Computers and Electronics in Agriculture 199 (2022) 107161

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity

Sébastien Dandrifosse [a,1], Elias Ennadifi [b,1], Alexis Carlier [a], Bernard Gosselin [b], Benjamin Dumont [c], Benoît Mercatoris [a,*]

[a] Biosystems Dynamics and Exchanges, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
[b] Information Signal and Artificial Intelligence Lab, Faculty of Engineering, University of Mons, 7000 Mons, Belgium
[c] Plant Sciences, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium

ARTICLE INFO

Keywords:
Wheat
RGB image
Counting
DeepMAC
YOLOv5

ABSTRACT

Recent deep learning methods have allowed important steps forward in the automatic detection of wheat ears in the field. Nevertheless, it was still lacking a method able to both count and segment the ears, validated at all the development stages from heading to maturity. Moreover, the critical step of converting the ear count in an image to an ear density, i.e. a number of ears per square metre in the field, has been widely ignored by most of the previous studies. For this research, wheat RGB images have been acquired from heading to maturity in two field trials displaying contrasted fertilisation scenarios. An unsupervised learning approach on the YOLOv5 model, as well as the cutting-edge DeepMAC segmentation method were exploited to develop a wheat ear counting and segmentation pipeline that necessitated only a limited amount of labelling work for the training. An additional label set including all the development stages was built for validation. The average F1 score of ear bounding box detection was 0.93 and the average F1 score of segmentation was 0.86. To convert the ear counts to ear densities, a second RGB camera was used so that the distance between the cameras and the ears could be measured by stereovision. That distance was exploited to compute the image footprint at ear level, and thus divide the number of ears by this footprint to get the ear density. The obtained ear densities were coherent regarding the fertilisation scenarios but, for a same fertilisation, differences were observed between acquisition dates. This highlights that the measurement was not able to retrieve absolute ear densities for all the development stages and conditions. The deep learning measurement considered the most reliable outperformed observations from three human operators.

1. Introduction

The ear density is one of the three yield components of wheat, which makes it an important trait to record. Its traditional measurement in the field relies on manual counting along a wooden stick. The method is slow, subject to human bias and may necessitate the operator to work in an uncomfortable position, leaning forward to isolate the tillers in a row, with the risk of back problems. Large uncertainties on that manual method were reported (Madec et al., 2019). For this reason, a major avenue of improvement in wheat phenotyping is to develop imaging methods to automate and objectify the ear counting task. To our knowledge, the first attempt of wheat ear detection in the field dates back to 1995 (Germain et al., 1995), followed by a period of sparse innovations dominated by classic image analysis and machine learning approaches (Chopinet and Cointault, 2006; Cointault et al., 2012, 2008a, 2008b; Zhu et al., 2016). From 2018, an impressive intensification of the wheat ear detection research was observed. Classic image analysis approaches continued to be proposed, but none of those studies was able to demonstrate a method robust for many wheat cultivars, development stages and light conditions. Indeed, most studies were validated on images from only one acquisition date (Alharbi et al., 2018; Dandrifosse et al., 2020; Tan et al., 2020; Zhou et al., 2018b, 2018a). Fernandez-Gallego et al. (2018) acquired images on two sites and at three development stages, but they discarded images acquired under

* Corresponding author.
E-mail addresses: sebastien.dandrifosse@uliege.be (S. Dandrifosse), elias.ennadifi@umons.ac.be (E. Ennadifi), alexis.carlier@uliege.be (A. Carlier), bernard.gosselin@umons.ac.be (B. Gosselin), benjamin.dumont@uliege.be (B. Dumont), benoit.mercatoris@uliege.be (B. Mercatoris).
[1] Authors share first-authorship.

https://doi.org/10.1016/j.compag.2022.107161
Received 30 November 2021; Received in revised form 8 June 2022; Accepted 17 June 2022

non optimal light conditions and development stages. Similarly, Fernandez-Gallego et al. (2020) acquired images at five dates but were able to validate their approach only on the images from the two dates of diffuse sky conditions. In parallel, the use of 3D sensors to detect ears has been studied to overcome the issue of illumination conditions, but only a few papers concern their implementation in the field (Saeys et al., 2009; Velumani et al., 2017), and these methods required expensive sensors.

In recent years, the field of image analysis has evolved due to the fast development of deep learning. Regarding wheat phenotyping, these algorithms were identified as the solution to propose a robust ear detection solution based on RGB images. A first deep learning approach was used by Pound et al. (2017) on wheat ears in controlled conditions. Then, in-field ear counting methods were proposed by Cao et al. (2020), Hasan et al. (2018), Lu and Cao (2020), Madec et al. (2019), Xiong et al. (2019), Xu et al. (2020), Yang et al. (2019) and Zhao et al. (2021). Each of those methods focused on a single or a few growth stages. In parallel, a major contribution occurred in 2020 when David et al. (2020) released their open access dataset: the Global Wheat Head Detection dataset (GWHD). Although other open datasets were released before, for example the SPIKE (Hasan et al., 2018) and the WED (Madec et al., 2019) datasets, the GWHD dataset was identified as the tipping point towards a new era of international data regrouping. This first version of the dataset gathered data from ten different places and obtained huge visibility thanks to an online competition. It was exploited for counting tasks by Ayalew et al. (2020), Fourati et al. (2021), Gong et al. (2020), Li et al. (2021a, 2021b), Wang et al. (2021), Wu et al. (2020) and Yang et al. (2021). A second version of the dataset was released to study the robustness of ear detection with respect to the development stage (David et al., 2021). New images were added and the development stages were organised according to four grades: post-flowering, filling, filling-ripening and ripening. Thanks to deep learning methods, the community started to master the ear counting task. It is however important to note that the ability to count the ears in images is not sufficient to get the ear density in the field, which is the variable of agronomic interest. To estimate the ear density, it is necessary to know the footprint of the image at ear height, as underpinned by David et al. (2020). As this height may vary from micro-plot to micro-plot the task is not so easy. Only a few partial solutions were proposed. Fernandez-Gallego et al. (2018) computed ear densities using the camera-canopy distance but is not detailed how this distance was measured. Madec et al. (2019) exploited a Light Detection And Ranging device (LiDAR) to measure that distance but without digging into the question of the reference height to compute the image footprint. Sadeghi-Tehran et al. (2019) placed a sheet of known dimensions in the image. This technique eliminates the camera-canopy distance issue but adds constraints to the acquisition system. Most of the other studies concerning ear detection did not include a way to measure the ear density. One of the reasons to explain this lack is that they often relied solely on existing datasets, such as the GWHD.

The interest of detecting ears goes beyond the counting and the estimation of density. It also implies the possibility to segment the ears in the images and build a mask that can be used for the extraction of further plant traits at the organ scale. Provided a mask of the ears, it is possible to study the wheat reflectance, related to its physiology, independently for the leaves and the ears. It is especially conceivable to exploit a mask built from RGB images to extract organ specific reflectance from registered multispectral images or even organ temperature from registered thermal images. Moreover, W. Li et al. (2021a, 2021b) demonstrated the impact of wheat ears on canopy reflectance, which supports the need for wheat ear segmentation methods. But ear counting and segmentation are two different tasks. The knowledge of ear bounding boxes, used for counting, does not directly provide the masks of the ears in those boxes. None of the previously mentioned ear counting studies tackled the segmentation step. However, as for counting, deep learning algorithms have been pointed out as a robust segmentation solution. A first deep learning segmentation was proposed on rice panicles by Xiong et al. (2017). For wheat, segmentation solutions were developed by Ma et al. (2020a, 2020b) and Wang et al. (2019). Methods allowing both counting and segmentation were detailed by Grbović et al. (2019), Sadeghi-Tehran et al. (2019) and Su et al. (2021). The limitation of all those studies is that each of them was validated for a few development stages. However, masks allowing to separate the ears from the leaves are needed throughout the season, especially if they aim at the extraction of physiological plant traits such as nitrogen status, senescence dynamics or water stress. Ear masks can also be used at various growth stages to look for the presence of diseases (Su et al., 2021). Additionally, they may provide ear morphological information such as ear width or length. In the laboratory, Pound et al. (2017) showed that it was even feasible to detect the spikelets. The measurement of such morphological information would probably exploit the ear masks only at late development stages, when the ears are bent from the weight of the grains, and thus imaged along their length.

This paper firstly proposes a robust deep learning approach able to count and segment the wheat ears. The performances for both tasks are evaluated at all the key development stages from heading to maturity. Secondly, it is detailed how the ear count in an image can be converted to an ear density in the field. Ear densities derived from the images are studied from heading to maturity and compared with (i) density measurements performed in the field at one date by three human operators and (ii) reference density measurements performed on harvested wheat plants.

2. Material and methods

2.1. In-field image acquisition

Images were acquired during the 2020 season in two field trials located in the Hesbaye area, Belgium (50° 32′ 40″ N and 4° 44′ 56″ E) on homogenous silty soil and a temperate climate. Trial 1 was planted with winter wheat (*Triticum aestivum L.* variety 'Mentor') on November 7th 2019. Trial 2 was planted with winter wheat (variety 'LG Vertikal') on November 5th 2019. Both trials were sowed with a density of 250 grains/m^2. The experimental micro-plots measured 1.95 × 6 m and the row spacing was 0.14 m. The micro-plots were fertilised three times, at BBCH stages 28, 30 and 39 (Meier, 2001), with 27 % ammonium nitrate. Trial 1 consisted of a randomised complete block design with eight replications of eight nitrogen fertilisation scenarios. Each scenario is described by three numbers, indicating the amount of nitrogen (kgN/ha) spread at each of the three inputs. The scenarios were: 0-0-0, 30-30-30, 40-40-40, 50-40-55, 60-60-60, 80-40-60, 90-30-60 and 105-105-105 kgN/ha. Trial 2 consisted of a randomised complete block design of four replications of sixteen regimes combining four fertilisation scenarios (40-40-40, 60-60-60, 80-60-60 and 100-80-80 kgN/ha) and four fungicide application scenarios (zero, one, two or three dates of fungicide treatment). Images of trial 1 were acquired on June 3rd, June 11th, June 15th, June 18th, June 23rd, July 7th and July 29th. Images of trial 2 were acquired on June 2nd, June 9th, June 16th, June 26th, July 7th, July 13th, July 22nd. Images were acquired by a pair of RGB cameras. At each date and for each camera, four images were taken by micro-plot except for half of trial 1 replicates, dedicated to punctual destructive measurements, for which only two images were taken.

The RGB cameras were GO-5000C-USB (JAI A/S, Copenhagen, Denmark), equipped with a CMOS sensor of 2560 × 2048 pixels and a LM16HC objective (Kowa GmbH, Düsseldorf, Germany). Their dynamic range was 60 dB. The focal length was 16 mm and the aperture was set to f/4.0. The two cameras were spaced from 50 mm. Their optical axes were theoretically parallel but small deviations were possible due to mechanical imperfections. The cameras were geometrically calibrated using 25 images of a 10 × 7 chessboard of 26 mm squares. The average error for the camera pair was 0.4 pixels.

The phenotyping platform was designed to capture nadir frames of wheat micro-plots. The camera pair was installed on a cantilever beam

to avoid shadows from the rest of the platform in the images. The height of the camera pair was adjusted at each acquisition date to keep a distance around 1.6 m from the top of the canopy. Using this distance, the theoretical footprint of the frames was 1.26 m^2 at the top of the canopy and the sampling distance was 0.49 mm/pixel. The images were recorded using a colour depth of 12 bits per pixel but reduced to 8 bits per pixel for this study. The auto-exposure algorithm was adapted to prevent image saturation. Starting from the exposure time computed by the manufacturer algorithm, the exposure time was gradually multiplied by a reduction factor, until no saturation was detected in the image or that the initial value was reduced by ten. The ISO and the aperture were not allowed to change.

2.2. Ear bounding box detection

2.2.1. Model

The deep learning model selected for the detection of ear bounding boxes was YOLOv5 (https://doi.org/10.5281/zenodo.3908559). YOLOv5 is the last born from the YOLO (You Only Look Once) family (Redmon et al., 2016). That model was picked because (i) it counted among the most recent and popular deep learning solutions, (ii) it is faster than other classical models such as models from the R-CNN family (Ren et al., 2017), and (iii) it already demonstrated good performances for wheat ear detection tasks (Yang et al., 2021; Zhao et al., 2021). The method to decide the final predictions among all the bounding box propositions was the weighted boxes fusion approach (Solovyev et al., 2021).

2.2.2. Image pre-processing and labelling

The images acquired during the 2020 season (Section 2.1) were complemented with the already labelled images of another dataset: the GWHD V2 (David et al., 2021). The images from the GWHD V2 were already in a square format (1024 × 1024 pixels), as required by the YOLOv5 algorithm. Since our images were acquired in the format of 2560 × 2048 pixels, they required some pre-processing steps. Considering the footprint of the image, if they were directly converted to 1024 × 1024 pixels the ears would have been too small compared to those in the GWHD V2. For this reason, the images were divided into four sub-images of 1280 × 1024 pixels, and then each sub-image was resized to 1024 × 1024 pixels. To label these sub-images, the LabelImg tool was used (https://github.com/tzutalin/labelImg). Two labelled sets were built from the 2020 sub-images: a set dedicated to validation and a set that aimed to complement the GWHD V2 for the model training. To build the validation set, at each acquisition date and for each trial, around forty sub-images were randomly selected and labelled, leading to a total of 64,091 labelled ears in 566 sub-images. Regarding the training set, images were picked at growth stages not present or under-represented in the GWHD V2: heading and maturity. One hundred and three sub-images were randomly picked among two dates at heading stage (June 3rd for trial 1, June 2nd for trial 2) and thirty-two sub-images were randomly selected among three dates at maturity stage (July 29th for trial 1, July 13th and July 22nd for trial 2). It led to a total of 13,566 labelled ears in 135 sub-images. All the labelled data, called the Contrasted-Fertilisation Wheat Ear Dataset 2020 (CFWED2020), are available online for future use in the community (https://doi.org/10.5281/zenodo.5709821).

2.2.3. Semi-supervised training

The training consisted of four steps. Firstly, the model was trained on the GWHD V2. Secondly, the trained model was used to predict ear bounding boxes on all the images acquired in 2020 (Section 2.1). All the predicted boxes were saved as pseudo-labels. Thirdly, both the labels from the GWHD V2 and the pseudo-labels were exploited to train the model again. Finally, the model was improved using transfer learning thanks to the labelled training set built from our 2020 images. That way, the last update of the model weights relied on trustworthy data.

2.2.4. Validation of bounding box detection

The performances of the bounding box detection were evaluated on the labelled validation set built from data acquired on trial 1 and 2 in 2020 (detailed in Section 2.3.2). The predictions of the model were compared with the labels to obtain a number of true positives, false positives and false negatives. A true positive (TP) is an ear that was correctly detected, a false positive (FP) is the detection of an ear that is not an ear, and a false negative (FN) is an ear that was not detected by the model but should have been. However, in that kind of object detection task, it is quite rare that the detected bounding box perfectly matches the labelled bounding box. For this, the definition of a correct or incorrect detection was based on the notion of Intersection over Union (IoU), i.e. the ratio between the area formed by the overlap of the detected box and the labelled box and the area formed by the set of these two boxes (Fig. 1). The TP, FP and FN are determined by choosing an IoU threshold. A value of 0.5 was chosen. This value is considered as the standard to evaluate a detection model. Moreover, it was pointed out as the optimal choice in the study of Madec et al. (2019).

From the TP, FP and FN, several other meaningful and widely-used indicators were built. The precision (Equation (1)) measures the fraction of correct detections among all the detections. The recall (Eq. (2)) measures the fraction of correct detections among all the objects that should have been detected. The accuracy (Eq. (3)) is an obvious performance metric. Its general formula contains the true negatives (TN) but in such object detection tasks, there is no TN. The F1 score (Eq. (4)) is the harmonic mean of precision and recall, which provides a robust model performance assessment.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F1score = 2\frac{Precision*Recall}{Precision + Recall} \tag{4}$$

Another useful indicator is the Average Precision (AP), which corresponds to the area under the precision-recall curve (PRC) (Equation (5)). This curve is obtained by plotting the precision versus the recall for various confidence levels of the network prediction. The PRC represents the influence of this confidence level on the relation between recall and precision.

$$AP@\alpha = \int_0^1 Precision(Recall)dRecall \tag{5}$$

where α is the IoU threshold for which precision and recall are determined, and *dRecall* is the differential of the recall. By averaging the AP obtained for each class of the object detection task, the mean average precision (mAP) is obtained (Equation (6)). However, in the case of this study, there was only one class and therefore the AP was identical to the mAP.

$$mAP@\alpha = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{6}$$

where n is the number of classes in the object detection problem.

Two mAP metrics were considered: the mAP@0.5 and the mAP@0.5:0.75. The mAP@0.5 is the AP with an IoU threshold of 0.5. The mAP@0.5:0.75 is the mean of AP values for thresholds ranging from 0.5 to 0.75 with a step of 0.05. One of the interests of building so many various indicators is to increase the possibility to compare on a common ground the performances with other studies.

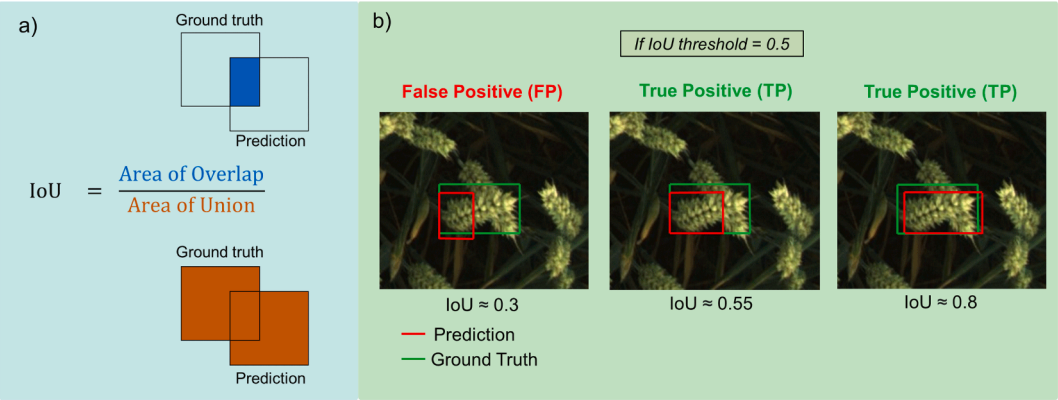

Fig. 1. a) Concept of Intersection over Union (IoU). b) Concept of false positive and true positive applied to bounding box detection.

2.3. Ear segmentation

2.3.1. Automatic segmentation using DeepMAC

The state-of-the-art DeepMAC (Deep Mask-heads Above CenterNet) neural network model (Birodkar et al., 2021) was used to segment the ears in the bounding boxes. The huge advantage of this approach is that it did not require manual construction of training masks, which is even more arduous and time-consuming work than building bounding boxes. Thanks to the strong generalisation ability of the pre-trained DeepMAC model, no specific training was needed to segment the wheat ears in their bounding boxes. Ear masks were generated for each square sub-image (Fig. 2). Then, the four sub-masks obtained associated with an image were transformed back to the original sub-image size of 1280 × 1024 pixels and brought together to form a mask of 2560 × 2048 pixels, having the same format as the original RGB image. The ears cut at the sub-image junctions were reconstituted when assembling the sub-masks (Fig. 3). However, it happened that the parts of such cut ears did not match to the pixel, especially when the mask predicted by DeepMAC did not extend to the edge of the image. To solve this issue, pixels lying in a 10 pixel range between two ear pixels from either side of a junction were considered as belonging to the ear mask. This filling algorithm has visually proven its usefulness. The risk was not excluded to regroup ear parts that did not belong to the same ear but the probability of such errors was judged small enough to be neglected.

2.3.2. Validation of segmentation

A custom annotation tool was created (Carlier et al., 2022). The method consisted in annotating eighteen pixels at fixed coordinates in each RGB image. The tool automatically zoomed on each pixel. Then, the operator pushed one of three buttons to attribute a class to this pixel: class 1 for the background (soil, leaves, stems, …), class 2 for the ears, and class 3 if it was not possible to decide between class 1 or 2, for example if the operator did not distinguish well what the pixel represented or if the pixel was located at the edge between an ear and the background. That procedure was executed for all the images of all the acquisition dates.

2.4. Ear counts and density

2.4.1. Automatic method

The sum of the bounding boxes in each of the four sub-images provided a biased number of ears for the whole image because some ears were cut between two sub-images and thus counted twice. Those problematic ears were however reconstituted when gathering the four sub-masks. The excess ear parts were counted by the difference between the sum of the objects in the sub-masks and in the whole mask. The total number of ears was corrected by subtracting the excess ear parts.

The ear density, expressed in ears per square metre, is the ratio of the number of ears in the image to the footprint of the image at ear height. That footprint was obtained by Equation (7):

$$footprint\ \ ears = 4(z + 0.05)^2 \tan\left(\frac{HFOV}{2}\right)\tan\left(\frac{VFOV}{2}\right) \quad (7)$$

where *footprint ears* (m^2) is the footprint at ear height, *HFOV* (°) is the horizontal field of view of the camera, *VFOV* (°) is the vertical field of view of the camera and z (m) is the distance between the camera and the tops of the ears. That distance was automatically measured by stereovision, exploiting the shift between the images from the two RGB cameras. Detailed explanations on the stereovision process can be found in Dandrifosse et al. (2020). For this study, the stereovision functions came from the OpenCV-Python library (version 4.5.3.56) (Bradski and Kaehler, 2008). Based on the geometrical calibration of the camera pair, the left and right images were rectified. To accelerate the matching and improve performances, they were binned from 2560 × 2048 pixels to 1280 × 1024 pixels. The stereo matching was then performed by the Semi-Global Block Matching algorithm (Hirschmüller, 2007) configured with a block size of 5. Post-filtering used a uniqueness value of 10, a speckle range of 4 and a speckle window size of 50. That matching step yielded for each pixel a disparity value or a value indicating that reliable disparity could not be computed. A second round of post-filtering relied on the Weighted Least Squares (WLS) filter (Min et al., 2014) from OpenCV-Python to smooth the disparity map and fill the gaps, helped by the content of the RGB image. For the last step of the stereovision process, disparity values were converted to depth values, knowing the focal length and the distance between the cameras. By subtracting the depth values to the camera height, it was also possible to compute a height map (Fig. 4). The ear mask from the segmentation step (Section 2.4.1) was applied on the depth map obtained by stereovision to produce a map of ear depths. As the ears were vertical most of the time, it was considered that the depth points were located at the tops of the ears. The distance z in Equation (7) was the median of ear depths. It was increased by 0.05 m to account for the size of the ears, and thus estimate the image footprint in the middle of the ear layer.

2.4.2. Manual methods

Ears were counted by humans on July 6th, 2020 for the 64 micro-plots of trial 1. All the micro-plots were assessed by three human operators. For each micro-plot, an operator performed three counts, at about a quarter, half and three-quarter the length of the micro-plot. The row and the exact spot of the count were randomly selected by blindly dropping a wooden stick. This selection never occurred in the two first rows of the plot in which the ear density could have been influenced by a border effect. The stick had a length of 50 cm and was positioned along a row, at the basis of the tillers. All fully grown tillers on the row were considered to carry an ear, and counted. Ears were also counted in the laboratory on wheat samples from trial 1 harvested on July 29th. Each

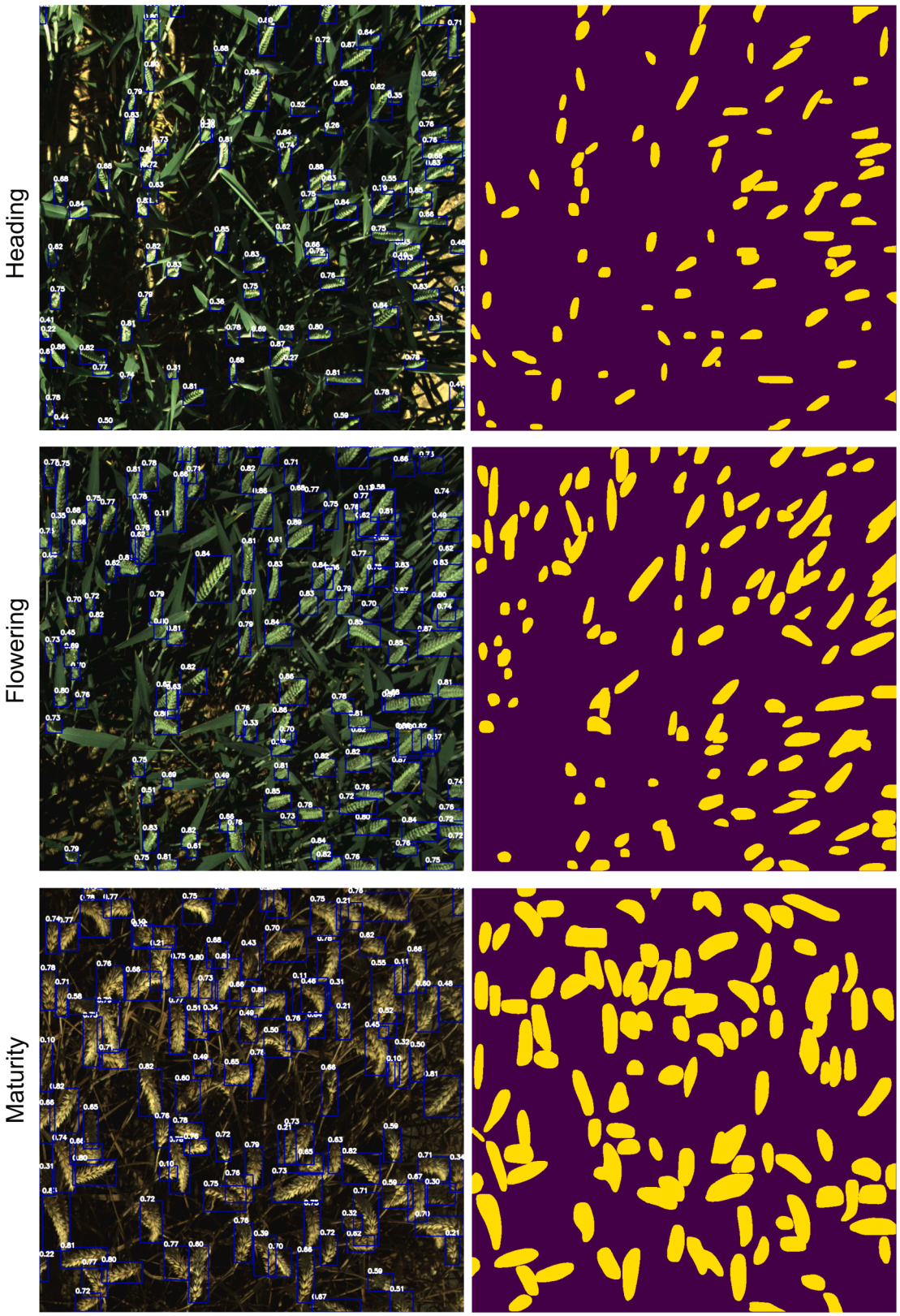

Fig. 2. Examples of ear bounding box detection and segmentation. The presented sub-images were randomly selected among the trial 1 images from three dates: June 3rd (heading), June 15th (flowering) and July 29th (maturity). The size of the sub-images is 1024 × 1024 pixels.

sample was collected over 50 cm along three rows located halfway across the width of the plot. Knowing the row spacing, both in-field and post-harvest counts were converted to ear densities.

2.4.3. Statistical analyses of differences in ear density

A two-way analysis of variance with repeated measures was used for each trial to investigate the effect of date and fertilisation on the ear density measured by the automatic method. The pingouin Python

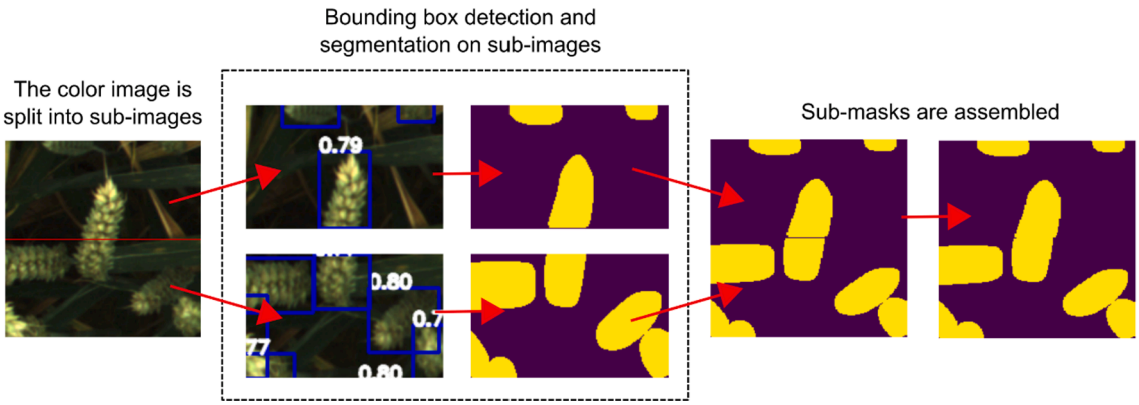

Fig. 3. Pipeline of sub-image treatment. The process is illustrated for a zone crossed by a sub-image delimitation line. For the steps illustrated in the dotted rectangle, the size of the sub-images is slightly different because they had been resized to a square format.

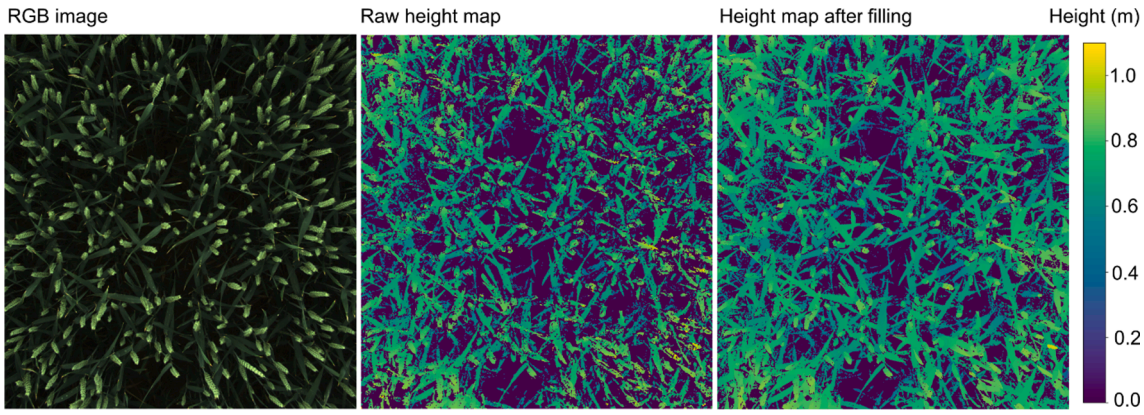

Fig. 4. Example of wheat height map. The scene was imaged on June 18th on a micro-plot of trial 1. This **Fig.** shows the RGB image cropped to the zone commonly observed by the two cameras, the height map from the raw disparity map, and the height map when the disparity map was filled with the WLS method. The dark blue zones, at 0 on the scale, corresponded either to soil or to plant pixels for which the height could be computed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

library (version 0.5.0) was used for the implementation (Vallat, 2018). The within-subjects factor was the date and the between-subjects factor was the fertilisation. The interaction between the factors was significant so a one-way analysis of variance was performed for each date to compare the fertilisation scenarios. Post-hoc Tukey HSD tests were used.

A two-way analysis of variance was also led to investigate the effect of the measurement method on the ear density of trial 1. The two factors were the fertilisation scenario and the measurement method. This last factor had five levels: the automatic method, the reference manual method and the three human observations in the field. The date was not a factor as each manual measurement was performed at one date, and the automatic measurement was considered only for July 7th, which is the date closest to the human field observations. The interaction between the factors was significant so a one-way analysis of variance was performed for each fertilisation scenario to compare the measurement methods. Post-hoc Tukey HSD tests were used. All the tests in this study had a significance level of 0.05.

As the fields were quite homogeneous and the disease pressure was weak (in particular no fusarium head blight was spotted), random blocks and fungicide treatments were not considered as factors for the analyses. The purpose of those analyses was to get some clues to evaluate the ear density measurement methods, not to explain the smallest variations of ear density in our fields.

3. Results and discussion

3.1. Evaluation of ear bounding box detection

The performances of ear bounding box detection for the YOLOv5 model are detailed in Table 1. Depending on the date, the mAP@0.5 varied between 0.79 and 0.94 and the F1 score between 0.89 and 0.97, with the best performances reached for both trials at flowering stage. One detail that could ease detection at that stage is the presence of anthers on the ears.

The selection of images to label was random at each date. An improvement would be to ensure labelling the same amount of images for each fertilisation scenario. That way, detection performances could be compared for different fertilisation practices.

3.2. Segmentation quality

The performances of ear segmentation are detailed in Table 2. As for bounding box detection (Section 3.1), the lowest performances were observed at heading and maturity stages, but even in these cases the F1 scores were close or superior to 0.75, which is considered good. The performances in this study were superior to the ones announced by Carlier et al. (2022) using a non deep learning segmentation technique on a part of the same dataset.

The DeepMAC model, thanks to its strong generalisation ability, performed ear segmentation without having been trained for this

Table 1
Evaluation of bounding box detection for all the labelled images from the acquired data set. The indicators are presented for each acquisition date, expressed as days after heading (DAH). The heading date for both trials was May 28th.

DAH	BBCH stage	Trial	TP (%)	FP (%)	FN (%)	Precision	mAP@0.5	mAP @0.5:0.75	Accuracy	F1 score
5	55: Heading	2	84.67	5.93	9.4	0.93	0.84	0.79	0.85	0.92
6	59: Heading	1	87.53	3.76	8.71	0.96	0.87	0.86	0.88	0.93
12	65: Flowering	2	94.25	2.24	3.51	0.98	0.94	0.93	0.94	0.97
14	67: Flowering	1	92.99	2.5	4.51	0.97	0.93	0.93	0.93	0.96
18	69: Flowering	1	90.95	2.92	6.13	0.97	0.91	0.90	0.91	0.95
19	69: Flowering	2	89.54	4.17	6.29	0.96	0.89	0.86	0.90	0.94
21	71: Watery ripe	1	91.47	3.35	5.18	0.96	0.92	0.91	0.91	0.96
26	75: medium milk	1	87.09	4.23	8.68	0.95	0.87	0.85	0.87	0.93
29	77: Late milk	2	91.68	3.31	5.01	0.97	0.92	0.88	0.92	0.96
40	83: Early dough	1	87.81	4.04	8.15	0.96	0.87	0.85	0.88	0.94
40	83: Early dough	2	87.66	4.13	8.21	0.96	0.88	0.87	0.88	0.93
46	85: Soft dough	2	84.84	4.75	10.41	0.95	0.84	0.82	0.85	0.92
55	89: Maturity	2	79.73	6.42	13.85	0.93	0.79	0.76	0.8	0.89
62	89: Maturity	1	81.61	5.35	13.04	0.94	0.81	0.79	0.82	0.9

Table 2
Evaluation of segmentation for all the images acquired in this study. The indicators are presented for each acquisition date, expressed as days after heading (DAH). The heading date for both trials was May 28th.

DAH	BBCH stage	Trial	TP (%)	FP (%)	FN (%)	TN (%)	Precision	Accuracy	F1 score
5	55:Heading	2	4.01	0.78	1.80	93.41	0.84	0.97	0.76
6	59: Heading	1	4.73	0.72	2.50	92.05	0.87	0.97	0.75
12	65: Flowering	2	7.14	0.72	0.18	91.95	0.91	0.99	0.94
14	67: Flowering	1	7.60	0.78	0.24	91.38	0.91	0.99	0.94
18	69: Flowering	1	10.98	1.39	1.07	86.57	0.89	0.98	0.90
19	69: Flowering	2	9.27	0.71	1.81	88.21	0.93	0.97	0.88
21	71: Watery ripe	1	10.91	0.94	1.50	86.66	0.92	0.98	0.90
26	75: medium milk	1	9.04	1.47	1.01	88.48	0.86	0.98	0.88
29	77: Late milk	2	10.46	2.94	0.39	86.21	0.78	0.97	0.86
40	83: Early dough	1	15.41	3.45	0.67	80.48	0.82	0.96	0.88
40	83: Early dough	2	17.21	0.86	2.46	78.98	0.95	0.97	0.91
46	85: Soft dough	2	9.21	5.83	0.79	84.17	0.61	0.93	0.74
55	89: Maturity	2	26.08	2.40	7.58	63.94	0.92	0.90	0.84
62	89: Maturity	1	22.01	2.40	4.61	70.98	0.90	0.93	0.86

specific task. There is however a downside to this ability: sometimes the algorithm segmented another object in the bounding box, rather than the ear. This phenomenon is illustrated in Fig. 5. To avoid those issues, future ear segmentation pipelines should include some training specific to the ear detection task, or at least some criteria to exclude the wrong bounding box segmentations.

3.3. From ear count to ear density: A non-trivial conversion

The ear densities were computed by dividing the number of ears detected in an image by the footprint of the image at ear height. The temporal evolution of the measured ear density is presented in Fig. 6 for the two trials and considering the different fertilisation scenarios. The two-way analysis of variance showed for both trials an interaction between the two factors: date and fertilisation scenario. For that reason, the fertilisation scenarios were statistically compared separately for each date. Almost all of the corresponding one-way variance analyses showed a significant effect of the fertilisation on ear density. The results of the post-hoc Tukey HSD test are presented in Table 3 and 4, respectively for trial 1 and 2.

The measured ear densities showed significant increases with the fertilisation level of the crop (Tables 3 and 4), which is an expected

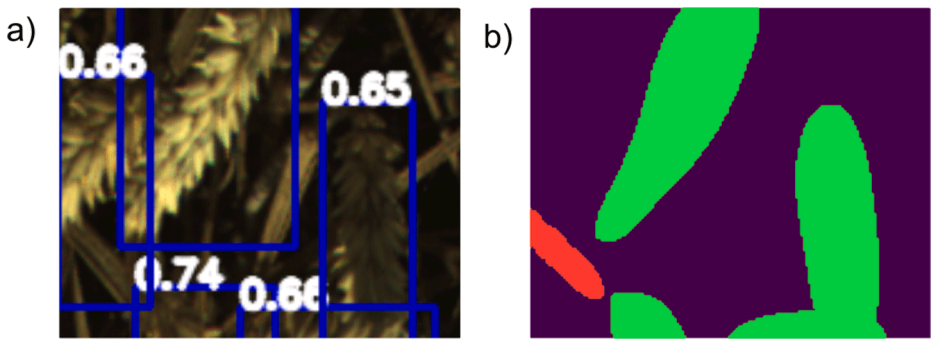

Fig. 5. Example of wrong segmentation. a) RGB image with the detected bounding boxes. b) Ear mask with correct mask parts in green and erroneous parts in red. In the bounding box on the left, DeepMAC isolated a leaf rather than the ear. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

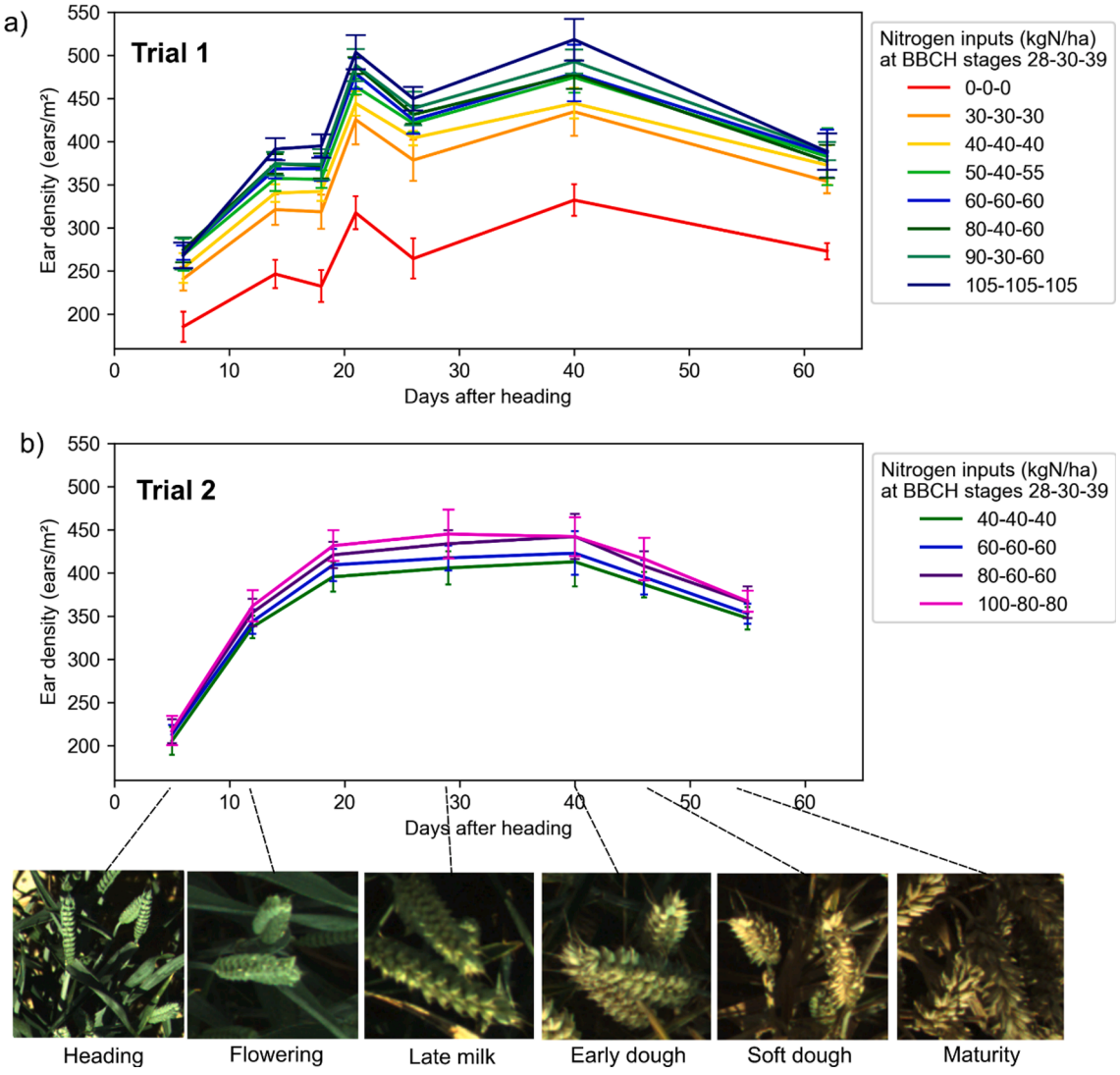

Fig. 6. Temporal evolution of the measured ear density for trial 1 and trial 2. The colour code indicates the different fertilisation levels. The standard deviation is indicated on each point. Various cap sizes are used to differentiate the standard deviation bars from different fertilisation levels.

Table 3
Tukey HSD test to investigate the effects of the fertilisation scenarios of trial 1 on the wheat ear density measured by the automatic method. The test was performed for each date, expressed in days after heading (DAH), corresponding to one row in the table. On each row, fertilisation scenarios marked with at least the same letter are not considered statistically different. The significance level is 0.05.

DAH	Nitrogen inputs (kgN/ha) at BBCH stages 28-30-39							
	0-0-0	30-30-30	40-40-40	50-40-55	60-60-60	80-40-60	90-30-60	105-105-105
6	c	b	ab	a	a	a	a	a
14	e	d	cd	bc	b	ab	ab	a
18	e	d	cd	bc	b	ab	ab	a
21	e	d	cd	bc	ab	ab	ab	a
26	e	d	cd	bc	abc	ab	ab	a
40	e	d	cd	bc	b	bc	ab	a
62	b	b	ab	ab	ab	ab	a	a

Table 4
Tukey HSD test to investigate the effects of the fertilisation scenarios of trial 2 on the wheat ear density measured by the automatic method. The test was performed for each date, expressed in days after heading (DAH), corresponding to one row in the table. On each row, fertilisation scenarios marked with at least the same letter are not considered statistically different. The significance level is 0.05.

DAH	Nitrogen inputs (kgN/ha) at BBCH stages 28-30-39			
	40-40-40	60-60-60	80-60-60	100-80-80
5	a	a	a	a
12	c	bc	a	a
19	c	bc	ab	a
29	c	bc	ab	a
40	b	ab	a	a
46	c	bc	ab	a
55	c	b	ab	a

behaviour (Oscarson, 2000): the first nitrogen input, at tillering, favours the tiller number while the second nitrogen input, at stem elongation, allows the tillers to grow and develop the ear. This effect of fertilisation highlights the importance of ear density measurement. At almost all of the growth stages, the method was useful to identify relative ear density differences between fertilisation scenarios. Nevertheless, for a given fertilisation object, the measured ear density varied between the dates (Fig. 6). Yet, no incident such as lodging was observed in the field and thus the number of ears was not supposed to change. That said, some of

the ear density variations were expected. The smaller densities measured at heading can be explained by the fact that not all ears did emerge from their sheaths yet. The smaller densities measured at maturity can be explained by the change of ear morphology and position: at this stage, the ears were larger and bent from the weight of the grains, which increased the number of ears totally hidden in the image. In addition, the bounding box detection at those stages was not as good as for the others (Table 1), probably because of many ear overlaps. That lower performance is translated by a number of ears that were not detected by the model, leading to an underestimation of the ear density.
A more troubling point for trial 1 is the variation of the measured ear density at the other development stages, where the ears were well visible and the bounding box detection presented excellent performances with mAP@0.5 close or superior to 0.9.

A first hypothesis to explain those ear density differences across dates is the effect of scene lighting. This may be hard to believe since the deep learning algorithm succeeded well in detecting the ears in both sunny and cloudy conditions. However, the bounding box detection performances could be a misleading track. Direct sunlight induces shadows and strong contrasts in wheat canopy images. In those conditions, some low ears could stand in dark areas and be missed at both the human annotation and the automatic detection step. An observation of this kind was made by Madec et al. (2019). Looking at the trial 1 data in Fig. 6, the ear densities measured at 26 days after heading (June 23rd) in sunny conditions are lower than the densities measured at the dates after and before, in cloudy conditions. The ear densities measured at 18 days after sowing are also lower than expected, and this coincides one again with sunny acquisitions. A solution to avoid ears missed by both human image annotators and the deep learning algorithm would be to deal with the high contrasts responsible for image zones either too dark or saturated. It could be achieved by acquiring the same scene using various exposure times to build high dynamic range images.

A second hypothesis to explain the ear density differences across dates is the variability inherent to the image footprint estimation. That footprint depends on the distance between the cameras and the ears (Fig. 7), which may vary depending on the wheat varieties, the fertilisation level or even the depth of the ruts between the micro-plots in which the image acquisition platform was placed. For this reason, the distance needs to be measured for each imaged wheat canopy zone. But, as highlighted by David et al. (2020), the definition of this distance is still an open question. Which depth in the ear layer should be considered to compute the image footprint? And how to measure that depth? In this study, it was considered the median depth of the tops of the ears plus 0.05 m, but this choice was quite arbitrary and did not account for the possible inclination of the ears. The curve in Fig. 7 (b) illustrates the variations of image footprint expected for a modification of the distance between the cameras and the ears.

3.4. Human and deep learning measurements of ear density

The comparison between the ear density measured by the algorithms, the human operators in the field and the reference measurement from samples counted in the laboratory is presented in Fig. 8 for the eight fertilisation scenarios of trial 1. The highest ear densities were recorded by the reference method. For all the scenarios, operator 3 recorded the lowest values. The two-ways variance analysis showed an interaction between the two factors: measurement method and fertilisation scenario. For that reason, the methods were statistically compared separately for each fertilisation level. Almost all of the corresponding one-way variance analyses showed a significant effect of the measurement method. The results of the post-hoc Tukey HSD tests are presented in Table 5. In most of the scenarios, operators 1 and 3 were significantly different from the reference, while operator 2 was only different from the reference in one scenario. The significant differences between the operators demonstrated an operator bias. Deep learning was never judged different from operator 2 or the reference. Both deep learning and operator 2 were reliable but, looking at their mean in each scenario (Fig. 8) and the number of scenarios in which they were assumed equal to the reference (Table 5), deep learning was considered a better estimator.

The ear densities provided by the reference method were the highest in seven fertilisation scenarios out of eight, but that so-called “reference” method may itself be questioned. The ear density was computed using the ratio of the number of ears counted after harvest to the harvested area in the field. This harvested area was delimited by the wheat rows and a 50-cm stick. It is possible that some or all the human operators tended to harvest a few additional tillers, planted at the boundaries of the harvest zone, leading to an overestimation of the ear density. It is also to be noted that the reference harvest and the human observations in the field included some small late ears that may not have been visible in the images.

It is important to consider the heterogeneity within the micro-plots. Both the counting in the field and the counting on harvested plants were based on small zones in the micro-plots. On the contrary, the images allowed measurements on larger zones and all along the micro-plots. Looking at Fig. 8, this may explain the smaller standard deviations observed for the deep learning method. To allow a better comparison between the deep learning and reference measurements, Madec et al. (2019) suggested working on the exact same zones in the micro-plots. They also proposed to increase the size of the sampled reference zones, which is a laborious task.

For deep learning, a critical question is the choice of the acquisition date to obtain the ear density values. In this study, the method did not yield the same ear densities at all the dates. July 7th, at early dough stage, was used for the comparison with other methods because the date

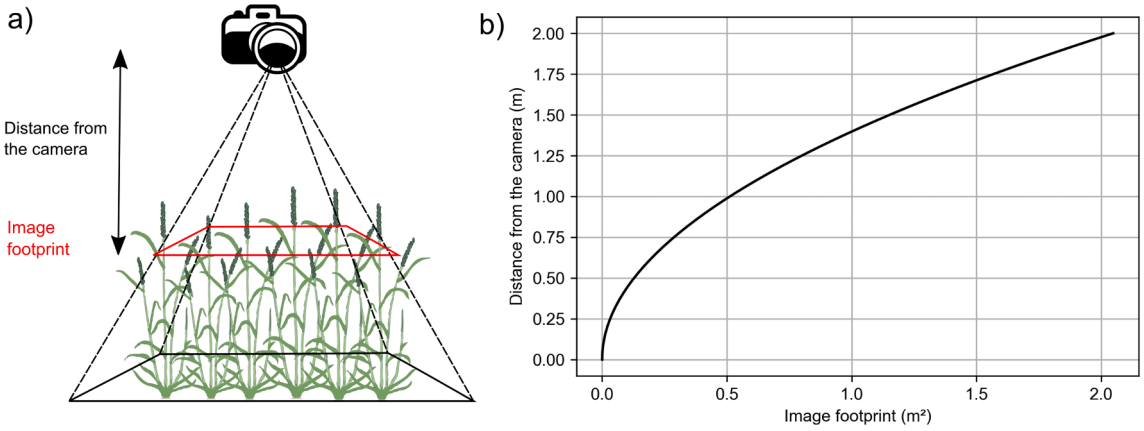

Fig. 7. a) Illustration of the notion of image footprint at ear height. b) Relation between image footprint at ear height and the distance between the ears and the camera.

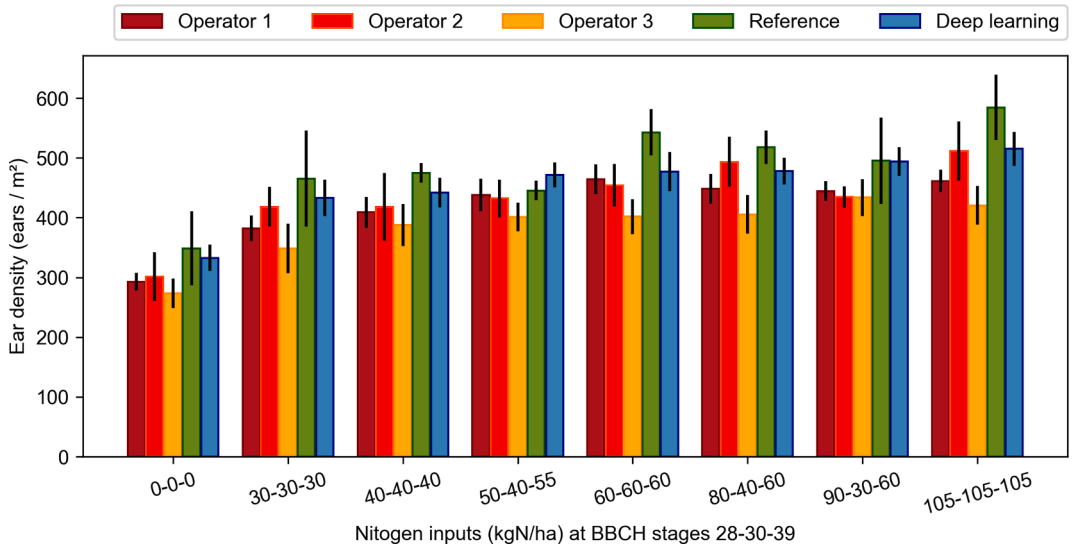

Fig. 8. Comparison of the trial 1 ear densities estimated by three operators in the field (July 6th), the deep learning algorithm (images from July 7th) and the reference measurement carried out on wheat samples after the harvest. The standard deviation is indicated on each bar.

Table 5
Tukey HSD test to investigate the effect of the measurement method on wheat ear density. The test was performed for each fertilisation level, corresponding to one row in the table. On each row, measurement methods marked with at least the same letter are not considered statistically different. The significance level is 0.05.

Nitrogen inputs (kgN/ha) at BBCH stages 28-30-39	Measurement method				
	Operator 1	Operator 2	Operator 3	Deep learning	Reference
0-0-0	ab	ab	b	a	a
30-30-30	ab	a	b	a	a
40-40-40	bc	abc	c	ab	a
50-40-55	ab	ab	b	a	ab
60-60-60	b	bc	c	ab	a
80-40-60	bc	ab	c	ab	a
90-30-60	a	a	a	a	a
105-105-105	bc	ab	c	ab	a

was close to the counting made by the human operators (July 6th). It was also considered as the most reliable date because performances of ear detection were excellent, the sky was cloudy and ear densities measured agreed with other reliable measurements sooner in the season. Nevertheless, values from another date would have modified the comparison with human measurements. This latter point highlights that, if human operator measurements can show ear density discrepancies on a same micro-plot, so does the automatic method. In the future, it will be important to work on the weak points of the method: the estimation of image footprint and the ears non visible because of shadows or overlaps. As only one trustworthy ear density measurement is needed for a season, the camera operators have the luxury to select convenient conditions or growth stages.

4. Conclusions

The combination of the YOLOv5 bounding box detection model and the DeepMAC segmentation model brought an innovative solution for both counting and segmenting the wheat ears. Thanks to the semi-supervised learning of the bounding box detection model, the use of a rich existing dataset and the strong generalisation ability of the segmentation model, few efforts were necessary to train the whole method and adapt it to reach robust performances at all development stages of wheat from heading to maturity. Most of the labelling work has been dedicated to build a strong validation dataset. All the labels were made available to the community. The average F1 score of bounding box detection was 0.93 and the average F1 score of segmentation was 0.86. A limitation of this study is the diversity of the acquired data set. It covered all the development stages of the crops from heading to maturity and included contrasted fertilisation scenarios, but it was limited to one growing season and two varieties. An improvement would be to acquire similar time series on a larger diversity of wheat varieties, especially including ears with awns. It would also be interesting to investigate the detection performances on ears from other cereal crops such as barley or rye.

A method was proposed to convert ear counts to ear density, i.e. the number of ears per square metre, which is the variable of high agronomic interest and which has been widely neglected in previous studies dedicated to ear detection, limited to the ear counts. The method relied on a second RGB camera to determine the distance between the cameras and the ears by stereovision and thus compute the image footprint at ear height. The most reliable deep learning ear density values outperformed observations from human operators in the field. They had the advantage to better capture the spatial heterogeneity of the micro-plots compared to human observations and reference harvests that were very localised. Further studies should not focus solely on ear counting methods based on existing datasets but also set field experiments and investigate the retrieval of ear density. A particular focus should be put to establish a robust reference distance between the camera and the ears to compute the image footprint.

Funding

This work was supported by the National Fund of Belgium FNRS-F.R.S. (FRIA grant), and the Agriculture, Natural Resources and Environment Research Direction of the Public Service of Wallonia (Belgium), project D31-1385 PHENWHEAT.

CRediT authorship contribution statement

Sébastien Dandrifosse: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Elias Ennadifi:** Conceptualization, Methodology, Software, Validation, Formal analysis,

Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Alexis Carlier:** Methodology, Validation, Resources, Writing – review & editing. **Bernard Gosselin:** Supervision, Project administration, Funding acquisition. **Benjamin Dumont:** Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Benoît Mercatoris:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

Acknowledgment

The authors thank the research and teaching support units Agriculture Is Life and Environment Is Life of TERRA Teaching and Research Centre, University of Liège for giving access to the field trials and providing the sensors. The authors are grateful to Jesse Jap, Sohaib Laraba, Arnaud Bouvry, Julien Kirstein, Amandine Collin, Abdelkbir Masrour, Françoise Thys and Rudy Schartz for their help.

References

Alharbi, N., Zhou, J., Wang, W., 2018. Automatic Counting of Wheat Spikes from Wheat Growth Images, in. In: Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, pp. 346–355. https://doi.org/10.5220/0006580403460355.

Ayalew, T.W., Ubbens, J.R., Stavness, I., 2020. Unsupervised Domain Adaptation for Plant Organ Counting. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 12540 LNCS, 330–346. https://doi.org/10.1007/978-3-030-65414-6_23.

Birodkar, V., Lu, Z., Li, S., Rathod, V., Huang, J., 2021. The surprising impact of mask-head architecture on novel class segmentation. arXiv abs/2104.0.

Bradski, G., Kaehler, A., 2008. Learning OpenCV. O'Reilly Media Inc, Newton, MA, USA. https://doi.org/10.1109/MRA.2009.933612.

Cao, L., Zhang, X., Pu, J., Xu, S., Cai, X., Li, Z., 2020. The Field Wheat Count Based on the Efficientdet Algorithm. In: Proc. 2020 IEEE 3rd Int. Conf. Inf. Syst. Comput. Aided Educ. ICISCAE 2020 557–561. https://doi.org/10.1109/ICISCAE51034.2020.9236918.

Carlier, A., Dandrifosse, S., Dumont, B., Mercatoris, B., 2022. Wheat Ear Segmentation Based on a Multisensor System and Superpixel Classification. Plant Phenom. 2022, 1–10. 10.34133/2022/9841985.

Chopinet, B., Cointault, F., 2006. Colour-texture image analysis for in-field wheat head counting. In: Second International Symposium on Communications, Control, and Signal Processing (ISCCSP 2006). pp. 13–15.

Cointault, F., Guerin, D., Guillemin, J.P., Chopinet, B., 2008a. In-field triticum aestivum ear counting using colour-texture image analysis. New Zeal. J. Crop Hortic. Sci. 36, 117–130. https://doi.org/10.1080/01140670809510227.

Cointault, F., Journaux, L., Miteran, O., Destain, M.F., 2008b. Improvements of image processing for wheat ear counting. Agricultural and Biosystems Engineering for a Sustainable World. International Conference on Agricultural Engineering.

Cointault, F., Journaux, L., Rabatel, G., Germain, C., Ooms, D., Destain, M.-F., Gorretta, N., Grenier, G., Lavialle, O., Marin, A., 2012. Texture, Color and Frequential Proxy-Detection Image Processing for Crop Characterization in a Context of Precision Agriculture. In: Aflakpui, G. (Ed.), Agricultural Science. INTech, London, pp. 49–70.

Dandrifosse, S., Bouvry, A., Leemans, V., Dumont, B., Mercatoris, B., 2020. Imaging wheat canopy through stereo vision : overcoming the challenges of the laboratory to field transition for morphological features extraction. Front. Plant Sci. 11, 1–15. https://doi.org/10.3389/fpls.2020.00096.

David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., Kirchgessner, N., Ishikawa, G., Nagasawa, K., Badhon, M.A., Pozniak, C., de Solan, B., Hund, A., Chapman, S.C., Baret, F., Stavness, I., Guo, W., 2020. Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high resolution RGB labelled images to develop and benchmark wheat head detection methods. Plant Phenom. 2020, 1–12. 10.34133/2020/3521852.

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., Wang, X., Pinto, F., Shafiee, S., Tahir, I.S.A., Tsujimoto, H., Nasuda, S., Zheng, B., Kirchgessner, N., Aasen, H., Hund, A., Sadhegi-Tehran, P., Nagasawa, K., Ishikawa, G., Dandrifosse, S., Carlier, A., Dumont, B., Mercatoris, B., Evers, B., Kuroki, K., Wang, H., Ishii, M., Badhon, M.A., Pozniak, C., LeBauer, D.S., Lillemo, M., Poland, J., Chapman, S., de Solan, B., Baret, F., Stavness, I., Guo, W., 2021. Global Wheat Head Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods. Plant Phenom. 2021, 1–9. 10.34133/2021/9846158.

Fernandez-Gallego, J.A., Kefauver, S.C., Gutiérrez, N.A., Nieto-Taladriz, M.T., Araus, J. L., 2018. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. Plant Methods 14, 1–12. https://doi.org/10.1186/s13007-018-0289-4.

Fernandez-Gallego, J.A., Lootens, P., Borra-Serrano, I., Derycke, V., Haesaert, G., Roldán-Ruiz, I., Araus, J.L., Kefauver, S.C., 2020. Automatic wheat ear counting using machine learning based on RGB UAV imagery. Plant J. 103, 1603–1613. https://doi.org/10.1111/tpj.14799.

Fourati, F., Mseddi, W.S., Attia, R., 2021. Wheat Head Detection using Deep, Semi-Supervised and Ensemble Learning. Can. J. Remote Sens. 47 (2), 198–208.

Germain, C., Rousseaud, R., Grenier, G., 1995. Non destructive counting of wheat ear with picture analysis. In: Fifth International Conference on Image Processing and Its Applications. IET, pp. 435–439. https://doi.org/10.1049/cp:19950696.

Gong, B., Ergu, D., Cai, Y., Ma, B., 2020. Real-Time Detection for Wheat Head Applying Deep Neural Network. Sensors 21, 191. https://doi.org/10.3390/s21010191.

Grbović, Z., Panić, M., Marko, O., Brdar, S., Crnojević, V., 2019. Wheat Ear Detection in RGB and Thermal Images Using Deep Neural Networks. In: International Conference on Machine Learning and Data Mining, MLDM 2019. ibai, Fockendorf, Germany.

Hasan, M.M., Chopin, J.P., Laga, H., Miklavcic, S.J., 2018. Detection and analysis of wheat spikes using Convolutional Neural Networks. Plant Methods 14, 1–13. https://doi.org/10.1186/s13007-018-0366-8.

Hirschmüller, H., 2007. Stereo Processing by Semi-Global Matching and Mutual Information. In: IEEE Transactions On Pattern Analysis and Machine Intelligence. IEEE, Manhattan, NY, USA, pp. 328–341.

Li, J., Li, C., Fei, S., Ma, C., Chen, W., Ding, F., Wang, Y., Li, Y., Shi, J., Xiao, Z., 2021a. Wheat ear recognition based on retinanet and transfer learning. Sensors 21 (14), 4845.

Li, W., Jiang, J., Weiss, M., Madec, S., Tison, F., Philippe, B., Comar, A., Baret, F., 2021b. Impact of the reproductive organs on crop BRDF as observed from a UAV. Remote Sens. Environ. 259, 112433.

Lu, H., Cao, Z., 2020. TasselNetV2+: A Fast Implementation for High-Throughput Plant Counting From High-Resolution RGB Imagery. Front. Plant Sci. 11, 1–15. https://doi.org/10.3389/fpls.2020.541960.

Ma, J., Li, Y., Du, K., Zheng, F., Zhang, L., Gong, Z., Jiao, W., 2020a. Segmenting ears of winter wheat at flowering stage using digital images and deep learning. Comput. Electron. Agric. 168, 105159. https://doi.org/10.1016/j.compag.2019.105159.

Ma, J., Li, Y., Liu, H., Du, K., Zheng, F., Wu, Y., Zhang, L., 2020b. Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation. Comput. Electron. Agric. 176, 105662. https://doi.org/10.1016/j.compag.2020.105662.

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., Heritier, E., Baret, F., 2019. Ear density estimation from high resolution RGB imagery using deep learning technique. Agric. For. Meteorol. 264, 225–234. https://doi.org/10.1016/j.agrformet.2018.10.013.

Meier, U., 2001. Growth Stages of Mono and Dicotyledonous Plants. BBCH Monograph 2nd Edition. Federal Biological Research Centre for Agriculture and Forestry, Quedlinburg, Germany. https://doi.org/10.5073/20180906-074619.

Min, D., Choi, S., Lu, J., Ham, B., Sohn, K., Do, M.N., 2014. Fast Global Image Smoothing Based on Weighted Least Squares. IEEE Trans. Image Process. 23, 5638–5653. https://doi.org/10.1109/TIP.2014.2366600.

Oscarson, P., 2000. The strategy of the wheat plant in acclimating growth and grain production to nitrogen availability. J. Exp. Bot. 51, 1921–1929. https://doi.org/10.1093/jexbot/51.352.1921.

Pound, M.P., Atkinson, J.A., Wells, D.M., Pridmore, T.P., French, A.P., 2017. Deep learning for multi-task plant phenotyping. In: Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017. pp. 2055–2063. https://doi.org/10.1109/ICCVW.2017.241.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 779–788. https://doi.org/10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Sadeghi-Tehran, P., Virlet, N., Ampe, E.M., Reyns, P., Hawkesford, M.J., 2019. DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. Front. Plant Sci. 10, 1–16. https://doi.org/10.3389/fpls.2019.01176.

Saeys, W., Lenaerts, B., Craessaerts, G., De Baerdemaeker, J., 2009. Estimation of the crop density of small grains using LiDAR sensors. Biosyst. Eng. 102, 22–30. https://doi.org/10.1016/j.biosystemseng.2008.10.003.

Solovyev, R., Wang, W., Gabruseva, T., 2021. Weighted boxes fusion: Ensembling boxes from different object detection models. Image Vis. Comput. 107, 104117.

Su, W.H., Zhang, J., Yang, C., Page, R., Szinyei, T., Hirsch, C.D., Steffenson, B.J., 2021. Automatic evaluation of wheat resistance to fusarium head blight using dual mask-rcnn deep learning frameworks in computer vision. Remote Sens. 13, 1–20. https://doi.org/10.3390/rs13010026.

Tan, C., Zhang, P., Zhang, Y., Zhou, X., Wang, Z., Du, Y., Mao, W., Li, W., Wang, D., Guo, W., 2020. Rapid Recognition of Field-Grown Wheat Spikes Based on a Superpixel Segmentation Algorithm Using Digital Images. Front. Plant Sci. 11, 1–13. https://doi.org/10.3389/fpls.2020.00259.

Vallat, R., 2018. Pingouin: statistics in Python. J. Open Source Softw. 3, 1026. 10.21105/joss.01026.

Velumani, K., Oude Elberink, S., Yang, M.Y., Baret, F., 2017. Wheat Ear Detection in Plots by Segmenting Mobile Laser Scanner Data. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. 4, 149–156. https://doi.org/10.5194/isprs-annals-IV-2-W4-149-2017.

Wang, D., Fu, Y., Yang, G., Yang, X., Liang, D., Zhou, C., Zhang, N., Wu, H., Zhang, D., 2019. Combined Use of FCN and Harris Corner Detection for Counting Wheat Ears in Field Conditions. IEEE Access 7, 178930–178941. https://doi.org/10.1109/ACCESS.2019.2958831.

Wang, Y., Qin, Y., Cui, J., 2021. Occlusion Robust Wheat Ear Counting Algorithm Based on Deep Learning. Front. Plant Sci. 12, 1–14. https://doi.org/10.3389/fpls.2021.645899.

Wu, Y., Hu, Y., Li, L., 2020. BTWD: Bag of Tricks for Wheat Detection. Springer International Publishing, New-York, NY, USA, pp. 450–460.

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., Shen, C., 2019. TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. Plant Methods 15. https://doi.org/10.1186/s13007-019-0537-2.

Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., Chen, G., Xiong, L., Yang, W., Liu, Q., 2017. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. Plant Methods 13, 1–15. https://doi.org/10.1186/s13007-017-0254-7.

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., Shen, S., Jiang, B., Ma, X., 2020. Wheat ear counting using K-means clustering segmentation and convolutional neural network. Plant Methods 16, 1–13. https://doi.org/10.1186/s13007-020-00648-8.

Yang, B., Gao, Z., Gao, Y., Zhu, Y., 2021. Rapid Detection and Counting of Wheat Ears in the Field Using YOLOv4 with Attention Module. Agronomy 11, 1–17.

Yang, Y., Huang, X., Cao, L., Chen, L., Huang, K., 2019. Field Wheat Ears Count Based on YOLOv3, in: Proceedings - 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019. IEEE, Manhattan, NY, USA, pp. 444–448. https://doi.org/10.1109/AIAM48774.2019.00094.

Zhao, J., Zhang, X., Yan, J., Qiu, X., Yao, X., Tian, Y., Zhu, Y., Cao, W., 2021. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. Remote Sens. 13, 1–16.

Zhou, C., Liang, D., Yang, X., Xu, B.o., Yang, G., 2018a. Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms. Remote Sens. 10 (2), 246.

Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., Yang, G., 2018b. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. Front. Plant Sci. 9 https://doi.org/10.3389/fpls.2018.01024.

Zhu, Y., Cao, Z., Lu, H., Li, Y., Xiao, Y., 2016. In-field automatic observation of wheat heading stage using computer vision. Biosyst. Eng. 143, 28–41. https://doi.org/10.1016/j.biosystemseng.2015.12.015.