Effect of the sun on the measurement of wheat ear density by deep learning

Sébastien Dandrifosse[1], Elias Ennadifi[2], Alexis Carlier[1], Bernard Gosselin[2], Benjamin Dumont[3], Benoît Mercatoris[1]

[1] Biosystems Dynamics and Exchanges, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium.; benoit.mercatoris@uliege.be

[2] Information Signal and Artificial Intelligence Lab, Faculty of Engineering, University of Mons, 7000 Mons, Belgium.; elias.ennadifi@umons.ac.be

[3] Plant Sciences, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium.; benjamin.dumont@uliege.be

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26 – 29, 2022
Minneapolis, Minnesota, United States

Abstract.

Ear density in the field, i.e. the number of ears per square meter, is one of the yield components of wheat and therefore a variable of high agronomic interest. Its traditional measurement necessitates laborious human observations in the field or destructive sampling. In recent years, deep learning based on RGB images has been identified as a low-cost, robust and high-throughput alternative to measure this variable. However, most of the studies were limited to the computer challenge of counting the ears in the images, without aiming to convert those counts into ear density. The aim of this study was to propose a method for automatic measurement of ear density, but also to evaluate the potential impact of the sun on the measurement.

A same zone of a wheat plot has been imaged by two nadir RGB cameras all over the daily course of the sun, this repeated at flowering, watery ripe, medium milk and hard dough development stages. The bounding boxes of the ears in the images were detected using the YOLOv5 deep learning model, trained on rich existing wheat ear datasets. The shifts between the same elements observed in the images from the two cameras were exploited to compute the image footprint by stereovision. The ear count divided by the image footprint yielded the ear density. To investigate the effect of the sun, a solar spectrum was recorded thanks to a spectrometer at the time of each image acquisition.

The F1 scores of ear bounding box detection at flowering, watery ripe, medium milk and hard dough were respectively 0.87, 0.92, 0.92 and 0.85. At watery ripe and medium milk, the measured ear density was robust during the day and between the two dates. At hard dough stage, increases of sunlight irradiance correlated with decreases of the number of ears detected by deep learning, but also with decreases of the number of ears labeled by humans. This demonstrates that, in some conditions, the wheat ear detection performance indicators based on labeled ear may be misleading regarding the capacity of the machine vision to measure the real ear density.

Keywords.
RGB image, wheat head, wheat spike, counting, phenotyping.

1. Introduction

The number of wheat ears per square meter, called the ear density, is one of the yield components and therefore a variable of high agronomic interest. Its traditional measurement necessitates laborious and possibly biased human observations in the field or destructive sampling. There is thus a high interest to develop high throughput, objective and automatic ear density measurement methods. From 1995 to 2018, sparse attempts were made, most of the time relying on computer vision and classic image analysis (Chopinet and Cointault 2006; Cointault et al. 2012; Germain et al. 1995; Zhu et al. 2016). None of those methods was validated in a robust way: in all the conditions and for several varieties. The variability of natural light and wheat development stage were especially pointed out as causing ear detection issues (Fernandez-Gallego et al. 2018).

The situation changed in 2018 with a first study exploiting deep learning to detect the wheat ears in field conditions (Hasan et al. 2018). Deep learning was advanced as a promising tool to develop robust ear detection methods from RGB images. Swiftly, it became a booming research topic and many studies proposed deep learning ear counting solutions (Cao et al. 2020; Lu and Cao 2020; Madec et al. 2019; Xiong et al. 2019; Xu et al. 2020; Y. Yang et al. 2019). This enthusiasm increased further thanks to the release of open access international datasets (David et al. 2020, 2021), allowing everyone to train and evaluate deep learning algorithms to detect the wheat ears (Ayalew et al. 2020; Fourati et al. 2021; Gong et al. 2020; Li et al. 2021; Ma et al. 2020; Wang et al. 2021; Wu et al. 2020; B. Yang et al. 2021). It became a popular challenge for computer scientists. Many of these studies relied solely on the existing datasets instead of setting field experiments. Although great ear counting performances are now possible, it leaves a number of unsolved issues regarding the ear density measurement.

Firstly, few studies tackled the conversion of the ear count in an image to an actual ear density in the field, which is the measure of agronomic interest. The task is not trivial. It necessitates an automatic determination of the image footprint at ear height, accounting that this height may vary. A possibility is to position a target of known dimensions in the image (Sadeghi-Tehran et al. 2019), but it limits the throughput of the acquisitions and reduces the visible canopy area. Another track was to use a Light Detection And Ranging (LiDAR) device to measure the ear height (Madec et al. 2019). LiDAR devices are however still expensive, require some scanning time and need to come in addition to the RGB camera.

Secondly, to our knowledge, no study investigated the effect of lighting conditions on the ear density measurement. Most of the studies claimed the robustness of the method because the ear detection algorithms achieved good performances in all the conditions when evaluated based on image labels drawn by human operators. A hypothesis is however that such ear detection performance indicators may be misleading. If direct sunlight induces a shadow that hides an ear for both the human image annotator and the deep learning model, no error will be detected but the real ear density will be underestimated (Madec et al. 2019).

To achieve a low-cost automatic measurement of ear density without the need of a reference target, Dandrifosse et al. (2022) proposed to add a second RGB camera so that the image footprint could be measured by stereovision. This second study aims to investigate the effect of varying sunlight on the measurement.

2. Material and methods

2.1 In field data acquisition

Images were recorded during the 2021 season in a plot located in the Hesbaye area, Belgium (50° 33' 50.7'' N and 4° 41' 57.3'' E) on homogeneous silty soil and a temperate climate. The plot was sown with winter wheat (*Triticum aestivum L.*, variety 'KWS Dorset') on November 13th, 2020 with a density of 400 grains/m². It was fertilized at BBCH stages 28, 30 and 39 (Meier 2001) with 27% ammonium nitrate: 60 kgN/ha at each input. The previous crop was winter wheat.

The idea to test the effect of the sun was to record data time series of the same wheat canopy zone during entire days. The position of the acquisition platform did not change during the day, but it changed between the dates, especially because the studied canopy area had been damaged. The platform, made of light metal profiles, was designed to capture nadir frames of wheat. At all the acquisition dates, two RGB cameras were placed around 1.5 m above the ears. They were screwed on a cantilever beam to avoid shadows of the structure in the images.

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 calibration provided the distortion coefficients, the intrinsic camera parameters and the extrinsic parameters linking the two cameras. The average error for the camera pair was 0.4 pixels.

Images were captured on June 10th, June 23rd, July 1st and July 22nd, corresponding respectively to flowering (BBCH 65), watery ripe (BBCH 71), medium milk (BBCH 75) and hard dough (BBCH87) stages. The acquisitions were performed between 9 a.m. and 5 p.m. Every quarter of an hour, two images were captured at an interval of 20 s. The images were recorded using a color depth of 12 bits per pixel but reduced to 8 bits per pixel for this study. To prevent saturation, a custom auto-exposure algorithm was designed. A dichotomous search was performed to find the highest exposure time for which no saturation was detected. The two limits of the search were the exposure time computed by the manufacturer algorithm and this time divided by five.

The sunlight spectrum was measured at the time of each image acquisition by a Avaspec-ULS2048 spectrometer (Avantes, Apeldoorn, Netherlands) equipped with a cosine corrector. Its signal-to-noise ratio was 200:1. The incident light was acquired using a 16-bit resolution. Each recorded spectrum was the average of three consecutive measurements. It was corrected for dark noise and non-linearity of pixel response to exposure time. Thanks to the factory calibration, digital values were converted to irradiance data. Each spectrum was integrated in the range of the cameras, from 485 to 910 nm, to obtain the irradiance value expressed in W/m².

Each acquisition of images and their associated solar spectrum took only a few seconds. It corresponded to the time necessary to average the spectra and ensure a proper exposure time for the cameras.

2.2 Ear bounding box detection

2.2.1 Image pre-processing and labeling

At each acquisition date, two images were captured every quarter of an hour, called the first and the second images. To create the training set, five images were randomly selected at each date among the first images. To create the validation set, all the second images were selected. The pre-process of the selected images was performed in two steps. Firstly, each image of 2560 × 2048 pixels was divided into four sub-images of 1280 × 1024 pixels. Secondly, each sub-image was resized to 1024 × 1024 pixels. That way, they were in the square format required by the deep learning algorithm and the ear had a size similar to those in the Global Wheat Head Dataset V2 (GWHDv2) (David et al. 2021), also used to train the model. After the pre-process, the sub-images were labeled using the LabelImg tool (https://github.com/tzutalin/labelImg).

2.2.2 Training of YOLOv5 model

The deep learning model for the ear bounding box detection was YOLOv5 (DOI: 10.5281/zenodo.3908559), the last born from the YOLO family (Redmon et al. 2016). The method to decide the final bounding box of an ear among all the box propositions was the weighted boxes fusion approach (Solovyev et al. 2021). The model was built by transfer learning: the 2021 training data (Section 2.1) were added to the model developed by Dandrifosse et al. (2022). This model was already trained on the GWHDv2, that is an international dataset with a wide diversity of ears,

and on a dataset acquired by the same cameras as in this study but on other varieties. The addition of 2021 images aimed to be sure that the model was trained for the studied variety. That way, the effect of the sun could be investigated in optimal training conditions.

2.2.3 Validation of bounding box detection

The performances of the bounding box detection were evaluated on the 2021 labeled validation data (Section 2.1). 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 rare that the detected box perfectly matches the labeled box. The definition of a correct or incorrect detection was based the Intersection over Union (IoU), i.e. the ratio between the area formed by the overlap of the detected box and the labeled box and the area formed by the set of the two boxes. The TP, FP and FN were determined by choosing an IoU threshold of 0.5. This value is the standard to evaluate a detection model. Moreover, it was pointed out as the optimal choice by Madec et al. (2019). From the TP, FP and FN, several other meaningful and widely-used indicators were built. The precision (Equation 1) is the fraction of correct detections among all the detections. The recall (Equation 2) is the fraction of correct detections among all the ears that should have been detected. The accuracy (Equation 3) is a well-known performance metric. Its general formula contains the true negatives (TN) but in such object detection tasks, there is no TN. The F1 score (Equation 4) is the harmonic mean of precision and recall, which gives a robust model performance assessment.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$F1\ score = 2\ \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

Another 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 several confidence levels of the network prediction. The PRC represents the influence of this confidence level on the relation between precision and recall.

$$AP@\alpha = \int_0^1 Precision(Recall)\ dRecall \quad (5)$$

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, for this work, there was only one class, so the AP was identical to the mAP.

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

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

Two mAP metrics were computed: 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 average of AP values for thresholds ranging from 0.5 to 0.75 with a step of 0.05. One interest of calculating so many indicators is to increase the capacity to compare on a common ground the performances with other studies.

2.3 Ear segmentation and count

The addition of the numbers of bounding boxes in the sub-images did not provide the number of ears in the full image. Indeed, some ears were cut at the junction of sub-images and parts of those ears appeared in several sub-images. To account for those ears, it was necessary to locate them at the pixel scale instead of the bounding box scale.

The state-of-the-art DeepMAC (Deep Mask-heads Above CenterNet) neural network model (Birodkar et al. 2022) was used to segment the ears in the bounding boxes. This approach did not require manual construction of training masks, which is a difficult and time-consuming work. Thanks to the strong generalization ability of the pre-trained DeepMAC model, no specific training was needed. Ear masks were generated for each square sub-image. Then, the four sub-masks 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, which is the format of the original RGB image. The ears cut at the sub-image junctions were reconstituted when assembling the sub-masks. However, the parts of such cut ears sometimes did not match to the pixel, especially when the mask predicted by DeepMAC did not reach the edge of the image. To solve this problem, 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 (Fig 1).

The difference between the sum of the objects in the sub-masks and in the full mask provided the number of excess ear parts. This number was subtracted to the sum of the bounding boxes in the four sub-images to compute the number of ears in the full image.

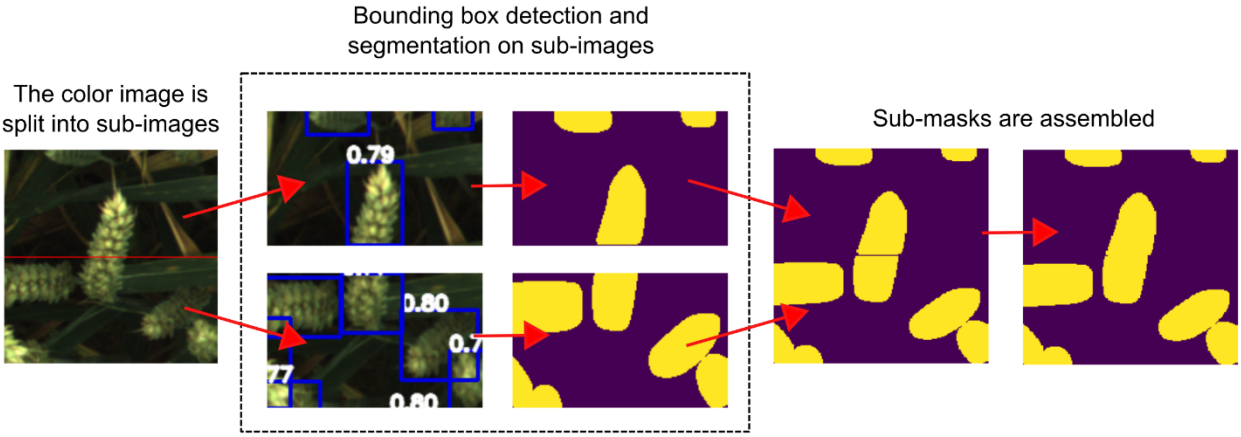

Fig 1. Pipeline of sub-image treatment. The process is illustrated for a zone crossed by a sub-image delimitation line. For the steps illustrated in the dotted rectangle, the size of the sub-images is slightly different because they had been resized to a square format. Figure taken from Dandrifosse et al. (2022).

2.4 Image footprint and ear density

The ear density, expressed in ears per square meter, was computed as 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 $footprint\ ears$ (m²) 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 the 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) parametrized 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 2). The ear mask from the segmentation step (Section 2.3) 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.

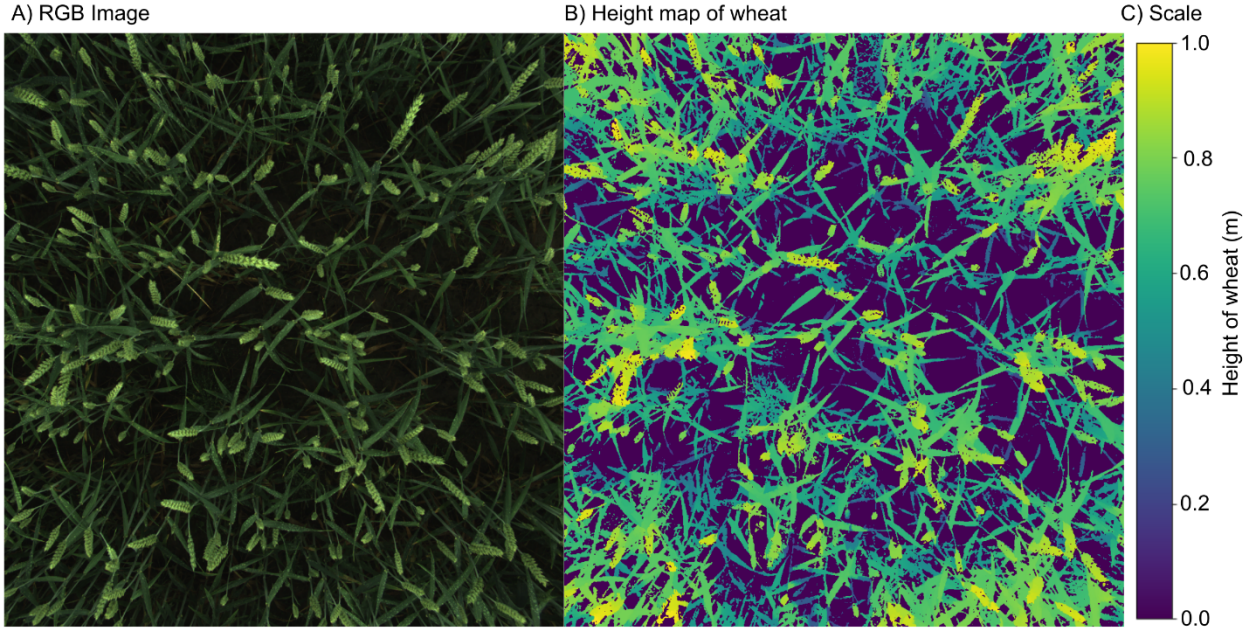

Fig 2. Wheat height map. The scene was imaged at watery ripe stage. The figure shows a) the RGB image, B) the height map and C) the color scale. The dark blue zones, at 0 on the scale, corresponded either to soil or to plant pixels for which the height could not be computed or interpolated through the WLS method.

3. Results and discussion

3.1 Ear bounding box detection

The performances of ear bounding box detection for each acquisition date are presented in Table 1 and some illustrations are provided in Fig 3. As expected, the detection was better at watery ripe and medium milk development stages. The ears were more difficult to detect at flowering stage because they were lower in the canopy and looked more like leaves, and at hard dough stage because they were bent from the weight of the grains and overlapped. Overall, ear detection performances are considered good. However, the training data contained, among others, images of the same scene as the validation images. It was expected near perfect performances, and yet there were still a number of false positives and false negatives.

Table 1. Performances of ear bounding box detection. Dates are expressed in days after sowing (DAS).

DAS	Stage	TP (%)	FP (%)	FN (%)	Precision	Recall	mAP@ 0.5	mAP@ 0.5:0.75	Accuracy	F1 score
209	Flowering	77.0	10.0	13.0	0.89	0.86	0.76	0.74	0.77	0.87
222	Watery ripe	85.5	8.0	6.5	0.91	0.93	0.85	0.85	0.85	0.92
230	Medium milk	85.8	7.0	7.2	0.92	0.92	0.85	0.85	0.86	0.92
251	Hard dough	73.4	9.2	17.4	0.89	0.80	0.72	0.71	0.73	0.85

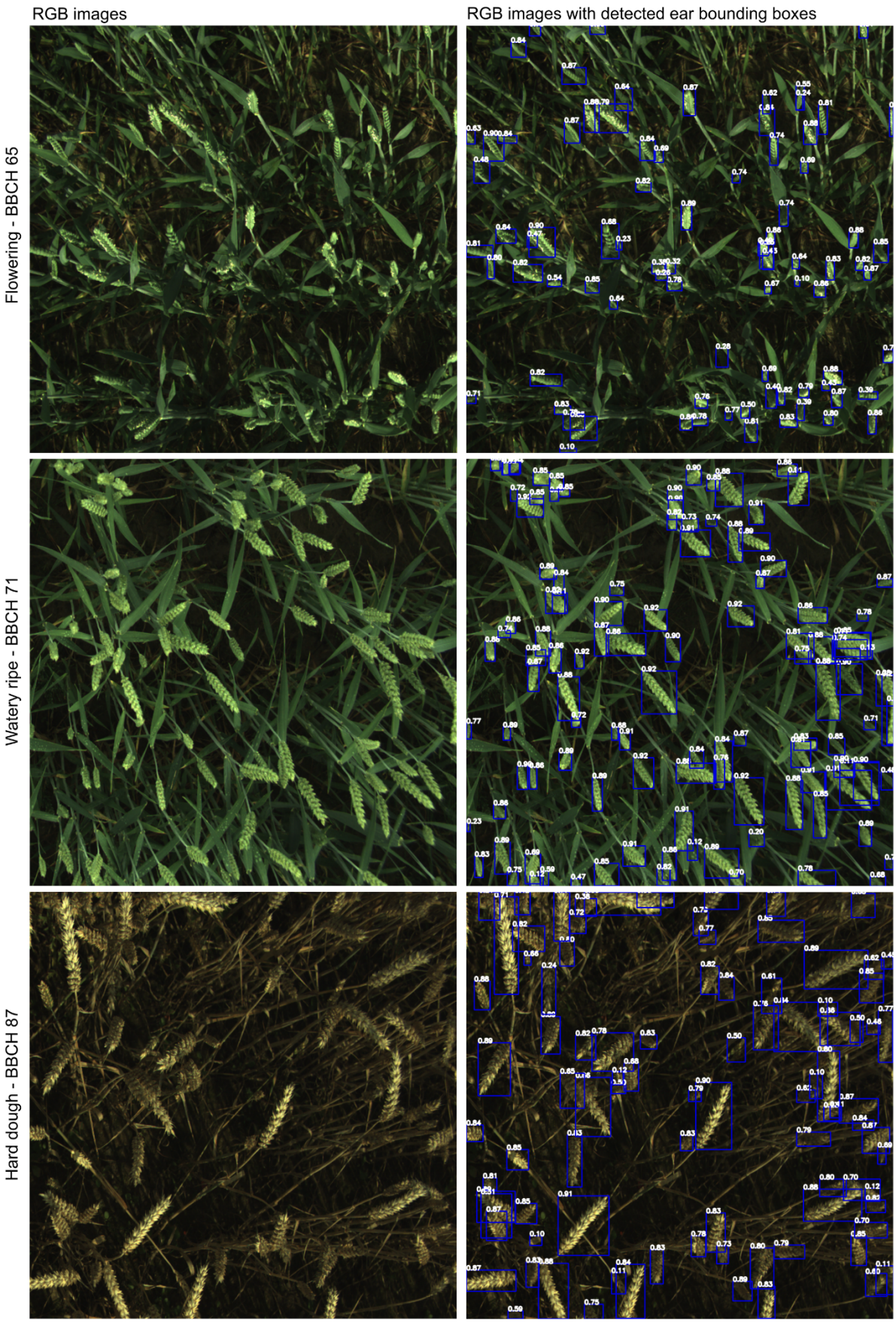

Fig 3. Ear bounding box detection.

3.2 Dynamics of measured variables during the day

The dynamics of the measured variables and the ear detection performance indicators during the day are presented in Fig 4 for the four acquisitions dates.

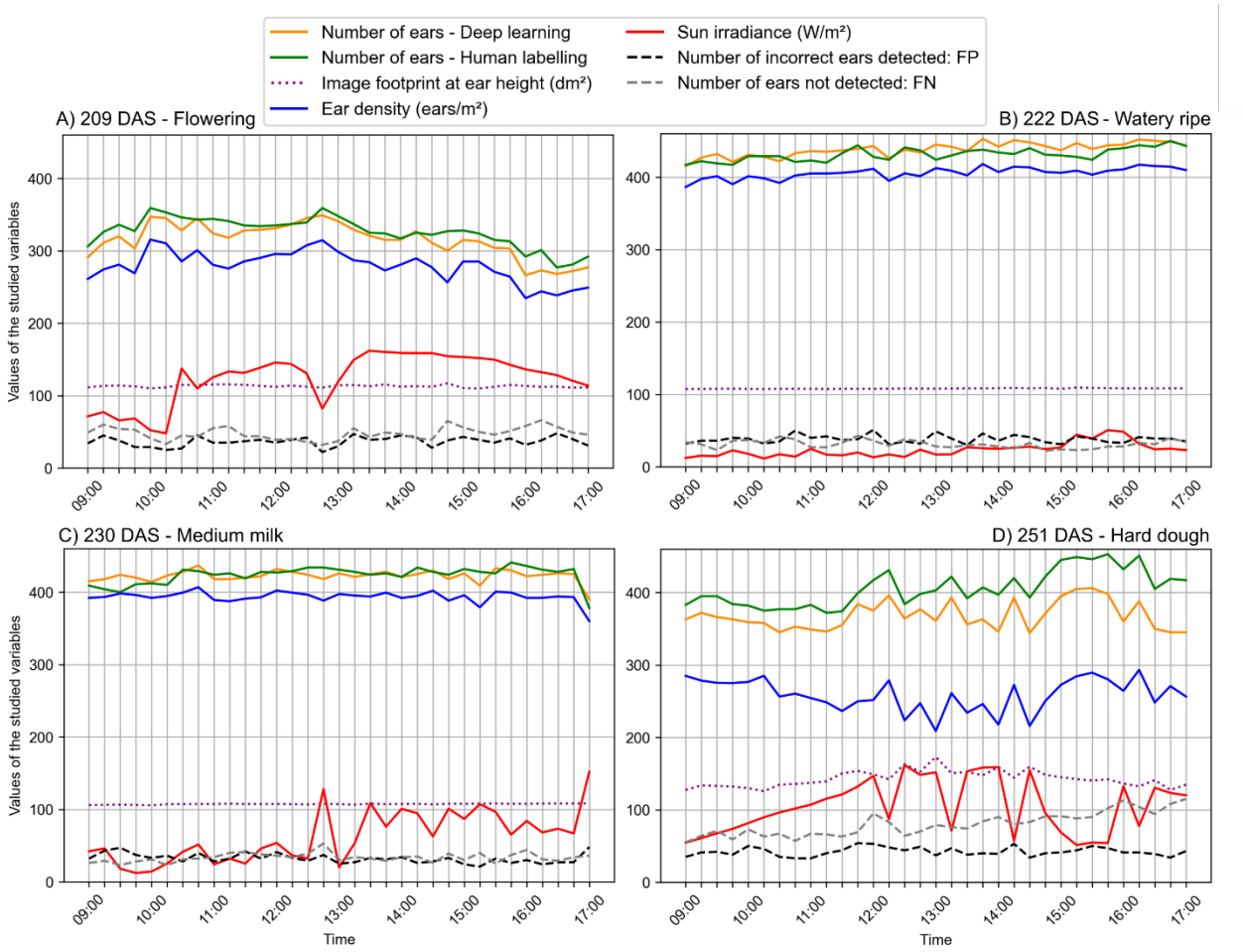

Fig 4. Dynamics of measured variables during the day presented for four development stages of wheat: A) flowering, B) watery ripe, C) medium milk and D) hard dough. The dates are expressed in days after sowing (DAS).

The first measure to examine is the ear density, which is the agronomic variable of interest. At watery ripe and medium milk stages, the measure was roughly the same throughout the day. This seems to indicate that both development stages are suitable for a robust measurement of ear density. At watery ripe, the sky was cloudy all the day, the diffuse light provided the ideal conditions for image analysis. At medium milk, the sun showed up in the afternoon but it was observed to have no major impact on the ear density measurement, except for the last quarter of an hour. At flowering and hard dough, lower ear densities were measured. Those growing stages are not suitable to measure ear density using nadir-viewing cameras if the goal is to derive the yield component. At flowering, not all the ears had emerged from their sheaths or the canopy, they were still green and thin like the leaves. At hard dough, the ears overlapped too much because they were bent from the weight of the grains.

At watery ripe and medium milk, the number of ears measured by deep learning was close to the number of ears labeled. An important point is that this was not due to near perfect detection performances, but rather to a compensation between the false positives and the false negatives.

Detecting the ears at hard dough may be of interest to derive information about their morphology or the number of spikelets. Nevertheless, the detection at this stage presented some difficulties. An interesting observation is the effect of the sun. During the day, when the sun irradiance increased, which can be associated with more direct light, the number of ears counted by both the algorithm and the human operator decreased. It seems to confirm the hypothesis that ears in

the shadow were missed also by the human annotators. This is due to the strong contrasts occurring in a wheat canopy under direct sunlight. To avoid pixel saturation on the ears, the auto-exposure algorithm decreased the exposure time, resulting in dark zones in the canopy. A solution to this issue would be to acquire the scene using multiple exposure times, and then reconstruct an image with a high dynamic range.

Considering the estimation of the image footprint at ear height by stereovision, the standard deviations at flowering, watery ripe, medium milk and hard dough were respectively 0.018, 0.005, 0.007 and 0.110 m². The footprint proved to be robust throughout the day for three of the four dates investigated. It remains however an open-question: which height in the ear layer should be considered to compute the image footprint? And how to identify that height? In this study, it was considered the height of the ear tops minus 50 mm. The height of the ear tops was easy to measure from nadir images containing erected ears. The distance of 50 mm to account for the depth of the ear layer was however arbitrary.

4. Conclusions

A method to measure wheat ear density from RGB cameras, combining deep learning and stereovision, was studied on the same wheat area throughout full days, at four development stages: flowering, watery ripe, medium milk and hard dough. The F1 scores of ear bounding box detection were respectively 0.87, 0.92, 0.92 and 0.85. At watery ripe and medium milk, the method measured an ear density that was robust throughout the day. At hard dough stage, increases of sunlight irradiance correlated with decreases of the number of ears detected by deep learning, but also with decreases of the number of ears labeled by humans. This demonstrates that, in some conditions, the wheat ear detection performance indicators based on labeled ear may be misleading regarding the capacity of the machine vision to measure the real ear density.

Acknowledgements

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.

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, Julien Kirstein, Amandine Collin, Abdelkbir Masrour, Françoise Thys and Rudy Schartz for their help.

References

Ayalew, T. W., Ubbens, J. R., & Stavness, I. (2020). Unsupervised Domain Adaptation for Plant Organ Counting. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in 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. (2022). The surprising impact of mask-head architecture on novel class segmentation. *arXiv*, *abs/2104.0*, 6995–7005. https://doi.org/10.1109/iccv48922.2021.00693

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV*. Newton, MA, USA: O'Reilly Media, Inc. 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. *Proceedings of 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education, ICISCAE 2020*, 557–561. https://doi.org/10.1109/ICISCAE51034.2020.9236918

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

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. *Frontiers in Plant Science*, *11*(February), 1–15. https://doi.org/10.3389/fpls.2020.00096

Dandrifosse, S., Ennadifi, E., Carlier, A., Gosselin, B., Dumont, B., & Mercatoris, B. (2022). Deep Learning for Wheat ear Segmentation and Ear Density Measurement: From Heading to Maturity. *Computers and Electronics in Agriculture.* Accepted under revisions.

David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., et al. (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 Phenomics*, *2020*. https://doi.org/10.34133/2020/3521852

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., et al. (2021). Global Wheat Head Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods. *Plant Phenomics*, *2021*(May), 1–9. https://doi.org/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), 1–12. https://doi.org/10.1186/s13007-018-0289-4

Fourati, F., Mseddi, W. S., & Attia, R. (2021). Wheat Head Detection using Deep, Semi-Supervised and Ensemble Learning. *Canadian Journal of Remote Sensing*, *0*(0), 1–13. https://doi.org/10.1080/07038992.2021.1906213

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* (pp. 435–439). Stevenage, UK: IET. 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*(1), 191. https://doi.org/10.3390/s21010191

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), 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* (Vol. 30, pp. 328–341). Manhattan, NY, USA: IEEE.

Li, J., Li, C., Fei, S., Shi, J., Xiao, Z., Ma, C., et al. (2021). Wheat ear recognition based on retinanet and transfer learning. *Sensors*, *21*(14). https://doi.org/10.3390/s21144845

Lu, H., & Cao, Z. (2020). TasselNetV2+: A Fast Implementation for High-Throughput Plant Counting From High-Resolution RGB Imagery. *Frontiers in Plant Science*, *11*(December), 1–15. https://doi.org/10.3389/fpls.2020.541960

Ma, J., Li, Y., Du, K., Zheng, F., Zhang, L., Gong, Z., & Jiao, W. (2020). Segmenting ears of winter wheat at flowering stage using digital images and deep learning. *Computers and Electronics in Agriculture*, *168*(November 2019), 105159. https://doi.org/10.1016/j.compag.2019.105159

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., et al. (2019). Ear density estimation

from high resolution RGB imagery using deep learning technique. *Agricultural and Forest Meteorology*, *264*(May 2018), 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*. Quedlinburg, Germany: Federal Biological Research Centre for Agriculture and Forestry. 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 Transactions on Image Processing*, *23*(12), 5638–5653. https://doi.org/10.1109/TIP.2014.2366600

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 779–788. https://doi.org/10.1109/CVPR.2016.91

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. *Frontiers in Plant Science*, *10*(September), 1–16. https://doi.org/10.3389/fpls.2019.01176

Solovyev, R., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, *107*. https://doi.org/10.1016/j.imavis.2021.104117

Wang, Y., Qin, Y., & Cui, J. (2021). Occlusion Robust Wheat Ear Counting Algorithm Based on Deep Learning. *Frontiers in Plant Science*, *12*(June), 1–14. https://doi.org/10.3389/fpls.2021.645899

Wu, Y., Hu, Y., & Li, L. (2020). BTWD: Bag of Tricks for Wheat Detection. In *Lecture Notes in Computer Science* (Vol. 12540 LNCS, pp. 450–460). New-York, NY, USA: Springer International Publishing. https://doi.org/10.1007/978-3-030-65414-6_31

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*(1), 0–14. https://doi.org/10.1186/s13007-019-0537-2

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods*, *16*(1), 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*(7), 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* (pp. 444–448). Manhattan, NY, USA: IEEE. https://doi.org/10.1109/AIAM48774.2019.00094

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