

Wheat yield estimation based on analysis of UAV images at low altitude

Mikhail Kozhekin^{1,2}, Mikhail Genaev^{1,2,3,4}, Vasily Koval^{1,2}, Andrey Slobodchikov¹ and Dmitry Afonnikov^{1,2,3,4,*}

¹Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

²Kurchatov Genomics Center, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

³Mathematical Center in Akademgorodok, 630090 Novosibirsk, Russia

⁴Department of Mathematics and Mechanics, Novosibirsk State University, 630090 Novosibirsk, Russia

Abstract. Information about the yield of wheat crops makes it possible to correctly assess their productivity and choose appropriate agronomic procedures to maximize yield. However, determining yields based on manual ear counts is labor intensive. Recently UAVs demonstrated high efficiency for rapid yield estimation. This paper presents a software package WDS (Wheat Detection System) for ears counting in wheat crops based on RGB images obtained from UAVs. WDS creates the flight plan, for the acquired images carries out automatic georeferencing to the appropriate fragment of the field, counts ears using the neural network models, reconstructs the density of ears in the crop and visualizes it as a heat map in the interactive web application. Based on the field experiment the accuracy of ears counting in plots was assessed: Spearman and Pearson correlation coefficients between the ears density counted manually and using WDS were 0.618 and 0.541, respectively (p -value < 0.05). WDS available at https://github.com/SI07h/wheat_detection.

1 Introduction

Wheat is one of the most important crops that feeds a significant portion of the world's population. In the process of its production it is necessary to constantly have information about the yield of crops, which allows to properly assess their productivity and to adapt agronomic procedures to maximize the yield. Protocols for manually counting the density of ears in crops (number of ears per square meter) have been the only way to estimate yields for a long time. However, this method is labour-intensive and time-consuming. An alternative is the development of automated systems operating in the field [1]. Most of such systems allow to obtain 2D images of crops and use computer vision methods for their automatic processing, in particular, for counting ears in the image. Modern methods of image analysis based on neural network algorithms and deep learning allow ears identification on the image of crops and counting their number with high accuracy [2-4]. The use of these technologies is justified due to the lower cost and acceptable accuracy compared to the labour costs of manual human

* Corresponding author: ada@bionet.nsc.ru

observation. In wheat yield estimation, stationary systems, ground-based robotic platforms and UAVs can be used to acquire images. The former allow high quality images to be obtained, but only for a small area of crops. The greatest degree of mobility can be achieved with UAVs, but the images are of poorer quality, often blurred due to wind and engine vibration. These effects are particularly significant at low UAV flight altitudes, which must be used to obtain images with sufficient resolution to analyze the spikes [5]. On the other hand, these interferences make it difficult to stitch the acquired images needed to relate them to the spatial coordinates of the field.

In this paper, we have created a software package designed to perform a wheat yield estimation in the field based on ears counts from UAV images.

2 Figures and tables

2.1 Field experiments

We studied wheat crops in a field of SibNIIRS located near Novosibirsk, Russia. Coordinates of the field: 54.875, 82.958. Sowing was performed on May 12, 2021; crop was photographed on July 29, 2021 in ear formation phase.

Manual counting of ears was performed after harvesting from the area inside square frames with a side of 0.5 m (area 0.25 m²) dug in at the beginning of the season. At the end of the season, all plants inside the frames were cut and the number of productive stems was counted. To reduce the influence of different soil composition, each experiment was performed in four replications; their sum gives an estimate of spike density per square meter.

2.2 Flight plan preparation

A DJI Mavic 2 Pro UAV with a 20 megapixel camera was used. The flight plan was composed based on the following conditions: flight height of 3 meters; flight speed of 1.5 m/s; camera pointing downward; frame frequency was determined from the condition of overlapping the images over the total area by at least 50%.

A python script was developed to automate the construction of the flight plan when imaging a section of the field, given by the coordinates of four vertices. The program takes as input image resolution and camera view angle, altitude and flight speed, degree of image overlap and coordinates of the four field vertices in the format: degree, minute and second. The output is a csv file in which the flight route is recorded. This file is input to the Litchi software (<https://flylitchi.com>), which is used to perform the UAV's flight. The algorithm for building a flight plan included the following steps:

1. The longest side of the field quadrilateral and its west point A are selected.
2. The remaining points are rotated around point A by the angle between the side and the horizon. As result, quadrilateral oriented with a base parallel to the equator.
3. The number of flight paths based on the necessary overlap of images is calculated.
4. The intersection of tracks is built taking into account the image overlaps.
5. Step 2 is repeated, but in the opposite direction.

2.3 Linking images to field coordinates

During the flight, information about the UAV position is recorded in the metadata of each image file and includes the actual values of coordinates, flight altitude, camera view angle, roll, pitch and yaw. Using these parameters the developed script determines which area of the field got into the frame and whether there was a protocol violation at the moment of shooting (a significant deviation of the drone from the shooting route by coordinates and

altitude). Images with protocol violations were excluded from further analysis during processing. The `exiftool` utility [6] and the `exif` library (<https://pypi.org/project/exif/>) were used to extract metadata. The metadata for all images of one overlap series were recorded in a single csv file.

The coordinates of the border of the area caught in the frame are calculated in two steps:

1. Based on the values of the camera view angle and the height of the copter we calculate the length of the diagonal of the fragment of the field, which gets on the image in meters. Then using the proportions of the image we calculate the lengths of the sides of the field quadrilateral, which corresponds to this fragment.

2. We perform a rotation of a quadrangle corresponding to a fragment of a field on the image concerning the azimuth value and shift it so the position of the center corresponds to coordinates of the copter. The obtained coordinates of the vertices determine the localization of the field fragment that corresponds to the image.

2.4 Ear detection and counting

We used the algorithms of deep learning neural networks with `resnet-50` [7] and `efficient-det d5` [8] as the basis for the extraction of spikes in the image. The parameters of the networks were taken from the website Global wheat head detection (GWHD) (https://www.kaggle.com/pestipeti/pytorch-starter-fasterfcn-train/data?select=fasterfcn_resnet50_fpn.pth, <https://www.kaggle.com/shonenkov/inference-efficientdet/data?select=fold0-best-all-states.bin>). Parameters are derived from competitions held in 2020 using the GWHD dataset (4700 RGB images of 1024x1024 pixels from 7 countries) [9].

To display the results in the form of HTML-pages, the `folium` library (<https://pypi.org/project/folium>) was used, which allows creating an interactive map with different layers. When calculating the density of ears, the coordinates of each ear were determined based on the spatial coordinates of the image projection and the magnitude of the scale. On this basis, the number of images in which an ear could fall was determined. After that for each ear the weight inversely proportional to number of images on which it was localized was calculated. Then a grid is built over the area around our field with the given size and for each cell the number of ears caught in it is calculated and divided by its area. The obtained data is visualized in the form of a heat map. To calculate the number of ears in individual plots, we can use manual marking of their coordinates in geojson format.

2.5 Ear counting performance estimation

Two techniques were used to estimate the accuracy of wheat ear counting.

The ear recognition accuracy estimation was used for neural network prediction results on an additional sample of images that were added to the 2021 GWHD dataset (not used in the network training) [10]. The average precision (AP) and mean average precision (mAP) for bounding boxes identified by neural networks with IoU over 50% were used as described in [11]. The metric implies how much overlap there is between the ear bounding boxes predicted by the model and those marked in the dataset; its value measured from 0 to 100%.

The performance estimation based on the actual yield determination was done as described in section 2.1. After image processing and ear counting in the WDS system, we marked the coordinates of the plots. The system counted the number of ears per plot. Knowing that the area of each plot is equal to 24.75 m², we obtain the ear density. Pearson and Spearman correlation coefficients were used to compare the values obtained by manual counting and our system.

3 Results and discussion

The structure of developed WDS package is shown in Fig. 1: (A) construction of the flight plan on the quadrangular field by 4 points, (B) marking of plot boundaries, (C) counting the number of ears on each plot.

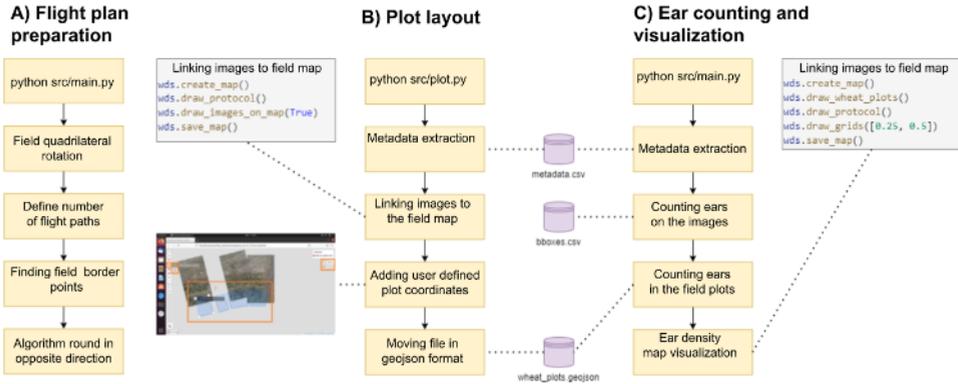


Fig. 1. The main steps of the program are: construction of the flight plan (A), marking the plot boundaries (B), counting the number of ears on each plot, and visualizing the density map (C).

Software package, instructions and scripts for installation are available at https://github.com/S107h/wheat_detection.

An example of visualizing the density of ears for an experimental field is shown in Fig. 2. Panel A shows the plot of image frames projections on the field with some images violating of protocol. The green frames correspond to the overlapping images of the field, which were taken at a height of 3 ± 0.3 m with the camera roll to the side by no more than 3 degrees. The red squares correspond to images that do not comply with the protocol (flight height 6 meters and camera tilt is not strictly down). In panel B, the intensity of the green color shows the densities of ears for 9 plots of the field, the coordinates of which were set by the user. Panels C and D show the density of ears in test crops at different sizes of the visualized grid.

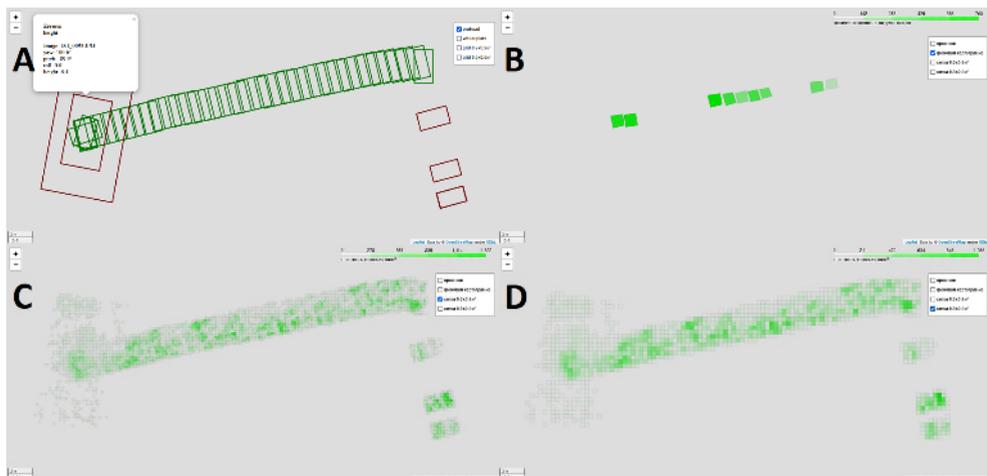


Fig. 2. Example of visualization of spike density estimation from UAV image analysis. (A) Correctness of the protocol. (B) Number of ears for individual plots. (C) Spike density at 0.25×0.25 m² grid. (D) Spike density in a 0.5×0.5 m² grid.

Table 1. The mAP performance metric for ear counting on the images for two network models evaluated using the 2021 GWHD image datasets.

	frcnn	effdet		frcnn	effdet		frcnn	effdet
ARC_1	39.81	34.63	CIMMYT_3	46.72	33.29	Terraref_1	19.29	14.60
Arvalis_10	50.53	69.74	KSU_1	22.03	38.33	Terraref_2	6.89	4.97
Arvalis_11	29.33	35.06	KSU_2	37.49	39.99	Ukyoto_1	34.25	36.62
Arvalis_12	51.37	40.97	KSU_3	30.52	37.33	ULiège-GxABT_1	25.14	33.68
Arvalis_7	62.79	43.13	KSU_4	13.61	26.00	UQ_10	49.70	39.84
Arvalis_8	57.31	47.16	NAU_2	72.45	69.98	UQ_11	30.18	24.95
Arvalis_9	44.60	36.09	NAU_3	73.26	73.54	UQ_7	50.64	46.49
CIMMYT_1	37.84	20.91	NMBU_1	55.72	37.84	UQ_8	41.23	32.99
CIMMYT_2	63.66	42.13	NMBU_2	33.13	22.17	UQ_9	38.28	30.33

The results of testing the models of recognition and spike counting in the image are shown in Table 1. Model testing was performed on the 27 sets of images added to GWHD in 2021 [10]. Each set is provided by one of 10 institutions. The variability of performance metrics and shooting conditions varies greatly. The best accuracy (73.54 on the mAP metric) is provided by the efficient-det model. The arithmetic mean mAP for these samples is 41.40 for faster rcnn and 37.51 for efficient-det.

A comparison of the ear density estimates made using our approach and those made manually showed that the Spearman and Pearson coefficient values between them are 0.6176, p-value=0.0013 and 0.5405, p-value=0.0064, respectively.

4 Conclusions

We have developed a software package to estimate wheat yield based on counting the number of ears in UAV images of wheat crops, which does not require image stitching. The software package allows to form a flight plan for low altitude flying over the crops (~3 m), to count the number of ears on each image by a deep learning neural network, to link the obtained images to the crop map, and to visualize the density of ears for the studied crops. We tested the packages on wheat crops in the summer of 2021 and showed that Spearman and Pearson correlation coefficients between the average values of ears density estimated by UAV and manually were more than 0.5.

Acknowledgement

The work was funded by the Kurchatov Genome Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, no. 075-15-2019- 1662.

References

1. X. Jin, et al., *IEEE Geosci*, **9** (1), 200-231 (2020)
2. S. Khaki, et al. "Wheatnet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting." *arXiv preprint arXiv:2103.09408* (2021)
3. Md Mehedi, et al., *Plant Methods*, **14**, 100 (2018)
4. J. Zhao, et al., *Remote Sensing*, vol. **13** (16), 3095 (2021)
5. X. Jin, et al., *Remote Sens. Environ.*, **198**, 105-114 (2017)

6. P. Harvey, and G. Körtner. "ExifTool." Kingston, Ontario, Canada.) Available at. <http://owl.phy.queensu.ca/~phil/exiftool/>[Verified 29 November 2018] (2016)
7. S. Ren, et al., *Faster r-cnn: Towards real-time object detection with region proposal networks*, in Proceedings of the 28th International Conference on Neural Information Processing Systems, NIPS'15, 7-12 Dec 2015, Montreal, Canada (2015)
8. M. Tan, R. Pang, and Q. V. Le., *Efficientdet: Scalable and efficient object detection*, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, CVPR, 14-19 Jun 2020, remote (2020)
9. E. David, et al., *Plant Phenomics*, **2020**, 3521852 (2020)
10. E. David, et al., *Plant Phenomics*, **2021**, 9846158 (2021)
11. J. Yu, W. Zhang, *Sensors*, **21(9)**, 3263 (2021)