






































## Database/Software Article

# Global Wheat Head Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods

**Etienne David** <sup>1,2</sup> **Mario Serouart** <sup>1,2</sup> **Daniel Smith** <sup>3</sup> **Simon Madec** <sup>1,3</sup>  
**Kaaviya Velumani** <sup>2,4</sup> **Shouyang Liu** <sup>5</sup> **Xu Wang** <sup>6</sup> **Francisco Pinto** <sup>7</sup>  
**Shahameh Shafiee** <sup>8</sup> **Izzat S. A. Tahir** <sup>9</sup> **Hisashi Tsujimoto** <sup>10</sup> **Shuhei Nasuda** <sup>11</sup>  
**Bangyou Zheng** <sup>12</sup> **Norbert Kirchgessner** <sup>13</sup> **Helge Aasen** <sup>13</sup> **Andreas Hund** <sup>13</sup>  
**Pouria Sadhegi-Tehran** <sup>14</sup> **Koichi Nagasawa** <sup>15</sup> **Goro Ishikawa** <sup>16</sup>  
**Sébastien Dandrifosse** <sup>17</sup> **Alexis Carlier** <sup>17</sup> **Benjamin Dumont** <sup>18</sup>  
**Benoit Mercatoris** <sup>17</sup> **Byron Evers** <sup>6</sup> **Ken Kuroki** <sup>19</sup> **Haozhou Wang** <sup>19</sup>  
**Masanori Ishii** <sup>19</sup> **Minhajul A. Badhon** <sup>20</sup> **Curtis Pozniak** <sup>21</sup> **David Shaner LeBauer** <sup>22</sup>  
**Morten Lillemo** <sup>8</sup> **Jesse Poland** <sup>6</sup> **Scott Chapman** <sup>3,12</sup> **Benoit de Solan** <sup>1</sup>  
**Frédéric Baret** <sup>2</sup> **Ian Stavness** <sup>20</sup> and **Wei Guo** <sup>19</sup>

<sup>1</sup>Arvalis, Institut du Végétal, 3 Rue Joseph et Marie Hackin, 75116 Paris, France

<sup>2</sup>UMR1114 EMMAH, INRAE, Centre PACA, Bâtiment Climat, Domaine Saint-Paul, 228 Route de l'Aérodrome, CS 40509, 84914 Avignon Cedex, France

<sup>3</sup>School of Food and Agricultural Sciences, The University of Queensland, Gatton, 4343 QLD, Australia

<sup>4</sup>Hiphen SAS, 120 Rue Jean Dausset, Agroparc, Bâtiment Technicité, 84140 Avignon, France

<sup>5</sup>Plant Phenomics Research Center, Nanjing Agricultural University, Nanjing, China

<sup>6</sup>Wheat Genetics Resource Center, Dep. of Plant Pathology, Kansas State Univ., 4024 Throckmorton Plant Sciences Center, Manhattan, Kansas, USA

<sup>7</sup>Global Wheat Program, International Maize and Wheat Improvement Centre (CIMMYT), Mexico, D.F., Mexico

<sup>8</sup>Faculty of Biosciences, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway

<sup>9</sup>Agricultural Research Corporation, Wheat Research Program, P.O. Box 126, Wad Medani, Sudan

<sup>10</sup>Arid Land Research Center, Tottori University, Tottori 680-0001, Japan

<sup>11</sup>Laboratories of Plant Genetics and Plant Breeding, Graduate School of Agriculture, Kyoto University, Japan

<sup>12</sup>CSIRO Agriculture and Food, Queensland Biosciences Precinct, 306 Carmody Road, St Lucia, 4067 QLD, Australia

<sup>13</sup>Institute of Agricultural Sciences, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland

<sup>14</sup>Plant Sciences Department, Rothamsted Research, Harpenden, UK

<sup>15</sup>Institute of Crop Science, National Agriculture and Food Research Organization, Japan

<sup>16</sup>Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Japan

<sup>17</sup>Biosystems Dynamics and Exchanges, TERRA Teaching and Research Center, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium

<sup>18</sup>Plant Sciences, TERRA Teaching and Research Center, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium

<sup>19</sup>Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Midori-cho, Nishitokyo City, Tokyo, Japan

<sup>20</sup>Department of Computer Science, University of Saskatchewan, Canada

<sup>21</sup>Department of Plant Sciences, University of Saskatchewan, Canada

<sup>22</sup>College of Agriculture and Life Sciences, University of Arizona, Tucson, Arizona, USA

Correspondence should be addressed to Etienne David; e.david@arvalis.fr

Received 31 May 2021; Accepted 11 August 2021; Published 22 September 2021

Copyright © 2021 Etienne David et al. Exclusive Licensee Nanjing Agricultural University. Distributed under a Creative Commons Attribution License (CC BY 4.0).

The Global Wheat Head Detection (GWHD) dataset was created in 2020 and has assembled 193,634 labelled wheat heads from 4700 RGB images acquired from various acquisition platforms and 7 countries/institutions. With an associated competition hosted in Kaggle, GWHD\_2020 has successfully attracted attention from both the computer vision and agricultural science communities. From this first experience, a few avenues for improvements have been identified regarding data size, head diversity, and label reliability. To address these issues, the 2020 dataset has been reexamined, relabeled, and complemented by adding 1722 images from 5 additional countries, allowing for 81,553 additional wheat heads. We now release in 2021 a new version of the Global Wheat Head Detection dataset, which is bigger, more diverse, and less noisy than the GWHD\_2020 version.

## 1. Introduction

Quality training data is essential for the deployment of deep learning (DL) techniques to get a general model that can scale on all the possible cases. Increasing dataset size, diversity, and quality is expected to be more efficient than increasing network complexity and depth [1]. Datasets like ImageNet [2] for classification or MS COCO [3] for instance detection are crucial for researchers to develop and rigorously benchmark new DL methods. Similarly, the importance of getting plant- or crop-specific datasets is recognized within the plant phenotyping community ([4–10], p. 2, [11–13]). These datasets allow benchmarking the algorithm performances used to estimate phenotyping traits while encouraging computer vision experts to further improvement ([10], p. 2, [14–17]). The emergence of affordable RGB cameras and platforms, including UAVs and smartphones, makes in-field image acquisition easily accessible. These high-throughput methods are progressively replacing manual measurement of important traits such as wheat head density. Wheat is a crop grown worldwide, and the number of heads per unit area is one of the main components of yield potential. Creating a robust deep learning model performing over all the situations requires a dataset of images covering a wide range of genotypes, sowing density and pattern, plant state and stage, and acquisition conditions. To answer this need for a large and diverse wheat head dataset with consistent and quality labeling, we developed in 2020 the Global Wheat Head Detection (GWHD\_2020) [18] that was used to benchmark methods proposed in the computer vision community and recommend best practices to acquire images and keep track of the metadata.

The GWHD\_2020 dataset results from the harmonization of several datasets coming from nine different institutions across seven countries and three continents. There are already 27 publications [19–45] (accessed July 2021) that have reported their wheat head detection model using the GWHD\_2020 dataset as the standard for training/testing data. A “Global Wheat Detection” competition hosted by Kaggle was also organized, attracting 2245 teams across the world [14], leading to improvements in wheat head detection models [23, 25, 31, 41]. However, issues with the GWHD\_2020 dataset were detected during the competition, including labeling noise and an unbalanced test dataset.

To provide a better benchmark dataset for the community, the GWHD\_2021 dataset was organized with the following improvements: (1) the GWHD\_2020 dataset was checked again to eliminate few poor-quality images, (2) images were re-labeled to avoid consistency issues, (3) a

wider range of developmental stages from the GWHD\_2020 sites was included, and (4) datasets from 5 new countries (the USA, Mexico, Republic of Sudan, Norway, and Belgium) were added. The resulting GWHD\_2021 dataset contains 275,187 wheat heads from 16 institutions distributed across 12 countries.

## 2. Materials and Methods

The first version of GWHD\_2020, used for the Kaggle competition, was divided into several subdatasets. Each subdataset represented all images from one location, acquired with one sensor while mixing several stages. However, wheat head detection models may be sensitive to the developmental stage and acquisition conditions: at the beginning of head emergence, a part of the head is barely visible because it is still not fully out from the last leaf sheath and possibly masked by the awns. Further, during ripening, wheat heads tend to bend and overlap, leading to more erratic labeling. A redefinition of the subdataset was hence necessary to help investigate the effect of the developmental stage on model performances. The new definition of a subdataset was then formulated as “a consistent set of images acquired over the same experimental unit, during the same acquisition session with the same vector and sensor.” A subdataset defines therefore a domain. This new definition forced to split the original GWHD\_2020 subdatasets into several smaller ones. The UQ\_1 was split into 6 much smaller subdatasets, Arvalis\_1 was split into 3 subdatasets, Arvalis\_3 into 2 subdatasets, and utokyo\_1 into 2 subdatasets. However, in the case of utokyo\_2 which was a collection of images taken by farmers at different stages and in different fields, the original subdataset was kept. Overall, the 11 original subdatasets in GWHD\_2020 were distributed into 19 subdatasets for GWHD\_2021.

Almost 2000 new images were added to GWHD\_2020, constituting a major improvement. A part of the new images comes from the institutions already contributing to GWHD\_2020 and was collected during a different year and/or at a different location. This was the case for Arvalis (Arvalis\_7 to Arvalis\_12), University of Queensland (UQ\_7 to UQ\_11), Nanjing Agricultural University (NAU\_2 and NAU\_3), and University of Tokyo (Utokyo\_1). In addition, 14 new subdatasets were included, coming from 5 new countries: Norway (NMBU), Belgium (Université of Liège [46]), United States of America (Kansas State University [47], TERRA-REF [7]), Mexico (CIMMYT), and Republic of Sudan (Agricultural Research Council). All these images were acquired at a ground sampling distance between 0.2

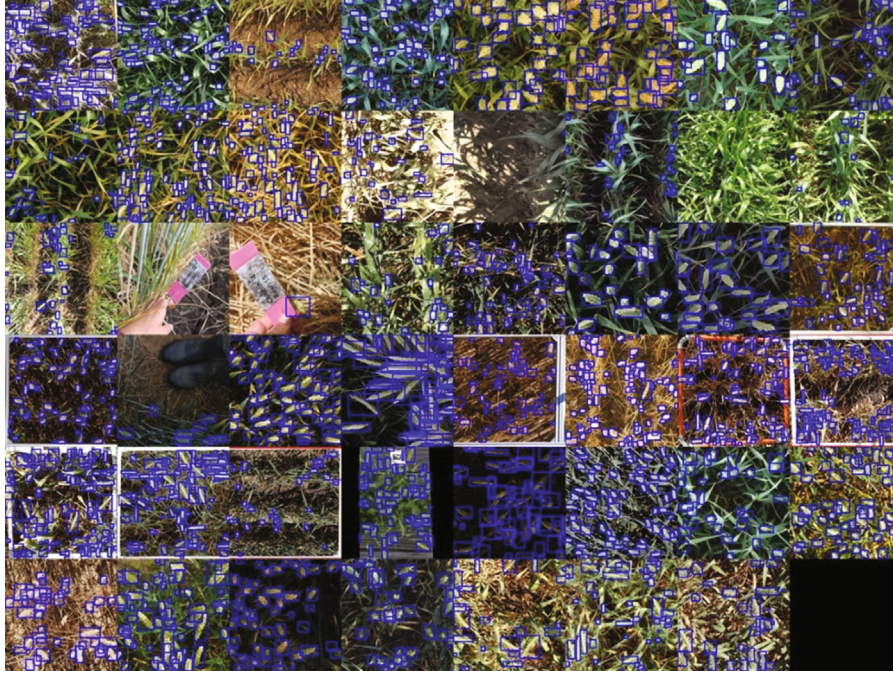


FIGURE 1: Sample images of the Global Wheat Head Detection 2021. The blue boxes correspond to the interactively labeled heads.

and 0.4 mm, i.e., similar to that of the images in GWHD\_2020. Because none of them was already labeled, a sample was selected by taking no more than one image per micro-plot, which was randomly cropped to  $1024 \times 1024$  px patches that will be called images in the following for the sake of simplicity.

With the addition of 1722 images and 86,000 wheat heads, the GWHD\_2021 dataset contains 6500 images and 275,000 wheat heads. The increase in the number of subdatasets from 18 to 47 leads to a larger diversity between them which can be observed on Figure 1. The subdatasets are described in Table 1. However, the new definition of a subdataset led also to more unbalanced subdatasets: the smallest (Arvalis\_8) contains only 20 images, while the biggest (ETHZ\_1) contains 747 images. This provides the opportunity to possibly take advantage of the data distribution to improve model training. Each subdataset has been visually assigned to several development stage classes depending on the respective color of leaves and heads (Figure 2): postflowering, filling, filling-ripening, and ripening. Examples of the different stages are presented in Figure 2. While being approximative, this metadata is expected to improve model training.

### 3. Dataset Diversity Analysis

In comparison to GWHD\_2020, the GWHD\_2021 dataset puts emphasis on metadata documentation of the different subdatasets, as described in the discussion section of David et al. [18]. Alongside the acquisition platform, each subdataset has been reviewed and a development stage was assigned to each, except for Utokyo\_3 (formerly utokyo\_2) as it is a

collection of images from various farmer fields and development stages. Globally, the GWHD\_2021 dataset covers well all development stages ranging from postanthesis to ripening (Figure 2).

The diversity between images within the GWHD\_2021 dataset was documented using the method proposed by Tolias et al. [48]. The deep learning image features were first extracted from the VGG-16 deep network pretrained on the ImageNet dataset that is considered representing well the general features of RGB images. We then selected the last layer which has a size of  $14 \times 14 \times 512$  and summed it into a unique vector of 512 channels, which is then normalized. Then, the UMAP dimensionality reduction algorithm [49] was used to project representations into a 2D space. The UMAP algorithm is used to keep the existing clusters during the projection to a low-dimension space. This 2D space is expected to capture the main features of the images. Results (Figure 3) demonstrate that the test dataset used for GWHD\_2020 was biased in comparison to the training dataset. The subdatasets added in 2021 populate more evenly the 2D space which is expected to improve the robustness of the models.

### 4. Presentation of Global Wheat Challenge 2021 (GWC 2021)

The results from the Kaggle challenge based on GWHD\_2020 have been analyzed by the authors [14]. The findings emphasize that the design of a competition is critical to enable solutions that improve the robustness of the wheat head detection models. The Kaggle competition was based on a metric that was averaged across all test images, without

TABLE 1: The subdatasets for GWHD\_2020 and GWHD\_2021. The column “2020 name” indicates the name given to the subdatasets for GWHD\_2020, which were split into several new subdatasets.

GWHD_2021 subdataset name	GWHD_2020 subdataset name	Owner	Country	Location	Acquisition date	Platform	Development stage	Number of images	Number of wheat head
<i>Ethz_1</i>	<i>ethz_1</i>	ETHZ	Switzerland	Usask	06/06/2018	Spidercam	Filling	747	49603
<i>Rres_1</i>	<i>rres_1</i>	Rothamsted	UK	Rothamsted	13/07/2015	Gantry	Filling-ripening	432	19210
<i>ULiège-GxABT_1</i>		Uliège/ Gembloux	Belgium	Gembloux	28/07/2020	Cart	Ripening	30	1847
<i>NMBU_1</i>		NMBU	Norway	NMBU	24/07/2020	Cart	Filling	82	7345
<i>NMBU_2</i>		NMBU	Norway	NMBU	07/08/2020	Cart	Ripening	98	5211
<i>Arvalis_1</i>	<i>arvalis_1</i>	Arvalis	France	Gréoux	02/06/2018	Handheld	Postflowering	66	2935
<i>Arvalis_2</i>	<i>arvalis_1</i>	Arvalis	France	Gréoux	16/06/2018	Handheld	Filling	401	21003
<i>Arvalis_3</i>	<i>arvalis_1</i>	Arvalis	France	Gréoux	07/2018	Handheld	Filling-ripening	588	21893
<i>Arvalis_4</i>	<i>arvalis_2</i>	Arvalis	France	Gréoux	27/05/2019	Handheld	Filling	204	4270
<i>Arvalis_5</i>	<i>arvalis_3</i>	Arvalis	France	VLB *	06/06/2019	Handheld	Filling	448	8180
<i>Arvalis_6</i>	<i>arvalis_3</i>	Arvalis	France	VSC *	26/06/2019	Handheld	Filling-ripening	160	8698
<i>Arvalis_7</i>		Arvalis	France	VLB *	06/2019	Handheld	Filling-ripening	24	1247
<i>Arvalis_8</i>		Arvalis	France	VLB *	06/2019	Handheld	Filling-ripening	20	1062
<i>Arvalis_9</i>		Arvalis	France	VLB *	06/2020	Handheld	Ripening	32	1894
<i>Arvalis_10</i>		Arvalis	France	Mons	10/06/2020	Handheld	Filling	60	1563
<i>Arvalis_11</i>		Arvalis	France	VLB *	18/06/2020	Handheld	Filling	60	2818
<i>Arvalis_12</i>		Arvalis	France	Gréoux	15/06/2020	Handheld	Filling	29	1277
<i>Inrae_1</i>	<i>inrae_1</i>	INRAe	France	Toulouse	28/05/2019	Handheld	Filling-ripening	176	3634
<i>Usask_1</i>	<i>usask_1</i>	USaskatchewan	Canada	Saskatchewan	06/06/2018	Tractor	Filling-ripening	200	5985
<i>KSU_1</i>		Kansas State University	US	KSU	19/05/2016	Tractor	Postflowering	100	6435
<i>KSU_2</i>		Kansas State University	US	KSU	12/05/2017	Tractor	Postflowering	100	5302
<i>KSU_3</i>		Kansas State University	US	KSU	25/05/2017	Tractor	Filling	95	5217
<i>KSU_4</i>		Kansas State University	US	KSU	25/05/2017	Tractor	Ripening	60	3285
<i>Terraref_1</i>		TERRA-REF project	US	Maricopa, AZ	02/04/2020	Gantry	Ripening	144	3360
<i>Terraref_2</i>		TERRA-REF project	US	Maricopa, AZ	20/03/2020	Gantry	Filling	106	1274
<i>CIMMYT_1</i>		CIMMYT	Mexico	Ciudad Obregon	24/03/2020	Cart	Postflowering	69	2843
<i>CIMMYT_2</i>		CIMMYT	Mexico	Ciudad Obregon	19/03/2020	Cart	Postflowering	77	2771
<i>CIMMYT_3</i>		CIMMYT	Mexico	Ciudad Obregon	23/03/2020	Cart	Postflowering	60	1561
<i>Utokyo_1</i>	<i>utokyo_1</i>	UTokyo	Japan	NARO-Tsukuba	22/05/2018	Cart - **	Ripening	538	14185
<i>Utokyo_2</i>	<i>utokyo_1</i>	UTokyo	Japan	NARO-Tsukuba	22/05/2018	Cart - **	Ripening	456	13010
<i>Utokyo_3</i>	<i>utokyo_2</i>	UTokyo	Japan	NARO-Hokkaido	Multi-years - ***	Handheld	Multiple	120	3085
<i>Ukyoto_1</i>		UKyoto	Japan	Kyoto	30/04/2020	Handheld	Postflowering	60	2670
<i>NAU_1</i>	<i>NAU_1</i>	NAU	China	Baima	n.a	Handheld	Postflowering	20	1240
<i>NAU_2</i>		NAU	China	Baima	02/05/2020	Cart	Postflowering	100	4918

TABLE 1: Continued.

GWHD_2021 subdataset name	GWHD_2020 subdataset name	Owner	Country	Location	Acquisition date	Platform	Development stage	Number of images	Number of wheat head
<u>NAU_3</u>		<u>NAU</u>	<u>China</u>	<u>Baima</u>	<u>09/05/2020</u>	<u>Cart</u>	<u>Filling</u>	<u>100</u>	<u>4596</u>
<i>UQ_1</i>	<i>uq_1</i>	<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>12/08/2015</i>	<i>Tractor</i>	<i>Postflowering</i>	<i>22</i>	<i>640</i>
<i>UQ_2</i>	<i>uq_1</i>	<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>08/09/2015</i>	<i>Tractor</i>	<i>Postflowering</i>	<i>16</i>	<i>39</i>
<i>UQ_3</i>	<i>uq_1</i>	<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>15/09/2015</i>	<i>Tractor</i>	<i>Filling</i>	<i>14</i>	<i>297</i>
<i>UQ_4</i>	<i>uq_1</i>	<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>01/10/2015</i>	<i>Tractor</i>	<i>Filling</i>	<i>30</i>	<i>1039</i>
<i>UQ_5</i>	<i>uq_1</i>	<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>09/10/2015</i>	<i>Tractor</i>	<i>Filling-ripening</i>	<i>30</i>	<i>3680</i>
<i>UQ_6</i>	<i>uq_1</i>	<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>14/10/2015</i>	<i>Tractor</i>	<i>Filling-ripening</i>	<i>30</i>	<i>1147</i>
<i>UQ_7</i>		<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>06/10/2020</i>	<i>Handheld</i>	<i>Ripening</i>	<i>17</i>	<i>1335</i>
<i>UQ_8</i>		<i>UQueensland</i>	<i>Australia</i>	<i>McAllister</i>	<i>09/10/2020</i>	<i>Handheld</i>	<i>Ripening</i>	<i>41</i>	<i>4835</i>
<i>UQ_9</i>		<i>UQueensland</i>	<i>Australia</i>	<i>Brookstead</i>	<i>16/10/2020</i>	<i>Handheld</i>	<i>Filling-ripening</i>	<i>33</i>	<i>2886</i>
<i>UQ_10</i>		<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>22/09/2020</i>	<i>Handheld</i>	<i>Filling-ripening</i>	<i>53</i>	<i>8629</i>
<i>UQ_11</i>		<i>UQueensland</i>	<i>Australia</i>	<i>Gatton</i>	<i>31/08/2020</i>	<i>Handheld</i>	<i>Postflowering</i>	<i>42</i>	<i>4345</i>
<u>ARC_1</u>		<u>ARC</u>	<u>Sudan</u>	<u>Wad Medani</u>	<u>03/2021</u>	<u>Handheld</u>	<u>Filling</u>	<u>30</u>	<u>888</u>
Total								6515	275187

\*VLB: Villiers le Bâcle; VSC: Villers-Saint-Christophe. \*\*Utokyo\_1 and Utokyo\_2 were taken at the same location with different sensors. \*\*\*Utokyo\_3 is a special subdataset made from images coming from a large variety of farmers in Hokaido between 2016 and 2019. Italic: Europe; bold: North America; underline: Asia; bold italic: Oceania; bold underline: Africa.

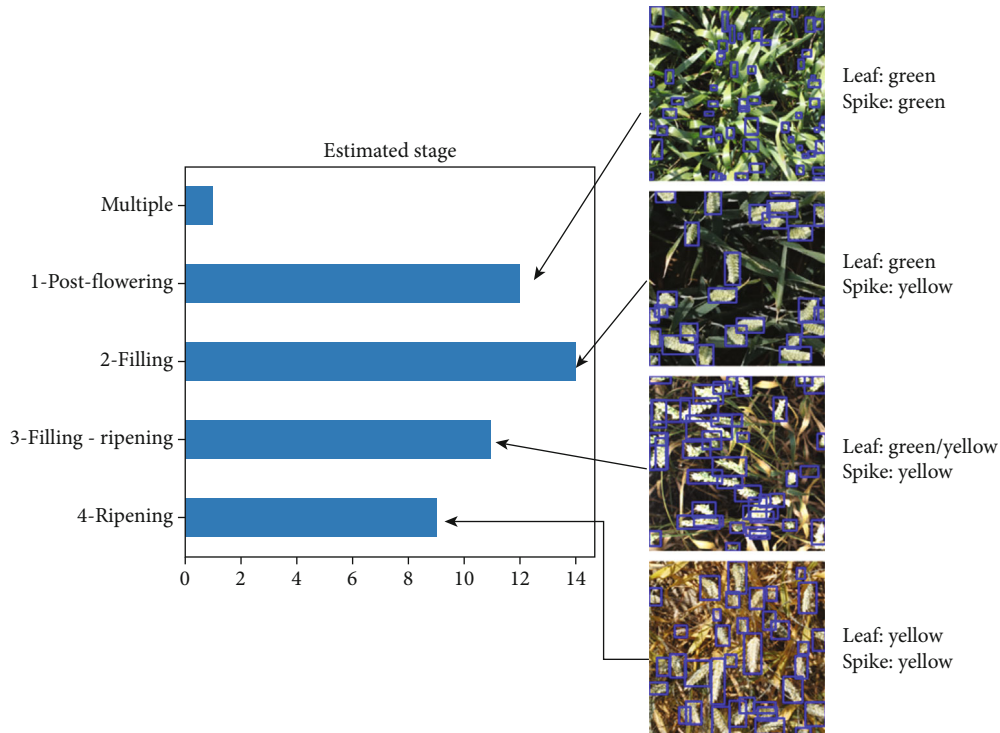


FIGURE 2: Distribution of the development stage. The x-axis presents the number of subdataset per development stage.

distinction for the subdatasets, and it was biased toward a strict match of the labelling. This artificially enhances the influence on the global score of the largest datasets such as utokyo\_1 (now split into Utokyo\_1 and Utokyo\_2). Further, the metrics used to score the agreement with the labeled

heads and largely used for big datasets, such as MS COCO, appear to be less efficient when some heads are labeled in a more uncertain way as it was the case in several situations depending on the development stage, illumination conditions, and head density. As a result, the weighted domain

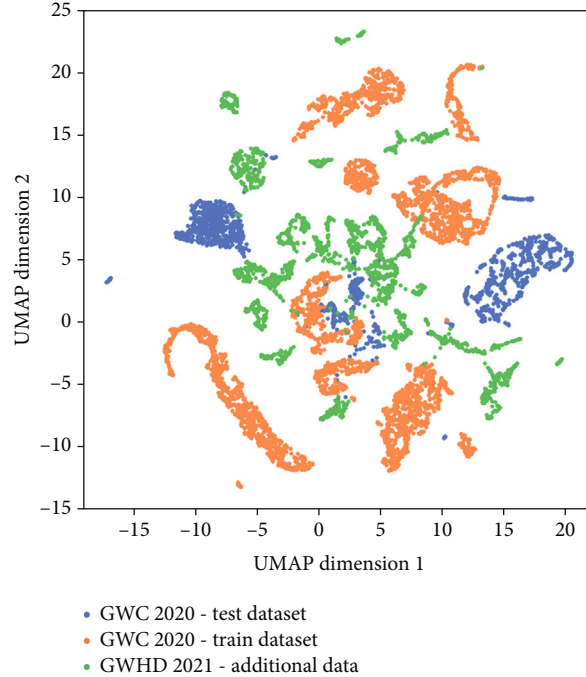


FIGURE 3: Distribution of the images in the two first dimensions defined by the UMAP algorithm for the GWHD 2021 dataset. The additional subdatasets as well as the training and test datasets from GWHD\_2020 are represented by colors.

TABLE 2: Presentation of the Global Wheat Challenge 2021 results.

Solution name	WDA
randomTeamName (1 <sup>st</sup> place)	0.700
David_jeon (2 <sup>nd</sup> place)	0.695
SMART (2 <sup>nd</sup> place)	0.695
Reference (faster-RCNN)	0.492

accuracy is proposed as a new metric [14]. The accuracy computed over image  $i$  belonging to domain  $d$ ,  $AI_d(i)$ , is classically defined as

$$AI_d(i) = \frac{TP}{TP + FN + FP}, \quad (1)$$

where TP, FN, and FP are, respectively, the number of true positive, false negative, and false positive found in image  $i$ . The weighted domain accuracy (WDA) is the weighted average of all domain accuracies:

$$WDA = \frac{1}{D} \sum_{d=1}^D \frac{1}{n_d} * \sum_{i=1}^{n_d} AI_{di}, \quad (2)$$

where  $D$  is the number of domains (subdatasets) and  $n_d$  is the number of images in domain  $d$ . The training, validation, and test datasets used are presented in Section 5.

The results of the Global Wheat Challenge 2021 are summarized in Table 2. The reference method is a faster-RCN with the same parameters than in the research paper GWHD\_2020 [18] and trained on the GWHD\_2021 (Global Wheat Challenge 2021 split) training dataset. The

full leaderboard can be found at <https://www.aicrowd.com/challenges/global-wheat-challenge-2021/leaderboards>.

## 5. How to Use/FAQ

- (i) How to download? The dataset can be download on Zenodo: <https://zenodo.org/record/5092309>
- (ii) What is the license of the dataset? The dataset is under the MIT license, allowing for reuse without restriction
- (iii) How to cite the dataset? The present paper can be cited when using the GWHD\_2021 dataset. However, cite preferentially [18] for wheat head detection challenges or when discussing the difficulty to constitute a large datasets
- (iv) How to benchmark? Depending on the objectives of the study, we recommend two sets of training, validation, and test (Table 3):
  - (a) The Global Wheat Challenge 2021 split when the dataset is used for phenotyping purpose, to allow direct comparison with the winning solutions
  - (b) The “GlobalWheat-WILDS” split is the one used for the WILDS paper [50]. We recommend to use the GlobalWheat-WILDS split when working on out-of-domain distribution shift problems

It is further recommended to keep the weighted domain accuracy for comparison with previous works.

TABLE 3: Presentation of the different splits which can be used with the GWHD\_2021.

Splits	Training	Validation	Test
Global Wheat Challenge 2021	Ethz_1, Rres_1, Inrae_1, Arvalis (all), NMBU (all), ULiège-GxABT (all)	UQ_1 to UQ_6, Utokyo (all), NAU_1, Usask_1	UQ_7 to UQ_12, Ukyoto_1, NAU_2 and NAU_3, ARC_1, CIMMYT (all), KSU (all), Terraref (all)
GlobalWheat-WILDS	Ethz_1, Rres_1, Inrae_1, Arvalis (all), NMBU (all), ULiège-GxABT (all)	UQ (all), Utokyo (all), Ukyoto_1, NAU (all)	CIMMYT (all), KSU (all); Terraref (all), Usask_1, ARC_1

## 6. Conclusion

The second edition of the Global Wheat Head Detection, GWHD\_2021, alongside the organization of a second Global Wheat Challenge is an important step for illustrating the usefulness of open and shared data across organizations to further improve high-throughput phenotyping methods. In comparison to the GWHD\_2020 dataset, it represents five new countries, 22 new subdatasets, 1200 new images, and 120,000 new-labeled wheat heads. Its revised organization and additional diversity are more representative of the type of images researchers and agronomists can acquire across the world. The revised metrics used to evaluate the models during the Global Wheat Challenge 2021 can help researchers to benchmark one-class localization models on a large range of acquisition conditions. GWHD\_2021 is expected to accelerate the building of robust solutions. However, progress on the representation of developing countries is still expected and we are open to new contributions from South America, Africa, and South Asia. We started to include nadir view photos from smartphones, to get a more comprehensive dataset and train reliable models for such affordable devices. Additional works are required to adapt such an approach to other vectors such as a camera mounted on unmanned aerial vehicle, or other high-resolution cameras working in other spectral domains. Further, it is planned to release wheat head masks alongside the bounding box given the very large number of boxes that already exists and provides more associated metadata.

## Data Availability

The dataset is available on Zenodo (<https://zenodo.org/record/5092309>).

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## Acknowledgments

We would like to thank the company “Human in the loop”, which corrected and labeled the new datasets. The help of Frederic Venault (INRAE Avignon) was also precious to check the labelled images. The work received support from ANRT for the CIFRE grant of Etienne David, cofunded by Arvalis for the project management. The labelling work was supported by several companies and projects, including Canada: The Global Institute Food Security, University of

Saskatchewan which supported the organization of the competition. France: This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004 PIA #Digitag. Institut Convergences Agriculture Numérique, Hiphen supported the organization of the competition. Japan: Kubota supported the organization of the competition. Australia: Grains Research and Development Corporation (UOQ2002-008RTX machine learning applied to high-throughput feature extraction from imagery to map spatial variability and UOQ2003-011RTX INVITA—a technology and analytics platform for improving variety selection) supported competition.

## References

- [1] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, *Everyone wants to do the model work, not the data work: data cascades in high-stakes AI*, New York, NY, USA, 2021.
- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: a large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.
- [3] T.-Y. Lin et al., “Microsoft coco: common objects in context,” *European conference on computer vision*, pp. 740–755, 2014.
- [4] J. A. Cruz, X. Yin, X. Liu et al., “Multi-modality imagery database for plant phenotyping,” *Machine Vision and Applications*, vol. 27, no. 5, pp. 735–749, 2016.
- [5] W. Guo, B. Zheng, A. B. Potgieter et al., “Aerial imagery analysis – quantifying appearance and number of sorghum heads for applications in breeding and agronomy,” *Frontiers in Plant Science*, vol. 9, p. 1544, 2018.
- [6] D. P. Hughes and M. Salathé, “An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing,” *CoRR*, 2015, <http://arxiv.org/abs/1511.08060>.
- [7] D. LeBauer et al., “Data from: TERRA-REF, an open reference data set from high resolution genomics, phenomics, and imaging sensors,” *Dryad*, p. 800302508 bytes, 2020.
- [8] S. Leminen Madsen, S. K. Mathiassen, M. Dyrmann, M. S. Laursen, L. C. Paz, and R. N. Jørgensen, “Open plant phenotype database of common weeds in Denmark,” *Remote Sensing*, vol. 12, no. 8, p. 1246, 2020.
- [9] H. Lu, Z. Cao, Y. Xiao, B. Zhuang, and C. Shen, “TasselNet: counting maize tassels in the wild via local counts regression network,” *Plant Methods*, vol. 13, no. 1, p. 79, 2017.
- [10] S. Madec, K. Irfan, E. David et al., *The P2S2 segmentation dataset: annotated in-field multi-crop RGB images acquired under various conditions*, Lyon, France, 2019 <https://hal.inrae.fr/hal-03140124>.

- [11] H. Scharr, M. Minervini, A. P. French et al., “Leaf segmentation in plant phenotyping: a collation study,” *Machine Vision and Applications*, vol. 27, no. 4, pp. 585–606, 2016.
- [12] R. Thapa, K. Zhang, N. Snaveley, S. Belongie, and A. Khan, “The Plant Pathology challenge 2020 data set to classify foliar disease of apples,” *Applications in Plant Sciences*, vol. 8, no. 9, article e11390, 2020.
- [13] T. Wiesner-Hanks, E. L. Stewart, N. Kaczmar et al., “Image set for deep learning: field images of maize annotated with disease symptoms,” *BMC Research Notes*, vol. 11, no. 1, p. 440, 2018.
- [14] E. David, F. Ogidi, W. Guo, F. Baret, and I. Stavness, *Global Wheat Challenge 2020: analysis of the competition design and winning models*, 2021.
- [15] N. Hani, P. Roy, and V. Isler, “MinneApple: a benchmark dataset for apple detection and segmentation,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 852–858, 2020.
- [16] M. Minervini, A. Fischbach, H. Scharr, and S. A. Tsafaris, “Finely-grained annotated datasets for image-based plant phenotyping,” *Pattern Recognition Letters*, vol. 81, pp. 80–89, 2016.
- [17] S. A. Tsafaris and H. Scharr, “Sharing the right data right: a symbiosis with machine learning,” *Trends in Plant Science*, vol. 24, no. 2, pp. 99–102, 2019.
- [18] E. David, S. Madec, P. Sadeghi-Tehran et al., “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*, vol. 2020, article 3521852, 12 pages, 2020.
- [19] G. Yu, Y. Wu, J. Xiao, and Y. Cao, “A novel pyramid network with feature fusion and disentanglement for object detection,” *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6685954, 13 pages, 2021.
- [20] T. W. Ayalew, J. R. Ubbens, and I. Stavness, “Unsupervised domain adaptation for plant organ counting,” *European conference on computer vision*, pp. 330–346, 2020.
- [21] M. N. Datta, Y. Rathi, and M. Eliazar, “Wheat heads detection using deep learning algorithms,” *Annals of the Romanian Society for Cell Biology*, pp. 5641–5654, 2021.
- [22] F. Fourati, W. S. Mseddi, and R. Attia, “Wheat head detection using deep, semi-supervised and ensemble learning,” *Canadian Journal of Remote Sensing*, vol. 47, no. 2, pp. 198–208, 2021.
- [23] F. Fourati, W. Souidene, and R. Attia, “An original framework for wheat head detection using deep, semi-supervised and ensemble learning within Global Wheat Head Detection (GWHD) dataset,” 2020, <https://arxiv.org/abs/2009.11977>.
- [24] A. S. Gomez, E. Aptoula, S. Parsons, and P. Bosilj, “Deep regression versus detection for counting in robotic phenotyping,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2902–2907, 2021.
- [25] B. Gong, D. Ergu, Y. Cai, and B. Ma, “Real-time detection for wheat head applying deep neural network,” *Sensors*, vol. 21, no. 1, p. 191, 2021.
- [26] M.-X. He, P. Hao, and Y. Z. Xin, “A robust method for wheat-ear detection using UAV in natural scenes,” *IEEE Access*, vol. 8, pp. 189043–189053, 2020.
- [27] Y. Jiang, C. Li, R. Xu, S. Sun, J. S. Robertson, and A. H. Pater-son, “DeepFlower: a deep learning-based approach to characterize flowering patterns of cotton plants in the field,” *Plant Methods*, vol. 16, no. 1, p. 156, 2020.
- [28] B. Jiang, J. Xia, and S. Li, “Few training data for objection detection,” in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*, pp. 579–584, November 2020.
- [29] A. Karwande, P. Kulkarni, P. Marathe, T. Kolhe, M. Wyawahare, and P. Kulkarni, “Computer vision-based wheat grading and breed classification system: a design approach,” *Machine Learning and Information Processing: Proceedings of ICMLIP 2020*, , Springer, p. 403, 2021.
- [30] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, “Review on convolutional neural networks (CNN) in vegetation remote sensing,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24–49, 2021.
- [31] S. Khaki, N. Safaei, H. Pham, and L. Wang, “WheatNet: a light-weight convolutional neural network for high-throughput image-based wheat head detection and counting,” 2021, <https://arxiv.org/abs/2103.09408>.
- [32] S. U. Kolhar and J. Jagtap, *Bibliometric Review on Image Based Plant Phenotyping*16.
- [33] J. Li, C. Li, S. Fei et al., “Wheat ear recognition based on RetinaNet and transfer learning,” *Sensors*, vol. 21, no. 14, p. 4845, 2021.
- [34] L. Lucks, L. Haraké, and L. Klingbeil, *Detektion von Weizenähren mithilfe neuronaler Netze und synthetisch erzeugter Trainingsdaten*, tm-Technisches Messen, 2021.
- [35] T. Misra, A. Arora, S. Marwaha et al., “Web-SpikeSegNet: deep learning framework for recognition and counting of spikes from visual images of wheat plants,” *IEEE Access*, vol. 9, pp. 76235–76247, 2021.
- [36] L. G. Riera, M. E. Carroll, Z. Zhang et al., “Deep multiview image fusion for soybean yield estimation in breeding applications,” *Plant Phenomics*, vol. 2021, article 9846470, 12 pages, 2021.
- [37] D. T. Smith, A. B. Potgieter, and S. C. Chapman, “Scaling up high-throughput phenotyping for abiotic stress selection in the field,” *Theoretical and Applied Genetics*, vol. 134, no. 6, pp. 1845–1866, 2021.
- [38] Y. Suzuki, D. Kuyoshi, and S. Yamane, “Transfer learning algorithm for object detection,” *Bulletin of Networking, Computing, Systems, and Software*, vol. 10, no. 1, pp. 1–3, 2021.
- [39] R. Trevisan, O. Pérez, N. Schmitz, B. Diers, and N. Martin, “High-throughput phenotyping of soybean maturity using time series UAV imagery and convolutional neural networks,” *Remote Sensing*, vol. 12, no. 21, p. 3617, 2020.
- [40] K. Velumani, R. Lopez-Lozano, S. Madec et al., “Estimates of maize plant density from UAV RGB images using Faster-RCNN detection model: impact of the spatial resolution,” 2021, <https://arxiv.org/abs/2105.11857>.
- [41] Y. Wu, Y. Hu, and L. Li, “BTWD: bag of tricks for wheat detection,” in *European Conference on Computer Vision*, pp. 450–460, Springer, 2020.
- [42] H. Wang, Y. Duan, Y. Shi, Y. Kato, S. Ninomiya, and W. Guo, “EasyIDP: a Python package for intermediate data processing in UAV-based plant phenotyping,” *Remote Sensing*, vol. 13, no. 13, p. 2622, 2021.
- [43] Y. Wang, Y. Qin, and J. Cui, “Occlusion robust wheat ear counting algorithm based on deep learning,” *Frontiers in Plant Science*, vol. 12, p. 1139, 2021.
- [44] B. Yang, Z. Gao, Y. Gao, and Y. Zhu, “Rapid detection and counting of wheat ears in the field using YOLOv4 with attention module,” *Agronomy*, vol. 11, no. 6, p. 1202, 2021.



- [45] H. Lu, L. Liu, Y. N. Li, X. M. Zhao, X. Q. Wang, and Z. G. Cao, "TasselNetV3: Explainable Plant Counting With Guided Upsampling and Background Suppression," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2021.
- [46] S. Dandrifosse, A. Carlier, B. Dumont, and B. Mercatoris, "Registration and fusion of close-range multimodal wheat images in field conditions," *Remote Sensing*, vol. 13, no. 7, p. 1380, 2021.
- [47] X. Wang, H. Xuan, B. Evers, S. Shrestha, R. Pless, and J. Poland, "High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat," *GigaScience*, vol. 8, no. giz120, 2019.
- [48] G. Tolas, R. Sicre, and H. Jégou, "Particular object retrieval with integral max-pooling of CNN activations," 2015, <https://arxiv.org/abs/1511.05879>.
- [49] L. McInnes, J. Healy, and J. Melville, *UMAP: Uniform manifold approximation and projection for dimension reduction.*, 2020.
- [50] P. W. Koh, S. Sagawa, H. Marklund et al., "WILDS: a benchmark of in-the-wild distribution shifts," 2021, April 2021, <https://arxiv.org/abs/2012.07421>.