence there are weeds counted as spinach plants. As all plants were annotated thinking that they were spinach plants, when they could be weeds, it can still have high theoretical accuracy. In addition, closely growing plants might not be distinguished, which leads to the undercounting of plants in the image.

All studies conducted until the moment, to the best of our knowledge, are crop-specific and cannot solve the challenge of applying the same algorithm to various crops from very distinct botanic families. Those experiments are also frame-specific, meaning that they can only be used for a certain pixel frame and their processing speed is low. Moreover, the algorithms are trained only with raw data (single UAV images), which include motion distortion and need to be rectified [17,25]. The current experiments that are carried out focus mainly on plant/fruit counting. However, they all struggle with missing detections and excluding weeds, which lead to undercounting and overcounting, respectively. In addition, the studies do not include multiple growth stages to train the algorithm, making them time-dependent and avoiding the opportunity of counting in an early stage. Furthermore, different phenotyping variables are not combined to predict yield and in the specific case of spinach, they only seek for correlation between yield components, not between a component and yield itself.

Based on the above rationale, the heart of the selected method for accomplishing the correlation of number of plants and phenotyping variables with seed yield is an algorithm that exhibits state-of-the-art accuracy, shows good performance when applying transfer learning, and is not computationally expensive. A known framework that combines all the desired characteristics is the already described Mask R-CNN.

This article aims to assess the correlation between the number of plants and two phenotyping variables (individual plant area and canopy cover) with two seed yield variables (count of seeds and average thousand seed weight). The information is extracted from UAV (Unmanned Aerial Vehicle) images and orthomosaics of a spinach field. This is achieved by applying the Mask R-CNN algorithm on a custom dataset as well as transfer learning.

The objectives of this article are (1) to solve the lack of knowledge on combining the number of plants and multiple phenotyping variables to predict seed yield in spinach; (2) to work with raw data (single UAV images, not orthomosaics) to train the neural network (NN), as Yeom et al. [44] suggest; (3) to work with distinct phenological stages (early and late stages) to widen the time window and predict spinach seed yield in a sooner date.

2. Material and methods

2.1. Study area and data collection

The experiment was conducted in a field in the province of North Holland, in the Netherlands, between April 19th and August 12th' 2019 (planting and harvesting dates). Because of the confidential origin of the datasets, Fig. 1 does not show the exact location and coordinates of the study area. The crop on the field was spinach (*Spinacia oleracea)* with one male variety and multiple female varieties to stimulate cross-pollination. The aim of this experiment was seed production and it was carried out with conventional cultivation. The soil type is calcareous polder soils with heavy silt. The area of the field is around 0.41 ha, with a 0.35 m spacing between rows and a 0.15 m spacing within rows. The length and width of this field are approximately 160 m long and 23.5 m wide.

DJI Phantom 4 Pro (Shenzhen, China) was flown over the field with a pre-designed flight plan controlled with the flight control software UgCS (SPH Engineering, Riga, Latvia). The specifications of the UAV and its flight parameters are shown in Table 1. The DJI Phantom 4 Pro carried an RGB camera, whose characteristics are presented in Table 2.

In this study, a total of 982 RGB images on three different dates (June 11th, July 8th' and August 12th' 2019) were collected. From the total of 982 RGB files, the ones from the same date were then stitched together into a unique ortho-mosaic image, one for each of the mentioned dates. The Structure for Motion (SfM) approach as implemented in the software Agisoft MetaShape (St. Petersburg, Russia, version 1.7.3) was used

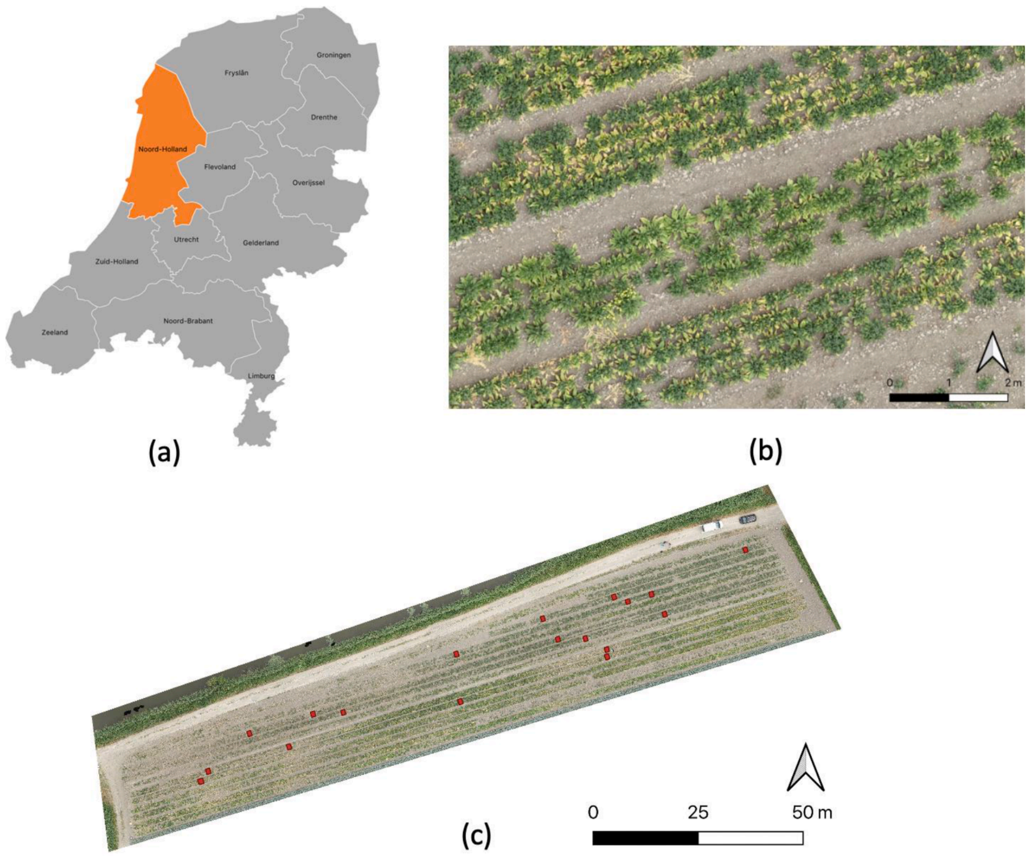

Fig. 1. Study area site: (a) Location of the province of Noord-Holland inside the Netherlands. (b) Close-up of spinach plants. (c) Distribution of the GT plots (red squares) around the spinach field.

Table 1
Characteristics of the UAV used (DJI Phantom 4 Pro) and the flight specifications.

	DJI Phantom 4 Pro
Weight (g)	1388
Number of rotors	4
Max flight time (min)	30
Overlap (front and side)	70–80%
Flight height (m)	20
Flight speed (m/s)	2–3
Mission time (min)	30

Table 2
Specifications of the RGB camera mounted on the DJI Phantom 4 Pro.

	DJI Phantom 4 Pro
Model	FC6310S
Spectrum range	RGB
Sensor resolution	5472 × 3648
Focal length (mm)	8.8
F-stop	f/4.5
Exposure time	1/120

for all the RGB images. All the settings for Agisoft MetaShape were set to "High Quality" in all processing steps. The ground sampling distance was 5.45 mm. A total of 440 images were taken in June, 254 in July, and 288 in August. The size of the created orthomosaics ranges from 512.4 to 677.8 MB and its spatial resolution is approximately 5 mm. At such a detailed resolution, the canopy size and shape can easily be identified (Fig. 1).

Ground Truth (GT) was observed by field specialists in 18 small plots (Fig. 1-c), that were the same for the three mentioned dates. They were approximately 105 × 90 cm and were disseminated all over the field at random distribution. The field observations for the GT dataset include the number of plants per plot on three different moments (June 11th, July 8th’ and August 12th all from 2019) and the height and width of each plant (in cm) in the plot for July 8th and August 6th. Moreover, two breeding seed yield variables were observed by field specialists. These variables include the count of seeds produced per each GT plot and the average thousand seed weight (TSW), in grams, per each GT plot. To measure the number of seeds produced, first the spinach plants were cut. After, specialized machines dried and sorted the seeds, and finally removed soiling and abnormal seeds. The last step was to introduce all the seeds in an advanced packing machine that counted and packed all the seeds to have the final seed count per GT plot. With respect to the average TSW (g), the thousand seed weight per each plant of the GT plot was computed and the final value was average among the number of plants of the GT plot to obtain the final average TSW (g).

The MS-COCO dataset [22], specially trained for object detection, counts with 80,000 images for training and 40,000 images for validation. The MS-COCO dataset contains images of ten food categories (banana, broccoli…) but does not include spinach images and therefore it was used as a coarsely annotated dataset. Consequently, extra training data for this specific case scenario was needed. For testing purposes, the eighteen GT plots of the orthomosaic of July 8th were introduced to the model to count the number of spinach plants and compare it with the number of plants that had already been annotated.

2.2. *Preprocessing*

Fig. 2 displays the whole flowchart of this study, from the inputs (raw

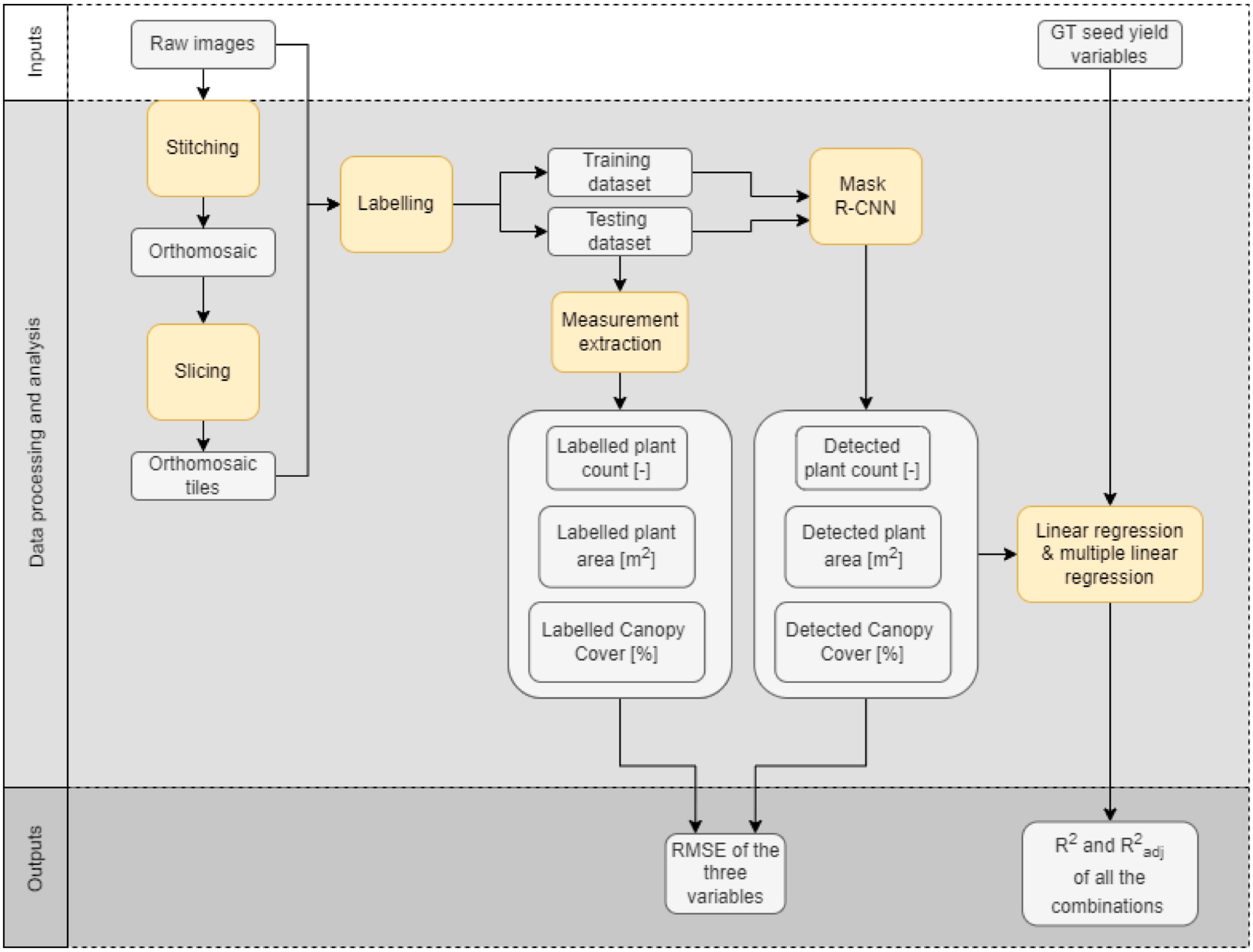

Fig. 2. Flowchart of the study. The initial inputs are the raw images and the Ground Truth seed yield variables measured on the field. The raw images are stitched to generate an orthomosaic, which is then sliced to obtain orthomosaic tiles. The raw images and the orthomosaic tiles are annotated to generate the training and testing dataset, which are used to train and test Mark R-CNN and obtain the detected plant count, plant area, and Canopy Cover. Those values are compared with the annotated measurements to obtain the RMSE. Finally, single linear and multiple linear regression are calculated from the three extracted variables and the seed yield variables (input) to compute the R2 and adjusted R2 of all the combinations.

images and GT seed yield variables) to the final outputs, which are the RMSE metrics and the R^2 and adjusted R^2 of the correlations between the variables measured with Mask R-CNN and the seed yield variables.

The first step was to stitch the images of July to generate the orthomosaic of July 8th, which is the one where spinach plants can be easily identified and distinguished. Moreover, the majority of the GT data correspond to that date. Afterward, the GT plots were identified with an ID number, that was introduced to the model and allowed us to compare the metrics of GT data and data extracted with the algorithm. Regarding their coordinates, polygons on top of the TIFF file were created on QGIS (version 3.16.7) to know the exact location of the GT plots. Apart from the 18 GT plots, 20 extra plots were generated. The purpose of these extra orthomosaic tiles, selected to be close to the GT plots, was to carry out a trial to check if including them as training and validating data increased the accuracy of the algorithm. The polygons (shapefiles) were masked later with the orthomosaic to obtain small tiles of the orthomosaic. This process was performed with the software R Core Team [32], version 1.2.5033, and saved as JPG files. The packages used were "raster", "rgdal", and "sf" [4,14,30]. The second step was to select which UAV images would be annotated. The selection criterium is described with the following questions, which can be answered in the provided order or in any other order, as long as all of them are answered.

(1) Is it a picture of the spinach field?
(2) Is it a blurry image for a human annotator, meaning that spinach plants can be distinguished?
(3) Is it a good representation of the field, meaning more than 70% of the image corresponds to the spinach field?

From the total number of images, 127 were selected because they perfectly fulfilled all the requirements.

The last preprocessing step was to label and manually delineate a polygon around each spinach plant of the selected RGB images with the software LabelMe (Massachusetts, USA) (Fig. 3). From the total number of images, 80% of the annotations were used for training, and the remaining 20% for validating to reach the number of needed annotations (around 500). The needed number of annotations was agreed upon based on literature, in the work from Liu et al. [23]; Machefer et al. [24]; Stewart et al. [36], and Yu et al. [46].

2.3. Training and validating Mask R-CNN

Detectron2 [41] was used to run the DL model. Detectron2 is a software developed by Facebook Research that implements state-of-the-art object detection algorithms. It has a user-friendly interface, which makes it easy to train and validate the algorithm on your custom dataset. Mask R-CNN was run on Google Colab (which comes with 12 GB of RAM that can be incremented to 25.5 GB) on a computer

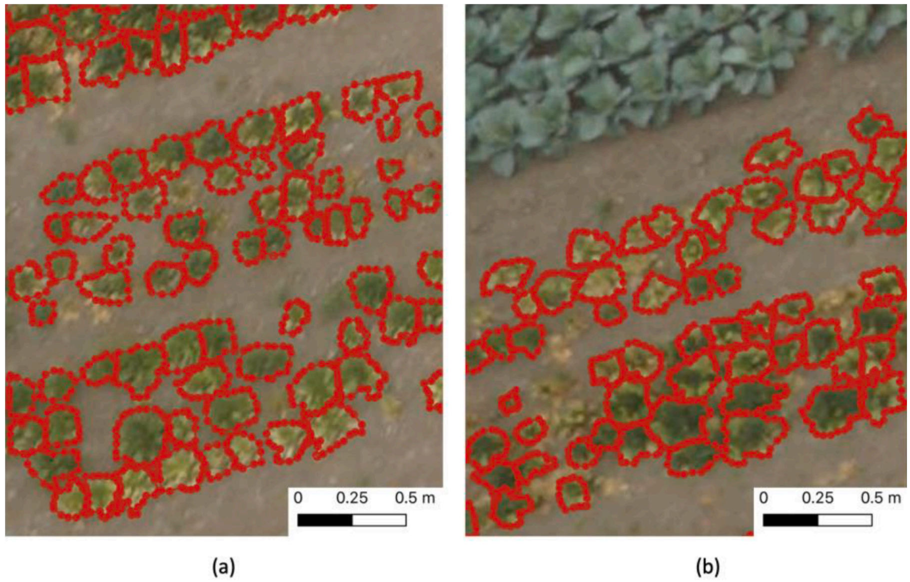

Fig. 3. Caption of the manually delineated spinach polygons using the software LabelMe. Each polygon shown in the image represents one spinach plant. (a) an image that includes only spinach plants and (b) an image that includes both spinach plants and another crop (cauliflower plants), which are not annotated.

with a 2 GHz Intel Core i5 processor, using 2.65 GB of RAM and 12.19 GB of GPU, running macOS Mojave version 10.14.6, GPU Intel Iris Graphic 540 1536 MB.

For training and validation, four different cases (Table 3) were considered to compare the obtained metrics and select the most accurate model. All of them use the 104 selected images as a base, with a resolution of 461 × 346 pixels. Case 3 and case 4 include also the 20 extra orthomosaic tiles for training and validation purposes. Data augmentation is implemented in half of the cases (Case 1 and Case 3) and it refers to a strategy to increase the diversity of training data without collecting new ones [15]. It includes for instance padding, rotating, flipping, and brightness effects, among many others.

Case 1 and case 2 include 104 images for training and 23 images for validating. The difference between case 1 and case 2 is that case 1 applies data augmentation and case 2 does not. The second trial (case 3 and case 4) consists of adding small tiles of the orthomosaic to check if the hypothesis of Yeom et al. [44] of not using orthomosaics to train the algorithm can also be generalized for this case scenario. The already mentioned twenty extra orthomosaic tiles are introduced in these cases (17 for training and 3 for validating). Case 3 uses data augmentation and case 4 does not. A summary of the characteristics of the datasets of the four cases is provided in Table 3.

The percentage of annotations was different for each month. July was the month with the highest number of annotations because the spinach plants were fully developed and were easily identified. August includes the least number of delineations (below 4%) since spinach plants had a high overlap and were complicated to differentiate. Table 4 compiles the percentage of annotations that were used for both cases 1 and 2 (without orthomosaic tiles), and cases 3 and 4 (adding orthomosaic tiles).

The training procedure started by implementing the default values of the hyper-parameters (Table 5) and was adapted by trial and error to the specific study case of spinach counting. The procedure consisted of modifying the values of the hyper-parameters increasing and decreasing the default values by 10% and checking if the metrics were higher than before. Then, the hyper-parameter value would be increased or decreased until the metrics were no longer higher. That same pattern was implemented with all hyper-parameters, implementing trial and error tests, following the guidelines of Bengio [3]. The adjustable parameters considered were the following: batch size, learning rate, number of training epochs, number of regions of interest proposed per image, and threshold to accept/reject annotations [35].

After validation, classification metrics (Precision, Recall, and F1-score) [8] were computed to the validation dataset to check how accurately spinach plants were detected. The metrics were determined by comparing the Intersection over Union (IoU) of spinach plants in the manually annotated images for validation and the ones measured with Mask R-CNN [29].

$$Precision = \frac{TP}{TP + FP} \tag{1.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{1.2}$$

Table 3
Summary of the characteristics of the four scenarios that were used for training and validating Mask R-CNN. The table shows the number of training and validating images used in every scenario, as well as if data augmentation and orthomosaic tiles were included. Finally, the last row displays the number of annotated plants that each case scenario had.

	Case 1	Case 2	Case 3	Case 4
Training images	104 RGB	104 RGB	104 RGB + 17 orthomosaic	104 RGB + 17 orthomosaic
Validating images	23 RGB	23 RGB	23 RGB + 3 orthomosaic	23 RGB + 3 orthomosaic
Data Augmentation	Yes	No	Yes	No
Orthomosaic tiles	No	No	Yes	Yes
Delineated plants	5556	5556	6048	6048

Table 4
Summary of the percentage of annotations per month for Cases 1 and 2, and Cases 3 and 4. It is observed that most of the annotations belong to June and July and very few to August. This is in accordance with the scope of the study to predict seed yield at a sooner date.

	Case 1 and Case 2	Case 3 and Case 4
June	44.5%	42.8%
July	52.3%	54.1%
August	3.2%	3.1%

Table 5
Default values of the hyper-parameters that were used to start training Mask R-CNN. Afterward, these values were modified to fine-tune Mask R-CNN and reach higher metrics' scores.

Batch size	Learning rate	Number of epochs	RoI per image	Threshold
8	0.0025	300	200	0.1

$$F1\,score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{1.3}$$

Where TP are the true positives (spinach plants detected correctly), FP the false positives (objects detected as spinach plants), and FN the false negatives (spinach plants wrongly detected).

2.4. Testing Mask R-CNN and computing plant count and phenotyping variables

After choosing the best model based on the metrics of the validating dataset, the model of Mask R-CNN with the best hyper-parameters scores was implemented for the test dataset (dataset with GT values), obtaining again Precision, Recall, and F1-score values. The outputs of Mask R-CNN are the number of detected spinach plants per GT plot and its mask coordinates. As the coordinates are in the projected coordinate system Amersfoort / RD New (EPSG: 28,992), the results are in meters. A polygon for each mask was created and the area of each mask was computed automatically and summed to have the spinach area per GT plot. The Canopy Cover was computed also at GT-plot level by implementing Eq. (1.4).

$$Canopy\,Cover\,(\%) = \frac{spinach\,area\,per\,GT\,plots}{GT\,plot\,area} * 100 \tag{1.4}$$

The number of spinach plants annotated could directly be obtained by running an algorithm to count them based on the JSON file created per each image. The Ground Truth dataset provided by the experts from the breeding company had already calculated the spinach area (m^2) per GT plot. By dividing the spinach area values per the area of the GT plot (computed with QGIS), the GT values of Cover Canopy were obtained.

Then, a comparison between the number of plants measured by the algorithm, spinach area, and canopy cover with their GT values was performed. As they are ratio variables, regression metrics were needed and hence, the root mean squared error (RMSE) was calculated.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y - \hat{y})^2}{n}} \tag{1.5}$$

Where n is the total sample size, y is the actual value, and $\hat{y}$ is the value obtained with Mask R-CNN.

2.5. Computing correlation between plant count and phenotyping variables with seed yield

As a final evaluation, the correlation between the number of plants and the phenotyping variables with seed yield was evaluated based on the correlations analyzed. The correlations that were performed were with the number of plants per GT plot and the two seed yield variables, the plant area (m^2) with the number of seeds (m^{-2}) and the TSW (g), and the Canopy Cover (%) with the two seed yield variables. Moreover, a combination of the previous variables was also performed to obtain the correlation with both the number of seeds (m^{-2}) and TSW (g). The correlations were performed by simple linear regression and multiple linear regression (when combining multiple variables). The adjusted coefficient of determination (R^2_{adj}, Eq. (1.7)) was computed in both cases.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \tag{1.6}$$

$$Adjusted\,R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \tag{1.7}$$

Where $\bar{y}$ is mean value of y, n is the total sample size, and p is the number of predictors.

3. Results

3.1. Training and validating

Ten trials were conducted to evaluate which were the most appropriate hyper-parameter's values for spinach detection. The trials consisted of modifying the already-mentioned hyper-parameter's values starting with default values, followed by trial and error tests [3], and examining if any changes corresponded to an increase in metrics' scores. The best configuration for all training cases resulted from trial 10. Their metrics can be observed in Table 6. The values of the hyper-parameters which conducted to these results were the following:

- Batch size: 8;
- Learning rate: 0.0025;
- Number of training epochs: 500;
- Number of RoI proposed per image: 200;
- Threshold to accept/reject annotations: 0.7.

Fig. 4 depicts the learning curves of the four different cases for both the validation and the test dataset. They consist of the change in accuracy and total loss through iterations. It can be observed that case 1, case 2, and case 3 show high accuracy and a low total loss. Moreover, models learn until the last epochs for both validating and test datasets (Fig. 4). However, in the three cases, there are appreciable differences between validation and test datasets. Lastly, case 4 performs with high accuracy, low total loss, and not many differences between datasets.

Comparing all the information that comes from Table 6 and Fig. 4, case 4 was adopted as the reference model for the rest of the project's research. Hence, all next Sections are only focused on this scenario to compute the testing metrics and correlation between the three variables and seed yield.

3.2. Testing

Fig. 5 represents how the most accurate model of Mask R-CNN (case 4, trial 10) detects spinach plants in unseen images (test dataset). The example on top (Ground Truth plot number 14) was the best case, with the same number of detected annotations as labelled ones, 16 in total. On the other hand, Ground Truth plot 12 was the worst scenario, where there was a big difference between the number of manually labelled spinach plants, 13, and the number of identified plants, 7.

With the outputs of Mask-RCNN (the detected plants), the plant area (m^2) and Canopy Cover (%) were calculated. Regarding the computation of both seed yield variables per GT plot, some assumptions were made. The company's ground truth dataset provided by the experts contained the average number of seeds per plant and the average TSW per plant

Table 6
Highest metric's scores for Cases 1 to 4, resulting from trial 10. The highest scores, highlighted in bold, are obtained for Case 4.

	Case 1	Case 2	Case 3	Case 4
Precision	0.582	0.596	0.645	0.782
Recall	0.079	0.086	0.091	0.543
F1-score	0.139	0.150	0.159	0.641

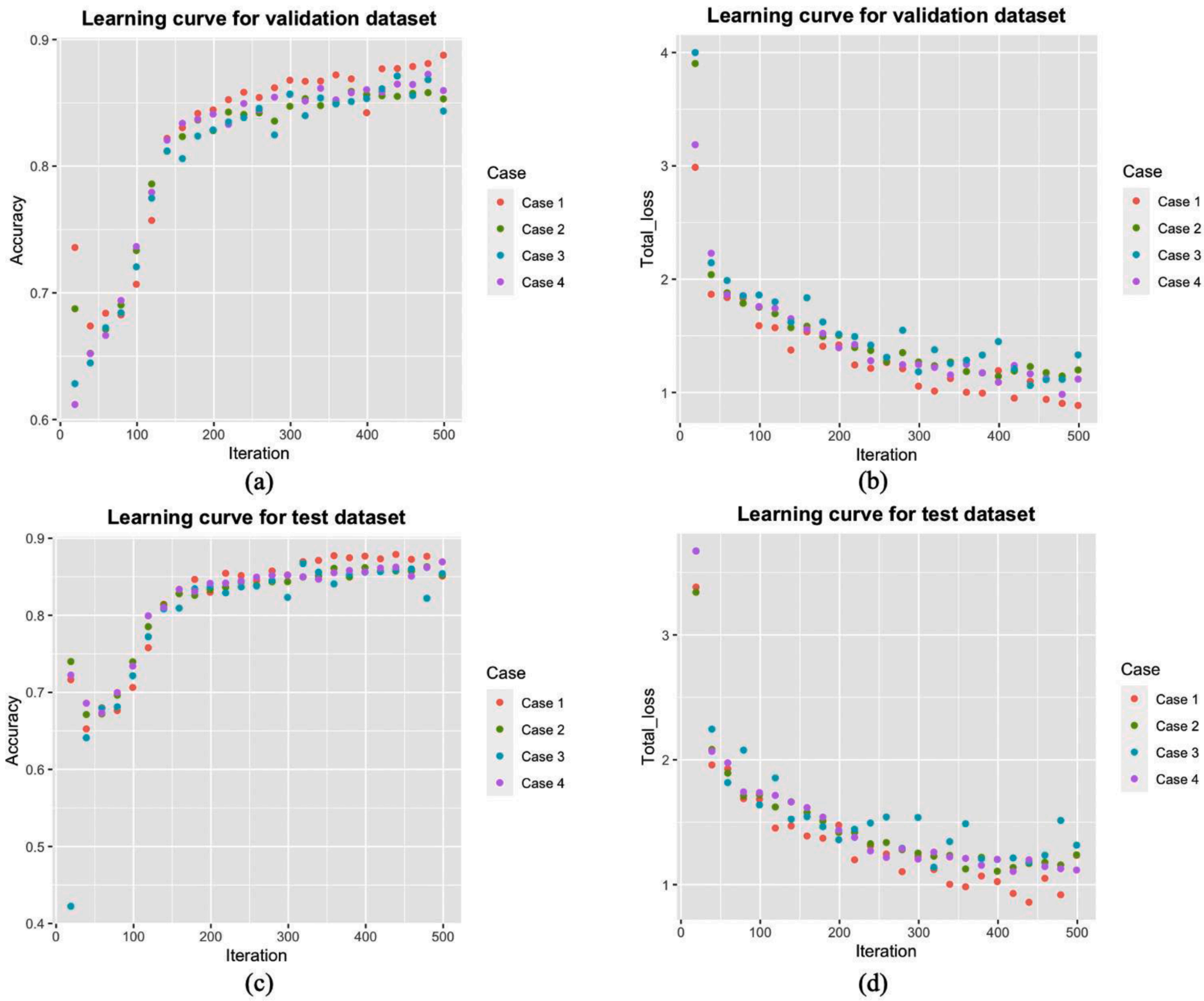

Fig. 4. Learning curves of the four different cases with the hyper-parameters values of trial 10, divided into validation and test datasets. On the left column (a) and (c), the evolution of accuracy through iterations. On the right side (b) and (d), the evolution of total loss through iterations.

subdivided per size (small, medium, and tall) and per variety (OS196 ♀, OS250 ♀, and OS648 ♀). To clarify, there were nine different values per seed yield variable, one for small plant from OS196 ♀ variety, another for medium plant from OS196 ♀ variety, along with others. The assumption made was to compute the average value of the three sizes per variety and have an estimated value of the number of seeds per plant and TSW (g) from a specific variety. Afterward, the previous value was multiplied by the number of plants per GT plot to obtain the estimated value of the number of seeds (m^{-2}) plot and TSW (g). Because of the assumptions made to obtain the seed yield values per GT plot, the number of seeds (m^{-2}) and thousand seed weight (g) per GT plot are estimated, not real values.

The regression metrics for the three computed variables are summarized in Table 7. The RMSE metrics are in the same units as the variables. It can be observed that all values are lower for plant area due to its small range of values (from 0.16 to 1.01 m^2). The RMSE for Canopy Cover is low (4%) and hence there is not much difference between the annotated and the values obtained with Mask R-CNN. Nevertheless, the RMSE for the number of spinach plants is intermediate, three plants, and it is mainly caused by three GT plots (GT 1, 9, and 12) that have differences of five, five, and six plants, respectively. Without these three GT plots, the RMSE would decrease to 2.6 plants, which is a very adequate value.

Fig. 6 displays the scatterplot representations of the three variables. It can be observed that for the number of plants (Fig. 6-a), there is a positive correlation between the annotated and the detected plants. Nevertheless, this correlation is not as high (as linear) as for the comparison of plant area and canopy cover (Fig. 6-b, c). It can be observed that Fig. 6-a contains 16 points instead of 18 since there were three GT plots with the same detected and annotated number of spinach plants.

3.3. Correlation between plant count and the two phenotyping variables with seed yield

The correlations between the number of plants, plant area (m^2), and canopy cover (%) with the number of seeds (m^{-2}) and thousand seed weight (g) plots are presented in Fig. 7. The highest correlations with seed yield are shown with the number of plants, with R^2 values of 0.70 and 0.77, being the largest correlation between the number of plants (m^{-2}) and TSW (g). The rest of the plots show flat lines indicating almost no correlation between the detected variables and seed yield variables. It can be observed in the two top plots of Fig. 7 (number of identified plants vs. number of seeds and TSW) that only 12 points are shown. As has been mentioned, some assumptions were made. Therefore, all GT plots that had the same number of plants and belonged to the same variety appear as one point in Fig. 7.

All the R^2_{adj} values of multiple linear regression correlations can be observed in Table 8. To predict the number of seeds, the addition of plant area increases the R^2_{adj} value compared to having only the number of plants, but not at significant levels, since the p-value of the plant area

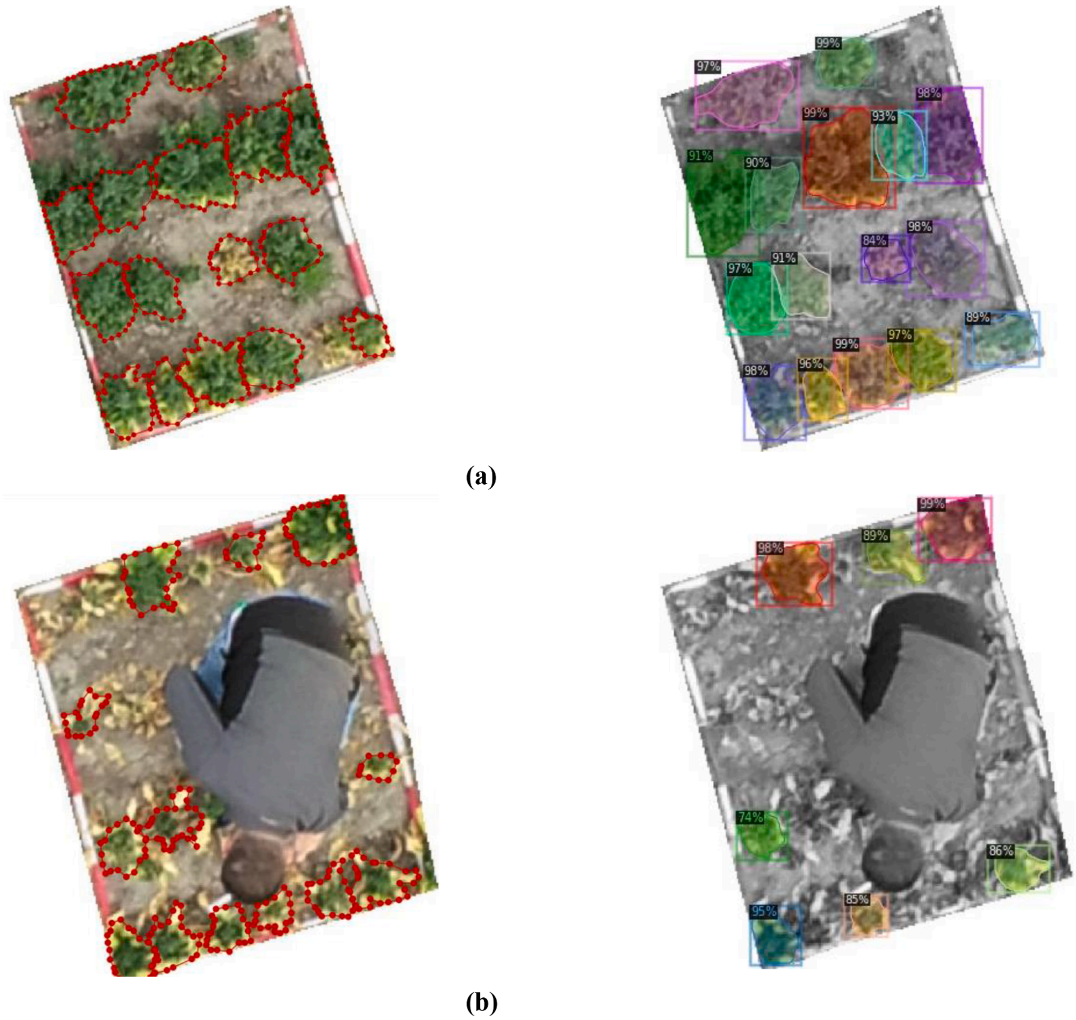

Fig. 5. Comparison between the annotated images (left) and detected images (right). (a) shows the best-detected example (16 annotated, 16 detected), whereas (b) illustrates the worst example found (13 annotated, 7 detected). Nevertheless, it can be observed that in (b) there is an operator in the field, which might confuse the algorithm. Still, the plot was kept to make the trained algorithm robust.

Table 7
RMSE regression metrics for the three variables (plant count, plant area, and canopy cover), by comparing the GT values with the detected ones.

	Number of plants[-]	Plant area[m^2]	Canopy Cover[%]
RMSE	3.4480	0.0537	4.4183

is higher than 0.05. Nevertheless, predicting the number of seeds with the three variables increases the coefficient of determination to 0.80, with both extra variables being significant. Differently, none of the combinations' R^2_{adj} is higher than the simple linear regression score, achieved by only the number of plants to predict TSW (g), with an R^2_{adj} of 0.77. In conclusion, both seed yield variables can be predicted with almost 80% accuracy.

4. Discussion

Mask R-CNN has proven to be a good algorithm for spinach plant detection (Table 6). The COCO-dataset [22] was chosen for this research because it is open-source and has an available configuration that can easily be implemented on custom cases, only a few parameters need to be modified for inference [22]. The adopted network was PyTorch, even if Keras and Tensorflow were the original networks for Mask R-CNN implementation [1]. The main reason for this decision was that PyTorch includes the software detectron2, which provides a high inference speed that allows the user to iterate more rapidly on model experiments [41].

Image labeling is a time-consuming task that took an approximate time of 181 h to be completed and is subject to human error [9]. The quality control was performed by first training the annotating skills on already labelled images downloaded from the internet [1] and afterward double checking all the manual delineations of the custom dataset to detect missing annotations or wrongly labeled ones.

The metrics of cases 1 and 2 were in all trials lower than the ones of cases 3 and 4 (Table 6). The unique difference between these datasets was the fact that the latter included 20 tiles of the orthomosaic that was used for testing. Contradicting the recommendation of Yeom et al. [44] of using raw data to train the NN, the model showed better metrics while adding orthomosaic data, as it already happened in other studies regarding plant phenotyping variables using UAV images [11,42]. The reason for this contradiction might be that part of the training and testing datasets belong to the same orthomosaic. By having the same pixel frame and brightness conditions, the algorithm learns better and consequently predicts more accurately. To prove that the hypothesis of Yeom et al. [44] cannot generalize, more experiments with different datasets should be conducted.

Regarding the distinct feature between cases 1 and 3 (with data augmentation), and cases 2 and 4 (without data augmentation), the results showed that the first group had lower metrics for all trials

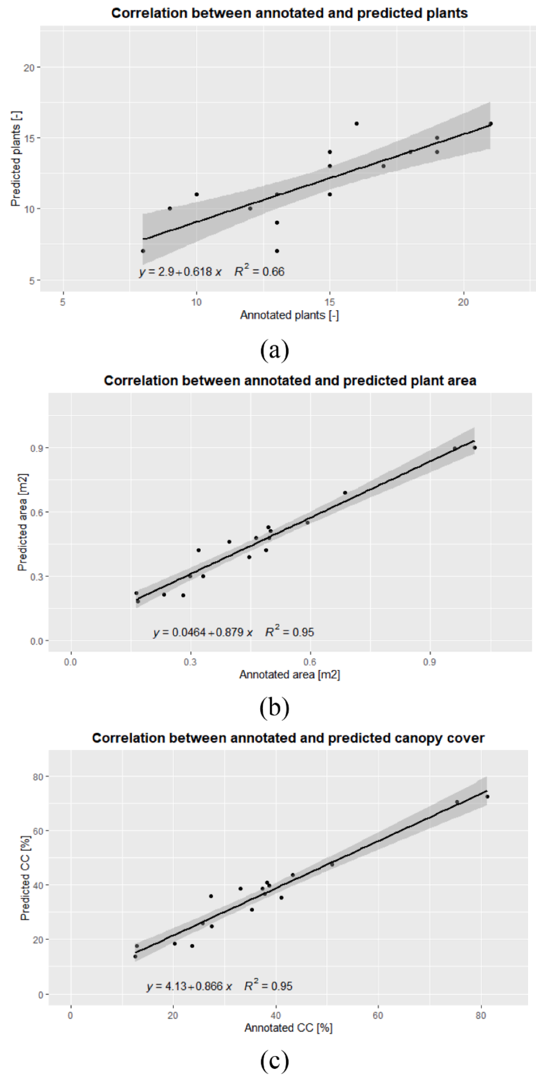

Fig. 6. Scatterplot representation of the comparison between annotated and measured variables for the test dataset (18 GT plots). There is a linear correlation for all annotated and measured variables, being the correlation higher for cases (b) and (c) with R^2 values of 0.95.

(Table 6). Interestingly, having extra data to train the model decreased the F1-score. The data augmentation methods applied were resizing, changing brightness, contrast, saturation, and lighting, and allowing flipping and rotations. All cases (e.g., flipped image, change in brightness) are valid images in the domain of spinach plants and cannot be the reason for the reduction in metrics. One motive for that could be that the model had a low capacity and could not learn patterns from the augmented dataset since it introduced noise to the real dataset. However, it is unlikely because the model detected accurately without data augmentation. It could be that the model learnt in a short amount of time. Nevertheless, increasing the number of epochs was not accompanied by an increase in the metrics' results ergo this hypothesis can be refused. Another cause could be that there were not enough annotations and when including data augmentation, the model was memorizing the training cases, causing overfitting [16]. To prove this last theory, more training annotations should be introduced to the model and in case the problem persists, the overfitting problem with the training dataset could be accepted.

The highest F1-score obtained in this study (64.1%) (Table 6) was 28% lower than the average F1-score calculated in similar studies of fruit detection with image segmentation [8,23], even if the RMSE values were all very accurate. The main difference between this study and the rest was the number of annotations present on each test image. Normally the number ranges from 1 to 5 plants/fruits per image and in the spinach case it ranges from 8 to 21 plants. Therefore, cropping the test datasets into smaller tiles (1 to 5 annotations each) could be considered in future work to find out if that was the reason for the low F1-score. Analyzing the decomposition of the F1-score formula (Eq. (1.3)), it can be seen that recall scores were lower than accuracy values for all cases and trials. The low values of the recall are coherent for the first trials and cases 1 to 3 since the difference between annotated and identified plants was remarkable. However, an important increase in recall rate for trial 10 case 4 is observed, turning recall into the main cause of the boosted F1-score.

The hyper-parameters values that were responsible for obtaining an F1 score of 0.64 are the ones mentioned in the results section. The batch size for the spinach case, 8, was lower than for other studies, 32 or 64 [12,46]. However, in the spinach scenario, the batch size could not be increased since CUDA was running out of memory. The threshold to accept/reject annotations used in the spinach case, 0.7, was the same as the one used by Liu et al., [23]. Regarding the learning rate, Häni et al. [12]; Liu et al. [23]; Yu et al. [46] showed that the value of 0.001 was better in their cases. Nevertheless, the best learning rate for the spinach study was 0.0025, even if 0.001 was included in one of the ten trials that were performed. Regarding the number of training epochs, the values implemented in the spinach study were inside the range of values used in similar studies for fruit detection with instance segmentation [8,12].

Concerning the RMSE values obtained for the comparison of the number of plants (Table 6), they are both close to the ones obtained by Prado Osco et al. [31] with citrus trees, which outperformed significantly object detection methods for plant counting. With respect to the two other variables (plant area and canopy cover), no studies are comparing the same metrics. Nevertheless, there are metrics for similar phenotyping variables measured, for instance, height and LAI, with RMSE values of 0.16 cm and 0.34 [-], respectively [45]. It can be observed that the obtained RMSE values for the spinach case are quite low and are comparable (in terms of absolute value) to the ones retrieved by Yu et al. [45]. Hence, it can be concluded that the metrics' values of the plant count and the two phenotyping variables are state-of-the-art scores. With reference to the three GT plots whose difference between annotated and detected plants was the highest (GT 1, 9, and 12), it can be observed that they are the plots with more plant density, with the most overlap and biggest size of the spinach plants. About GT 12, there was an operator in the middle of the GT plot, that could have confused the algorithm. Nevertheless, it was not removed to make the trained algorithm robust to any unexpected object. Therefore, more GT plots with denser and bigger spinach plants should be added to train the algorithm and exclude the test plots where there are people or random objects.

As has been mentioned, some assumptions were made to estimate the seed yield variables' values for each GT plot. For future studies, it would be recommended to gather real measurements per GT plot to confirm that the obtained values were properly estimated and that the obtained correlations can be generalized to other spinach fields. Furthermore, for future studies, plant height, which is the most mentioned variable to predict yield [7,44,45], should also be included as a GT measurement for all measured dates to correlate this variable with seed yield.

The high values of the coefficient of determination obtained to predict the number of seeds and TSW (0.80 and 0.77, respectively) (Fig. 7), are comparable to state-of-the-art scores achieved by Bendig et al. [2] to correlate barley's height with biomass. Nevertheless, in the spinach study, the height was not measured with the 2D UAV images but could be calculated in future work with 3D point clouds collected from laser scanners mounted on a UAV or with the Digital Surface Model and Digital Terrain Models obtained from the photogrammetric process. To predict the number of seeds, the combination of the number of plants, plant area, and canopy cover was required to achieve the R^2_{adj} of 0.80. On the other hand, only the number of plants was needed to predict the TSW

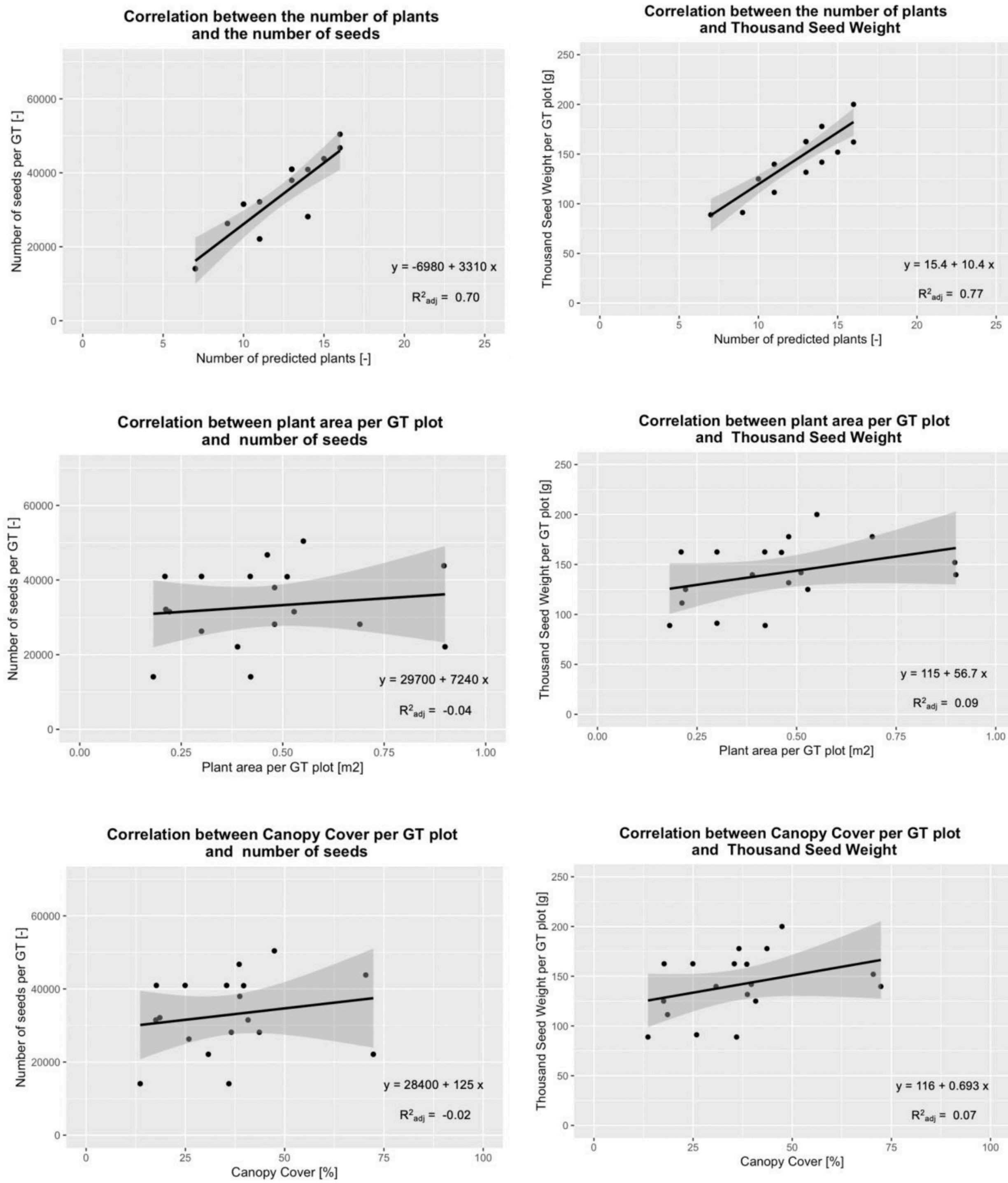

Fig. 7. Correlations between the measured variables and seed yield variables, with their regression equations and R^2 score. The points display each observation, the gray shadow represents the 95% confidence interval, and the black line is the regression equation. A clear correlation is observed for the comparison of number of plants and both seed yield variables, whereas no clear correlation is shown for the plant area and canopy cover with the seed yield variables.

with an R^2_{adj} of 0.77. Therefore, depending on the seed yield variable to be predicted, only one variable or the three mentioned need to be calculated. With this study, the correlation between some yield components and seed yield itself was provided, which was lacking in all the studies performed until the date on spinach [13,33].

5. Conclusions

The main objective of this research was to assess the correlation between plant count and two phenotyping variables (individual plant area and canopy cover) derived from UAV images orthomosaics of a spinach field with seed yield measured in the field. The number of plants and the phenotyping variables were derived by applying Mask R-CNN on a custom dataset to count the number of plants per GT plot and to provide a mask per each spinach plant detected.

This study showed that Mask R-CNN can be applied to count spinach plants by training the algorithm with a combination of UAV images and orthomosaic tiles. The algorithm achieved the first purpose by applying

Table 8
R^2_{adj} values and *p*-values for multiple linear regression correlations. Predicting the number of seeds by combining the three variables (number of plants, plant area, and canopy cover), highlighted in bold, is the unique combinate of multiple linear regression that increases the coefficient of determination, with both all variables being significant.

	Number of seeds	Thousand seed weight
Number of plants + plant area	R^2adj = 0.7427 *p*-value of plant area > 0.05	R^2adj = 0.7591 *p*-value of plant area > 0.05
Number of plants + canopy cover	R^2adj = 0.7165 *p*-value of canopy cover > 0.05	R^2adj = 0.7595 *p*-value of canopy cover > 0.05
Number of plants + plant area + canopy cover	R^2adj = 0.8031 *p*-value of plant area & CC < 0.05	R^2adj = 0.7558 *p*-value of plant area & CC > 0.05

transfer learning using the COCO-dataset. The hyper-parameters modified were the batch size, the learning rate, the number of training epochs, the number of regions of interest proposed per image, the threshold to accept/reject annotations. The accuracy obtained (F1-score = 0.641) is lower than the state-of-the-art ones, but the model detected the number of plants with a relatively low RMSE of 3.4480 plants. Mask R-CNN could also obtain individual plant area as part of the outputs of the algorithm, with a very low RMSE of 0.0537 m^2.

There was a linear correlation between the plant count and phenotyping variables with the seed yield variables. The highest coefficient of determination for the number of seeds was found using multiple linear regression with three different variables (the number of plants, the plant area, and the Canopy Cover). With this, 80.31% of the variation of the number of seeds (m^{-2}) plot could be explained. The largest correlation found to predict the thousand seed weight (g) was with single linear regression, directly correlating the number of plants with TSW. Employing that, 77.42% of the variation of TSW (g) could be explained by the number of plants per GT plot.

This research explored a method to predict spinach seed yield at an earlier date by working with multiple phenological stages. It solved the lack of knowledge on combining the number of plants and multiple phenotyping variables to predict seed yield more accurately and refused the hypothesis of Yeom et al. [44] to work only with raw data to train the algorithm.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that has been used is confidential.

Acknowledgments

This work was supported by the SPECTORS project (143081) which is funded by the European cooperation program INTERREG Deutschland-Nederland. The authors would also like to thank Enrique Apolo-Apolo and Johan de Groot for their help during the development of this research.

References

[1] Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. Mask R-CNN for Object Detection and Segmentation.
[2] J. Bendig, A. Bolten, S. Bennertz, J. Broscheit, S. Eichfuss, G. Bareth, Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging, Remote Sens. 6 (11) (2014) 10395–10412, https://doi.org/10.3390/rs61110395 (Basel).
[3] Y. Bengio, Gradient-based optimization of hyperparameters, Neural Comput. 12 (8) (2000) 1889–1900, https://doi.org/10.1162/089976600300015187.
[4] Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Baston, D., Roualt, E., Warmerdam, F., Ooms, J., & Rundel, C. (2021). rgrdal: bindings for the "geospatial" data abstraction library (1.5-23). Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=rgdal.
[5] C. Camaschella, Iron-deficiency anemia, N. Engl. J. Med. 372 (19) (2015) 1832–1843, https://doi.org/10.1056/NEJMra1401038.
[6] C. Chen, J. Pan, S.K. Lam, A review of precision fertilization research, Environ. Earth Sci. 71 (2014) 4073–4080, https://doi.org/10.1007/s12665-013-2792-2.
[7] A. Feng, J. Zhou, E.D. Vories, K.A. Sudduth, M. Zhang, Yield estimation in cotton using UAV-based multi-sensor imagery, Biosyst. Eng. 193 (2020) 101–114, https://doi.org/10.1016/j.biosystemseng.2020.02.014.
[8] P. Ganesh, K. Volle, T.F. Burks, S.S. Mehta, Deep orange: Mask R-CNN based orange detection and segmentation, IFAC PapersOnLine 52 (30) (2019) 70–75, https://doi.org/10.1016/j.ifacol.2019.12.499.
[9] S. Ghosal, D. Blystone, A. Singh, B. Ganapathysubramanian, A. Singh, S. Sarkar, An explainable deep machine vision framework for plant stress phenotyping, Proc. Natl. Acad. Sci. USA. 11 (18) (2018) 4613–4618.
[10] H.C.J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, C. Toulmin, Food security: the challenge of feeding 9 billion people, Science 327 (5967) (2010) 812–818, https://doi.org/10.1126/science.1185383. American Association for the Advancement of Science.
[11] W. Guo, Y. Fukano, K. Noshita, S. Ninomiya, Field-based individual plant phenotyping of herbaceous species by unmanned aerial vehicle, Ecol. Evol. 10 (21) (2020) 12318–12326, https://doi.org/10.1002/ECE3.6861.
[12] N. Häni, P. Roy, V. Isler, A comparative study of fruit detection and counting methods for yield mapping in apple orchards, J. Field Rob. 37 (2) (2020) 263–282, https://doi.org/10.1002/rob.21902.
[13] R. Heyduck, S. Guldan, I. Guzman, Effect of sowing date and harvest schedule on organic spinach grown during the winter in high tunnels, Horttechnology 29 (3) (2019) 320–329. https://journals.ashs.org/horttech/abstract/journals/horttech/29/3/article-p320.xml.
[14] Hijmans, R.J., van Etten, J., Sumner, M., Cheng, J., Baston, J., Bevan, A., Bivand, R., Busetto, L., Canty, M., Fasoli, B., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J.A., Hiemstra, P., Hingee, K., Karney, C., Mattiuzzi, M., … Wueest, R. (2021). raster: geographic data analysis and modeling (3.4-13). Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=raster.
[15] Ho, D., Liang, E., & Liaw, R. (2019, June 7). Faster data augmentation. The Berkeley Artificial Intelligence Research. https://bair.berkeley.edu/blog/2019/06/07/data_aug/.
[16] Inoue, H. (2018). Data augmentation by pairing samples for images classification. In Data augmentation by pairing samples for images classification. arXiv. http://arxiv.org/abs/1801.02929.
[17] J.P. Jhan, J.Y. Rau, N. Haala, Robust and adaptive band-to-band image transform of UAS miniature multi-lens multispectral camera, ISPRS J. Photogramm. Remote Sens. 137 (2018) 47–60, https://doi.org/10.1016/J.ISPRSJPRS.2017.12.009.
[18] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, Y. Zheng, Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot, Comput. Electron. Agric. 172 (2020), https://doi.org/10.1016/j.compag.2020.105380.
[19] Kaplan, K. (2019, November 4). World's first true red spinach variety released. USDA Red. https://www.ars.usda.gov/news-events/news/research-news/2019/worlds-first-true-red-spinach-variety-released/.
[20] C.H.H. Le, The prevalence of anemia and moderate-severe anemia in the US population (NHANES 2003-2012), PLoS One 11 (11) (2016), e0166635, https://doi.org/10.1371/journal.pone.0166635.
[21] B. Li, X. Xu, J. Han, L. Zhang, C. Bian, L. Jin, J. Liu, The estimation of crop emergence in potatoes by UAV RGB imagery, Plant Methods 15 (1) (2019) 1–13, https://doi.org/10.1186/s13007-019-0399-7.
[22] T.Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C.L. Zitnick, P. Dolí, Microsoft COCO: common objects in context, in: Proceedings of the Computer Vision – ECCV, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48, 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham.
[23] X. Liu, D. Zhao, W. Jia, W. Ji, C. Ruan, Y. Sun, Cucumber fruits detection in greenhouses based on instance segmentation, IEEE Access 7 (2019) 139635–139642, https://doi.org/10.1109/ACCESS.2019.2942144.
[24] M. Machefer, F. Lemarchand, V. Bonnefond, A. Hitchins, P. Sidiropoulos, Mask R-CNN refitting strategy for plant counting and sizing in UAV imagery, Remote Sens. 12 (18) (2020) 3015, https://doi.org/10.3390/rs12183015 (Basel).
[25] H. Maître, From Photon to Pixel: the Digital Camera Handbook, John Wiley & Sons, Inc, 2017, https://doi.org/10.1002/9781119402442. From Photon to Pixel.
[26] R. Messner, H. Johnson, C. Richards, From surplus-to-waste: a study of systemic overproduction, surplus and food waste in horticultural supply chains, J. Clean. Prod. 6526 (20) (2020) 959–997, https://doi.org/10.1016/j.jclepro.2020.123952.
[27] E. Monsen, Iron nutrition and absorption: dietary factors which impact iron bioavailability, J. Am. Diet. Assoc. 88 (7) (1988) 786–790. https://pubmed.ncbi.nlm.nih.gov/3290310/.
[28] K. Osorio, A. Puerto, C. Pedraza, D. Jamaica, L. Rodríguez, A deep learning approach for weed detection in lettuce crops using multispectral images,

AgriEngineering 2 (3) (2020) 471–488, https://doi.org/10.3390/agriengineering2030032.
[29] R. Padilla, S. Netto, E. da Silva, A survey on performance metrics for object-detection algorithms, in: Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP), 2020, https://doi.org/10.1109/IWSSIP48289.2020.
[30] E. Pebesma, Simple features for R: standardized support for spatial vector data, R J. 10 (1) (2018) 439–446, https://doi.org/10.32614/RJ-2018-009.
[31] L. Prado Osco, M. Dos Santos de Arruda, J. Marcato Junior, N. Buceli da Silva, A. Marques Ramo, E. Akemi Saito Moryia, N. Nobuhiro Imai, D. Pereira, J. Creste, E. Takashi Matsubara, J. Li, W. Nunes Gonçalves, A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral images, ISPRS J. Photogramm. Remote Sens. 160 (2020) 97–106.
[32] R Core Team, R: A language and Environment For Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.r-project.org/.
[33] S.E.A. Raza, H.K. Smith, G.J.J. Clarkson, G. Taylor, A.J. Thompson, Automatic detection of regions in spinach canopies responding to soil moisture deficit using combined visible and thermal imagery, PLoS One 9 (6) (2014) 97612, https://doi.org/10.1371/journal.pone.0097612.
[34] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2015) 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031.
[35] Smith, L.N. (2018). A disciplined approach to neural network hyper-parameters: part 1 – learning rate, batch size, momentum, and weight decay. In arXiv. arXiv. https://github.com/lnsmith54/hyperParam1.
[36] E.L. Stewart, T. Wiesner-Hanks, N. Kaczmar, C. DeChant, H. Wu, H. Lipson, R. J. Nelson, M.A. Gore, Quantitative phenotyping of northern leaf blight in UAV images using deep learning, Remote Sens. 11 (19) (2019) 2209, https://doi.org/10.3390/rs11192209 (Basel).
[37] W.H. Su, J. Zhang, C. Yang, R. Page, T. Szinyei, C.D. Hirsch, B.J. Steffenson, Evaluation of mask RCNN for learning to detect fusarium head blight in wheat images, in: Proceedings of the ASABE Annual International Virtual Meeting, 2020, https://doi.org/10.13031/aim.202000816.
[38] U. S. Department of Agriculture. (2019). FoodData central - raw spinach. FoodData Central. https://fdc.nal.usda.gov/fdc-app.html#/food-details/168462/nutrients.
[39] United Nations. (2020). FAOSTAT. value of agricultural production. http://www.fao.org/faostat/en/#data/QV.
[40] J. Valente, B. Sari, L. Kooistra, H. Kramer, S. Mücher, Automated crop plant counting from very high-resolution aerial imagery, Precis. Agric. (2020), https://doi.org/10.1007/s11119-020-09725-3, 0123456789.
[41] Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., & Girshick, R. (2019). Detectron2. Detectron2. https://github.com/facebookresearch/detectron2.
[42] Q. Yang, L. Shi, J. Han, J. Yu, K. Huang, A near real-time deep learning approach for detecting rice phenology based on UAV images, Agric. For. Meteorol. 287 (2020), 107938, https://doi.org/10.1016/J.AGRFORMET.2020.107938.
[43] Q. Yang, L. Shi, J. Han, Y. Zha, P. Zhu, Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images, Field Crops Res. 235 (2019) 142–153, https://doi.org/10.1016/j.fcr.2019.02.022.
[44] J. Yeom, J. Jung, A. Chang, M. Maeda, J. Landivar, Automated open cotton boll detection for yield estimation using unmanned aircraft vehicle (UAV) data, Remote Sens. 10 (12) (2018) 1895, https://doi.org/10.3390/rs10121895 (Basel).
[45] D. Yu, Y. Zha, L. Shi, X. Jin, S. Hu, Q. Yang, K. Huang, W. Zeng, Improvement of sugarcane yield estimation by assimilating UAV-derived plant height observations, Eur. J. Agron. 121 (2020), 126159, https://doi.org/10.1016/j.eja.2020.126159.
[46] Y. Yu, K. Zhang, L. Yang, D. Zhang, Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN, Comput. Electron. Agric. 163 (2019), 104846, https://doi.org/10.1016/j.compag.2019.06.001.
[47] T. Zhao, Y. Yang, H. Niu, Y. Chen, D. Wang, Comparing U-Net convolutional networks with fully convolutional networks in the performances of pomegranate tree canopy segmentation, In A. M. Larar, M. Suzuki, & J. Wang (Eds.), in: Proceedings of the Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VII 10780, 2018, https://doi.org/10.1117/12.2325570. Vol.SPIE.
[48] X. Zhou, H.B. Zheng, X.Q. Xu, J.Y. He, X.K. Ge, X. Yao, T. Cheng, Y. Zhu, W.X. Cao, Y.C. Tian, Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery, ISPRS J. Photogramm. Remote Sens. 130 (2017) 246–255, https://doi.org/10.1016/j.isprsjprs.2017.05.003.