# Deep learning algorithms for high-throughput cereal plant and organ identification

Kaaviya Velumani

**THÈSE DE DOCTORAT D'AVIGNON UNIVERSITÉ**

École Doctorale N° 536
Agrosciences et Sciences

**Spécialité / Discipline de doctorat : Télédétection de la Végétation**

**INRAE UMR 1114 EMMAH - UMT CAPTE (CAPteurs et TElédétection)**

**Présentée par**

**Kaaviya Velumani**

---

# Deep learning algorithms for high-throughput cereal plant and organ identification

---

**Soutenance publique** : 5 Juillet 2021

**Jury composé de** :

| | | |
|---|---|---|
| Mme. Evelyn Costes | Directrice de Recherche, INRAE UMR AGAP, Montpellier | Rapporteur |
| M. David Rousseau | Enseignant-Chercheur, LARIS, Université d'Angers | Rapporteur |
| M. Frederic Baret | Directeur de Recherche, EMMAH-CAPTE, INRAE Avignon | Directeur de thèse |
| M. Raul Lopez-Lozano | Chargé de Recherche, EMMAH-CAPTE, INRAE Avignon | Co-directeur de thèse |

**Encadrement de la thèse : Dr Frederic Baret, Dr Raul Lopez-Lozano**

# Acknowledgements

My heartfelt thanks to the directors of my thesis, Fred and Raul, for their guidance, for sharing their experiences, ideas and a considerable amount of their time during these three years. A special mention to Raul who was always open to discussions and very encouraging. Without their constant support, this work would not have been possible.

I also take this opportunity to extend my thanks to Alexis, Joss and the members of Hiphen.

I would also like to thank Fred Venault, Franck Tison, Adrien Vielix, Guy Deshayes, Wenjuan Li, Myriam Bounaffa, Micheline Debroux, Shouyang Liu, Jeremy Labrosse, Rafael Benicio, Daniel Smith, Simon Madec, Gaetan Daubige and Etienne David for their support for the collection, annotations and processing of various datasets which were important to complete this study!

I am grateful to the members of INRAe, Hiphen and Arvalis for the wonderful memories and experiences shared during these three years.

Finally, I would like to thank my family and friends who were always supportive through the ups and downs.

# List of abbreviations and acronyms

| | |
|---|---|
| UAV | Unmanned Aerial Vehicles |
| RGB | Red Green Blue |
| GSM | Global System for Mobile communication |
| LiDAR | Light Detection And Ranging |
| NIR | Near Infra-Red |
| SWIR | Short Wave Infra-Red |
| GAI | Green Area Index |
| GF | Green Fraction |
| SVM | Support Vector Machines |
| CNN | Convolutional Neural Network |
| RT | Radiative Transfer |
| TLR | Technological Readiness Level |
| GSD | Ground Sampling Distance |
| GPU | Graphical Processing Unit |
| SfM | Structure from Motion |
| Faster R-CNN | Faster Regions with CNN |
| rRMSE | relative Root Mean Squared Error |
| GAN | Generative Adversarial Network |
| UGV | Unmanned Ground Vehicle |
| ESRGAN | Enhanced Super-Resolution Generative Adversarial Network |
| SR | Super Resolution |
| HR | High Resolution |
| LR | Low Resolution |
| AP | Average Precision |
| TP | True Positive |
| FP | False Positive |
| FN | False Negative |
| IOU | Intersection Over Union |
| GWHD | Global Wheat Head Detection |
| RTK GPS | Real Time Kinematic Global Positioning System |
| GCP | Ground Control Point |
| SFC$^2$Net | Scale Fusion Counting Classification Network |
| rMAE | relative Mean Absolute Error |
| mAP | mean Average Precision |
| rRMSE | relative Root Mean Squared Error |
| rBias | Relative Bias |
| $R^2$ | Correlation co-efficient |
| OA | Overall Accuracy |
| $H^2$ | Broad-sense Heritability |
| IoTA | Internet of Things for Agriculture |
| PNG | Portable Network Graphics |
| FOV | Field of View |
| Grad-CAM | Gradient Class Activation Maps |

# Table des matières

# 1 Introduction

## 1.1 Role of cereal crops and food security, climate change

Cereal crops are plants belonging to the family Poaceae (also known as Gramineae), and are the most critical source of food for the world population (FAO 2002). They are cultivated worldwide for their edible grains which have high nutritional content in terms of energy, protein, carbohydrates, fiber as well as a variety of macronutrients (McKevith 2004). The cereals thus make an important part of the human diet and livestock feed. The most widely cultivated cereals are wheat, rice, maize, barley, sorghum, millets, oats and rye with an all-time high total production of 2.742 billion tones in 2019 (FAO 2020). However, the number of under-nourished people in the world has increased in the recent years and it is projected to reach 841.4 million by 2030, not considering the impact of the Covid-19 pandemic (FAO et al. 2020). Even though the total cereal yield is expected to increase by 1.1% per annum (OECD and Food and Agriculture Organization of the United Nations 2020), it might not be sufficient to achieve zero hunger and food security targeted for 2030 (FAO et al. 2020), especially if we take into account the regional disparity regarding the yield. In addition, the agriculture industry will be challenged by the following key factors in the coming years –

- An increase in global food demand triggered by a growing population.
- Economic growth and increase in purchasing power of developing countries which would hasten a dietary transition with higher consumption of meat, fish and vegetables, adding pressure to natural resources (FAO 2017).
- Climate change studies indicate a reduction in crop yields due to extreme weather conditions resulting in drought, floods, cyclones, intense pest attacks which is expected to affect predominantly the tropical regions (FAO 2017; Maggio, Van Criekinge, and Malingreau 2016; Rosenzweig et al. 2014; Rosenzweig and Parry 1994).
- Shift towards sustainable and organic farming practices: the growing concerns over global warming and environmental degradation from conventional agriculture has led to the popularity of organic produce, with a doubling in the area under organic farming in the last ten years (Lernoud and Willer 2019). However, at present, the yield from sustainable farming practices is relatively less than from conventional (Lal 2016; Muller et al. 2017).
- Depletion of natural resources and reduction in cultivable land area.

To sum up, the agricultural production systems need to reinvent themselves to produce more in an economically, environmentally, and socially viable manner. The use of technology is favored for tackling these challenges by adopting smart crop breeding programs to develop resistant and appropriate cultivars (Parent et al. 2018); and informed farm management practices to optimize the yield with minimal exploitation

of the available natural resources (Basso and Antle 2020; Gebbers and Adamchuk 2010). In the next section we will see how the use of technology in agriculture can help to overcome these challenges.

## 1.2 Plant Phenomics as a tool to support cereal breeding programs

The recent advances in plant genomics have generated new opportunities to increase plant genetic variability, with tremendous potentials for crop improvement. However, the effective contribution of these advances to increase crop productivity depends on how tightly genotypic traits can be linked with those eco-physiological mechanisms that produce a distinguishable response of the genotype to the environment (Tardieu et al. 2017). The result of that response is known as phenotype.

Plant phenomics –the observation of plant phenotypic traits– is the discipline that must fill the gap between genotype and phenotype (Fiorani and Schurr 2013). Traditionally, field phenotyping has relied on manual or destructive, low-throughput, observations of phenotypic traits such as plant height, crop development stage, and yield components. The development, in the recent years, of high-throughput field phenotyping platforms (Figure 1) and instruments –capable of acquiring and processing efficiently massive volumes of *in situ* observations over field experiments– has opened a new era of plant phenomics. This has an enormous potential impact on the efficiency of breeding programs, as it would enable plant breeders to phenotype large number of genotypes accurately, thus allowing them to evaluate and identify the best ones (Araus and Cairns 2014).

*Figure 1 Aerial view of wheat phenotyping platforms at Queensland, Australia and Gréoux les Bains, France used in this study.*

### 1.2.1 Vectors and sensors for field high-throughput phenotyping

Field trials for plant breeding require the monitoring of a large number of cultivars under multiple environmental conditions. To satisfy this need, field phenotyping platforms are designed to conduct factorial experiments–genotypes x treatments x replicates– over thousands of microplots (with size usually between 10 and 20 $m^2$). In order to efficiently monitor such large-scale experiments, high-throughput methods are required. The type of vectors used in phenotyping platforms are classified roughly in two groups: ground level or *in situ* vectors, and aerial vectors (Table 1).

Unmanned aerial vehicles (UAVs) are the most widely used aerial vector (Xie and Yang 2020; Yang et al. 2017), and have a higher throughput compared to the ground vectors, requiring less than a second to acquire observations for a given microplot. The main drawback of UAVs is the low autonomy of the batteries, which prevents their use to cover very large areas.

Ground vectors, by contrast, have a higher autonomy but a lower throughput compared to UAVs. In the last decade, within the group of ground vectors, the development of autonomous ground-based robots (Deery et al. 2014; Madec et al. 2017; Quaglia et al. 2020; Ruckelshausen et al. 2009) have constituted a major evolution in the field of plant phenotyping. This is mainly due to their ability to carry multiple sensors (Table 1) and their relatively higher throughput. Handheld instruments were developed as a cost-efficient alternative to autonomous ground robots (Kaur, Donis-Gonzalez, and St. Clair 2020; Khanna et al. 2015; Reynolds, Baret, et al. 2019), that integrate small-size and lightweight cameras or portable spectrometers. With a throughput of about 1-2 minutes per microplot, handheld sensors are a suitable option for small-size phenotyping platforms. Finally, fixed systems –in most cases equipped with downward-looking RGB camera and/or a portable spectrometer– can monitor a relatively small footprint (about 10 $m^2$) with a sub-daily frequency (Reynolds, Ball, et al. 2019; Velumani et al. 2020; Yalcin 2018), which is especially useful for phenological traits among others. These fixed cameras are *a priori* not suitable for large phenotyping platforms, but they can be used to create a network of sensors on distant experimental fields, as they can communicate by GSM with a cloud storage infrastructure in real time. Sensor networks can be very useful to provide specific phenotypic traits of certain genotypes over contrasted environments, at a relatively low cost.

*Table 1. Characteristics of the main phenotyping platforms and vectors.*

| Type | Vector | | Sensors | Autonomy | Area covered | Throughput | Ground Sampling Distance |
|---|---|---|---|---|---|---|---|
| **Aerial** | Unmanned Air Vehicle (UAV) |  | Multispectral camera<br>RGB camera<br>Thermal camera | 15-30 min | 3-5 ha | 0.7 s / microplot | 0.1 – 1 cm |
| **Ground** | Autonomous ground-based robots |  | Multispectral camera<br>RGB camera<br>Thermal camera<br>LiDAR | 7 h | 1.5-2 ha | 30 s / microplot | 0.1 – 1 mm |
| | Handheld instruments |  | RGB camera<br>Portable spectrometer | 4 h | 0.2-0.4 ha | 1 – 2 min / microplot | 0.1 – 0.5 mm |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Fixed instruments |  | RGB camera<br>Portable spectrometer | 6 months | 10 $m^2$ | - | 0.3 – 0.5 mm |

Over the years, numerous works have been published using different sensors which observe at different wavelengths of the electromagnetic spectrum to study specific aspects of the plant physiology and canopy structure. Roughly, we can differentiate five main groups:

- RGB cameras which are one of the most widely used instruments to access plant physiological traits due their affordability (Araus and Kefauver 2018; L. Li, Zhang, and Huang 2014). They are versatile, lightweight and they can be mounted in any phenotyping vector, providing optical data at a very high spatial resolution.
- Multispectral cameras, which provide information in different channels within the visible, near infrared (NIR, from at 0.75 to 1.4 µm) and short-wave infrared (SWIR, 1.4 to 2.4 µm) domains of the electromagnetic spectrum. The reflectance on different spectral bands allows a better characterization of crop biophysical variables like green leaf area index (GAI), or leaf and canopy chlorophyll content, (Blancon et al. 2019; Daughtry 2000; Hunt et al. 2010; Jay et al. 2019; Laliberte et al. 2011; Verger et al. 2014).
- Portable spectrometers: Hyperspectral sensors are usually expensive systems capable of observing across several hundred wavelengths with fine bandwidth simultaneously. They are highly useful to characterize the crop biochemistry (F. Li et al. 2014; Yendrek et al. 2017) and for disease identification and quantification (Mahlein 2010; Mahlein et al. 2019; Nagasubramanian et al. 2019), similar to the applications of multi-spectral sensors.
- LiDAR (Light detection and ranging): LiDAR are active sensors initially developed to measure distance, based on the phase delay between an emitted light beam at a given trajectory and the reflected beam after hitting an object within the trajectory. They have gained popularity for plant phenotyping as the 3D point clouds generated by LiDAR provide access to canopy structure, height estimations and individual plant detections (Lin 2015; Lumme et al. 2008; Paproki et al. 2011; Paulus et al. 2014; Velumani et al. 2017). On field phenotyping applications, LiDAR are often restricted to ground vehicles as the range of the instrument is usually below 5 meters. In the recent years, UAV-mounted LiDAR systems are gaining popularity due to the introduction of advanced and light-weight systems (Lin and Habib 2021; Shendryk et al. 2020; Zhou et al. 2020)
- Thermal sensors allow the measurement of canopy temperature. Observing the dynamics of the canopy temperature allows to evaluate the altered rates of photosynthesis and transpiration which could be indirectly linked to crop responses to biotic and abiotic stresses (Alchanatis et al. 2010; Deery et al. 2016; Gómez-Candón et al. 2016; Jones et al. 2009; Ludovisi et al. 2017).

The choice of imaging system and the acquisition vector is highly dependent on the spatial/temporal/spectral resolution required for the application, extent of the study area, operational cost and acceptable error rate. In this study, we will exclusively work with RGB cameras, that permit observations under the visible range, providing high spatial resolution at an affordable cost. In the following section we will detail different traits and the throughputs at which they are accessible.

### 1.2.2 Methods and algorithms to derive phenotypic traits from RGB sensors

The advances in image analysis have made possible the estimation of a variety of plant traits. These traits may be broadly grouped into two categories – a) traits that characterize the canopy structure b) traits that are associated to the individual plant/organ characteristics. In Table 2, we resume the traits accessible under field conditions, the methods currently used and their TRL, technological readiness level (Mankins 1995), which indicates the robustness and limitations of these methods.

*Table 2 A summary of the canopy-level and organ-level traits accessible from RGB sensors along with their technology level readiness.*

| | Trait | Method | Resolution | | Technological Readiness Level | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Small (Wheat) | Small (Maize) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Canopy-level | Vegetation Index | Band comination | 20 cm | 20 cm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| | Canopy height | Structure from motion | 1 cm | 2 cm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| | Green fraction (GF) | SVM | 1 mm | 5 mm | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| | | CNN | 0.2 mm | 0.5 mm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | |
| | GAI | 1D RT inversion (GF) | 20 cm | 20 cm | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| | | 3D RT inversion (GF) | 20 cm | 20 cm | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| | Leaf orientation | RT inversion | 1 mm | 5 mm | ■ | ■ | ■ | ■ | ■ | | | | |
| | | 3D inversion | 1 mm | 5 mm | ■ | ■ | ■ | ■ | ■ | | | | |
| | Senescene | SVM | 1 mm | 5 mm | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| | Leaf rolling | DHP | x | 5 mm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | |
| | Lodging | Height / (SfM) | 1 cm | 2 cm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| Plant\Organ level | Plant density at emergence | CNN | | 5 mm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| | | Hand-crafted | 0.2 mm | | ■ | ■ | ■ | ■ | ■ | | | | |
| | Stem density | CNN (After harvest) | 0.2 mm | 1 mm | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| | Stem diameter | CNN (After harvest) | 0.2 mm | 2 mm | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| | Organ density | CNN | 0.5 mm | 0.5 mm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | |
| | Organ size | CNN | 0.5 mm | 0.5 mm | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | |
| | Phenological stage | CNN | 0.5 mm | 0.5 mm | ■ | ■ | ■ | ■ | | | | | |
| | Leaf size orientation | Stereo | 0.5 mm | 5 mm | ■ | ■ | ■ | | | | | | |
| | Leaf glaucousness | Glaucousness | 20 cm | 20 cm | ■ | ■ | ■ | | | | | | |
| | Weed infestation | CNN | | 1-5 mm | ■ | ■ | ■ | ■ | ■ | | | | |
| | Disease | Northern leaf blight | | 0.2-0.5 mm | ■ | ■ | ■ | ■ | ■ | | | | |
| | | Jaunisse / Fusariose | 0.1 mm | | ■ | ■ | ■ | | | | | | |

This table has been compiled by Dr Frederic Baret from the literature and the works conducted within the CAPTE unit (refer to Madec (2019) for a summary extended to all types of sensors). More details on traits extracted from images may be found in the literature (Fiorani and Schurr 2013; Li et al. 2020; L. Li et al. 2014; Xie and Yang 2020).

Most of the canopy level traits are accessible at a relatively higher TLR. These traits may be linked to the observed reflectance values via empirical methods and hence can be easily accessed using model inversions, band ratioing or image segmentation. On the other hand, for the plant/organ-level traits, the technological readiness is generally low. This is because the state-of-the-art solutions to access these traits involve convolutional neural networks (CNN) which outperform traditional hand-crafted methods (Li et al. 2020). Unfortunately, the lack of large, diverse labelled datasets for the CNN model training hinders the readiness of these traits. Besides, while some of these methods (e.g., organ density) are at an already mature level, more work is needed to increase their throughput.

Among the traits presented, those requiring a ground sampling distance (GSD) of more than 2mm, such as canopy height, vegetation index, crop lodging, etc., are now accessible at high throughput thanks to UAV observations. It should be noted that the GSD requirements and acquisition procedures normally differ between small plants (e.g., wheat, rice) and big plants (e.g., maize, sunflower). While the plant density at emergence for big plants is accessible now from UAV images, the TLR of this trait for small plants is not yet mature enough. Also, certain important traits such as phenological stage, organ density and disease identification are not yet high throughput for all crops and still require manual scoring in the field. The detection of phenological stage and disease onset is important both for plant breeding and crop management and requires frequent revisits to the field. Similarly, plant and plant organ density are important indicators of emergence rate and crucial for yield prediction.
The estimation of these traits poses the following challenges which are difficult to overcome using classical image analysis methods: a) tasks like phenology detection and disease appearance and quantification require data with high spatial as well as temporal resolution. b) even though a few methods have been proposed for specific disease detection in certain crops and a few phenological stages, we lack comprehensive pipelines covering all scenarios. c) The developed methods should be robust with reproducible results to be made operational across different sites. Thus, we require methods that can handle time series data and complexity of field conditions – varied illumination and soil conditions, ever-growing and changing architecture of crops and generalize over cultivars (Minervini, Scharr, and Tsaftaris 2015). On the other hand, for precision farming, certain traits need to be computed across large production farms to identify crop stress, disease and weed infestation at initial appearance. This would require regular monitoring of the crops and near real-time analysis of the data collected. For all of these reasons, it is important to develop

automatic image processing pipelines that are accurate enough and cost-efficient (Roitsch et al. 2019). This also explains why we will focus on the throughput of the methodologies while describing the traits accessible from RGB images.

## 1.3 Phenotypic traits from visual identification of plants and organs: potentials of deep learning methods for plant phenomics

As mentioned, an important number of essential phenotypic traits related with plant phenology, yield components or diseases are based on the visual identification of plants and organs –e.g., fruits, anthers, spikes/heads, or infected leaves. During the last decade, the development of pattern and object recognition algorithms from computer vision has paved the way to retrieve such traits from optical images acquired on high-throughput phenotyping platforms. Particularly, those methods based on machine learning and, especially, on convolutional neural networks (CNN), also known as deep learning methods, have garnered interest recently thanks to their versatility to solve a wide variety of problems. Progressively, deep learning methods have replaced hand-crafted methods –based on mathematical morphology (Jin et al. 2017) or hand-crafted descriptors (Jin et al. 2017; Pan, Kudo, and Toyama 2009) and conventional machine learning (Aydın and Uğur 2011; Niu et al. 2014) – providing state-of-the-art results for several computer vision applications including image classification, object detection and tracking, object counting, segmentation, image enhancement, among others.

In the following section, we will see how the advances in computer vision and the introduction of deep learning is transforming several traits previously accessible only through manual sampling into high throughput ones.

### 1.3.1 A brief history of convolution neural networks

The potential of neural networks composed of convolutional layers for pattern recognition was first showcased in LeNet (LeCun et al. 1989). Following this, Krizhevsky, Sutskever, and Hinton (2012) developed AlexNet, a CNN-based image classification model that achieved a remarkable 10% improvement in classification accuracy on ImageNet, a popular benchmark dataset for real-world object classification (Russakovsky et al. 2015). This was a significant event that led to an exponential increase in the use of CNNs within the computer vision community. The availability of large, annotated databases and the increase of the graphical processing units (GPUs) along with their affordability are the two main component that can explain this success. Indeed to train their models, which was made of 62.3 million of parameters Krizhevsky, Sutskever, and Hinton (2012) used two GPUs for 5 to 6 days on the 1.3 million images of ImageNet .

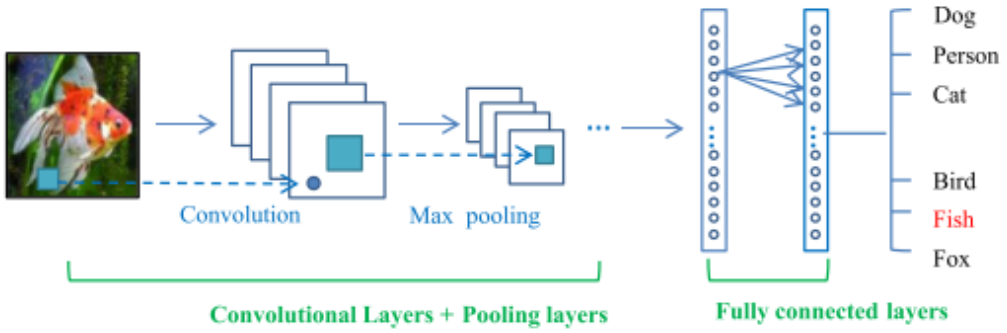


*Figure 2 An example of a typical convolutional neural network architecture containing its three base components – Convolutional layers, pooling layers and fully connected layers.* (Guo et al. 2016)

Today, CNNs are the most widely used group of deep learning models and are particularly suited for computer vision tasks. In general, a CNN is composed of three main neural layers –

a) Convolutional Layers – These are the basic component of CNNs and are capable of automatically abstracting low-level image features. Despite classical neurons, a convolutional layer consists of a series of convolutional filters with varying kernel sizes. Each neuron in the feature map is connected to a network of neurons in the previous layer through a set of trainable weights. As a result, the first convolutional layer convolves the input image into an intermediate feature map which in turn is subject to consecutive convolutions by the deeper layers producing various feature maps (Guo et al. 2016). Each feature map is passed through a non-linear activation function which allows for the extraction of non-linear features. One of the main advantages of convolutional operations is that they retain the spatial connectivity by learning correlation among neighboring pixels and are invariant to the location of the object within the image. In addition, the number of parameters is lesser owing to the weight-sharing mechanism within feature maps (Zeiler and Fergus 2012).
b) Pooling layers – A convolutional layer is usually followed by a pooling layer in order to reduce the spatial resolution of the feature maps and the number of trainable parameters. Like convolutional layers, pooling layers are also spatially invariant since they aggregate the values over neighboring pixels. There exist several strategies of aggregation during the pooling operations. The most widely used ones are max pooling which takes the maximum value among the neighboring pixels; and average pooling which takes the average value of the neighborhood pixels (Boureau, Ponce, and Lecun 2010). The drawbacks of these strategies have led to the development of other approaches such as: spatial pyramid pooling (He et al. 2015) to handle objects of different scales, sizes and aspect ratios; stochastic pooling (Zeiler and Fergus 2013) to overcome overfitting to the training data by randomly selecting an activation within the pooling regions and so on.

c) Fully connected layers – The fully connected layers are placed at the bottom of the network, following several stacks of convolutional and pooling layers. The objective of this last part of the network is to interpret the features extracted by the previous layers and convert them into high-level representations. These layers are computationally intensive, containing almost 90% of the parameters in a CNN (Guo et al. 2016). Depending on the purpose of the network, an appropriate activation is used before the last layer. For example, for a classification task, a softmax function would be used to have a probability value between 0 and 1.

Thus, the number of convolutional, pooling, fully connected layers and activation functions are selected considering the type and complexity of the task and size of the objects of interest. Another important factor to be considered while designing the network architecture is the size of the labelled dataset available for model training. Higher the number of network layers, higher are the number of network parameters that need to be trained which makes the model convergence challenging.
In the case of traditional machine learning methods, the performance of the model saturates when the size of the training data is increased beyond a certain point. On the contrary, it has been showcased that the performance of deep learning methods increases with respect to increments in the training data (Alom et al. 2019). This has led to the enlargement of existing datasets such as ImageNet and collaborative efforts among organizations to introduce new open datasets (Lin et al. 2014) in the last years. Thus, the availability of large publicly annotated datasets and easy access to high-power graphical processing units has accelerated the progress of deep learning methods. In the recent years, several different architectures have been proposed improving the state-of-the-art results on the ImageNet classification challenge. For a detailed review on the state-of-the-art models in computer vision refer to (Alom et al. 2019; Guo et al. 2016; Khan et al. 2020; Voulodimos et al. 2018).

### 1.3.2 Deep learning for plant phenomics: a domain in expansion

Over the last few years, deep learning-based methods have been incorporated within the data processing pipelines of plant phenotyping. The organization of challenges, conferences, and availability of open labelled datasets under controlled conditions have eased the initial transition from traditional machine learning approaches towards deep learning methods. This would include the popular benchmark dataset of rosette plants (Minervini et al. 2016) with successive challenges on leaf instance segmentation and counting; the PlantVillage dataset (Hughes and Salathe 2015) to classify healthy and infected leaves and characterize the type of stress. Not surprisingly, it was showcased that deep learning methods outperformed previous state of the art methods on these tasks with attempts to promote open-source codes and cross-domain applicability (Dobrescu, Giuffrida, and Tsaftaris 2017; Praveen Kumar and Domnic 2020; Ubbens et al. 2018).

The introduction of deep learning has been slower for phenotyping under field conditions. An important reason would be the additional challenges of imaging under field conditions, e.g., variation in illumination conditions, and complexities in data processing, e.g., to differentiate between individual plants within the canopy and assign organs to the respective plants to achieve phenotyping plant level (Kelly et al. 2016). One of the first applications of deep learning for field phenotyping was found for the real-time identification of weed from farm robots (McCool, Perez, and Upcroft 2017). In their study, the images were acquired using a pulse lighting system, which allowed the authors to control the illumination conditions. The authors present impressive results for the task of differentiating between vegetation and weed against soil background via image segmentation. Following this, other researches have been conducted to tackle this problem (Fawakherji et al. 2019; Lottes, Behley, Milioto, et al. 2018; Olsen et al. 2019; Sa et al. 2018). Besides segmentation, the task of detection and counting has also been widely applied in field crop environment. For example, the detection of rice panicle (Xiong et al. 2017; Zhou et al. 2019) and fruits detection (Hani, Roy, and Isler 2020; Sa et al. 2016). W. Guo et al. (2018) and Oh, Olsen, & Ramamurthy (2019) show the possibility to detect sorghum heads from UAV imageries. Lu and Cao (2020) and Madec et al. (2019) have addressed the task of wheat heads and maize tassels counting using convolutional neural networks. Both highlight the domain shift issues faced when the models were applied in the wild, due to differences in phenological stages and growing conditions.

A general drawback of most of these studies is the lack of a benchmark dataset, independent of the training dataset, to evaluate the generalization ability of the models. This is crucial to understand the technological readiness of the developed methods under operational conditions for agricultural missions. However, the construction of such large, and diverse labelled datasets covering different field scenarios is quite expensive. This has encouraged organizations to come forward and publish open datasets such as CropDeep with around 20 species taken in greenhouse conditions (Zheng et al. 2019), Sorghum-Head dataset (Guo et al. 2018), maize tassels dataset (Lu et al. 2016), cauliflower and broccoli dataset (Bender, Whelan, and Sukkarieh 2020). More recently, a collaboration across several research institutes led to the constitution of the largest open wheat head dataset (David et al. 2020) of 4,700 RGB images acquired using hand-held poles and ground-based robots in the field. The publication of such collaborative open datasets over a range of crop species is crucial to promote deep learning for traits extraction within the field phenotyping community. Other approaches such as the use of transfer learning and creation of synthetic datasets are possible alternatives to consider tackling the challenge of constituting a diverse dataset. Table 3 lists several open-access labelled datasets for field phenotyping and precision agriculture applications.

*Table 3 A non-exhaustive list of open labelled datasets available for precision agriculture and field phenotyping. Columns: Number of instances indicates the number of occurrences of the objects within the images; Label type expresses if the labelling was done at the image-level (typically classification tasks), pixel-level (for segmentation tasks) or as a bounding box or circles.*

| Type | Dataset Name | Number of images | Number of instances | Number of classes | Label Type | | |
|---|---|---|---|---|---|---|---|
| | | | | | Image-level | Pixel-level | Bounding box/ Circle |
| Fruits and Vegetable identification | CropDeep (Zheng et al. 2019) | 31147 | 49765 | 31 | | | |
| | DeepFruits (Sa et al. 2016) | 587 | | 6 | | | |
| | Orchard Fruit (Bargoti and Underwood 2017) | 3704 | | 3 | | | |
| | Date Fruit (Altaheri et al. 2019) | 8079 | | 5 | | | |
| | KFuji RGB-D (Gené-Mola et al. 2019) | 967 | 12,839 | 1 | | | |
| | MangoNet (Kestur, Meduri, and Narasipura 2019) | 49 | | | | | |
| | MangoYOLO (Koirala et al. 2019) | 1730 | | 1 | | | |
| | WSU Apple Dataset (Bhusal, Karkee, and Zhang 2019) | 2298 | | 1 | | | |
| | Fuji-SFM (Gené-Mola et al. 2020) | 288 | | | | | |
| | MinneApple (Hani et al. 2020) | 981 | 41,325 | | | | |
| | | 70,865 | | | | | |
| | Capsicum Annum – Synthetic + Empirical images (Barth et al. 2018) | 10,500 + 750 | | 8 | | | |
| | Fruit Flower dataset (Dias, Tabb, and Medeiros 2018) | 190 | | 4 | | | |
| | Cauliflower/Brocolli dataset (Bender et al. 2020) | 1 248 | 2030 | 2 | | | |
| | Oil radish growth dataset (Mortensen et al. 2016) | 129 | | 6 | | | |
| Cereals | Sorghum Head Dataset (Guo et al. 2018) | 92 | 18893 | 1 | | | |
| | Maize Tassel Dataset (Lu et al. 2017) | 361 | 20000 | 1 | | | |
| | Global Wheat Head Dataset (David et al. 2020) | 4698 | 188445 | 1 | | | |
| Stress/ disease | Plant Village (Hughes and Salathe 2015) | 54 309 | - | 37 | | | |
| | Maize disease (Wiesner-Hanks et al. 2018) | 18,222 | 105,735 | 1 | | | |
| | Rice disease (Prajapati, Shah, and Dabhi 2017) | 120 | | 3 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Apple foliar disease (Thapa et al. 2020) | 3651 | | 4 | | | |
| Weed – Crop detection/segmentation | Open Plant Phenotype Database (Leminen Madsen et al. 2020) | 7590 | 315,038 | 47 | | | |
| | P2S2 (Madec, Irfan, et al. 2019) | 75 | | 2 | | | |
| | Sugar beet-weed dataset (Chebrolu et al. 2017) | 300 | | 10 | | | |
| | Carrot-weed dataset (Lameski et al. 2017) | 39 | | 3 | | | |
| | Grass-broadleaf (dos Santos Ferreira et al. 2017) | 15,336 | | 4 | | | |
| | DeepWeeds (Olsen et al. 2019) | 17,509 | | 8 | | | |
| | Joint Stem detection (Lottes, Behley, Chebrolu, et al. 2018) | 1321 | | 5 | | | |
| | Plant seedlings (Giselsson et al. 2017) | 407 | | 12 | | | |
| | Synthetic Sugarbeet-weed dataset (Di Cicco et al. 2016) | 8518 | | 3 | | | |
| | Early stage weed dataset (Espejo-Garcia et al. 2020) | 508 | | 4 | | | |
| | Weed growth stage dataset (Teimouri et al. 2018) | 9372 | | 18 | | | |

## 1.4 Open research questions in deep learning methods for plant phenotyping

The popularity of deep learning algorithms and their expected role in traits estimation in plant phenotyping was discussed in the previous section. Thanks to their impressive performance, the rapid adoption of these techniques for field plant phenotyping has progressed rapidly in the last five years. As mentioned above, the main challenge for the use of deep learning in operational conditions are linked with the lack of generalization where CNNs are applied over datasets that differ to some extent –i.e. the belong to a different domain– from the dataset used for training them. Compared to the identification of real-world objects, the implementation of deep learning in field phenotyping still has some specific issues that have not been fully addressed by the existing literature:

1. What is the impact of object size and image spatial resolution on the generalization ability of deep learning methods, and how to optimize it? A particular feature of phenotyping applications is that object size is often restricted to a small site- or vector-specific interval. While in general real-world detection applications the training data contains instances of the same object at different sizes, in phenotyping applications the object size is heavily determined by the vector used for data acquisition or the specific operation model (e.g., flight or camera height). Moreover, while attempting to maximize

the sampled area and throughput, the size of the objects observed is in general rather small compared to the desired object size (for instance, plant or organ counting). This constitutes an important challenge for deep neural networks (Tong, Wu, and Zhou 2020) that usually perform better on large objects (Kisantal et al. 2019). Consequently, CNNs trained on specific datasets may not generalize well when they are applied to completely independent ones where the object size differs from the latter. Different possibilities have been proposed in real-world applications to overcome this problem, such as the use of data augmentation strategies on the training set (Kisantal et al. 2019), multi-scale network architecture (Hu et al. 2018) and image super-resolution using generative adversarial networks (González et al. 2019; Magoulianitis et al. 2019). On phenotyping initiatives, like the Global Wheat Head Dataset (David et al. 2020), the dataset from different sites were harmonized to a pre-determined object size. The object size domain, however, remains an open question in phenotyping applications that have not been fully addressed in the current state of the art. A strategy to construct CNNs that generalize well across object sizes is still needed to understand how to integrate images from different vectors (e.g.; UAV and ground instruments, etc.) in the same pipeline.
Partially linked to the object size problem, understanding the potential impact of image resolution on the performance of object detection/counting algorithms is essential from the perspective of UAV observations. Due to the low autonomy of UAV, increasing flight height is an important factor to increase throughput and minimize the acquisition costs per sampled area. However, the number of textural features contained in the images decreases with resolution, perhaps impacting the ability of CNNs to correctly identify plants/organs. Currently, there are only few studies who have addressed specifically the link between image resolution/object size and CNN performance on plant phenotyping applications (Kitano et al. 2019; Madec, Jin, et al. 2019).

2. What kind of deep learning algorithm is better suited for density estimations - object counting or object detection methods? For applications such as plant/plant-organ density estimations, different methods have been proposed: object detection (Ghosal et al. 2019; Jin et al. 2019; Madec, Jin, et al. 2019); counting by regression (Dobrescu et al. 2017; Liu et al. 2020; Xiong et al. 2019) or instance segmentation (Kitano et al. 2019; Machefer et al. 2020). Among these, the preparation of annotations for instance segmentation task is extremely time consuming in comparison to the other approaches. This is the main reason why the object-based detection and counting by regression methods are, in general, more widely adopted. For wheat head counting from ground-level images, Madec et al. (2019) compared the performance of an object detection and counting by regression algorithm over a range of different object sizes. They concluded that object detection provided the best results if

object size is enough ( > 5000 pixels). Lu et al., (2017) indicated that for small object sizes counting by regression method were better suited. Therefore, counting by regression methods can be a priori better suited for density estimations from UAV images, but further work is required to understand whether data augmentation on the object size domain can help to overcome the limitations of object detection algorithms.

3. What is the contribution of temporal resolution and sampling area in the performance of deep learning models to derive phenological traits? The use of indirect methods to identify plant/organs based on deep learning techniques may introduce some uncertainties in the targeted phenotypic traits. For specific traits as those related with the estimation of phenological dates, requiring the identification of specific plant organs, the acquisition of frequent observations may help to solve partially this problem. The analysis, in relative terms, of the dynamics of the number of plant organs identified can help to mitigate the possible bias of object detection algorithms. However, vectors providing frequent images, like the fixed sensors, sample relatively smaller areas (Table 1), which may not be representative of the whole microplot or plot studied. What is the trade-off between the sampling area and temporal resolution for retrieving phenological traits? Only few studies have addressed this problem of estimating phenotyping traits (Wang et al. 2019; Yalcin 2018).

## 1.5 Objectives and organization of the study

The present thesis studied the use of deep learning techniques for the estimation of three essential traits for plant phenotyping: plant density at early stages for maize, wheat head density, and wheat heading date. The thesis is structured into three chapters that take the form of scientific papers, each one dealing with a specific phenotypic trait, and using a specific vector and detection/counting algorithm. Each chapter answers one or more of the research questions enumerated in Section 1.4:

- **Velumani, K.**, Lopez-Lozano, R., Madec, S., Guo, W., Gillet, J., Comar, A., Baret, F., 2021. Estimates of maize plant density from UAV RGB images using Faster-RCNN detection model: impact of the spatial resolution. **Submitted (under review).**
    - **Maize plant counting** at early stage is an important trait required for crop breeding experiments and farm management. It is now possible to access this plant density trait at high throughput owing to the rapid popularity and affordability of UAV systems. In this study, Faster-RCNN, an object detection method, is used for the estimation of plant density from RGB imageries acquired from UAVs. We further analyze the performance of the detection method for small object detection on low resolution imageries, linked to research question #1. The study shows how data augmentation, and the use of super resolution approaches enables us to access plant density from low-resolution imageries.

- **Velumani, K.**, Lopez-Lozano, R., Lu, H., Madec, S., David, E., Li, W., Liu, S., Smith, D., Chapman, S., Comar, A., Baret, F., 2021. Wheat head density estimation from UAVs: data augmentation and data preparation strategies to exploit labelled ground-based imagery. **Draft.**
    - **Wheat head density** is a crucial trait for plant breeders and precision agriculture as it is an important component used in yield estimation. We propose an automatic density estimation method from UAV which promises high-throughput and overcomes the sampling issues faced by the current state-of-the-art methods that use ground imageries. In this study, we use an existing high-resolution annotated dataset acquired at the ground-level to train deep learning to be applied to low-resolution UAV imageries. The study also analyses the suitability of two approaches of deep learning - object detection and regression-based object counting methods to achieve wheat head counting on UAV images Further, a comparison between the density estimated from the UAV images and the in-situ manual density, along with their broad-sense heritability, is presented. **This paper addresses the research questions #1 and #2**

- **Velumani, K.,** Madec, S., de Solan, B., Lopez-Lozano, R., Gillet, J., Labrosse, J., Jezequel, S., Comar, A., Baret, F., 2020. An automatic method based on daily in situ images and deep learning to date wheat heading stage. F. Crop. Res. 252, 107793. https://doi.org/10.1016/j.fcr.2020.107793
    - **Wheat heading date** from daily RGB imageries: The development of wireless IOT (internet of thing) fixed sensors now allows for continuous monitoring of crops in the field using RGB cameras. In this chapter, we propose an automatic method using CNNs to estimate wheat heading date through daily time series images taken throughout the growing cycle. We compare the performance of the method developed, relying on frequent observations on a relatively low sampling area, with reference heading date by manual scoring in the field and directly on the images. The method proves to be robust –despite the relatively small area observed– over an independent test dataset acquired during a different growing cycle and over cultivars not used during the model training. **This paper addresses the research question #3.**

# References

Alchanatis, V., Y. Cohen, S. Cohen, M. Moller, M. Sprinstin, M. Meron, J. Tsipris, Y. Saranga, and E. Sela. 2010. "Evaluation of Different Approaches for Estimating and Mapping Crop Water Status in Cotton with Thermal Imaging." *Precision Agriculture* 11(1):27–41. doi: 10.1007/s11119-009-9111-7.

Alom, Md Zahangir, Tarek M. Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C. Van Essen, Abdul A. S. Awwal, and Vijayan K. Asari. 2019. "A State-of-the-Art Survey on Deep Learning Theory and Architectures." *Electronics* 8(3):292. doi:

10.3390/electronics8030292.

Altaheri, Hamdi, Mansour Alsulaiman, Ghulam Muhammad, Syed Umar Amin, Mohamed Bencherif, and Mohamed Mekhtiche. 2019. "Date Fruit Dataset for Intelligent Harvesting." *Data in Brief* 26:104514. doi: 10.1016/j.dib.2019.104514.

Araus, José L., and J. E. Cairns. 2014. "Field High-Throughput Phenotyping: The New Crop Breeding Frontier." *Trends in Plant Science* 19(1):52–61. doi: 10.1016/j.tplants.2013.09.008.

Araus, José L., and Shawn C. Kefauver. 2018. "Breeding to Adapt Agriculture to Climate Change: Affordable Phenotyping Solutions." *Current Opinion in Plant Biology* 45:237–47. doi: 10.1016/j.pbi.2018.05.003.

Aydın, Doğan, and Aybars Uğur. 2011. "Extraction of Flower Regions in Color Images Using Ant Colony Optimization." *Procedia Computer Science* 3:530–36. doi: 10.1016/j.procs.2010.12.088.

Bargoti, Suchet, and James Underwood. 2017. "Deep Fruit Detection in Orchards." Pp. 3626–33 in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.

Barth, R., J. IJsselmuiden, J. Hemming, and E. J. Van Henten. 2018. "Data Synthesis Methods for Semantic Segmentation in Agriculture: A Capsicum Annuum Dataset." *Computers and Electronics in Agriculture* 144:284–96. doi: 10.1016/j.compag.2017.12.001.

Basso, Bruno, and John Antle. 2020. "Digital Agriculture to Design Sustainable Agricultural Systems." *Nature Sustainability* 3(4):254–56. doi: 10.1038/s41893-020-0510-0.

Bender, Asher, Brett Whelan, and Salah Sukkarieh. 2020. "A High-resolution, Multimodal Data Set for Agricultural Robotics: A Ladybird 's-eye View of Brassica." *Journal of Field Robotics* 37(1):73–96. doi: 10.1002/rob.21877.

Bhusal, Santosh, Manoj Karkee, and Qin Zhang. 2019. "Apple Dataset Benchmark from Orchard Environment in Modern Fruiting Wall." Retrieved February 10, 2021 (https://research.libraries.wsu.edu/xmlui/handle/2376/17721?show=full).

Blancon, Justin, Dan Dutartre, Marie-Hélène Tixier, Marie Weiss, Alexis Comar, Sébastien Praud, and Frédéric Baret. 2019. "A High-Throughput Model-Assisted Method for Phenotyping Maize Green Leaf Area Index Dynamics Using Unmanned Aerial Vehicle Imagery." *Frontiers in Plant Science* 10:685. doi: 10.3389/fpls.2019.00685.

Boureau, Y. Lan, Jean Ponce, and Yann Lecun. 2010. "A Theoretical Analysis of Feature Pooling in Visual Recognition." Pp. 111–18 in *Proceedings of the 27th international conference on machine learning (ICML-10)*.

Chebrolu, Nived, Philipp Lottes, Alexander Schaefer, Wera Winterhalter, Wolfram Burgard, and Cyrill Stachniss. 2017. "Agricultural Robot Dataset for Plant Classification, Localization and Mapping on Sugar Beet Fields." *International Journal of Robotics Research* 36(10):1045–52. doi: 10.1177/0278364917720510.

Di Cicco, Maurilio, Ciro Potena, Giorgio Grisetti, and Alberto Pretto. 2016. "Automatic Model Based Dataset Generation for Fast and Accurate Crop and Weeds Detection." *IEEE International Conference on Intelligent Robots and Systems* 2017-September:5188–95. doi: 10.1109/IROS.2017.8206408.

Daughtry, C. 2000. "Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance." *Remote Sensing of Environment* 74(2):229–39. doi: 10.1016/S0034-4257(00)00113-9.

David, Etienne, Simon Madec, Pouria Sadeghi-Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchgessner, Goro Ishikawa, Koichi Nagasawa, Minhajul A. Badhon, Curtis Pozniak, Benoit de Solan, Andreas Hund, Scott C. Chapman, Frédéric Baret, Ian Stavness, and Wei Guo. 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* 2020:1–12. doi: 10.34133/2020/3521852.

Deery, David, Jose Jimenez-Berni, Hamlyn Jones, Xavier Sirault, and Robert Furbank. 2014. "Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping." *Agronomy* 4(3):349–79. doi: 10.3390/agronomy4030349.

Deery, David M., Greg J. Rebetzke, Jose A. Jimenez-Berni, Richard A. James, Anthony G. Condon, William D. Bovill, Paul Hutchinson, Jamie Scarrow, Robert Davy, and Robert T. Furbank. 2016. "Methodology for High-Throughput Field Phenotyping of Canopy Temperature Using Airborne Thermography." *Frontiers in Plant Science* 7(DECEMBER2016):1808. doi: 10.3389/fpls.2016.01808.

Dias, Philipe A., Amy Tabb, and Henry Medeiros. 2018. "Multispecies Fruit Flower Detection Using a

Refined Semantic Segmentation Network." (1). doi: 10.1109/LRA.2018.2849498.

Dobrescu, Andrei, Mario Valerio Giuffrida, and Sotirios A. Tsaftaris. 2017. "Leveraging Multiple Datasets for Deep Leaf Counting." Pp. 2072–79 in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Espejo-Garcia, Borja, Nikos Mylonas, Loukas Athanasakos, Spyros Fountas, and Ioannis Vasilakoglou. 2020. "Towards Weeds Identification Assistance through Transfer Learning." *Computers and Electronics in Agriculture* 171:105306. doi: 10.1016/j.compag.2020.105306.

FAO. 2002. *World Agriculture: Towards 2015/2030: Summary Report*. edited by J. Bruinsma. Rome: Food and Agriculture Organization of the United Nations (FAO).

FAO. 2017. *The Future of Food and Agriculture - Trends and Challenges*. Rome.

FAO. 2020. *World Food and Agriculture - Statistical Yearbook 2020*. Rome: FAO.

FAO, IFAD, UNICEF, WFP, and WHO. 2020. *The State of Food Security and Nutrition in the World 2020*. Rome: FAO, IFAD, UNICEF, WFP and WHO.

Fawakherji, Mulham, Ali Youssef, Domenico Bloisi, Alberto Pretto, and Daniele Nardi. 2019. "Crop and Weeds Classification for Precision Agriculture Using Context-Independent Pixel-Wise Segmentation." Pp. 146–52 in *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE.

Fiorani, Fabio, and Ulrich Schurr. 2013. "Future Scenarios for Plant Phenotyping." *Annual Review of Plant Biology* 64(1):267–91. doi: 10.1146/annurev-arplant-050312-120137.

Gebbers, Robin, and Viacheslav I. Adamchuk. 2010. "Precision Agriculture and Food Security." *Science* 327(5967):828–31. doi: 10.1126/science.1183899.

Gené-Mola, J., Verónica Vilaplana, Joan R. Rosell-Polo, Josep Ramon Morros, Javier Ruiz-Hidalgo, and Eduard Gregorio. 2019. "KFuji RGB-DS Database: Fuji Apple Multi-Modal Images for Fruit Detection with Color, Depth and Range-Corrected IR Data." *Data in Brief* 25:104289. doi: 10.1016/j.dib.2019.104289.

Gené-Mola, Jordi, Ricardo Sanz-Cortiella, Joan R. Rosell-Polo, Josep Ramon Morros, Javier Ruiz-Hidalgo, Verónica Vilaplana, and Eduard Gregorio. 2020. "Fruit Detection and 3D Location Using Instance Segmentation Neural Networks and Structure-from-Motion Photogrammetry." *Computers and Electronics in Agriculture* 169:105165. doi: 10.1016/j.compag.2019.105165.

Ghosal, Sambuddha, Bangyou Zheng, Scott C. Chapman, Andries B. Potgieter, David R. Jordan, Xuemin Wang, Asheesh K. Singh, Arti Singh, Masayuki Hirafuji, Seishi Ninomiya, Baskar Ganapathysubramanian, Soumik Sarkar, and Wei Guo. 2019. "A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting." *Plant Phenomics* 2019:1525874. doi: 10.34133/2019/1525874.

Giselsson, Thomas Mosgaard, Rasmus Nyholm Jørgensen, Peter Kryger Jensen, Mads Dyrmann, and Henrik Skov Midtiby. 2017. "A Public Image Database for Benchmark of Plant Seedling Classification Algorithms." *ArXiv Preprint ArXiv:1711.05458*.

Gómez-Candón, David, Nicolas Virlet, Sylvain Labbé, Audrey Jolivot, and Jean Luc Regnard. 2016. "Field Phenotyping of Water Stress at Tree Scale by UAV-Sensed Imagery: New Insights for Thermal Acquisition and Calibration." *Precision Agriculture* 17(6):786–800. doi: 10.1007/s11119-016-9449-6.

González, Daniel, Miguel A. Patricio, Antonio Berlanga, and José M. Molina. 2019. "A Super-Resolution Enhancement of UAV Images Based on a Convolutional Neural Network for Mobile Devices." *Personal and Ubiquitous Computing*. doi: 10.1007/s00779-019-01355-5.

Guo, Wei, Bangyou Zheng, Andries B. Potgieter, Julien Diot, Kakeru Watanabe, Koji Noshita, David R. Jordan, Xuemin Wang, James Watson, Seishi Ninomiya, and Scott C. Chapman. 2018. "Aerial Imagery Analysis – Quantifying Appearance and Number of Sorghum Heads for Applications in Breeding and Agronomy." *Frontiers in Plant Science* 9:1544. doi: 10.3389/fpls.2018.01544.

Guo, Yanming, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. 2016. "Deep Learning for Visual Understanding: A Review." *Neurocomputing* 187:27–48. doi: 10.1016/j.neucom.2015.09.116.

Hani, Nicolai, Pravakar Roy, and Volkan Isler. 2020. "MinneApple: A Benchmark Dataset for Apple Detection and Segmentation." *IEEE Robotics and Automation Letters* 5(2):852–58. doi: 10.1109/LRA.2020.2965061.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Spatial Pyramid Pooling in Deep

Convolutional Networks for Visual Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(9):1904–16. doi: 10.1109/TPAMI.2015.2389824.

Hu, Guo X., Zhong Yang, Lei Hu, Li Huang, and Jia M. Han. 2018. "Small Object Detection with Multiscale Features." *International Journal of Digital Multimedia Broadcasting* 2018:1–10. doi: 10.1155/2018/4546896.

Hughes, David. P., and Marcel Salathe. 2015. "An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics." *ArXiv Preprint ArXiv:1511.08060*.

Hunt, E. Raymond, W. Dean Hively, Stephen Fujikawa, David Linden, Craig S. Daughtry, and Greg McCarty. 2010. "Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring." *Remote Sensing* 2(1):290–305. doi: 10.3390/rs2010290.

Jay, Sylvain, Frédéric Baret, Dan Dutartre, Ghislain Malatesta, Stéphanie Héno, Alexis Comar, Marie Weiss, and Fabienne Maupas. 2019. "Exploiting the Centimeter Resolution of UAV Multispectral Imagery to Improve Remote-Sensing Estimates of Canopy Structure and Biochemistry in Sugar Beet Crops." *Remote Sensing of Environment* 231:110898. doi: 10.1016/j.rse.2018.09.011.

Jin, Xiuliang, Shouyang Liu, Frédéric Baret, Matthieu Hemerlé, and Alexis Comar. 2017. "Estimates of Plant Density of Wheat Crops at Emergence from Very Low Altitude UAV Imagery." *Remote Sensing of Environment* 198:105–14. doi: 10.1016/j.rse.2017.06.007.

Jin, Xiuliang, Simon Madec, Dan Dutartre, Benoit de Solan, Alexis Comar, and Frédéric Baret. 2019. "High-Throughput Measurements of Stem Characteristics to Estimate Ear Density and Above-Ground Biomass." *Plant Phenomics* 2019:1–10. doi: 10.34133/2019/4820305.

Jones, Hamlyn G., Rachid Serraj, Brian R. Loveys, Lizhong Xiong, Ashley Wheaton, and Adam H. Price. 2009. "Thermal Infrared Imaging of Crop Canopies for the Remote Diagnosis and Quantification of Plant Responses to Water Stress in the Field." doi: 10.1071/FP09123.

Kaur, Amanjot, Irwin R. Donis-Gonzalez, and Dina A. St. Clair. 2020. "Evaluation of a Hand-held Spectrophotometer as an In-field Phenotyping Tool for Tomato and Pepper Fruit Quality." *The Plant Phenome Journal* 3(1):e20008. doi: 10.1002/ppj2.20008.

Kelly, Derek, Avimanyou Vatsa, Wade Mayham, Linh Ngô, Addie Thompson, and Toni Kazic. 2016. "An Opinion on Imaging Challenges in Phenotyping Field Crops." *Machine Vision and Applications* 27(5):681–94. doi: 10.1007/s00138-015-0728-4.

Kestur, Ramesh, Avadesh Meduri, and Omkar Narasipura. 2019. "MangoNet: A Deep Semantic Segmentation Architecture for a Method to Detect and Count Mangoes in an Open Orchard." *Engineering Applications of Artificial Intelligence* 77:59–69. doi: 10.1016/j.engappai.2018.09.011.

Khan, Asifullah, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. 2020. "A Survey of the Recent Architectures of Deep Convolutional Neural Networks." *Artificial Intelligence Review* 53(8):5455–5516. doi: 10.1007/s10462-020-09825-6.

Khanna, Raghav, J. Rehder, M. Moeller, E. Galceran, and R. Siegwart. 2015. "Studying Phenotypic Variability in Crops Using a Hand-Held Sensor Platform." *IROS Work, Agri-Food Robot*.

Kisantal, Mate, Zbigniew Wojna, Jakub Murawski, Jacek Naruniec, and Kyunghyun Cho. 2019. "Augmentation for Small Object Detection." *ArXiv Preprint ArXiv:1902.07296*.

Kitano, Bruno T., Caio C. T. Mendes, Andre R. Geus, Henrique C. Oliveira, and Jefferson R. Souza. 2019. "Corn Plant Counting Using Deep Learning and UAV Images." *IEEE Geoscience and Remote Sensing Letters* 1–5. doi: 10.1109/LGRS.2019.2930549.

Koirala, A., K. B. Walsh, Z. Wang, and C. McCarthy. 2019. "Deep Learning for Real-Time Fruit Detection and Orchard Fruit Load Estimation: Benchmarking of 'MangoYOLO.'" *Precision Agriculture* 20(6):1107–35. doi: 10.1007/s11119-019-09642-0.

Krizhevsky, A., I. Sutskever, and GE Hinton. 2012. *Imagenet Classification with Deep Convolutional Neural Networks*.

Lal, Rattan. 2016. "Feeding 11 Billion on 0.5 Billion Hectare of Area under Cereal Crops." *Food and Energy Security* 5(4):239–51. doi: 10.1002/fes3.99.

Laliberte, Andrea S., Mark A. Goforth, Caitriana M. Steele, and Albert Rango. 2011. "Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments." *Remote Sensing* 3(11):2529–51. doi: 10.3390/rs3112529.

Lameski, Petre, Eftim Zdravevski, Vladimir Trajkovik, and Andrea Kulakov. 2017. "Weed Detection Dataset with RGB Images Taken Under Variable Light Conditions." Pp. 112–19 in *Communications in Computer and Information Science*. Vol. 778. Springer Verlag.

LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. "Backpropagation Applied to Handwritten Zip Code Recognition." *Neural Computation* 1(4):541–51. doi: 10.1162/neco.1989.1.4.541.

Leminen Madsen, Simon, Solvejg Kopp Mathiassen, Mads Dyrmann, Morten Stigaard Laursen, Laura-Carlota Paz, and Rasmus Nyholm Jørgensen. 2020. "Open Plant Phenotype Database of Common Weeds in Denmark." *Remote Sensing* 12(8):1246. doi: 10.3390/rs12081246.

Lernoud, Julia, and Helga Willer. 2019. "Current Statistics on Organic Agriculture Worldwide: Area, Operators, and Market." Pp. 35–125 in *The World of Organic Agriculture. Statistics and Emerging Trends 2019*, edited by H. Willer and J. Lernoud. Frick, Suisse: Research Institute of Organic Agriculture FiBL and IFOAM Organics International.

Li, Fei, Bodo Mistele, Yuncai Hu, Xinping Chen, and Urs Schmidhalter. 2014. "Reflectance Estimation of Canopy Nitrogen Content in Winter Wheat Using Optimised Hyperspectral Spectral Indices and Partial Least Squares Regression." *European Journal of Agronomy* 52:198–209. doi: 10.1016/j.eja.2013.09.006.

Li, Lei, Qin Zhang, and Danfeng Huang. 2014. "A Review of Imaging Techniques for Plant Phenotyping." *Sensors (Switzerland)* 14(11):20078–111. doi: 10.3390/s141120078.

Li, Zhenbo, Ruohao Guo, Meng Li, Yaru Chen, and Guangyao Li. 2020. "A Review of Computer Vision Technologies for Plant Phenotyping." *Computers and Electronics in Agriculture* 176(July):105672. doi: 10.1016/j.compag.2020.105672.

Lin, Tsung Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. "Microsoft COCO: Common Objects in Context." Pp. 740–55 in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8693 LNCS. Springer Verlag.

Lin, Yi-Chun, and Ayman Habib. 2021. "Quality Control and Crop Characterization Framework for Multi-Temporal UAV LiDAR Data over Mechanized Agricultural Fields." *Remote Sensing of Environment* 256:112299. doi: 10.1016/j.rse.2021.112299.

Lin, Yi. 2015. "LiDAR: An Important Tool for next-Generation Phenotyping Technology of High Potential for Plant Phenomics?" *Computers and Electronics in Agriculture* 119:61–73. doi: 10.1016/j.compag.2015.10.011.

Liu, Liang, Hao Lu, Yanan Li, and Zhiguo Cao. 2020. "High-Throughput Rice Density Estimation from Transplantation to Tillering Stages Using Deep Networks." *Plant Phenomics* 2020:1–14. doi: 10.34133/2020/1375957.

Lottes, Philipp, Jens Behley, Nived Chebrolu, Andres Milioto, and Cyrill Stachniss. 2018. "Joint Stem Detection and Crop-Weed Classification for Plant-Specific Treatment in Precision Farming." Pp. 8233–38 in *IEEE International Conference on Intelligent Robots and Systems*. Institute of Electrical and Electronics Engineers Inc.

Lottes, Philipp, Jens Behley, Andres Milioto, and Cyrill Stachniss. 2018. "Fully Convolutional Networks With Sequential Information for Robust Crop and Weed Detection in Precision Farming." *IEEE Robotics and Automation Letters* 3(4):2870–77. doi: 10.1109/LRA.2018.2846289.

Lu, Hao, and Zhiguo Cao. 2020. "TasselNetV2+: A Fast Implementation for High-Throughput Plant Counting From High-Resolution RGB Imagery." *Frontiers in Plant Science* 11:541960. doi: 10.3389/fpls.2020.541960.

Lu, Hao, Zhiguo Cao, Yang Xiao, Yanan Li, and Yanjun Zhu. 2016. "Joint Crop and Tassel Segmentation in the Wild." Pp. 474–79 in *Proceedings - 2015 Chinese Automation Congress, CAC 2015*. Institute of Electrical and Electronics Engineers Inc.

Lu, Hao, Zhiguo Cao, Yang Xiao, Bohan Zhuang, and Chunhua Shen. 2017. "TasselNet: Counting Maize Tassels in the Wild via Local Counts Regression Network." *Plant Methods* 13(1):79. doi: 10.1186/s13007-017-0224-0.

Ludovisi, Riccardo, Flavia Tauro, Riccardo Salvati, Sacha Khoury, Giuseppe Mugnozza Scarascia, and Antoine Harfouche. 2017. "UAV-Based Thermal Imaging for High-Throughput Field Phenotyping of Black Poplar Response to Drought." *Frontiers in Plant Science* 8:1681. doi: 10.3389/fpls.2017.01681.

Lumme, J., M. Karjalainen, H. Kaartinen, A. Kukko, J. Hyyppä, H. Hyyppä, A. Jaakkola, and J. Kleemola. 2008. "Terrestrial Laser Scanning of Agricultural Crops." Pp. 563–566 in *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sciences*. Vol. 37.

Machefer, Mélissande, François Lemarchand, Virginie Bonnefond, Alasdair Hitchins, and Panagiotis

Sidiropoulos. 2020. "Mask R-CNN Refitting Strategy for Plant Counting and Sizing in Uav Imagery." *Remote Sensing* 12(18):3015. doi: 10.3390/RS12183015.

Madec, Simon. 2019. "Phenotyping Wheat Structural Traits from Millimetric Resolution RGB Imagery in Field Conditions." Université d'Avignon.

Madec, Simon, Fred Baret, Benoît de Solan, Samuel Thomas, Dan Dutartre, Stéphane Jezequel, Matthieu Hemmerlé, Gallian Colombeau, and Alexis Comar. 2017. "High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates." *Frontiers in Plant Science* 8(November):1–14. doi: 10.3389/fpls.2017.02002.

Madec, Simon, Kamran Irfan, Etienne David, Kaaviya Velumani, Gaetan Daubige, Jeremy Labrosse, Wei Guo, Marie Weiss, Frederic Baret, and Kamran Ifran. 2019. "The P2S2 Segmentation Dataset: Annotated in-Field Multi-Crop RGB Images Acquired under Various Conditions." in *7th International Workshop on Image Analysis Methods in the Plant Sciences (IAMPS)*. Lyon.

Madec, Simon, Xiuliang Jin, Hao Lu, Benoit De Solan, Shouyang Liu, Florent Duyme, Emmanuelle Heritier, and Frédéric Baret. 2019. "Ear Density Estimation from High Resolution RGB Imagery Using Deep Learning Technique." *Agricultural and Forest Meteorology* 264(May 2018):225–34. doi: 10.1016/j.agrformet.2018.10.013.

Maggio, Albino, Tine Van Criekinge, and Jean-Paul Malingreau. 2016. "Global Food Security: Assessing Trends in View of Guiding Future EU Policies." *Foresight* 18(5):551–60. doi: 10.1108/FS-07-2015-0040.

Magoulianitis, Vasileios, Dimitrios Ataloglou, Anastasios Dimou, Dimitrios Zarpalas, and Petros Daras. 2019. "Does Deep Super-Resolution Enhance UAV Detection?" *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019* 1–6. doi: 10.1109/AVSS.2019.8909865.

Mahlein, A. K. 2010. "Detection, Identification, and Quantification of Fungal Diseases of Sugar Beet Leaves Using Imaging and Non-Imaging Hyperspectral Techniques." Rheinischen Friedrich-Wilhelms-Universität, Bonn (Germany).

Mahlein, Anne-Katrin, Matheus Thomas Kuska, Stefan Thomas, Mirwaes Wahabzada, Jan Behmann, Uwe Rascher, and Kristian Kersting. 2019. "Quantitative and Qualitative Phenotyping of Disease Resistance of Crops by Hyperspectral Sensors: Seamless Interlocking of Phytopathology, Sensors, and Machine Learning Is Needed!" *Current Opinion in Plant Biology* 50:156–62. doi: 10.1016/j.pbi.2019.06.007.

Mankins, John C. 1995. "TECHNOLOGY READINESS LEVELS." *A White Paper*.

McCool, Chris, Tristan Perez, and Ben Upcroft. 2017. "Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics." *IEEE Robotics and Automation Letters* 2(3):1344–51. doi: 10.1109/LRA.2017.2667039.

McKevith, Brigid. 2004. "Nutritional Aspects of Cereals." *Nutrition Bulletin* 29(2):111–42. doi: 10.1111/j.1467-3010.2004.00418.x.

Minervini, Massimo, Andreas Fischbach, Hanno Scharr, and Sotirios A. Tsaftaris. 2016. "Finely-Grained Annotated Datasets for Image-Based Plant Phenotyping." *Pattern Recognition Letters* 81:80–89. doi: 10.1016/j.patrec.2015.10.013.

Minervini, Massimo, Hanno Scharr, and Sotirios Tsaftaris. 2015. "Image Analysis: The New Bottleneck in Plant Phenotyping [Applications Corner]." *IEEE Signal Processing Magazine* 32(4):126–31. doi: 10.1109/MSP.2015.2405111.

Mortensen, A. K., M. Dyrmann, H. Karstoft, R. N. Jørgensen, and R. Gislum. 2016. "Semantic Segmentation of Mixed Crops Using Deep Convolutional Neural Network." Pp. 1–6 in *CIGR-AgEng Conference, 26-29 June 2016, Aarhus, Denmark. Abstracts and Full papers*. Organising Committee, CIGR 2016.

Muller, Adrian, Christian Schader, Nadia El-Hage Scialabba, Judith Brüggemann, Anne Isensee, Karl-Heinz Erb, Pete Smith, Peter Klocke, Florian Leiber, Matthias Stolze, and Urs Niggli. 2017. "Strategies for Feeding the World More Sustainably with Organic Agriculture." *Nature Communications* 8(1):1290. doi: 10.1038/s41467-017-01410-w.

Nagasubramanian, Koushik, Sarah Jones, Asheesh K. Singh, Soumik Sarkar, Arti Singh, and Baskar Ganapathysubramanian. 2019. "Plant Disease Identification Using Explainable 3D Deep Learning on Hyperspectral Images." *Plant Methods* 15(1):98. doi: 10.1186/s13007-019-0479-8.

Niu, Xiaojing, Meili Wang, Xianqiang Chen, Shihui Guo, Hongming Zhang, and Dongjian He. 2014. "Image Segmentation Algorithm for Disease Detection of Wheat Leaves." Pp. 270–73 in

*Proceedings of the 2014 International Conference on Advanced Mechatronic Systems*. IEEE.

OECD, and Food and Agriculture Organization of the United Nations. 2020. *OECD-FAO Agricultural Outlook 2020-2029*. FAO/Paris, Rome: OECD Publishing.

Oh, Min-hwan, Peder Olsen, and Karthikeyan Natesan Ramamurthy. 2019. "Counting and Segmenting Sorghum Heads." *ArXiv*.

Olsen, Alex, Dmitry A. Konovalov, Bronson Philippa, Peter Ridd, Jake C. Wood, Jamie Johns, Wesley Banks, Benjamin Girgenti, Owen Kenny, James Whinney, Brendan Calvert, Mostafa Rahimi Azghadi, and Ronald D. White. 2019. "DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning." *Scientific Reports* 9(1):2058. doi: 10.1038/s41598-018-38343-3.

Pan, Shen, Mineichi Kudo, and Jun Toyama. 2009. "Edge Detection of Tobacco Leaf Images Based on Fuzzy Mathematical Morphology." Pp. 1219–22 in *2009 First International Conference on Information Science and Engineering*. IEEE.

Paproki, Anthony, Jurgen Fripp, Olivier Salvado, Xavier Sirault, Scott Berry, and Robert Furbank. 2011. "Automated 3D Segmentation and Analysis of Cotton Plants." Pp. 555–60 in *2011 International Conference on Digital Image Computing: Techniques and Applications*. Noosa, QLD: IEEE.

Parent, Boris, Margot Leclere, Sébastien Lacube, Mikhail A. Semenov, Claude Welcker, Pierre Martre, and François Tardieu. 2018. "Maize Yields over Europe May Increase in Spite of Climate Change, with an Appropriate Use of the Genetic Variability of Flowering Time." *Proceedings of the National Academy of Sciences of the United States of America* 115(42):10642–47. doi: 10.1073/pnas.1720716115.

Paulus, Stefan, Jan Behmann, Anne-Katrin Mahlein, Lutz Plümer, and Heiner Kuhlmann. 2014. "Low-Cost 3D Systems: Suitable Tools for Plant Phenotyping." *Sensors (Basel, Switzerland)* 14(2):3001–18. doi: 10.3390/s140203001.

Prajapati, Harshadkumar B., Jitesh P. Shah, and Vipul K. Dabhi. 2017. "Detection and Classification of Rice Plant Diseases." *Intelligent Decision Technologies* 11(3):357–73. doi: 10.3233/IDT-170301.

Praveen Kumar, J., and S. Domnic. 2020. "Rosette Plant Segmentation with Leaf Count Using Orthogonal Transform and Deep Convolutional Neural Network." *Machine Vision and Applications* 31(1–2):6. doi: 10.1007/s00138-019-01056-2.

Quaglia, Giuseppe, Carmen Visconte, Leonardo Sabatino Scimmi, Matteo Melchiorre, Paride Cavallone, and Stefano Pastorelli. 2020. "Design of a UGV Powered by Solar Energy for Precision Agriculture." *Robotics* 9(1):13. doi: 10.3390/robotics9010013.

Reynolds, Daniel, Joshua Ball, Alan Bauer, Robert Davey, Simon Griffiths, and Ji Zhou. 2019. "CropSight: A Scalable and Open-Source Information Management System for Distributed Plant Phenotyping and IoT-Based Crop Management." *GigaScience* 8(3):1–11. doi: 10.1093/gigascience/giz009.

Reynolds, Daniel, Frederic Baret, Claude Welcker, Aaron Bostrom, Joshua Ball, Francesco Cellini, Argelia Lorence, Aakash Chawade, Mehdi Khafif, Koji Noshita, Mark Mueller-Linow, Ji Zhou, and François Tardieu. 2019. "What Is Cost-Efficient Phenotyping? Optimizing Costs for Different Scenarios." *Plant Science* 282(August):14–22. doi: 10.1016/j.plantsci.2018.06.015.

Roitsch, Thomas, Llorenç Cabrera-Bosquet, Antoine Fournier, Kioumars Ghamkhar, José Jiménez-Berni, Francisco Pinto, and Eric S. Ober. 2019. "Review: New Sensors and Data-Driven Approaches—A Path to next Generation Phenomics." *Plant Science* 282:2–10. doi: 10.1016/j.plantsci.2019.01.011.

Rosenzweig, Cynthia, Joshua Elliott, Delphine Deryng, Alex C. Ruane, Christoph Müller, Almut Arneth, Kenneth J. Boote, Christian Folberth, Michael Glotter, Nikolay Khabarov, Kathleen Neumann, Franziska Piontek, Thomas A. M. Pugh, Erwin Schmid, Elke Stehfest, Hong Yang, and James W. Jones. 2014. "Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison." *Proceedings of the National Academy of Sciences of the United States of America* 111(9):3268–73. doi: 10.1073/pnas.1222463110.

Rosenzweig, Cynthia, and Martin L. Parry. 1994. "Potential Impact of Climate Change on World Food Supply." *Nature* 367(6459):133–38. doi: 10.1038/367133a0.

Ruckelshausen, A., P. Biber, M. Dorna, H. Gremmes, R. Klose, A. Linz, R. Rahe, R. Resch, M. Thiel, D. Trautz, and U. Weiss. 2009. "BoniRob–an Autonomous Field Robot Platform for Individual Plant Phenotyping." *Precision Agriculture* 9(841):1.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. "ImageNet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision* 115(3):211–52. doi: 10.1007/s11263-015-0816-y.

Sa, Inkyu, Zetao Chen, Marija Popovic, Raghav Khanna, Frank Liebisch, Juan Nieto, and Roland Siegwart. 2018. "WeedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming." *IEEE Robotics and Automation Letters* 3(1):588–95. doi: 10.1109/LRA.2017.2774979.

Sa, Inkyu, Zongyuan Ge, Feras Dayoub, Ben Upcroft, Tristan Perez, and Chris McCool. 2016. "DeepFruits: A Fruit Detection System Using Deep Neural Networks." *Sensors* 16(8):1222. doi: 10.3390/s16081222.

dos Santos Ferreira, Alessandro, Daniel Matte Freitas, Gercina Gonçalves da Silva, Hemerson Pistori, and Marcelo Theophilo Folhes. 2017. "Weed Detection in Soybean Crops Using ConvNets." *Computers and Electronics in Agriculture* 143:314–24. doi: 10.1016/j.compag.2017.10.027.

Shendryk, Yuri, Jeremy Sofonia, Robert Garrard, Yannik Rist, Danielle Skocaj, and Peter Thorburn. 2020. "Fine-Scale Prediction of Biomass and Leaf Nitrogen Content in Sugarcane Using UAV LiDAR and Multispectral Imaging." *International Journal of Applied Earth Observation and Geoinformation* 92:102177. doi: 10.1016/j.jag.2020.102177.

Tardieu, François, Llorenç Cabrera-Bosquet, Tony Pridmore, and Malcolm Bennett. 2017. "Plant Phenomics, From Sensors to Knowledge." *Current Biology* 27(15):R770–83. doi: 10.1016/j.cub.2017.05.055.

Teimouri, Nima, Mads Dyrmann, Per Nielsen, Solvejg Mathiassen, Gayle Somerville, and Rasmus Jørgensen. 2018. "Weed Growth Stage Estimator Using Deep Convolutional Neural Networks." *Sensors* 18(5):1580. doi: 10.3390/s18051580.

Thapa, Ranjita, Kai Zhang, Noah Snavely, Serge Belongie, and Awais Khan. 2020. "The Plant Pathology Challenge 2020 Data Set to Classify Foliar Disease of Apples." *Applications in Plant Sciences* 8(9):e11390. doi: 10.1002/aps3.11390.

Tong, Kang, Yiquan Wu, and Fei Zhou. 2020. "Recent Advances in Small Object Detection Based on Deep Learning: A Review." *Image and Vision Computing* 97:103910. doi: 10.1016/j.imavis.2020.103910.

Ubbens, Jordan, Mikolaj Cieslak, Przemyslaw Prusinkiewicz, and Ian Stavness. 2018. "The Use of Plant Models in Deep Learning: An Application to Leaf Counting in Rosette Plants." *Plant Methods* 14(1). doi: 10.1186/s13007-018-0273-z.

Velumani, Kaaviya, Simon Madec, Benoit de Solan, Raul Lopez-Lozano, Jocelyn Gillet, Jeremy Labrosse, Stephane Jezequel, Alexis Comar, and Frédéric Baret. 2020. "An Automatic Method Based on Daily in Situ Images and Deep Learning to Date Wheat Heading Stage." *Field Crops Research* 252(December 2019):107793. doi: 10.1016/j.fcr.2020.107793.

Velumani, Kaaviya, S. Oude Elberink, M. Y. Yang, and F. Baret. 2017. "Wheat Ear Detection in Plots by Segmenting Mobile Laser Scanner Data." Pp. 149–56 in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vols. IV-2/W4. Wuhan, China.

Verger, Aleixandre, Nathalie Vigneau, Corentin Chéron, Jean Marc Gilliot, Alexis Comar, and Frédéric Baret. 2014. "Green Area Index from an Unmanned Aerial System over Wheat and Rapeseed Crops." *Remote Sensing of Environment* 152:654–64. doi: 10.1016/j.rse.2014.06.006.

Voulodimos, Athanasios, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. 2018. "Deep Learning for Computer Vision: A Brief Review." doi: 10.1155/2018/7068349.

Wang, Xu, Hong Xuan, Byron Evers, Sandesh Shrestha, Robert Pless, and Jesse Poland. 2019. "High-Throughput Phenotyping with Deep Learning Gives Insight into the Genetic Architecture of Flowering Time in Wheat." *GigaScience* 8(11):1–11. doi: 10.1093/gigascience/giz120.

Wiesner-Hanks, Tyr, Ethan L. Stewart, Nicholas Kaczmar, Chad DeChant, Harvey Wu, Rebecca J. Nelson, Hod Lipson, and Michael A. Gore. 2018. "Image Set for Deep Learning: Field Images of Maize Annotated with Disease Symptoms." *BMC Research Notes* 11(1):440. doi: 10.1186/s13104-018-3548-6.

Xie, Chuanqi, and Ce Yang. 2020. "A Review on Plant High-Throughput Phenotyping Traits Using UAV-Based Sensors." *Computers and Electronics in Agriculture* 178:105731.

Xiong, Haipeng, Zhiguo Cao, Hao Lu, Simon Madec, Liang Liu, and Chunhua Shen. 2019. "TasselNetv2: In-Field Counting of Wheat Spikes with Context-Augmented Local Regression Networks." *Plant Methods* 15(1):150. doi: 10.1186/s13007-019-0537-2.

Xiong, Xiong, Lingfeng Duan, Lingbo Liu, Haifu Tu, Peng Yang, Dan Wu, Guoxing Chen, Lizhong Xiong, Wanneng Yang, and Qian Liu. 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):104. doi: 10.1186/s13007-017-0254-7.

Yalcin, Hulya. 2018. "Phenology Recognition Using Deep Learning: DeepPheno." Pp. 1–4 in *2018 26th Signal Processing and Communications Applications Conference (SIU)*. IEEE.

Yang, Guijun, Jiangang Liu, Chunjiang Zhao, Zhenhong Li, Yanbo Huang, Haiyang Yu, Bo Xu, Xiaodong Yang, Dongmei Zhu, Xiaoyan Zhang, Ruyang Zhang, Haikuan Feng, Xiaoqing Zhao, Zhenhai Li, Heli Li, and Hao Yang. 2017. "Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives." *Frontiers in Plant Science* 8:1111. doi: 10.3389/fpls.2017.01111.

Yendrek, Craig R., Tiago Tomaz, Christopher M. Montes, Youyuan Cao, Alison M. Morse, Patrick J. Brown, Lauren M. McIntyre, Andrew D. B. Leakey, and Elizabeth A. Ainsworth. 2017. "High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance." *Plant Physiology* 173(1):614–26. doi: 10.1104/pp.16.01447.

Zeiler, Matthew D., and Rob Fergus. 2012. "Differentiable Pooling for Hierarchical Feature Learning."

Zeiler, Matthew D., and Rob Fergus. 2013. "Stochastic Pooling for Regularization of Deep Convolutional Neural Networks." in *1st International Conference on Learning Representations, ICLR 2013 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR.

Zheng, Yang-Yang, Jian-Lei Kong, Xue-Bo Jin, Xiao-Yi Wang, and Min Zuo. 2019. "CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture." *Sensors* 19(5):1058. doi: 10.3390/s19051058.

Zhou, Chengquan, Hongbao Ye, Jun Hu, Xiaoyan Shi, Shan Hua, Jibo Yue, Zhifu Xu, and Guijun Yang. 2019. "Automated Counting of Rice Panicle by Applying Deep Learning Model to Images from Unmanned Aerial Vehicle Platform." *Sensors (Switzerland)* 19(14). doi: 10.3390/s19143106.

Zhou, Longfei, Xiaohe Gu, Shu Cheng, Guijun Yang, Meiyan Shu, and Qian Sun. 2020. "Analysis of Plant Height Changes of Lodged Maize Using UAV-LiDAR Data." *Agriculture (Switzerland)* 10(5). doi: 10.3390/agriculture10050146.

# 2. Estimates of maize plant density from UAV RGB images using Faster-RCNN detection model: impact of the spatial resolution

Early-stage plant density is an important trait to discern the rate of seed emergence, optimal sowing density, and potential yield. It is thus an attractive trait for plant breeders as well as agriculture practitioners that was previously measured by visual counting in the field. The affordability of UAVs and popularization of deep learning methods now allow the accurate localization and counting of early-stage plants through high-resolution RGB images. The study aims at understanding the spatial resolution requirements of a commonly used object detection algorithm, Faster-RCNN, and how the model is affected by differences in object sizes. In particular, the efficiency of Faster-RCNN while handling small objects or low-resolution UAV images is analyzed. Using lower resolution images would considerably increase the acquisition and processing throughputs of this trait by reducing the cost and time involved in data acquisition as well as for data processing. The study was conducted on early-stage maize datasets, with a range of object sizes, acquired over eight sites where the UAV was flying at different altitudes. This chapter corresponds to a journal article currently under review.

## Title

Full title**: Estimates of maize plant density from UAV RGB images using Faster-RCNN detection model: impact of the spatial resolution**
Short title: **Impact of spatial resolution on plant density estimation from UAV using Faster-RCNN**

## Authors

K. Velumani,[1,2]* R. Lopez-Lozano,[2]* S. Madec,[3] W. Guo,[4] J. Gillet,[1] A. Comar,[1] F. Baret[2]

[1] Hiphen SAS, 120 rue Jean Dausset, Agroparc, Bâtiment Technicité, 84140 Avignon, France

[2] INRAE, UMR EMMAH, UMT CAPTE, 228 route de l'Aérodrome, Domaine Saint Paul - Site Agroparc CS 40509, 84914 Avignon Cedex 9, France

[3] Arvalis, 228, route de l'Aérodrome – CS 40509, 84914 Avignon Cedex 9, France

[4] International Field Phenomics Research Laboratory, Institute for Sustainable Agro-Ecosystem Services, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

*Corresponding authors. Email : kaaviya.velumani@inrae.fr, raul.lopez-lozano@inrae.fr

## Abstract

Early-stage plant density is an essential trait that determines the fate of a genotype under given environmental conditions and management practices. The use of RGB images taken from UAVs may replace the traditional visual counting in fields with improved throughput, accuracy and access to plant localization. However, high-resolution images are required to detect the small plants present at the early stages. This study explores the impact of image ground sampling distance (GSD) on the performances of maize plant detection at three-to-five leaves stage using Faster-RCNN object detection algorithm. Data collected at high-resolution (GSD≈0.3cm) over six contrasted sites were used for model training. Two additional sites with images acquired both at high and low (GSD≈0.6cm) resolution were used to evaluate the model performances. Results show that Faster-RCNN achieved very good plant detection and counting (rRMSE=0.08) performances when native high-resolution images are used both for training and validation. Similarly, good performances were observed (rRMSE=0.11) when the model is trained over synthetic low-resolution images obtained by down-sampling the native training high-resolution images, and applied to the synthetic low-resolution validation images. Conversely, poor performances are obtained when the model is trained on a given spatial resolution and applied to another spatial resolution. Training on a mix of high- and low-resolution images allows to get very good performances on the native high-resolution (rRMSE=0.06) and synthetic low-resolution (rRMSE=0.10) images. However, very low performances are still observed over the native low-resolution images (rRMSE=0.48), mainly due to the poor quality of the native low-resolution images. Finally, an advanced super-resolution method based on GAN (generative adversarial network) that introduces additional textural information derived from the native high-resolution images was applied to the native low-resolution validation images. Results show some significant improvement (rRMSE=0.22) compared to bicubic up-sampling approach, while still far below the performances achieved over the native high-resolution images.

## 1. Introduction

Plant density at emergence is an essential trait for crops since it is the first yield component that determines the fate of a genotype under given environmental conditions and management practices [1]–[5]. Competition between plants within the canopy depends on the sowing pattern and its understanding requires reliable observations of the plant localization and density [6]–[9]. An accurate estimation of actual plant density is also necessary to evaluate the seed vigor by linking the emergence rate to the environmental factors [10]–[13].

Maize plant density was measured by visual counting in the field. However, this method is labor-intensive, time consuming and prone to sampling errors. Several higher throughput methods based on optical imagery have been developed in the last twenty years. This was permitted by the technological advances with the increasing availability of small, light and affordable high spatial resolution cameras and autonomous vehicles. Unmanned ground vehicles (UGV) provide access to detailed phenotypic traits [14]–[16] while being generally expensive and associated with throughputs of the order of few hundreds of microplots per hour. Conversely, unmanned aerial vehicles (UAV) are very affordable with higher acquisition throughput than UGVs. When carrying very high-resolution cameras, they can access potentially several traits [17], [18] including plant density [19], [20].

Image interpretation methods used to estimate plant density can be classified into three main categories. The first one is based on machine learning where the plant density measured over a small set of sampling area is related to other canopy level descriptors including vegetation indices derived from RGB and multispectral data [21]–[23]. However, this type of method may lead to significant errors due to the lack of representativeness of the training data set as well as the effect of possible confounding factors including changes in background properties or plant architecture under genetic control. The second category of methods is based on standard computer vision techniques, where the image is first binarized to identify the green objects that are then classified into plants according to the geometrical features defined by the operator (e.g. [24], [25]). The last category of methods is based on deep learning algorithms for automatic object detection [26]–[28]. The main advantage of deep learning methods is their ability to automatically extract low-level features from the images to identify the targeted objects. While deep learning methods appear very promising, their generalization capacity is determined by the volume and diversity of the training dataset [29]. While large collections of images can now be easily acquired, labeling the images used to train the deep models represents a significant effort that is the main limiting factor to build very large training datasets. Few international initiatives have been proposed to share massive labelled datasets that will contribute to maximize the performances of deep learning models [30]–[34], with however questions regarding the consistency of the acquisition conditions and particularly the ground sampling distance (GSD).

The use of UAV images for plant detection at early stages introduces important requirements on image resolution, as deep learning algorithms are sensitive to object scales with the identification of small objects being very challenging [35], [36]. For a given camera, low altitude flights are therefore preferred to get the desired GSD. However, low altitude flights decrease the acquisition throughput because of a reduced camera swath forcing to complete more tracks to cover the same experiment, and requires additionally to slow down the flying speed to reduce motion blur. An optimal altitude should therefore be selected to compromise between the acquisition throughput and the image GSD. Previous studies reporting early-stage maize plant detection from UAVs from deep learning methods did not addressed specifically this important scaling issue [20], [26], [27]. One way to address this scaling issue is to transform the low-resolution images into higher resolution ones using super-resolution techniques. Dai [37] have demonstrated the efficiency of super-resolution techniques to enhance segmentation and edge detection. Later, Fromm [38] and Magoulianitis [39] showed improvements in object

detection performances when using the super-resolution methods. The more advanced super-resolution techniques use deep convolutional networks trained over paired high- and low-resolution images [40]–[42]. Since the construction of a real-world paired high- and low-resolution dataset is a complicated task, the high-resolution images are often degraded using a bicubic kernel or less frequently using gaussian noise to constitute the low-resolution images [43]. However more recent studies have shown the drawbacks of the bicubic down-sampling approaches as it smoothens sensor noise and other compression artifacts, thus failing to generalize while applied to real world images [41]. More recent studies propose the use of unsupervised domain translation techniques to generate realistic paired datasets for training the super-resolution networks [44].

We propose here to explore the impact of image GSD on the performances of maize plant detection at stages from three to five leaves using deep learning methods. More specifically, three specific objectives are targeted: (1) to assess the accuracy and robustness of deep learning algorithms for detecting maize plants with high-resolution images used both, in the training and validation datasets; (2) to study the ability of these algorithms to generalize in the resolution domain, i.e. when applied to images with higher and lower resolution compared to the training dataset; and (3) to evaluate the efficiency of data augmentation and preparation techniques in the resolution domain to improve the detection performances. Special emphasis was put here on assessing the contribution of two contrasting methods to up-sample low-resolution images: a simple bicubic up-sampling algorithm, and a more advanced super-resolution model based on GAN (generative adversarial network) that introduces additional textural information. Data collected over several sites across France with UAV flights completed at several altitudes providing a range of GSDs were used.

## 2. Materials and Methods

### 2.1 Study sites

This study was conducted over 8 sites corresponding to different field phenotyping platforms distributed across the west of France and sampled from 2016 to 2019 (Figure 3). The list of sites and their geographic coordinates are given in Table 4. Each platform included different maize microplots with size 20 to 40 square meters. Depending on the experimental design of the platform, the microplots were sown with two to seven rows of maize of different cultivars and row spacing varying from 30 to 110 cm. The sowing dates were always between mid-April and mid-May.

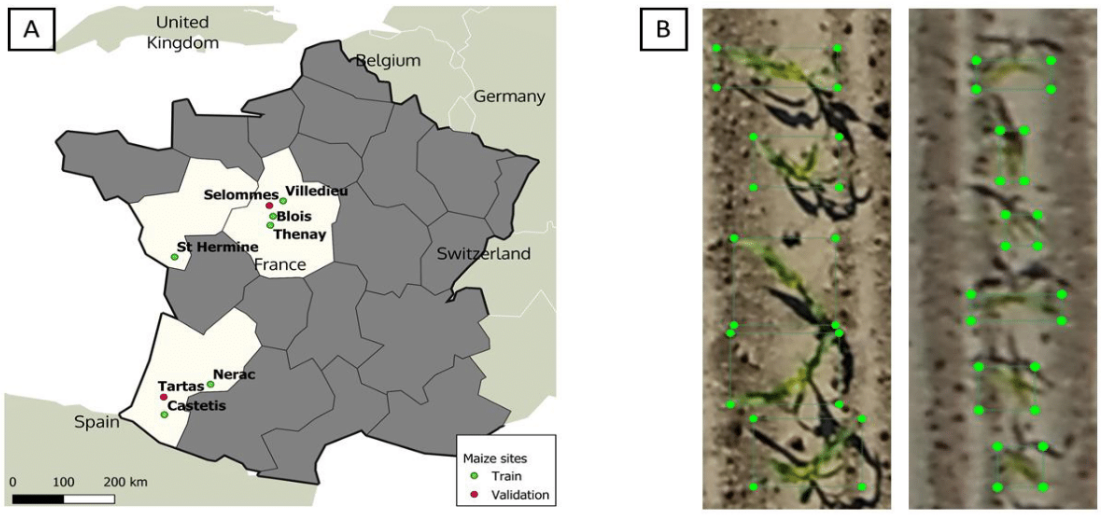


**Figure 3 Location of the study sites with example extracts of the maize microplots acquired from UAV (A)** A map displaying the location of the eight maize phenotyping platforms located in the west of France used in this

study. **(B)** An illustration of the bounding boxes drawn around the maize plants. The examples shown are from the Tartas site (GSD=0.27cm) on the left and Tartas site (GSD=0.63cm) on the right.

**Table 4 The dataset used for the training and validation of the object detection models are listed here.** $T^h$, is the training high-resolution dataset, $V^h$ is the validation high-resolution dataset and $V^l$ is the validation low-resolution dataset. * For this site, the microplot extracts were resampled to GSD=0.25 cm before annotating.

| Dataset | # | Site Name | Latitude (°) | Longitude (°) | Acquisition Date | Camera | Flight Altitude (m) | Focal length (mm) | GSD (cm) | Nb. microplots labelled | Nb. plants labelled | Average plant size (pixels) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T^h$ | 1 | Castetis | 43.46 | -0.71 | 06-06-2019 | FC6540 | 26 | 24 | 0.35 | 20 | 2239 | 939 |
| | 2 | St Hermine | 46.54 | -1.06 | 23-05-2019 | FC6540 | 26 | 35 | 0.33* | 20 | 2674 | 806 |
| | 3 | Nerac | 44.16 | 0.3 | 01-06-2017 | ILCE-6000 | 25 | 30 | 0.32 | 44 | 3338 | 2253 |
| | 4 | Thenay | 47.38 | 1.28 | 18-05-2018 | ILCE-6000 | 22 | 30 | 0.25 | 72 | 7454 | 1505 |
| | 5 | Villedieu | 47.88 | 1.53 | 28-06-2016 | n/a | n/a | n/a | 0.27 | 26 | 2390 | 2159 |
| | 6 | Blois | 47.56 | 1.32 | 18-05-2018 | ILCE-6000 | 25 | 30 | 0.33 | 20 | 1746 | 1419 |
| $V^h$ | 7 | Tartas | 43.80 | -0.79 | 08-06-2019 | FC6540 | 20 | 24 | 0.32 | 22 | 2151 | 1336 |
| | 8 | Selommes | 47.76 | 1.19 | 17-05-2019 | L1D-20c | 16.2 | 10.26 | 0.27 | 14 | 1105 | 891 |
| $V^l$ | 9 | Tartas | 43.80 | -0.79 | 08-06-2019 | FC6540 | 40 | 24 | 0.63 | 24 | 2151 | 437 |
| | 10 | Selommes | 47.76 | 1.19 | 17-05-2019 | L1D-20c | 30 | 10.26 | 0.66 | 14 | 1105 | 156 |

### 2.2 Data acquisition and processing

UAV flights were carried out on the eight sites approximately one month after the sowing date, between mid-May and mid-June (Table 4). Maize plants were in between three to five leaves stage, ensuring that there is almost no overlap among individual plants from near nadir viewing. Three different RGB cameras were used for the data acquisition: Sony Alpha (ILCE-6000) with a focal length of 30 mm, DJI X7 (FC6540) with focal lengths of 24 mm and 30 mm and the default camera with DJI Mavic 2 pro (L1D-20c) with a focal length of 10.26 mm mounted on AltiGator Mikrokopter (Belgium) and DJI Mavic 2 pro (China). To geo-reference the images, ground control points (GCPs) were evenly distributed around the sites and their geographic coordinates were registered using a Real Time Kinematic GPS.
The flights were conducted at an altitude above the ground ranging between 15 and 22 meters, providing a ground sampling distance (GSD) between 0.27 and 0.35 cm (Table 4). For Tartas and Selommes sites, an additional flight was done at a higher altitude on the same day providing a GSD between 0.63 and 0.66 cm.

The flights were planned with a lateral and front overlap of 60/ 80% between individual images. Each dataset was processed using Photoscan Professional (Agisoft LLC, Russia) to align the overlapping images by automatic tie point matching, optimize the aligned camera positions and finally geo-reference the results using the GCPs. The steps followed are similar to the data processing detailed by Madec [15]. Once ortho-rectified, the multiple instances of the microplot present in the overlapping images were extracted using Phenoscript, a software developed within the CAPTE research unit. Phenoscript allows to select, among the individual images available for each microplots, those with full coverage of the microplot, minimum blur and view direction closer to the nadir one. Only these images were used in this study.

### 2.3 Manual labelling of individual plants

From each site, the microplots were labelled with an offline tool, LabelImg [45]: bounding boxes around each maize plant were interactively drawn (Figure 3B). The available sites (Table 4) were divided into three groups: (1) the first group ($T^h$) composed of six sites was used to train the plant detection models. It includes a total of 202 microplots corresponding to 19,841 plants. (2) The second group ($V^h$) corresponding to the Tartas and Selommes with low altitude

flights was used to evaluate the model performance at high-resolution. It includes a total of 36 microplots corresponding to 3256 plants. (3) The third group ($V^l$) corresponds to the high-altitude flights in Tartas and Selommes was used to evaluate the model performance at low-resolution. It includes a total of 36 microplots corresponding to 3256 plants. An example of images extracted from the three groups is shown in Figure 4.

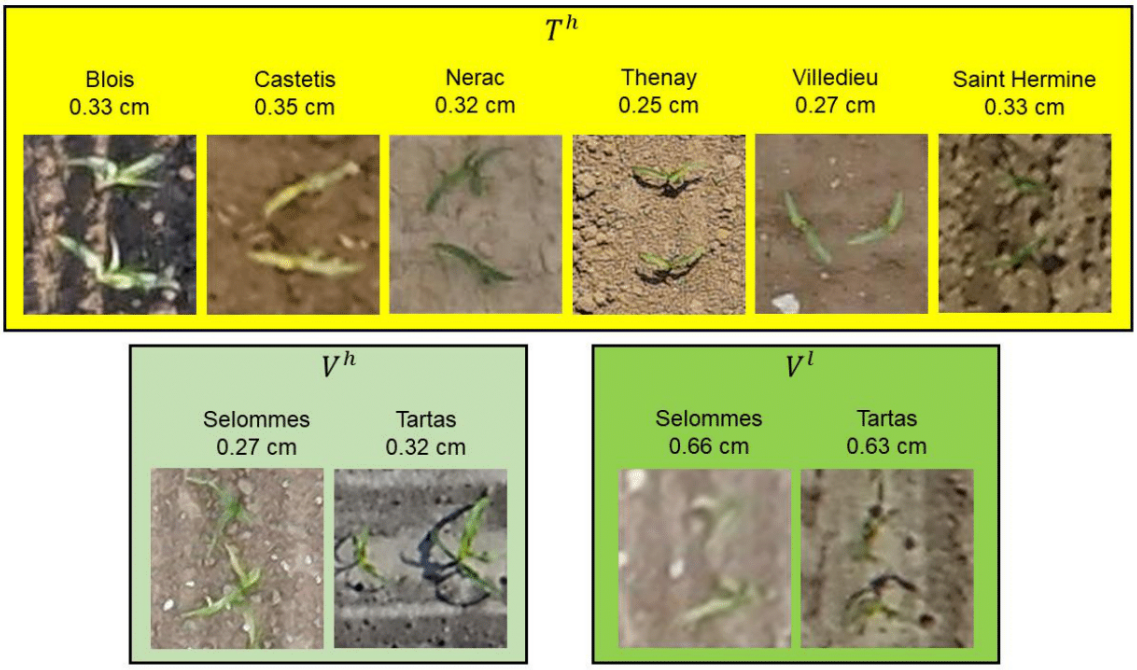


**Figure 4 Examples of maize plants extracted from the in the eight sites used in this study**. The image titles indicate the location of the sites. $T^h$, $V^h$ and $V^l$ are the training high-resolution dataset, validation high-resolution dataset and the validation low-resolution dataset, respectively.

### 2.4 The Faster RCNN object detection model

Faster-RCNN [46], a convolutional neural network designed for object detection was selected to identify maize plants in the image. Besides its wide popularity outside the plant phenotyping community, Faster-RCNN has also been proved to be suitable for various plant and plant-organ detection tasks [47]–[49]. We used the implementation of Faster RCNN in the open-source MMDetection Toolbox [50], written in PyTorch, with pre-trained weights on ImageNet. The Faster-RCNN model with a ResNet50 backbone was trained for 12 epochs with a batch size of 2. The weights were optimized using an SGD optimizer (Stochastic Gradient Descent) with a learning rate of 0.02. For the model training, ten patches of 512 x 512 pixels were randomly extracted from each microplot in the training sites. Standard data augmentation strategies such as rotate, flip, scale, brightness/contrast and jpeg compression were applied.

### 2.5 Experimental Plan

To evaluate the effect of the resolution on the reliability of maize plant detection, we compared Faster RCNN performances over training and validation datasets made of images of high (GSD≈0.30 cm) and low (GSD≈0.60 cm) resolution. Three training datasets built from $T^h$ (Table 4) were considered: (1) the original $T^h$ dataset with around 0.32 cm GSD; (2) A dataset, $T_{gm}^{h\to l}$ where the images from $T^h$ were down-sampled to 0.64 cm GSD using a gaussian filter and motion blur that mimics the actual low-resolution imagery acquired at higher altitude as described later (section 2.6.1); (3) A dataset, where the original $T^h$ high-resolution dataset was merged with its low-resolution transform, $T_{gm}^{h\to l}$. This $T^h + T_{gm}^{h\to l}$ is expected to provide

robustness of the model towards changes in GSD. Note that we did not investigated the training with the native low-resolution images because the labeling of low-resolution images is often difficult because plants are not easy to identify visually and to draw accurately the corresponding bounding box. Further, only two flights were available at the high altitudes (Table 4) that were reserved for the validation. A specific model was trained over each of the three training datasets considered (Table 5) and then evaluated over independent high- and low-resolution validation datasets.

We considered three validation datasets for the high-resolution images: (1) the native high-resolution validation dataset, $V^h$ acquired at low altitude with GSD around 0.30 cm (Table 4). (2) a synthetic high-resolution dataset of GSD around 0.30 cm obtained by up-sampling the native low-resolution dataset, acquired at high altitude, using a bicubic interpolation algorithm as described in section 2.6.2. It will be called $V_{bc}^{l\to h}$. (3) A synthetic high-resolution dataset, $V_{sr}^{l\to h}$, obtained by applying a super-resolution algorithm (see Section 2.6.3) to the native low-resolution dataset $V^l$ and resulting in images with a GSD around 0.30 cm. Finally two low-resolution datasets will be also considered: (1) The native low-resolution validation dataset, $V^l$ (Table 4), with a GSD around 0.60 cm. (2) A synthetic low-resolution dataset, $V_{gm}^{h\to l}$, obtained by applying a Gaussian filter to down-sample (see Section 2.6.1) the original high-resolution dataset, $V^h$, and get a GSD around 0.60 cm.

**Table 5 Description of the training and validation datasets.**

| | Dataset name | Nb. microplots | Nb. plants | Comment |
|---|---|---|---|---|
| Training | $T^h$ | 202 | 19,841 | Native high-resolution training dataset |
| | $T_{gm}^{h\to l}$ | 202 | 19,841 | Down-sampling $T^h$ with gaussian filter and motion blur |
| | $T^h + T_{gm}^{h\to l}$ | 404 | 39,682 | Merging $T^h$ and $T^{h\to l}$ |
| Validation | $V^h$ | 36 | 3256 | Native high-resolution validation dataset |
| | $V_{bc}^{l\to h}$ | 36 | 3256 | Up-sampling $V^l$ with bi-cubic algorithm |
| | $V_{sr}^{l\to h}$ | 36 | 3256 | Up-sampling $V^l$ with Cycle-ESRGAN super-resolution |
| | $V^l$ | 36 | 3256 | Native low-resolution validation dataset |
| | $V_{gm}^{h\to l}$ | 36 | 3256 | Down-sampling $V^h$ with gaussian filter and motion blur |

## 2.6 Methods for image up- and down-sampling

### 2.6.1 Gaussian filter down-sampling

To create the synthetic low-resolution datasets $T_{gm}^{h\to l}$ and $V_{gm}^{h\to l}$, a Gaussian filter with a sigma=0.63 and a window size=9 followed by a motion blur with a kernel size=3 and angle=45 were applied to down-sample the native high-resolution datasets $T^h$ and $V^h$ by a factor of 2. This solution was preferred to the commonly used bicubic down-sampling method because it provides low-resolution images more similar to the native low-resolution UAV images (Figure 5). This was confirmed by comparing the image variance over the Selommes and Tartas sites where both native high and low-resolution images were available: the variance of the $V_{gm}^{h\to l}$ was closer to that of $V^l$ whereas the bicubic down-sampled dataset had a larger variance corresponding to sharper images. This is consistent with [38] and [51] who used the same method to realistically downsample high-resolution images.

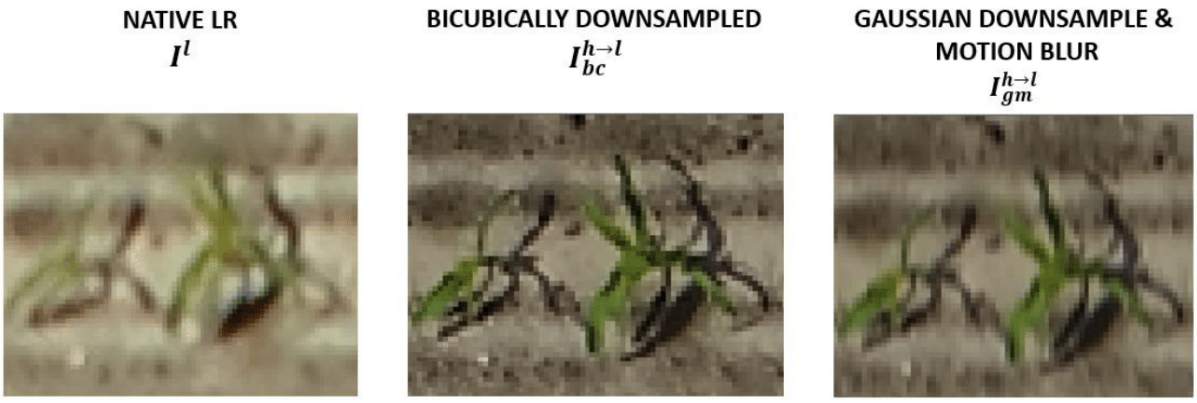


**Figure 5 Visual comparison of the extract of the same plant from the Tartas site between different versions of low-resolution.** Native low-resolution, synthetic low-resolution from bicubic down-sampling, and synthetic low-resolution from Gaussian down-sampling (sigma=0.63, window=9) followed by a motion blur (kernel size=3 and angle=45).

### 2.6.2 Bicubic up-sampling

The bicubic interpolation algorithm was used to generate $V_{bc}^{l\to h}$ by up-sampling the native low-resolution UAV images, $V^l$. The bicubic interpolation available within Pillow, the Python Imaging Library [52] was used to resample the images.

### 2.6.3 Super-resolution images derived from Cycle-ESRGAN

The super-resolution (SR) is an advanced technique that artificially enhances the textural information while up-sampling images. We used a SR model inspired from [53]. It is a two-stage network composed of a CycleGAN network that generates synthetic paired data and a ESRGAN network capable of image upsampling. The CycleGAN [54] performs unsupervised domain mapping between the native low-resolution and bicubic downsampled low-resolution domains. Thus, for any given input image, CycleGAN is trained to add realistic image noise typical of low-resolution images. The ESRGAN-type super-resolution network [42] up-samples by a factor of two the low-resolution images.

The paired high-resolution and "realistic" low-resolution dataset generated by the CycleGAN was used in the simultaneous training of the ESRGAN-stage of the network. The CycleGAN stage of the network was initially trained for a few epochs following which the two stages (CycleGAN + ESRGAN) were trained together simultaneously. It should be noted that during inference, only the ESRGAN stage of the network would be activated. The training parameters and losses reported by Han [53] were used for the model training. The model weights were initialized over the Div2k dataset [55] and finetuned on the UAV dataset detailed below. The Cycle-ESRGAN network was implemented using Keras [56] deep learning library in Python. The codes will be made available on Github at the following link: https://github.com/kaaviyave/Cycle-ESRGAN

A dedicated training dataset for the super-resolution network was prepared using UAV imagery belonging to the following two domains:

- **Native high-resolution domain**: 2234 microplot extractions from four sites with an average GSD of less than 0.33 cm. Some of the sites belonging to the $T^h$dataset was used as a part of the training.
- **Native low-resolution domain**: 1713 microplot extractions from three sites with an average GSD of 0.46 cm per site.

None of the validation sites ($V^l$ and $V^h$ in Table 4) were used in the training of the super resolution model. The synthetic downsampled dataset used to train the CycleGAN was prepared by bicubic downsampling the native high-resolution domain by a factor of 2. The

images were split into patches of size 256 x 256 pixels for the high-resolution domain and into 128 x 128 pixels for the low-resolution domain.

### 2.7 Evaluation metrics

In this study, the Average Precision (AP), Root Mean Squared Error (RMSE) and Accuracy will be utilized for the evaluation of the Faster-RCNN models for the purpose of maize plant detection and counting.

**AP**: The AP is a frequently used metric for the evaluation of object detection models and can be considered as the area under the precision-recall curve.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

Where TP is the number of True Positive, FP is the number of False Positive and the FN is the number of False Negative. For the calculation of AP, a predicted bounding box is considered True Positive (TP) if its intersection area over union area (IoU) with the corresponding labelled bounding box is larger than a given threshold. Depending on the objective of the study, different variations exist in the AP metric calculation and the choice of IOU threshold used to qualify a predicted bounding box as TP. After considering several IoU threshold values, we decided to use an IoU threshold of 0.25 to compute AP. This will be later justified. The Python COCO API was used for the calculation of the AP metric [57].

**Accuracy** evaluates the model's performance by calculating the ratio of correctly identified plants to all the predictions made by the model. A predicted bounding box is considered true positive if it has a confidence score of more than 0.5 and an IoU threshold of 0.25. Accuracy is then calculated as:

$$Ac = \frac{TP}{TP + FP + FN}$$

The **relative root mean square error (rRMSE)** between the number of labeled and detected plants across all images belonging to the same dataset:

$$rRMSE = \frac{\sqrt{\frac{\sum_i^n (P_{o,i} - P_{p,i})^2}{n}}}{\overline{P_{o,i}}}$$

where $P_{o,i}$ is the number of plants labeled on image $i$ and $P_{p,i}$ is the number of images predicted by the CNN (confidence score > 0.5 and an IoU > 0.25) and $\overline{P_{o,i}}$.is the average number of labeled plants per image.

## 3. Results and Discussion

### 3.1 Faster RCNN detects plants with high accuracy at high spatial resolution

Very good performances (rRMSE=0.08; Ac=0.88, AP=0.95) are achieved when the model is trained over the high-resolution images ($T^h$) and applied on high-resolution images taken on independent sites ($V^h$). The good performances are explained by the high rate of true positives (Figure 6a). However, the detector performs slightly differently on the two sites used for the

validation: in Selommes, an over-detection (false positives, FP) is observed for a small number of plants, when the detector splits a plant into two different objects (Figure 6b). Conversely, in the Tartas site, some under-detection (false negatives, FN) is observed, with a small number of undetected plants (Figure 6).

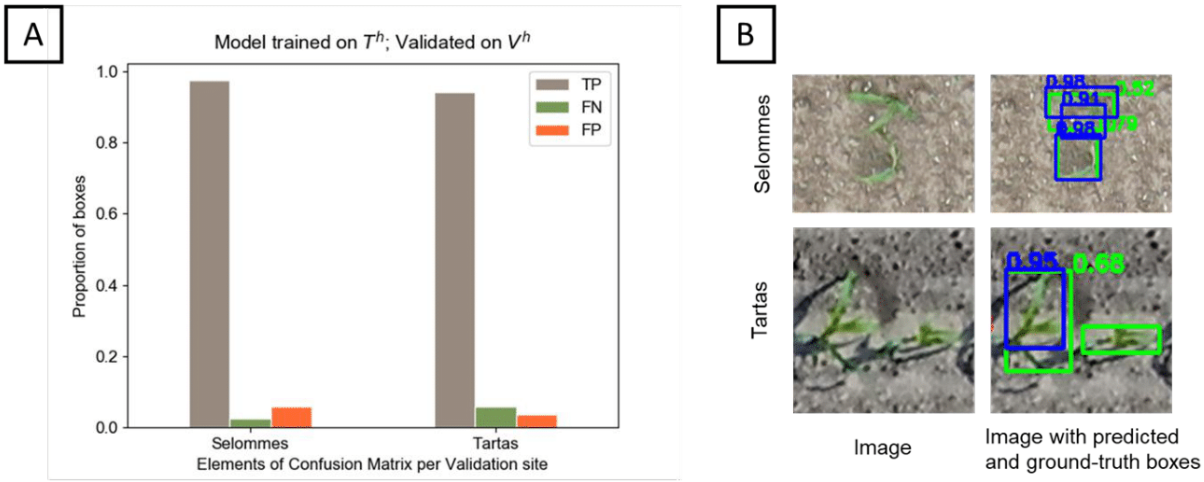


**Figure 6 Results of the model trained on the native HR dataset, $T^h$, and applied to the HR validation dataset, $V^h$. A)** Elements of the confusion matrix - True Positives (TP), False Negatives (FN) and False Positives (FP) for Selommes and Tartas sites. **B)** An example of false positive and false negative observed in the two validation sites. The ground truth bounding boxes are shown in green and the predicted bounding boxes are shown in blue. The green text indicates the IoU of the predicted bounding box with the ground truth and the blue text indicates the confidence score of the predictions.

A detailed analysis of the Precision-Recall curves for the configuration [$T^h$, $V^h$] at different IoU (Figure 7) shows a drastic degradation of the detector performances when the IoU is higher than 0.3. This indicates that the model is not accurate when determining the exact dimensions of maize plants. This is partly explained by the difficulty of separating the green from the ground in the shadowed parts of the images. As a consequence, some shaded leaves are excluded from the bounding boxes proposed by the detector and, conversely, some shadowed ground are wrongly included in the bounding boxes proposed (Figure 6b). Further, when a single plant is split into two separate objects by the detector, the resulting bounding boxes are obviously smaller than the corresponding plant (Figure 6b). As a consequence, we proposed to use an IoU threshold of 0.25 to evaluate the model performance to better account for the smaller size of the detected bounding boxes. This contrasts from most object detection applications where an IoU threshold of 0.5 or 0.75 is commonly used to evaluate the performance of the methods [58], [59]. The observed degradation of the model performance for IoU above 0.3 indicates that the method presented provides less accurate localization than in other object detection studies, including both, real world objects and phenotyping applications [49], [60], [61]. An inaccurate estimation of plant dimensions is not critical for those applications assessing germination or emergence rates and uniformity, where plant density is the targeted phenotypic trait. If the focus is to additionally assess the plant size in early developmental stages as well, mask-based RCNN [62], [63] could be used instead. In contrast to algorithms trained on rectangular regions like Faster-RCNN, mask-based algorithms have the potentials to more efficiently manage the shadow projected on the ground by plants, limiting therefore the possible confusion between shaded leaves and ground during the training. However, generating mask annotations is time-consuming, increasing the effort needed to generate a diverse training dataset.

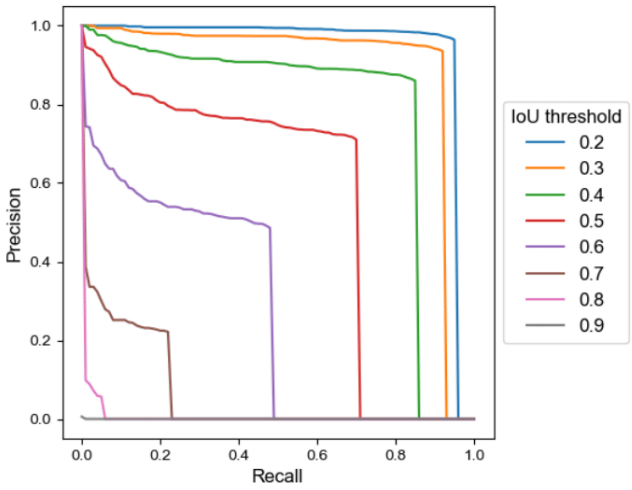


**Figure 7 Precision-Recall curve at different IoU thresholds for the plant detection model trained and applied with high-resolution images ($T^h, V^h$).**

These results provide slightly better performances as those reported by David [20] with Ac≈0.8 and rRMSE≈0.1 when using the "out-domain" approach as the one used in this study, i.e. when the training and validation sites are completely independent. They used images with a spatial resolution around 0.3cm as in our study. This is also consistent with the results of Karami [26] who obtained an accuracy of 0.82 with a spatial resolution of around 1cm. They used the anchor-free Few Shot Leaning (FSL) method which identifies and localizes the maize plants by estimating the central position. They claim that their method is little sensitive to object size and thus to the spatial resolution of the images. The accuracy increases up to 0.89 when introducing few images from the validation sites in the training dataset. Kitano [27] proposed a two-step method: they first, segment the images using a CNN-based method and then count the segmented objects. They report an average rRMSE of 0.24 over a test dataset where many factors including image resolution vary (ranging from GSD≈0.3cm to 0.56 cm). They report that their method is sensitive to the size and density of the objects. In the following, we will further investigate the dependency of the performances to image resolution.

### 3.2 The Faster-RCNN model is sensitive to image resolution and apparent plant size

The performances of the model were evaluated when it is trained and validated over images with different resolution. When Faster-RCNN is trained on the high-resolution domain ($T^h$) and applied to a dataset with low-resolution ($V^l$), both AP and Ac decrease almost by 30% (Table 6) compared to the results where the model is trained and applied over high-resolution images. The rate of true positive drops because of the drastic increase of false negatives indicating a high rate of misdetection (Figure 8, $[T^h, V^l]$). This degradation of the detection performances impacts highly the rRMSE that increases up to 0.48. This indicates that the model is sensitive to the resolution of the images. We further investigated if this was linked to the apparent size of the plants and therefore up-sampled the validation low-resolution images with a bicubic interpolation method ($V_{bc}^{l\to h}$) to get plants with the same size as in the native high-resolution images ($V^h$). Results show that Ac increases from 0.54 to 0.63 and AP from 0.64 to 0.77. However, because of the high imbalance between FN and FP (Figure 8, $\left[T^h, V_{bc}^{l\to h}\right]$), the counting performances remains poor with rRMSE=0.49.

**Table 6 Comparison of the performance of the Faster-RCNN models trained and validated over datasets with different resolution.** The colors indicate the goodness of the performances for the three metrics (green: best; red: worst).

| | | Validation | | | | | |
|---|---|---|---|---|---|---|---|
| | | High Resolution ($V^h$) | | | Low Resolution ($V^l$) | | |
| | | rRMSE | Ac | AP | rRMSE | Ac | AP |
| Training | $T^h$ | 0.08 | 0.88 | 0.95 | 0.48 | 0.54 | 0.64 |
| | $T_{gm}^{h\rightarrow l}$ | 0.52 | 0.56 | 0.71 | 0.29 | 0.76 | 0.81 |

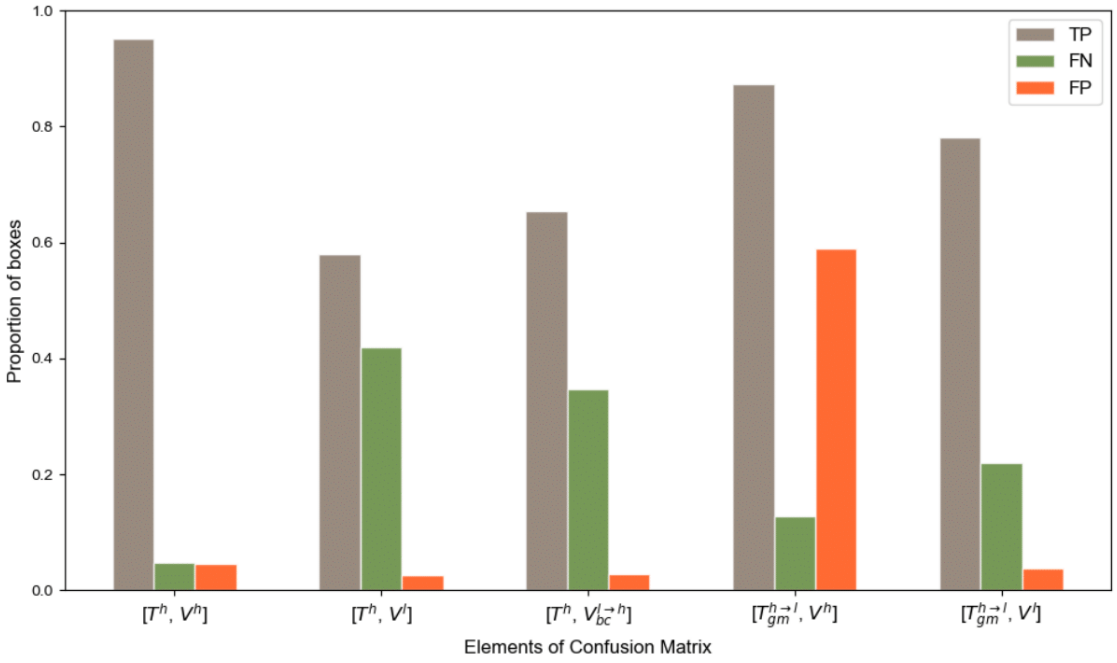


**Figure 8 Results of maize plant detection when trained and evaluated across different resolution domains. $T^h$ native high-resolution training dataset; $T_{gm}^{h\rightarrow l}$ low-resolution training dataset by down-sampling $T^h$ using gaussian motion blur; $V^h$ native high-resolution validation dataset; $V^l$ native low-resolution validation dataset; $V_{bc}^{l\rightarrow h}$ high-resolution dataset by up-sampling $V^l$ using bicubic interpolation.**

When the model is trained over simulated low-resolution images ($T_{gm}^{h\rightarrow l}$), the detection and counting performances evaluated on high-resolution images ($V^h$) also degrades drastically (Table 6). The rate of true positive is relatively high, but the rate of false positive increases drastically (Figure 8$\left[T_{gm}^{h\rightarrow l}, V^h\right]$). We observe that the average number of predicted bounding boxes overlapping each labelled box increases linearly with its size (Figure 9). For example, the model identifies on average two plants inside plants larger than 4000 pixels. The imbalance between FN and FP explains the very poor counting performances with rRMSE=0.52 (Table 6). This result confirms the importance to keep consistent the resolution and plant size between the training and the application datasets since Faster-RCNN tends to identify objects that have a similar size to the objects used during the training.

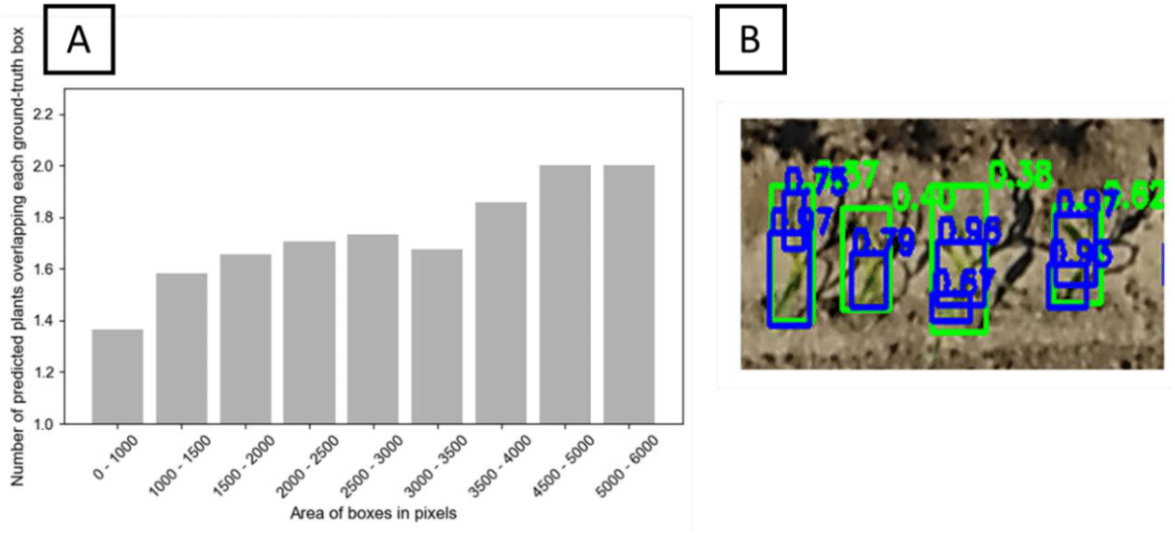


**Figure 9 Effects of the hyper-specialization of Faster-RCNN trained with synthetic low-resolution images ($T_{gm}^{h\to l}$) and applied to a high-resolution datatset ($V^l$): A)** Relationship between the size of the ground–truth bounding boxes and the average number of predicted bounding boxes intersecting with them; **B)** Example of over-detection of maize plants due to different object size. The ground truth bounding boxes are shown in green and the predicted boxes are shown in blue. The green text indicates the IoU of the predicted boxes with the ground truth and the blue text indicates the confidence score of the predictions.

We thus evaluated whether data-augmentation may improve the performances on the low-resolution images ($V^l$): the Faster-RCNN model trained on the simulated low-resolution images ($T_{gm}^{h\to l}$) shows improved detection performances as compared to the training over the native high-resolution images (Table 6) with a decrease of the rRMSE down to 0.29 (Table 6). When this model trained with synthetic low-resolution images ($T_{gm}^{h\to l}$) is applied to a dataset downscaled to a similar resolution ($V_{gm}^{h\to l}$), the performances improve dramatically with Ac increasing from 0.56 to 0.89 and AP from 0.71 to 0.90 while the rRMSE drops to 0.10. However, when this model trained with synthetic low-resolution images ($T_{gm}^{h\to l}$) is applied to the native low-resolution images ($V^l$), moderate detection performances are observed which degrades the counting estimates with rRMSE=0.29 (Table 6).

The performances of the model trained over the synthetic low-resolution images ($T_{gm}^{h\to l}$) are quite different when evaluated over the native images ($V^l$) or the synthetic ones ($V_{gm}^{h\to l}$) with the latter yielding results almost comparable to the high-resolution configurations with AP=0.90 (Table 6). This indicates that the low-resolution synthetic images contain enough information to detect accurately the maize plants. Conversely, the native low-resolution image, $V^l$, have probably lost part of the textural information. In addition, the model trained on the synthetic low-resolution images is not able to extract the remaining pertinent plant descriptors from the native low-resolution images. We can observe that the native low-resolution images contain less details as compared to the synthetic ones (Figure 10): some plants are almost not visible in the $V^l$ images, as the textural information vanishes and even the color of maize leaves cannot be clearly distinguished from the soil background. This explains why the model was not able to detect the plants, even when it is trained with the synthetic low-resolution images ($T_{gm}^{h\to l}$). Contrary to vectors that operate at an almost constant height like ground vehicles [16], [64]–[66] or fixed cameras [67]–[70], camera settings (aperture, focus and integration time) in UAVs need to be adapted to the flight conditions, especially flight altitude, to maximize image quality. Further, the jpg recording format of the images may also significantly impact image quality.

Recording the images in raw format would thus improve the detection capability at the expense of increased data volume and sometimes image acquisition frequency.

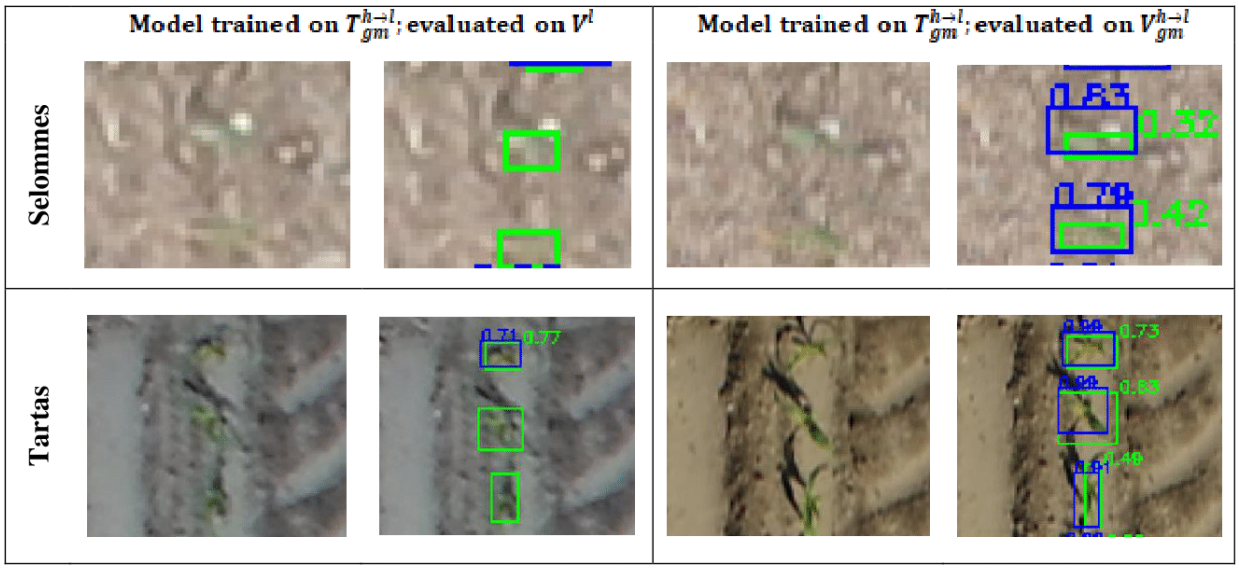


**Figure 10 An example showing the same plants extracted from the exact same locations in two versions of the validation dataset: native LR ($V^l$) and the synthetic LR obtained from gaussian downsampling ($V_{gm}^{h\to l}$).** The first and third column show the raw images while the second and fourth column shows the detector predictions. The ground truth bounding boxes are shown in green and the predicted bounding boxes are shown in blue. The green text indicates the IoU of the predicted box with the ground truth and the blue text indicates the confidence score of the predictions.

### 3.3 Data Augmentation makes the model more resistant to changes in image resolution

We finally investigated whether mixing high and low-resolution images in the training dataset would make the model more resistant to changes in the image resolution. Results show that merging native high-resolution with synthetic low-resolution images ($T^h + T_{gm}^{h\to l}$) provides (Table 7) performances similar to those observed when the model is trained only over high ($T^h$) or synthetic low ($T_{gm}^{h\to l}$) and validated on the same resolution ($V^h$ or $V_{gm}^{h\to l}$) (Table 6). This proves that data augmentation could be a very efficient way to deal with images having different resolution. Further, this model trained on augmented data ($T^h + T_{gm}^{h\to l}$) (Table 7) surprisingly beats the performances of the model trained only on the high-resolution images ($T^h$) as displayed in Table 6. This is probably a side effect of the increase of the size of the training dataset (Table 5). Nevertheless, when validating on the native low-resolution images ($V^l$) (Table 7) the performances are relatively poor as compared to the model trained only on the synthetic low-resolution images ($T_{gm}^{h\to l}$). This is explained by the lower quality of the native low-resolution images as already described in the previous section.

**Table 7 Comparison of the performance of the Faster-RCNN models trained over the augmented data ( $T^h + T_{gm}^{h\to l}$) and validated over datasets with different resolution.** The colors indicate the goodness of the performances for the three metrics (green: best; red: worst).

| | rRMSE | Ac | AP |
|---|---|---|---|
| $V^h$ | 0.06 | 0.91 | 0.96 |
| $V_{gm}^{h\to l}$ | 0.11 | 0.92 | 0.91 |
| $V^l$ | 0.48 | 0.64 | 0.72 |

### 3.4 Up-sampling with the super-resolution method improves the performances of plant detection on the native low-resolution images

If the training is difficult with the native low-resolution images because plants are visually difficult to identify and label, the training should be done over low-resolution images derived from the high-resolution images using a more realistic up-sampling method than the standard bicubic interpolation one. Alternatively, the training could be done using the high-resolution images and the low-resolution dataset may be up-sampled to a synthetic high-resolution domain using bicubic interpolation or super-resolution techniques.

Results show that the super-resolution technique improved plant detection very significantly as compared to the native low-resolution ($V^l$) and bi-cubic up-sampled ($V_{bc}^{l\to h}$) images (Table 8). This impacts positively the counting performances while not reaching the performances obtained with the high-resolution images ($V^h$). The super-resolution reduces drastically the under-detection of maize plants particularly on the Tartas site (Figure 11) where, as mentioned in Section 3.2, these native low-resolution images have lower textural information and green fraction per plant.

**Table 8 Comparison of the performance of the Faster-RCNN models trained over high-resolution images and applied to the native low-resolution images ($V^l$), the synthetic high-resolution images that is up-sampled/transformed using either bi-cubic ($V_{bc}^{l\to h}$) or super-resolution ($V_{sr}^{l\to h}$) techniques.** The colors indicate the goodness of the performances for the three metrics (green: best; red: worst).

| | rRMSE | Ac | AP |
|---|---|---|---|
| $V^l$ | 0.58 | 0.54 | 0.64 |
| $V_{bc}^{l\to h}$ | 0.43 | 0.63 | 0.77 |
| $V_{sr}^{l\to h}$ | 0.22 | 0.80 | 0.85 |

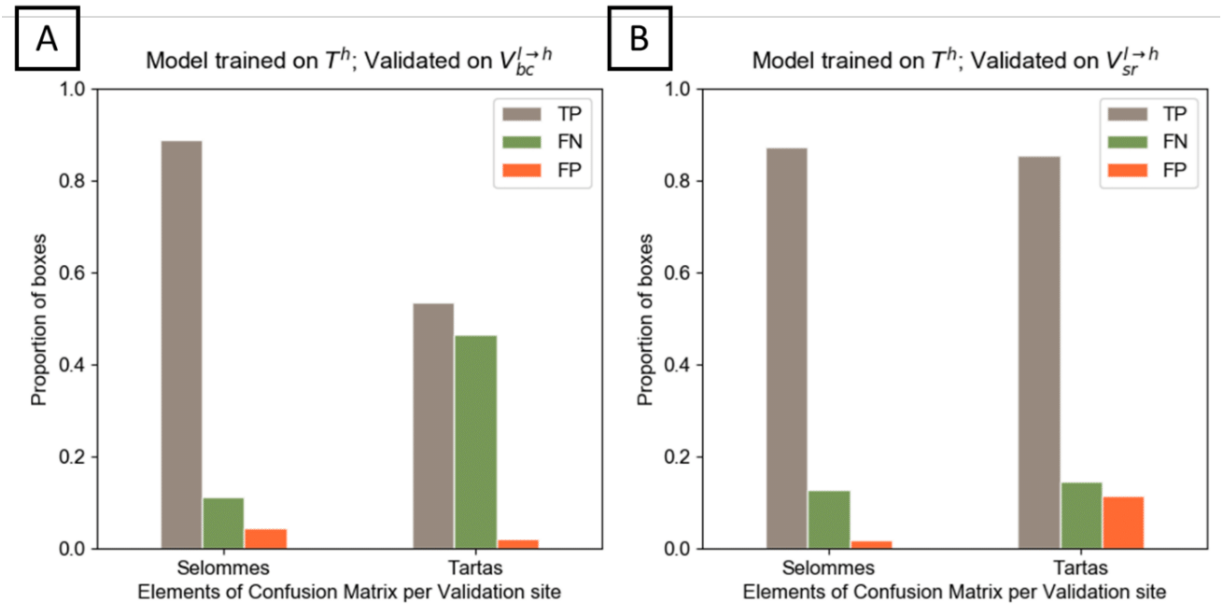


**Figure 11 A comparison between the performance of the Faster-RCNN model trained on the HR dataset, $T^h$, and applied to the synthetic high-resolution datasets. A)** Model trained on $T^h$ and evaluated on the synthetic HR dataset $V_{bc}^{l\to h}$, from the bicubic up-sampling; **B)** Model trained on $T^h$ applied to the synthetic HR dataset $V_{sr}^{l\to h}$, from the super-resolution technique.

The super resolution approach enhances the features used to identify maize plants, with colors and edges more pronounced than in the corresponding native LR images (Figure 12). Maize plants are visually easier to recognize in the super-resolved images as compared to both the native low-resolution and the bicubically up-sampled images.

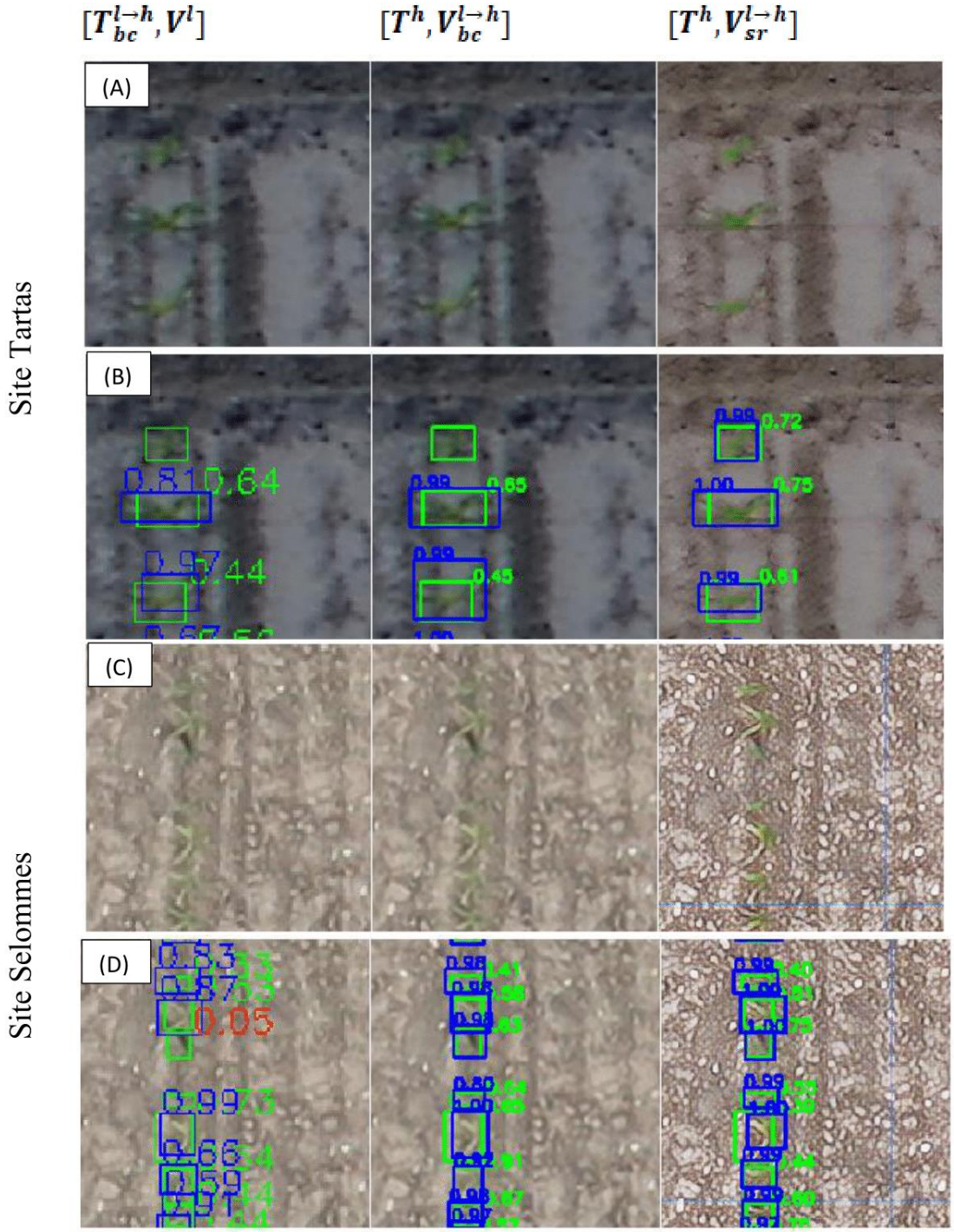


**Figure 12 Illustration of the performance of the Faster-RCNN model on the synthetic high-resolution and native low-resolution datasets. (A) and (C):** Images belonging to three datasets - native low-resolution $V^{l}$, bicubically up-sampled $V_{bc}^{l\to h}$ and finally the up-sampling by super-resolution technique $V_{sr}^{l\to h}$. **(B) and (D):** The results predicted by the model trained on $T_{gm}^{h\to l}$ applied to $V^{l}$ (first column) and the model trained on $T^{h}$ applied to

the synthetic high-resolution datasets $V_{hr}^{l\to h}$ and $V_{sr}^{l\to h}$ (second and third column). The ground truth bounding boxes are shown in green and the predicted bounding boxes are shown in blue. The green text indicates the IoU of the predicted bounding box with the ground truth and the blue text indicates the confidence score of the predictions.

Nevertheless, although easier to interpret, the images generated by super-resolution do not appear natural with some exaggerated textural features of the soil background (Figure 12 c, d). In few cases, super-resolution images show new features – e.g. coloring some pixels in green– in leaf-shaped shadows or tractor tracks in the background leading to an increase in the proportion of false positives in certain microplots of the Tartas site (Figure 11b). Training the super-resolution model with a larger dataset might help the generator network to limit those artifacts. Alternatively, some studies [39], [71], [72] have proposed to integrate the training of the super-resolution model with the training of the Faster-RCNN. The use of a combined detection loss would provide additional information on the location of the plants, thus forcing the super-resolution network to differentiate between plants and background while up-sampling the images.

## 4. Conclusion

We evaluated the performances of automatic maize plant detection from UAV images using deep learning methods. Our results show that the Faster-RCNN model achieved very good plant detection and counting (rRMSE=0.08) performances when high-resolution images (GSD≈0.3cm) are used both for training and validation. However, when this model is applied to the low-resolution images acquired at higher altitudes, the detection and counting performances degrade drastically with rRMSE=0.48. We demonstrated that this was mostly due to the hyper-specialization of Faster-RCNN that is expecting plants of similar size as in the training dataset. The sensitivity of the detection method to the object size is a critical issue for plant phenotyping applications, where datasets can be generated from different platforms (UAVs, ground vehicles, portable imaging systems, etc.) each one of them providing images within at a specific ground resolution. Concurrently, it would be optimal to share labeled images to get a wide training dataset. Data augmentation techniques where high and low-resolution images populate the training dataset were proved to be efficient and provides performances similar to the ones achieved when the model is trained and validated over the same image resolution. However, the native low-resolution images acquired from the UAV have significant low quality that prevents accurate plant detection. In some cases, the images are difficult to visually interpret which poses a problem both for their labeling and for the detector to localize plants due to the lack of pertinent information. These low-quality images were characterized by a loss of image texture that could come from camera intrinsic performances, inadequate settings and the jpg recording format. It is thus recommended to pay a great attention to the camera choice, settings and recording format when the UAV is flying at altitudes that provides resolution coarser than 0.3 cm for maize plant counting.

Finally, we evaluated a super-resolution Cycle-ESRGAN based method to partially overcome the problem of sub-optimal image quality. The super-resolution method significantly improved the results on the native low-resolution dataset compared to the classic bicubic up-sampling strategies. However, the performances when applied to the native low-resolution images were moderate and far poorer than those obtained with the native high-resolution images with simulated super-resolved images showing sometimes artifacts. A future direction to reduce the artifacts of such super-resolution algorithms can be to integrate the GAN training along with the training of the plant detection network. Another direction would be to introduce some labeled low-resolution images in the training dataset to possibly integrate their features in model.

## Acknowledgments

**General**: We would like to thank the CAPTE team for contributing to the construction of the labelled dataset used in this study.

**Author contributions:** KV, RL, SM and FB designed the study. KV implemented the super-resolution pipeline and conducted the analysis. KV and RL contributed extensively to the writing of the article. RL and FB supervised the study. AC and WG participated in discussions and in writing the manuscript. All authors read, revised and approved the final manuscript.

**Funding:** We received support from ANRT for the CIFRE grant of KV, co-funded by Hiphen.

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

## References

[1] R. F. Holt and D. R. Timmons, "Influence of Precipitation, Soil Water, and Plant Population Interactions on Corn Grain Yields," *Agron. J.*, vol. 60, no. 4, pp. 379–381, Jul. 1968.

[2] W. Xu *et al.*, "Adjusting maize plant density to different climatic conditions across a large longitudinal distance in China," *F. Crop. Res.*, vol. 212, pp. 126–134, Oct. 2017.

[3] Y. Zhao, S. Xing, Q. Zhang, F. Zhang, and W. Ma, "Causes of maize density loss in farmers' fields in Northeast China," *J. Integr. Agric.*, vol. 18, no. 8, pp. 1680–1689, Aug. 2019.

[4] Y. Gan, E. H. Stobbe, and J. Moes, "Relative Date of Wheat Seedling Emergence and Its Impact on Grain Yield," *Crop Sci.*, vol. 32, no. 5, pp. 1275–1281, Sep. 1992.

[5] J. S. Graybill, W. J. Cox, and D. J. Otis, "Yield and Quality of Forage Maize as Influenced by Hybrid, Planting Date, and Plant Density," *Agron. J.*, vol. 83, no. 3, pp. 559–564, May 1991.

[6] R. A. Fischer, O. H. Moreno Ramos, I. Ortiz Monasterio, and K. D. Sayre, "Yield response to plant density, row spacing and raised beds in low latitude spring wheat with ample soil resources: An update," *F. Crop. Res.*, vol. 232, pp. 95–105, Feb. 2019.

[7] G. . Maddonni, M. . Otegui, and A. . Cirilo, "Plant population density, row spacing and hybrid effects on maize canopy architecture and light attenuation," *F. Crop. Res.*, vol. 71, no. 3, pp. 183–193, Jul. 2001.

[8] L. Li, J. Sun, F. Zhang, X. Li, S. Yang, and Z. Rengel, "Wheat/maize or wheat/soybean strip intercropping I. Yield advantage and interspecific interactions on nutrients," *F. Crop. Res.*, vol. 71, no. 2, pp. 123–137, Jun. 2001.

[9] J. Olsen, L. Kristensen, and J. Weiner, "Influence of sowing density and spatial pattern of spring wheat (Triticum aestivum) on the suppression of different weed species," *Weed Biol. Manag.*, vol. 6, no. 3, pp. 165–173, Sep. 2006.

[10] D. B. Egli and M. Rucker, "Seed Vigor and the Uniformity of Emergence of Corn Seedlings," *Crop Sci.*, vol. 52, no. 6, pp. 2774–2782, Nov. 2012.

[11] N. W. Hopper, J. R. Overholt, and J. R. Martin, "Effect of Cultivar, Temperature and Seed Size on the Germination and Emergence of Soya Beans (Glycine max (L.) Merr.)," *Ann. Bot.*, vol. 44, no. 3, pp. 301–308, Sep. 1979.

[12] P. V. V. Prasad, K. J. Boote, J. M. G. Thomas, L. H. Allen, and D. W. Gorbet, "Influence of Soil Temperature on Seedling Emergence and Early Growth of Peanut Cultivars in Field Conditions," *J. Agron. Crop Sci.*, vol. 192, no. 3, pp. 168–177, Jun.

2006.
[13] Z. Berzsenyi and I. S. Tokatlidis, “Density Dependence Rather Than Maturity Determines Hybrid Selection in Dryland Maize Production,” *Agron. J.*, vol. 104, no. 2, pp. 331–336, Mar. 2012.
[14] D. Deery, J. Jimenez-Berni, H. Jones, X. Sirault, and R. Furbank, “Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping,” *Agronomy*, vol. 4, no. 3, pp. 349–379, 2014.
[15] S. Madec *et al.*, “High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates,” *Front. Plant Sci.*, vol. 8, no. November, pp. 1–14, Nov. 2017.
[16] A. Ruckelshausen *et al.*, “BoniRob–an autonomous field robot platform for individual plant phenotyping,” *Precis. Agric.*, vol. 9, no. 841, p. 1, 2009.
[17] Y. Shi *et al.*, “Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research,” *PLoS One*, vol. 11, no. 7, p. e0159781, Jul. 2016.
[18] G. Yang *et al.*, “Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives,” *Front. Plant Sci.*, vol. 8, p. 1111, Jun. 2017.
[19] X. Jin, S. Liu, F. Baret, M. Hemerlé, and A. Comar, “Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery,” *Remote Sens. Environ.*, vol. 198, pp. 105–114, Sep. 2017.
[20] E. David *et al.*, “Plant detection and counting from high-resolution RGB images acquired from UAVs: comparison between deep-learning and handcrafted methods with application to maize, sugar beet, and sunflower crops,” *Submitt. to Front. Plant Sci.*, 2021.
[21] D. Stroppiana *et al.*, “ESTIMATING CROP DENSITY FROM MULTI-SPECTRAL UAV IMAGERY IN MAIZE CROP,” *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-2/W13, pp. 619–624, Jun. 2019.
[22] P. Ranđelović *et al.*, “Prediction of Soybean Plant Density Using a Machine Learning Model and Vegetation Indices Extracted from RGB Images Taken with a UAV,” *Agronomy*, vol. 10, no. 8, p. 1108, Jul. 2020.
[23] B. Li *et al.*, “The estimation of crop emergence in potatoes by UAV RGB imagery,” *Plant Methods*, vol. 15, no. 1, p. 15, Feb. 2019.
[24] D. S. Shrestha and B. L. Steward, “Shape and size analysis of corn plant canopies for plant population and spacing sensing,” *Appl. Eng. Agric.*, vol. 21, no. 2, pp. 295–303, 2005.
[25] S. Liu, F. Baret, B. Andrieu, P. Burger, and M. Hemmerlé, “Estimation of Wheat Plant Density at Early Stages Using High Resolution Imagery,” *Front. Plant Sci.*, vol. 8, p. 739, May 2017.
[26] A. Karami, M. Crawford, and E. J. Delp, “Automatic Plant Counting and Location Based on a Few-Shot Learning Technique,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 5872–5886, 2020.
[27] B. T. Kitano, C. C. T. Mendes, A. R. Geus, H. C. Oliveira, and J. R. Souza, “Corn Plant Counting Using Deep Learning and UAV Images,” *IEEE Geosci. Remote Sens. Lett.*, pp. 1–5, Aug. 2019.
[28] J. Ribera, Y. Chen, C. Boomsma, and E. J. Delp, “Counting plants using deep learning,” in *2017 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2017 - Proceedings*, 2018, vol. 2018-January, pp. 1344–1348.
[29] M. Z. Alom *et al.*, “A State-of-the-Art Survey on Deep Learning Theory and Architectures,” *Electronics*, vol. 8, no. 3, p. 292, Mar. 2019.
[30] E. David *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, pp. 1–12, Aug. 2020.
[31] A. Dobrescu, M. V. Giuffrida, and S. A. Tsaftaris, "Leveraging Multiple Datasets for Deep Leaf Counting," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2072–2079.
[32] S. Ghosal *et al.*, "A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting," *Plant Phenomics*, vol. 2019, p. 1525874, Jun. 2019.
[33] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv Prepr. arXiv1511.08060*, Nov. 2015.
[34] M. Minervini, A. Fischbach, H. Scharr, and S. A. Tsaftaris, "Finely-grained annotated datasets for image-based plant phenotyping," *Pattern Recognit. Lett.*, vol. 81, pp. 80–89, Oct. 2016.
[35] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," *arXiv Prepr. arXiv1902.07296*, Feb. 2019.
[36] K. Tong, Y. Wu, and F. Zhou, "Recent advances in small object detection based on deep learning: A review," *Image Vis. Comput.*, vol. 97, p. 103910, May 2020.
[37] D. Dai, Y. Wang, Y. Chen, and L. Van Gool, "Is Image Super-resolution Helpful for Other Vision Tasks?," Sep. 2015.
[38] M. Fromm, M. Berrendorf, E. Faerman, Y. Chen, and B. Sch, "XD-STOD : Cross-Domain Superresolution for Tiny Object Detection," pp. 142–148, 2019.
[39] V. Magoulianitis, D. Ataloglou, A. Dimou, D. Zarpalas, and P. Daras, "Does deep super-resolution enhance UAV detection?," *2019 16th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2019*, pp. 1–6, 2019.
[40] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[41] A. Lugmayr, M. Danelljan, and R. Timofte, "Unsupervised Learning for Real-World Super-Resolution," Sep. 2019.
[42] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11133 LNCS, pp. 63–79.
[43] K. Zhang, W. Zuo, and L. Zhang, "Learning a Single Convolutional Super-Resolution Network for Multiple Degradations," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3262–3271, Dec. 2017.
[44] M. Fritsche, S. Gu, and R. Timofte, "Frequency Separation for Real-World Super-Resolution," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 3599–3608.
[45] Tzutalin, "LabelImg," *Git code*, 2015. [Online]. Available: https://github.com/tzutalin/labelImg.
[46] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[47] X. Jin, S. Madec, D. Dutartre, B. de Solan, A. Comar, and F. Baret, "High-Throughput Measurements of Stem Characteristics to Estimate Ear Density and Above-Ground Biomass," *Plant Phenomics*, vol. 2019, p. 4820305, May 2019.
[48] Y. Liu, C. Cen, Y. Che, R. Ke, Y. Ma, and Y. Ma, "Detection of Maize Tassels from UAV RGB Imagery with Faster R-CNN," *Remote Sens.*, vol. 12, no. 2, p. 338, Jan. 2020.
[49] S. Madec *et al.*, "Ear density estimation from high resolution RGB imagery using deep learning technique," *Agric. For. Meteorol.*, vol. 264, no. May 2018, pp. 225–234, Jan.

2019.
[50] K. Chen *et al.*, “MMDetection: Open MMLab Detection Toolbox and Benchmark,” *arXiv*, Jun. 2019.
[51] J. Shermeyer and A. Van Etten, “The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery,” 2018.
[52] A. Clark, “Pillow (PIL Fork) Documentation.” https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf, 2015.
[53] Z. Han *et al.*, “Unsupervised Image Super-Resolution with an Indirect Supervised Path,” Oct. 2019.
[54] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 2242–2251, Mar. 2017.
[55] R. Timofte Eirikur *et al.*, “NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017, pp. 126–135.
[56] F. Chollet, “Keras,” *https://keras.io*. GitHub, 2015.
[57] T. Y. Lin *et al.*, “Microsoft COCO: Common objects in context,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8693 LNCS, no. PART 5, pp. 740–755.
[58] R. Padilla, S. L. Netto, and E. A. B. da Silva, “A Survey on Performance Metrics for Object-Detection Algorithms,” in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, vol. 2020-July, pp. 237–242.
[59] T. Kong, A. Yao, Y. Chen, and F. Sun, “HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 845–853.
[60] H. Jiang and E. Learned-Miller, “Face Detection with the Faster R-CNN,” in *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017 - 1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heteroge*, 2017, pp. 650–657.
[61] S. Zhang, R. Wu, K. Xu, J. Wang, and W. Sun, “R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery,” *Remote Sens.*, vol. 11, no. 6, p. 631, Mar. 2019.
[62] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.
[63] M. Machefer, F. Lemarchand, V. Bonnefond, A. Hitchins, and P. Sidiropoulos, “Mask R-CNN refitting strategy for plant counting and sizing in uav imagery,” *Remote Sens.*, vol. 12, no. 18, p. 3015, Sep. 2020.
[64] A. Comar, P. Burger, B. de Solan, F. Baret, F. Daumard, and J.-F. Hanocq, “A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results,” *Funct. Plant Biol.*, vol. 39, no. 11, p. 914, Nov. 2012.
[65] J. W. White and M. M. Conley, “A Flexible, Low-Cost Cart for Proximal Sensing,” *Crop Sci.*, vol. 53, no. 4, p. 1646, 2013.
[66] G. Quaglia, C. Visconte, L. S. Scimmi, M. Melchiorre, P. Cavallone, and S. Pastorelli, “Design of a UGV Powered by Solar Energy for Precision Agriculture,” *Robotics*, vol. 9, no. 1, p. 13, Mar. 2020.
[67] R. Khanna, J. Rehder, M. Moeller, E. Galceran, and R. Siegwart, “Studying Phenotypic Variability in Crops using a Hand-held Sensor Platform,” 2015.
[68] J. L. Crain *et al.*, “Development and Deployment of a Portable Field Phenotyping Platform,” *Crop Sci.*, vol. 56, no. 3, pp. 965–975, May 2016.
[69] S. Wu *et al.*, “MVS-Pheno: A Portable and Low-Cost Phenotyping Platform for Maize

Shoots Using Multiview Stereo 3D Reconstruction," *Plant Phenomics*, vol. 2020, pp. 1–17, Mar. 2020.
[70] N. Virlet, K. Sabermanesh, P. Sadeghi-Tehran, and M. J. Hawkesford, "Field Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring," *Funct. Plant Biol.*, vol. 44, no. 1, p. 143, Jan. 2017.
[71] M. Haris, G. Shakhnarovich, and N. Ukita, "Task-Driven Super Resolution: Object Detection in Low-resolution Images," Mar. 2018.
[72] H. Ji, Z. Gao, T. Mei, and B. Ramesh, "Vehicle Detection in Remote Sensing Images Leveraging on Simultaneous Super-Resolution," *IEEE Geosci. Remote Sens. Lett.*, pp. 1–5, Aug. 2019.

# 3. Wheat head density estimation from UAVs: data augmentation and data preparation strategies to exploit labelled ground-based imagery

Wheat head density is a direct indicator of yield and hence is extremely useful for plant breeders to evaluate the performance of cultivars under different environmental conditions and management practices. The possibility to access this trait from RGB images acquired in-situ, at 1 or 2 meters above the canopy using object detection methods has already been well established. This had also led to the creation of a large, diverse labelled dataset of wheat heads in the field, covering different stages of maturity and illumination conditions. However, these in-situ image-based methods, are still low throughput and suffer from sampling issues. This study aims to overcome this problem by developing a methodology to estimate wheat head density from UAV images by exploiting the existing high-resolution ground annotations. Since the wheat canopy observed from UAV is characterized by small, overlapping objects with high density per image, the suitability of an object detection method, Faster-RCNN previously used for maize plant counting, and a regression-based object counting method, $SFC^2Net$, was investigated. The models were evaluated on three ground and UAV datasets acquired during the 2020 growing campaign over three sites (two in France and one in Australia). This chapter corresponds to a draft article.

# Exploiting ground-level images for training deep models to estimate wheat head density from UAV images

Kaaviya Velumani, Raul Lopez Lozano, Hao Lu, Simon Madec, Etienne David, Wenjuan Li, Shouyang Liu, Daniel Smith, Scott Chapman, Alexis Comar, Frederic Baret

## Abstract

Wheat is one of the most widely cultivated cereal crops in the world. The wheat head density is an important indicator of yield and a trait of interest for plant breeders as well as farmers. The recent advances in low-cost data acquisition platforms, deep learning algorithms and publication of diverse, open datasets now allows for automated wheat head density estimation from high resolution RGB images acquired from ground-based platforms. Unmanned aerial vehicles (UAV) show great potential to replace these ground-based platforms, providing a higher throughput at a lower cost per $m^2$. This study evaluates the suitability of Global Wheat Head Dataset, a labelled dataset acquired from ground-level platforms, for the estimation of wheat head density from UAV imageries on three new sites. The study also compares the suitability of two deep learning approaches for wheat head density estimation – object detection using Faster-RCNN and regression-based object counting method using SFC$^2$Net. For images taken at the ground level, the GWHD was demonstrated to provide an efficient training database allowing to get performances comparable to those of other studies (9%<rRMSE<16%). We found that the SFC$^2$Net outperforms Faster-RCNN for both ground level and UAV images, with unbiased estimates ($\mathrm{rRMSE} \approx 14\%$ and $\mathrm{rBias} \approx \pm 2\%$) over the three study sites. We further compared the image-based SFC$^2$Net predicted density estimated from the UAV images with the in-situ manual density collected in the fields. We observed a trend of under-estimation, with an overall $\mathrm{rRMSE}=23.2\%$ and $\mathrm{rBias}=-16.3\%$ averaged on the three sites. Particularly higher rates of under-estimation were noted in the site with a heterogenous canopy development, causing the smaller wheat heads to be unseen from UAV images. In addition, it was found that the wheat head density estimates from the UAV images were more repeatable than the in-situ ground density.

## 1 Introduction

Wheat is the third most cultivated crop in the world and is grown under a wide range of climate, soil and management conditions (FAO, 2002). The wheat head density is directly related to the plant population and the number of tillers per plant which in turn are important sub-components of the grain yield (Slafer, Calderini, & Miralles, 1996). It is hence an appealing trait for plant breeders to assess the genetic gains within a target environment (Reynolds, Manes, Izanloo, & Langridge, 2009). However, it is currently a trait tedious to measure by manual counting which is also prone to human and sampling errors.

In the recent years, the advances in low-cost data acquisition platforms and computer vision algorithms have paved the way for automated wheat head density estimation from high resolution RGB images, slowly replacing manual counting in the field. The state-of-the-art results for plant/organ density estimations are currently being reported using deep learning algorithms that promise robust results under operational conditions (Chandra, Desai, Guo, & Balasubramanian, 2020). For the task of wheat head density estimations, there are two main categories of deep learning approaches applicable

– object detection and regression-based object counting. The object detection methods provide localization and identification of all the instances of an object type within the image via bounding boxes. These methods perform well on large-scale objects and are frequently used for wheat head detection from ground images (Gong, Ergu, Cai, & Ma, 2020; Hasan, Chopin, Laga, & Miklavcic, 2018; Khaki, Safaei, Pham, & Wang, 2021; Madec et al., 2019). However, most of these detection-based methods have problems adapting to small objects and are penalized by occlusions and crowded scenes (Gomez, Aptoula, Parsons, & Bosilj, 2021; Tong, Wu, & Zhou, 2020; Zhao, Zheng, Xu, & Wu, 2019) which are typical within a field phenotyping scenario. The regression-based object methods, on the other hand are reportedly robust towards crowded and occluded small objects (Lempitsky & Zisserman, 2010) since they were specifically developed for crowd counting tasks. However, they provide only an approximate localization of objects without their actual sizes and are more susceptible to changes in source-target domain shift (Wang, Han, Gao, & Yuan, 2021).

Nevertheless, the robustness and generalization of these deep learning algorithms is highly dependent on the availability of large and diverse labelled datasets. This has led to the construction of public dataset such as the Global Wheat Head Detection dataset (David et al., 2020) that compiled 11 datasets acquired using different ground-acquisition platforms over different places of the world. This Global Wheat Head Detection (GWHD) dataset contains roughly 4700 images and approximately 190,000 labelled wheat heads with a wide range of genotypes and growth stages from seven countries and has been doubled in 2021[1]. This project had attracted the participation of several researchers and teams worldwide to develop and test their algorithms on the benchmark dataset. The results suggest that deep learning can now reasonably identify wheat heads within high-resolution ground-level images considering that the domain shift between the training and testing domains are not significant (Fourati, Mseddi, & Attia, 2021; Gong et al., 2020; Khaki et al., 2021; Wu, Hu, & Li, 2020). The appearance variation due to illumination conditions, phenological and genotypic differences, presence/absence of awns can be exploited by deep learning networks which when trained over large and diverse datasets such as the GWHD dataset, can extract and combine high- and low-level image features.

The ground-acquisition platforms such as autonomous robots (Deery, Jimenez-Berni, Jones, Sirault, & Furbank, 2014; Quaglia et al., 2020), hand-held poles and carts (Khanna, Rehder, Moeller, Galceran, & Siegwart, 2015; Mahlein, 2010) have lower throughput and could constitute a bottleneck over large phenotyping experiments. Conversely, the introduction of easy-to-use and affordable unmanned aerial vehicles (UAVs) allows image acquisition at a considerably higher throughput and lower cost per microplot. Thus, UAVs show great potential to replace ground-acquisition platforms for several phenotyping traits like height (Holman et al., 2016; Madec et al., 2017), NDVI (Assmann, Kerby, Cunliffe, & Myers-Smith, 2019; W. Li et al., 2021), green area index (Jay, Maupas, Bendoula, & Gorretta, 2017), plant counting (David, Daubige, Joudelat, Burger, Comar, De Solan, et al., 2021; Jin, Liu, Baret, Hemerlé, & Comar, 2017; Karami, Crawford, & Delp, 2020) including wheat head density estimation (Fernandez-Gallego et al., 2020). However, the spatial resolution achievable using standard UAVs with low-cost RGB sensors is significantly lower than that of the ground-based acquisitions. It is thus quite challenging to identify and label the individual wheat heads considering the relatively small size of the wheat spike in the images, occlusions, and the appearance variation. It is thus important to study if and how the existing labelled ground datasets such as the GWHD dataset could be exploited to estimate wheat head density from UAV imagery rather than constructing a new labelled UAV dataset from scratch.

[1] https://www.aicrowd.com/challenges/global-wheat-challenge-2021

It is well known that deep learning algorithms are sensitive to variations in object scale and hence to the image spatial resolution (Fan, Brown, & Smith, 2016; Velumani et al., 2021). Thus, a model trained directly on the native spatial resolution of GWHD would fail when applied to UAV images. It would thus be important to downscale the GWHD dataset using an appropriate downsampling strategy to successfully bridge the source-target domain shift and detect the wheat heads present in the lower resolution UAV images. Thus, the purpose of this paper is to investigate if an algorithm trained on ground datasets could be used to detect and count wheat heads from UAV images. It is based on three new datasets where both ground and UAV level images were concurrently acquired. We will address four main questions: (1) are the models trained on the original GWHD labelled images performing equally on the ground level images of our three new sites? (2) How to transform the high-resolution images for training a deep learning model dedicated to estimate wheat head density from UAV images? (3) Are the performances obtained from UAV comparable to those associated to the high-resolution ground images? (4) How regression methods perform as compared to object identification ones? (5) How UAV derived head density agrees with visual counting in the field and what is the associated heritability?

# 2 Materials and Methods

## 2.1 Dataset Acquisition and Processing

The dataset used in this study was collected from three wheat phenotyping experiments – two of them located in France (Gréoux and Estrées) and one in Australia (Gatton) during the 2020 growing campaign. The experiments included a few hundred of microplots measuring from 12 to 20 square meters and were sown with different cultivars of winter wheat. In each of these sites, RGB images were collected from the ground with hand-held system and from a UAV (Figure 1). Both systems were operated within less than two days on each site, when the wheat was between stage Z75 (milk/dough) and Z88 (maturity) according to Zadok's scale (Zadoks, Chang, & Konzak, 1974).

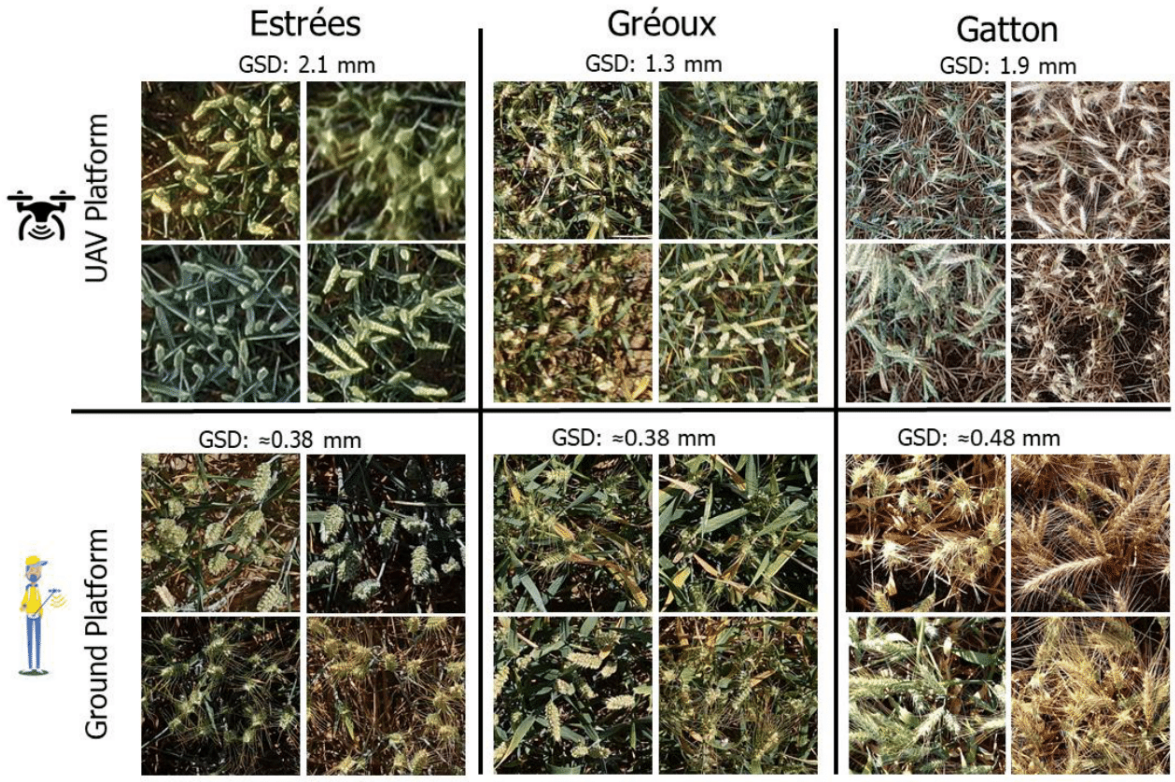


*Figure 1 Extracts from the ground-level and aerial datasets acquired in our three study sites.*

The ground acquisition system (Figure 2) is composed of a hand-held pole equipped with nadir-looking Sony RX0 cameras (https://electronics.sony.com/imaging/compact-cameras/p/dscrx0m2-b) with a one-inch CMOS sensor providing 4800 x 3200 pixels and a field of view of 84° in the diagonal. The operator passed through each microplot to acquire 3 to 4 images from 2.0 to 2.2 m above the ground. Only the center of the image was used, corresponding approximately to a 0.6 m² footprint and an approximate ground sampling distance (GSD) ranging between 0.58 mm and 0.72 mm. It should be noted that this approximate GSD was computed using the height of the camera. Since the height of the canopies within the microplot was not measured, an accurate GSD at the canopy level could not be computed. Hence, the wheat head density from the ground-images could not be accessed since the accurate footprint of these images was not available.

*Figure 2 An operator acquiring images using the ground system consisting in a hand-held pole on the extremity of which the camera was fixed.*

*Table 1 Characteristics of the experiments and UAV image acquisition.*

| Site name | Estrées | Gréoux | Gatton |
|---|---|---|---|
| Latitude (°N) | 50.29 | 43.77 | 27.55 |
| Longitude (°E) | 3.07 | 5.85 | 152.27 |
| Acquisition date | 10/06/2020 | 16/06/2020 | 08/10/2020 |
| Zadoks stage | Z75-80 | Z75-80 | Z77-88 |
| Number of microplots | 218 | 120 | 606 |
| UAV platform | DJI-Mavic 2 Pro | DJI-Matrix 600 | DJI-Phantom4 |
| Camera model on UAV | DJI-L1D-20c | Sony-ILCE-6000 | DJI-FC6310S |
| Image size | 5472 x 3648 | 6000 x 4000 | 5472 x 3648 |
| focal length (mm) | 10.26 | 60 | 8.8 |
| Flight altitude (m) | 8 | 8 | 20 |
| GSD (mm) | 2.1 | 1.3 | 1.9 |

The UAV flights were carried out with different UAV platforms and cameras (Table 1), while the image size was kept about the same. The flight plan was designed to provide a GSD between 1.3 to 2.1 mm with a side and front overlap of 70 to 80%. Gray plastic disks of 60cm diameter were evenly distributed in the field to serve as ground control points (GCP) used to precisely geo-reference the images. The coordinates of the GCPs were measured with a RTK GPS. The microplots were extracted from the UAV

images following the procedure described in Madec et al., (2017) using Agisoft MetaShape Professional (Agisoft LLC, Russia).

## 2.2 In-situ Manual Counting

Wheat head density was measured in the field using standard manual counting protocols. In Gatton, an area of 0.25 $m^2$ was sampled in 18 microplots by placing a frame in the middle of these microplots and counting the number of wheat heads within the frame. In Estrées, the heads were counted in 148 microplots along 2 rows of 1m, which were averaged to represent an area of 0.25 $m^2$ per microplot. Among these 148 microplots sampled, 78 belonged to two repetitions of 8 modalities (3 cultivars x 3 modalities of sowing densities and 6 cultivars x 5 modalities of sowing densities). In Gréoux, 36 microplots (6 cultivars x 3 repetitions x 2 modalities of irrigation management) were sampled using three rows of 1m representing a 0.5 $m^2$ area.

## 2.3 Dataset Labelling

In each of the three experiments, about 30 microplots (Table 2) were randomly selected for labelling both the ground level and UAV images, representing 68 different cultivars. Ground level and UAV images were then prepared differently (Table 2). For each microplot, one of the available ground level images was randomly selected and resized by a factor of 1.5 to get approximately 5700 pixels per head as it was done for the GWHD (David et al., 2020). The center 1024 x 1024 pixels of the rescaled image was then extracted for drawing bounding boxes around each head using Datatorch, an online labelling platform (Brooks, 2020). Within the 92 microplots selected, 4 937 heads with a size between 3 500 and 6 500 pixels were labelled (Table 2). For the UAV, the image with the maximum coverage of the microplot, minimum blur and closest to the nadir direction was selected over each of the 92 microplots. A 620 x 620 pixels patch located in the center of the microplot was cropped for labelling. A total of 36 044 heads with a mean area of 224 pixels to 400 pixels were labelled (Table 2).

*Table 2 Characteristics of the labelled dataset available. The GWHD dataset is described in detail in David et al., (2020)*

| | Ground-level | | | | UAV | | |
|---|---|---|---|---|---|---|---|
| Dataset | GWHD | Estrées | Gréoux | Gatton | Estrées | Gréoux | Gatton |
| GSD (mm) | 0.43 | ≈0.38 | ≈0.38 | ≈0.48 | 2.1 | 1.3 | 1.9 |
| Number of images | 4 698 | 33 | 29 | 30 | 33 | 29 | 30 |
| Number of heads | 184 034 | 1 614 | 1 277 | 2 046 | 20 671 | 6 049 | 9 324 |
| Footprint ($m^2$) | 0.19 | 0.15 | 0.15 | 0.24 | 1.57 | 0.65 | 1.33 |
| Mean size of heads (pixels) | 80 x 80 | 65 x 65 | 62 x 62 | 80 x 80 | 12 x 12 | 18 x 18 | 20 x 20 |
| Mean number of heads per image | 40 | 50 | 44 | 68 | 626 | 209 | 311 |
| Image size (pixels) | 1024 x 1024 | | | | 620 x 620 | | |

## 2.4 Downsampling GWHD images to UAVs resolution

We downsampled the GWHD images by a factor randomly varying between four to six to prepare a low-resolution dataset to train a deep model for application to the UAV images. This factor was computed from the size of the heads in the GWHD and that of the UAV images (Table 2). Two downsampling techniques were used: the standard bicubic interpolation and a gaussian filter followed by motion blur. This later simulates in a more realistic way the degradation of image quality when the camera is flying with the UAV. The characteristics of the Gaussian filter (sigma=2, window size=15) and the motion blur (degree=2, angle=45) were selected after trials and error to get similar visual impression as that provided by the actual UAV images. Thus, during the training phase, a part of the training dataset was randomly selected and downsampled using the above described gaussian-motion

blur filter while the rest were downsampled using the classic bicubic interpolation. Given the large size of the training dataset (approx. 4700 images) and iterative nature of the training phase, this random selection ensures that each image is seen several times by the model under different scaling and downsampling strategy.

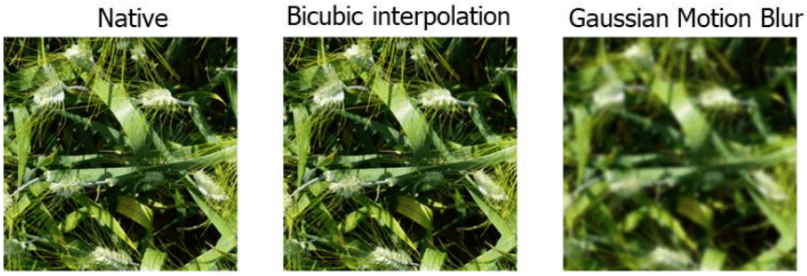


*Figure 3 Example of the native resolution, downsampling using bicubic interpolation and downsampling Gaussian filter followed by motion blur applied to an image from the GWHD dataset.*

## 2.5 The deep models

### 2.5.1 Object detection using Faster-RCNN model

Faster-RCNN (Ren, He, Girshick, & Sun, 2017) is a two-stage object-detection network that detects and localizes multiple instances of objects of interest within an image. The effectiveness of the Faster-RCNN method to successfully localize and count plants and organs (David, Daubige, Joudelat, Burger, Comar, Solan, et al., 2021; Jin et al., 2019) and more specifically for wheat heads (Gomez et al., 2021; Hasan et al., 2018; Madec et al., 2019; Wu et al., 2020) has already been demonstrated. Hence, we selected this method for our comparison study to serve as a benchmark for object-detection methods. A PyTorch implementation of Faster-RCNN available in the open-source MMDetection Toolbox (Chen et al., 2019) was used. We trained the Faster-RCNN model with ResNet50 backbone pretrained on ImageNet (Russakovsky et al., 2015) for 60 epochs, with a batch size of 4. The model was optimized using stochastic gradient descent (SGD) with a learning rate of 0.02 and momentum of 0.9. Data augmentation on brightness, contrast, jpeg compression along with random flip and rotate from the Albumentations library (Buslaev et al., 2020) was used.

The input images were 1024 x 1024 in size when Faster-RCNN was used for the ground level images, with head size in the range of 3500-6500 pixels. Conversely, when the Faster-RCNN was used for the UAV images, the head size in the range of 224-400 pixels (Table 2) was too small for extracting pertinent features required for object detection (Hu, Yang, Hu, Huang, & Han, 2018; Le, Zheng, Zhu, Luu, & Savvides, 2016). For this reason, we rescaled the input images to the UAV model using bicubic interpolation to ensure an average head size in the range of 1600 pixels, with a rescale factor adapted for each site. Finally, the downsampled GWHD images were upsampled to a standard 512 x 512 size. During the test phase, the upscaled UAV images were split into patches of 512 x 512 to be input to the model.

### 2.5.2 Object counting using SFC$^2$Net model

The Scale-Fusion Counting Classification Network, SFC$^2$Net (Liu, Lu, Li, & Cao, 2020), integrates various state-of-the-art computer vision ideas to estimate the number of objects present within an image by splitting it into patches. The network utilizes a multi-layer fusion module accounting for the low-level features and a block-wise classification module to achieve the patch-level counting. Liu et al., (2020) have shown that this method outperforms the recent state-of-the-art counting approaches for the estimation of in-field rice seedlings density. We used the original PyTorch implementation of the model (https://github.com/poppinace/sfc2net). The density maps required for training the SFC$^2$Net model were generated by fitting a 2-dimensional Gaussian function along the height and width of each

annotated bounding box. Around 20% of the training dataset was reserved for train-time validation. We trained the SFC²Net with a MixNet-L backbone with weights pre-trained on ImageNet, for 900 epochs. The model was optimized using stochastic gradient descent (SGD) with an initial learning rate of 0.01, which was decayed by a factor of 0.1 at epochs 250 and 600.

For the ground-level images, patches of 512 x 512 pixels were extracted from the 1024x1024 pixels images and used as input to SFC²Net with a batch size of 8. Conversely, for the UAV images, the training dataset was prepared by downscaling the GWHD dataset by random factors within [4, 6], resulting in an average head size around 256 pixels. The model was trained with a batch size of 16 and by randomly cropping patches of 256 x 256 pixels from the downsampled image. In cases where the downsampled image was smaller than 256 x 256 pixels, the image size was increased by tiling i.e., by repeating the image across the two dimensions. For application to the actual UAV images, the 620 x 620 pixel images were rescaled to get an average head size close to 256 pixels as those in the training dataset and split into patches of 256 x 256 to be input to the model.

## 2.6 Evaluation Metrics

The performances of the methods considered to estimate the head number per image are computed over each of the three sites and evaluated based on the relative root mean square error (rRMSE), relative bias (rBias) and the coefficient of determination, $R^2$.

$$rRMSE = \frac{\sqrt{\frac{\sum_i^n (N_i - \widehat{N_i})^2}{n}}}{\bar{N}} * 100 \quad ; \quad rBias = \sum_i^n \frac{\frac{(\widehat{N_i} - N_i)}{N_i}}{n} * 100 \quad ; \quad R^2 = 1 - \frac{\sum_i^n (N_i - \widehat{N_i})^2}{\sum_i^n (N_i - \bar{N})^2}$$

where $N_i$ is the number of heads labelled on image $i$, $\widehat{N_i}$ is the number of heads predicted by the model on the same image, and $\bar{N}$ is the average number of labelled heads per image for the considered site. Using the relative values for rRMSE and rBias was preferred to limit the possible scaling impact. For the Faster-RCNN model, the number of predicted heads was computed by considering a confidence score > 0.5 as commonly used for large object detection (Kong, Yao, Chen, & Sun, 2016; Padilla, Netto, & da Silva, 2020). For the SFC²Net, the number of predicted heads was calculated by summing up all the pixels within the count map output by the model.

The predicted head density values were used for comparisons with the head density measured in the field by the operators. The same metrics (rRMSE, rBias and $R^2$) adapted to the head density were used in this case. The head density values, $D_i$, were computed from the number of ears per image according to:

$$D_i = \frac{N_i}{GSD_c^2 * S_i}$$

where $N_i$ is the number of heads in image $i$, $S_i$ is the size of image $i$ in pixels, and $GSD_c$ is the ground sampling distance at the canopy-level. $GSD_c$ was derived from the distance between the camera and the height of the canopy as described by Madec et al., (2017). Finally, the broad-sense heritability, $H^2$, was computed using the wheat-head density estimates as follows:

$$H^2 = \frac{V_g}{V_t} \quad , \qquad \text{where } V_t = V_g + V_e + V_r$$

where $V_g$ is the genetic variance, $V_t$ is the total variance, $V_e$ is the environmental or phenotypic variance which could be attributed to the modalities and the repetitions and $V_r$ is the residual variance. This $H^2$ value will hence allow to quantify the repeatability of the proposed image-based head density estimates for the various trials where several replicates of a genotype are available.

# 3 Results and Discussion

## 3.1 Robust training with the GWHD dataset is achieved for the ground-level images

We first verified that the GWHD provides a robust training dataset for wheat head counting at the ground-level. The Faster-RCNN and SFC$^2$Net models trained on the native resolution of the GWHD dataset were therefore evaluated over the corresponding ground-level labelled images available for the three sites. Results show that the performances vary strongly between sites and models. On the Estrées site, good performances are observed for both models, with slightly lower rRMSE for Faster-RCNN (Table 3) and a slight negative bias (Figure 4). The highest errors are observed on the Gatton site (Table 3) particularly with the Faster-RCNN model that shows a noticeable underestimation (Figure 4). This could be attributed to the advanced growth stage at the time of image acquisition (Table 2) with a possible higher rate of overlap because of the more inclined orientation of the heads as compared to the GWHD dataset situation. Further, the Gatton site shows also a significantly higher number of heads per image (Table 2) which was less frequently represented in the GWHD dataset. However, additional observations should be required to verify these possible explanations.

The errors obtained on the ground-level images acquired in the three new sites are slightly higher than in other works reported over the GWHD benchmark test dataset with an rRMSE≈13% (Gomez et al., 2021; Khaki et al., 2021) for object detection methods and an rRMSE≈8% (Gomez et al., 2021) using regression methods. Since the results here are reported over a wild dataset, further tests on the GWHD benchmark dataset are required to validate the performances of the two methods.

*Table 3 Performances (rRMSE, Bias) of the Faster-RCNN and the SFC$^2$Net models over the ground-level dataset acquired in the 2020 campaign. The models were trained over the GWHD dataset. The metrics were computed over the three test sites along with their mean values. The color indicates the goodness of the metrics (dark green very good; dark red, very poor).*

| Sites | Faster-RCNN | | | SFC2Net | | |
|---|---|---|---|---|---|---|
| | rRMSE | rBias | $R^2$ | rRMSE | rBias | $R^2$ |
| **Estrées** | **9.61** | -6.53 | **0.78** | 10.54 | **0.59** | 0.72 |
| **Gréoux** | 19.24 | -15.56 | -0.13 | **12.75** | **1.88** | **0.56** |
| **Gatton** | 22.04 | -16.10 | 0.71 | **15.78** | **4.91** | **0.86** |
| **Overall** | 19.66 | -12.50 | 0.78 | **14.52** | **2.41** | **0.89** |

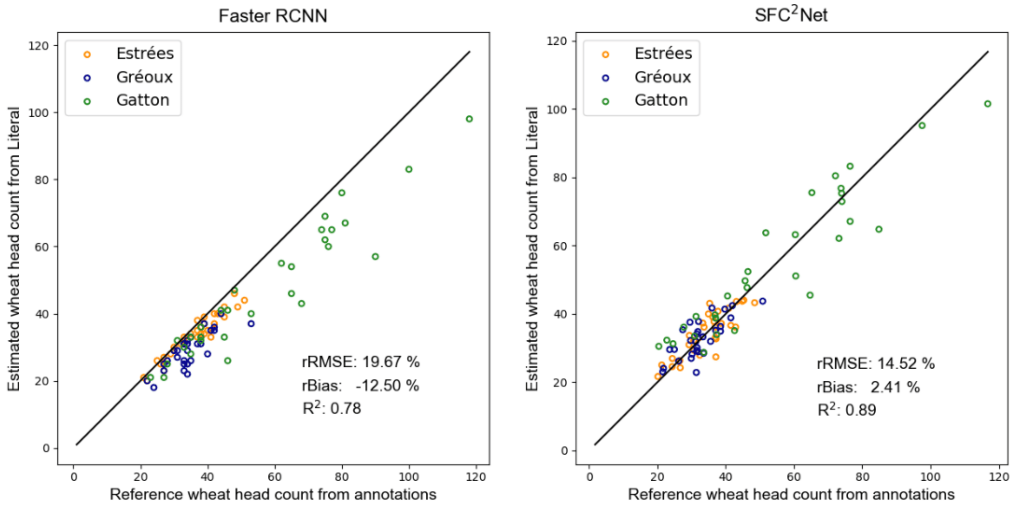


*Figure 4 Comparison between wheat head count estimated by Faster-RCNN (left) and SFC$^2$-Net (right) trained with the GWHD dataset with the counts of the labeled ground-level images from the three test sites represented by a distinct color. The solid line is the 1:1 line.*

## 3.2 The Gaussian-Motion Blur is an efficient data augmentation strategy for UAV images

The performances of bicubic and gaussian filter followed by motion blur (called Gaussian-blur in the following) data augmentation techniques allowing to transform the ground level images into UAV level ones were evaluated for the Faster-RCNN and SFC$^2$Net deep models. We compare the model performances when trained on two versions of low-resolution dataset – one prepared by using only bicubic interpolation and the second prepared by using a combination of Gaussian-blur and bicubic interpolation as explained in Section 2.4. The models were trained using the training configurations mentioned in Sections 2.5.1 and 2.5.2. The performance was evaluated over the UAV images as summarized in Table 4.

*Table 4 Impact of the downsampling technique (bicubic interpolation or Gaussion filter and motion blur (Gaussian blur) on the performances (rRMSE, rBias and R²) of Faster-RCNN and SFC2Net models to estimate the number of heads per image in the three sites.*

| Sites | Faster-RCNN | | | | | | SFC$^2$Net | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bicubic | | | Gaussian Blur | | | Bicubic | | | Gaussian Blur | | |
| | *rRMSE* | *rBias* | $R^2$ | *rRMSE* | *rBias* | $R^2$ | *rRMSE* | *rBias* | $R^2$ | *rRMSE* | *rBias* | $R^2$ |
| **Estrées** | 18.9 | -16.5 | -0.33 | **13.6** | **-10.5** | **0.31** | 24.4 | -20.3 | -1.02 | **10.8** | **-1.4** | **0.60** |
| **Gréoux** | 10.0 | -6.4 | 0.47 | **7.3** | **-0.89** | **0.72** | 26.9 | -24.8 | -2.7 | **10.1** | **4.5** | **0.48** |
| **Gatton** | 32.5 | -27.0 | -0.83 | **27.2** | **-20.5** | **-0.29** | 46.5 | -43.5 | -2.49 | **21.5** | **-11.0** | **0.25** |
| **Overall** | 23.6 | -16.7 | 0.79 | **18.0** | **-10.7** | **0.88** | 32.4 | -29.3 | 0.59 | **14.4** | **-1.7** | **0.92** |

We observe that using a combination of the gaussian-motion blur and bicubic interpolation improves the overall generalization ability of the models to the UAV dataset, achieving significantly better results with consistent performances over the three sites. Especially, in the case of SFC$^2$Net, the gaussian-blur strategy improves the average error rates (both $rRMSE$ and $rBias$) by ≈20% compared to using only the bicubic downsampling approach. From this, we may infer that the SFC$^2$Net is more sensitive towards the domain shift between the training and the inference datasets. On the other hand, the Faster-RCNN model shows an overall improvement of 6% ($rRMSE$ and rBias) on using the gaussian-motion blur data augmentation strategy. This is in line with the observations of Jiang & Wang, (2016) on improving the performance of object-detection models to low quality video frames. However, Faster-RCNN's performances without the addition of gaussian-blur is significantly better compared to those of SFC$^2$Net, suggesting that the former is more robust towards domain shift.

Hence, it was decided to include the gaussian-blur downsampling within the data augmentation strategy to downsample the GWHD dataset to the spatial resolution of the UAV dataset. Moreover, this experiment highlights the importance of appropriate data augmentation strategies to bridge the domain shift between the ground-level training images and the inference images from aerial platforms to achieve optimal model performances.

## 3.3 SFC$^2$Net performs better than Faster-RCNN both for ground-level and UAV images

The performances of the object detection and object counting models trained on the downsampled GWHD dataset were evaluated on the annotated UAV images from 92 microplots in the three test sites. SFC$^2$Net provides more robust and generally lower rRMSE and rBias as compared to the Faster-RCNN model (Table 5 Table 3and Figure 5). However, poorer performances are observed for the Gréoux site (Table 5 Table 3and Figure 5). Table 3Table 3The general superiority of regression methods

for head counting problem may be partly explained by the presence of heads fully or partly masked by other ones, making the identification more complex, while the regression-based methods may partly learn to account for these artefacts. A study comparing the performances of object-detection and regression-based counting methods for phenotyping activities reported a similar drop in performances of object-detection based methods when the number of objects per image increases (Gomez et al., 2021).

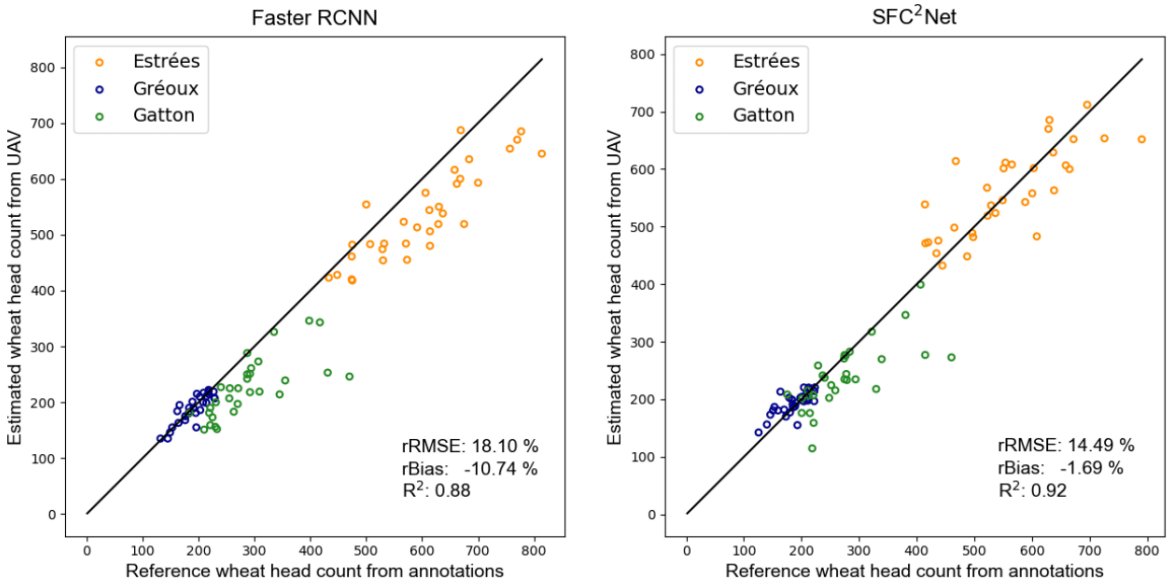


*Figure 5. Comparison between wheat head count estimated by Faster-RCNN (left) and SFC²-Net (right) trained with the downsampled GWHD images with the counts of the labeled UAV-level images from the three test sites represented by a distinct color. The solid line is the 1:1 line.*

*Table 5 Performance (rRMSE and Bias) of the Faster-RCNN and the SFC²Net models over the UAV dataset*

| Sites | Faster-RCNN | | | SFC²Net | | |
|---|---|---|---|---|---|---|
| | Gaussian Blur | | | Gaussian Blur | | |
| | *rRMSE* | *rBias* | $R^2$ | *rRMSE* | *rBias* | $R^2$ |
| **Estrées** | 13.6 | -10.5 | 0.31 | **10.8** | **1.4** | **0.60** |
| **Gréoux** | **7.3** | **-0.89** | **0.72** | 10.1 | 4.5 | 0.48 |
| **Gatton** | 27.2 | -20.5 | -0.29 | **21.5** | **-11.0** | **0.25** |
| **Overall** | 18.10 | -10.7 | 0.88 | **14.4** | **-1.7** | **0.92** |

We then evaluated if the above reported results obtained over the UAV dataset is comparable to the ground level results.

The performance of the Faster-RCNN and SFC²Net models are consistent over the UAV and the ground-level datasets acquired in the 2020 campaign, with the SFC²Net outperforming the Faster-RCNN model in both the scenarios. In addition, it is observed that the Faster-RCNN tends to systematically under-estimate the number of wheat heads in both the UAV and the ground-level images. We may hence infer that the diversity of the GWHD dataset, combined with the right data augmentation strategy, allows the models to generalize considerably well to the domains of the new sites. Further, the estimations from the SFC²Net are unbiased, with an overall $\mathrm{rRMSE} \approx 14\%$ in both scenarios. Hence, we may conclude that it is possible to acquire wheat head counts from UAV images, with almost the same

accuracy as from the ground-level images. This would substantially increase the throughput and lower the cost per microplot involved with accessing wheat head density trait.

## 3.4 Comparing the wheat head density estimated from UAV images to reference density from in-situ manual counting

We used here the SFC²Net model that was demonstrated to perform better than Faster-RCNN. The estimated density was compared with the in-situ density obtained from manual counting over the three sites. Results show that the agreement is loose with an overall rRMSE=23.1% (Table 6). These poor performances are mainly explained by a large scatter of the points and a negative bias that is particularly high for Gréoux (Figure 6a and Table 6). The bias is not expected to come from the SFC²Net model since this was verified previously (Table 5 and Figure 5). This is further confirmed in Table 6 and Figure 6b where a similar underestimation of the measured in-situ head density is observed with the labeled UAV images. However, the reduced object size and image quality leads to wheat heads not visible or identifiable in the UAV images owing to occlusion, shadow, and small size (smaller wheat heads represented by fewer pixels, resulting in their poor visibility). This was confirmed by our observations in Gatton where the in-situ sampling was done within frames positioned in the middle of the microplots, which were also visible in the UAV images. On comparing the wheat heads counted within the frame directly in the field with that of the annotations made on the UAV images, we found a negative bias of approximately 12.5% due to wheat heads not clearly visible in the UAV images. Moreover, in Gréoux, due to heavy rains during the sowing season and subsequent delay in sowing, the crop development was heterogenous. It was noted during the in-situ counting that there were two strata within the canopy with small, under-developed wheat heads on the shorter plants. Such wheat heads would typically be occluded by the taller plants and would not be visible in the UAV images, thus being missed by the models and the human labelers. This would explain the higher rate of under-estimations in Gréoux in the order of ≈-30% (Table 6).

Finally, the variability associated with the representativeness of the samples used for the in-situ density measurement as compared to the exhaustive UAV spatial sampling may also explain the scatter observed in Figure 6. To better understand the influence of the spatial heterogeneity on the in-situ sampling, we replicated the in-situ sampling protocols using the full-microplot density map predicted by SFC²Net on the UAV images. Random sub-samples were extracted from this density map and a corresponding density was computed from these sample counts. While repeating this exercise, we observed that such computed densities varied with a relative standard deviation of up to 8% among the different samples. This indicates that the in-situ density estimations would in fact be affected by their limited sampling (0.5 $m^2$ in Gréoux and 0.25 $m^2$ in Estrées and Gatton) and the choice of the sampling location. This might explain a part of the deviation observed between the image-based densities and the in-situ manual density (Figure 6 a and b).

*Table 6 Comparison of the image-based densities against the in-situ density: the SFC²Net predictions were evaluated over 202 microplots whereas the image-based annotations were evaluated on 48 microplots.*

| Sites | SFC²Net (UAV) | | | | | Image annotations (UAV) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample area at canopy ($m^2$) | Nb. micro plots | *rRMSE* | *rBias* | $R^2$ | Sample area at canopy ($m^2$) | Nb. micro plots | *rRMSE* | *rBias* | $R^2$ |
| **Estrées** | 6.4 | 148 | 19.4 | -13.5 | -0.37 | 1.6 | 22 | 14.1 | -3.9 | -2.37 |
| **Gréoux** | 8.1 | 36 | 34.4 | -29.7 | -4.01 | 0.6 | 9 | 36.3 | -31.9 | -6.51 |
| **Gatton** | 4.5 | 18 | 22.4 | -16.2 | 0.15 | 1.3 | 17 | 23.7 | -7.4 | 0.09 |
| **Overall** | - | 202 | 23.19 | -16.34 | -0.30 | - | 48 | 23.11 | -10.39 | 0.18 |

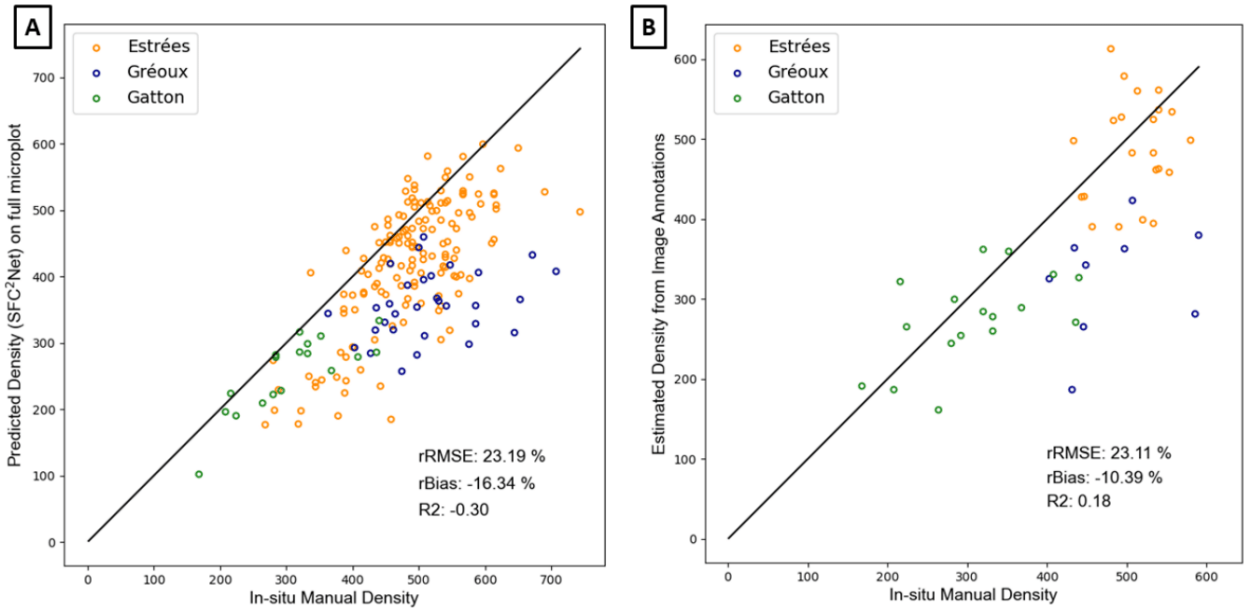


*Figure 6 Scatter plots comparing the manual in-situ density against the image-based density in the three sites. A) by in situ-counting in the ground to the estimated density from the wheat head count from the SFC²Net method. B) In-situ manual density vs image-based density estimated from the labelled annotations on 48 microplots.*

Thus, we conclude that the under-estimations in the model predicted densities is mainly due to the reduced visibility of heads in the UAV images, with a lower but still significant contribution to the scatter from the limited sampling of the in-situ densities.

## 3.5 Image-based estimates are more heritable than the ground-based ones

The broad sense heritability of the wheat head density was then computed for the Gréoux and the Estrées sites, where we had in-situ manual counting over all the repetitions in the modalities. Gatton was not included because we had in-situ sampling on only one of the repetitions. It was found that the wheat head density estimated from the UAV had significantly higher heritability, than the in-situ density measurements for both the sites (Table 7). This indicates that the wheat head density derived from UAV images is more repeatable than manual measurements and might have higher potential to interpret the interaction of a genotype to the environment.

*Table 7 Comparison of the broad-sense heritability of the wheat head density derived from the in-situ manual counting and the UAV images using SFC²Net, in two sites.*

| Site | # Modalities | Reps | $H^2$ Manual | $H^2$ UAV-SFC²Net |
|---|---|---|---|---|
| **Estrées** | 39 | 2 | 0.62 | **0.91** |
| **Gréoux** | 10 | 3 | 0.45 | **0.53** |

# 4 Conclusion

This study evaluated the suitability of Global Wheat Head Dataset, a labelled dataset acquired from ground-level platforms, for the estimation of wheat head density from UAV imageries on three new sites. For images taken at the ground level, the GWHD was demonstrated to provide an efficient training database allowing to get performances comparable to those of other studies (9%<rRMSE<16%). We found that the SFC²Net outperforms Faster-RCNN for both ground level and UAV images, with unbiased estimates (rRMSE ≈14% and rBias ≈ ±2%) over the three study sites. This suggests that object counting algorithms might be more robust for wheat head counting since this

method directly estimates the count whereas, in case of the detection methods, the count is derived from the identified objects. It was also found to be significantly faster than the Faster-RCNN in terms of inference time. However, the SFC$^2$Net model trained without data augmentation strategies tended to have poor performances over the low-resolution UAV dataset whereas there were no significant differences for Faster-RCNN. This suggests that counting by regression and density-based approaches is more sensitive to domain shift. Appropriate domain adaptation/data augmentation strategies, such as the Gaussian-motion blur used in this study should then be employed to ensure optimal performance of the object-counting methods. More advanced domain adaptation strategies can also be explored (Zhuang et al., 2019).

However, one drawback of such density-map based object counting approaches is that they do not allow to access the size of the identified objects. For applications where the accurate localization of the object and its size are important, a possible alternative would be to explore techniques that allow a guided object detection through a predicted density map (C. Li, Yang, Zhu, Chen, & Guan, 2020; Lian, Li, Zheng, Luo, & Gao, 2019). This would ensure optimal performances of the detection model by splitting a high-density image into equal density sub-patches and rescaling the patches containing small objects.

We further compared the image-based SFC$^2$Net predicted density estimated from the UAV images with the in-situ manual density collected in the fields. We observed a trend of under-estimation, with an overall $\mathrm{rRMSE}$=23.2% and $\mathrm{rBias}$=-16.3% averaged on the three sites. Particularly higher rates of under-estimation were noted in the site with a heterogenous canopy development, causing the smaller wheat heads to be unseen from UAV images. In addition, it was found that the wheat head density estimates from the UAV images were more repeatable than the in-situ ground density estimated.

The continuous increase in the availability of labelled dataset, the advancements in low-cost high-resolution sensors offering higher spatial resolution, with in-built GPS RTK offer a good avenue for wheat head counting by UAV. They are able to successfully leverage the trade-off between the spatial resolution and sampling area, by acquiring multiple overlapping images at high throughput. It was demonstrated that the performances of the wheat head counting methods on the UAV dataset is comparable to that on the ground-level dataset. Even though, the wheat head density estimated from the UAV images suffer from an under-estimation, they were found to be more heritable than the in-situ density measurements. We also show an efficient strategy to leverage on existing ground-image annotations which could open the door to other such applications of bridging existing multi-platform and multi-scale datasets, thus saving a considerable amount of labelling time and cost.

# 5 References


Assmann, J. J., Kerby, J. T., Cunliffe, A. M., & Myers-Smith, I. H. (2019). Vegetation monitoring using multispectral sensors — best practices and lessons learned from high latitudes. *Journal of Unmanned Vehicle Systems*, *7*(1), 54–75. https://doi.org/10.1139/juvs-2018-0018

Brooks, J. (2020). *Datatorch.io*. Retrieved from https://github.com/datatorch/datatorch

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information*, *11*(2). https://doi.org/10.3390/info11020125

Chandra, A. L., Desai, S. V., Guo, W., & Balasubramanian, V. N. (2020). Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey. *ArXiv Preprint ArXiv:2006.11391*, 1–26.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., … Lin, D. (2019). MMDetection: Open MMLab Detection Toolbox and Benchmark. *ArXiv*. Retrieved from http://arxiv.org/abs/1906.07155

David, E., Daubige, G., Joudelat, F., Burger, P., Comar, A., De Solan, B., & Baret, F. (2021). Plant detection and counting from high-resolution RGB images acquired from UAVs: comparison between deep-learning and handcrafted methods with application to maize, sugar beet, and sunflower crops. *BioRxiv*, 2021.04.27.441631. https://doi.org/10.1101/2021.04.27.441631

David, E., Daubige, G., Joudelat, F., Burger, P., Comar, A., Solan, B. de, & Baret, F. (2021). Plant detection and counting from high-resolution RGB images acquired from UAVs: comparison between deep-learning and handcrafted methods with application to maize, sugar beet, and sunflower crops. *Submitted to Frontiers in Plant Science*.

David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., … 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 Phenomics*, *2020*, 1–12. https://doi.org/10.34133/2020/3521852

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., & Furbank, R. (2014). Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. *Agronomy*, *4*(3), 349–379. https://doi.org/10.3390/agronomy4030349

Fan, Q., Brown, L., & Smith, J. (2016). A closer look at Faster R-CNN for vehicle detection. *2016 IEEE Intelligent Vehicles Symposium (IV)*, *2016-Augus*, 124–129. https://doi.org/10.1109/IVS.2016.7535375

FAO. (2002). *World Agriculture: Towards 2015/2030: Summary report* (J. Bruinsma, Ed.). Rome: Food and Agriculture Organization of the United Nations (FAO).

Fernandez-Gallego, J. A., Lootens, P., Borra-Serrano, I., Derycke, V., Haesaert, G., Roldán-Ruiz, I., … Kefauver, S. C. (2020). Automatic wheat ear counting using machine learning based on RGB UAV imagery. *The Plant Journal*, *103*(4), 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. *Canadian Journal of Remote Sensing*, 1–13. https://doi.org/10.1080/07038992.2021.1906213

Gomez, A. S., Aptoula, E., Parsons, S., & Bosilj, P. (2021). Deep Regression Versus Detection for Counting in Robotic Phenotyping. *IEEE Robotics and Automation Letters*, *6*(2), 2902–2907. https://doi.org/10.1109/LRA.2021.3062586

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), 100. https://doi.org/10.1186/s13007-018-0366-8

Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., & Hawkesford, M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sensing*, *8*(12). https://doi.org/10.3390/rs8121031

Hu, G. X., Yang, Z., Hu, L., Huang, L., & Han, J. M. (2018). Small Object Detection with Multiscale Features. *International Journal of Digital Multimedia Broadcasting*, *2018*, 1–10. https://doi.org/10.1155/2018/4546896

Jay, S., Maupas, F., Bendoula, R., & Gorretta, N. (2017). Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Research*, *210*, 33–46. https://doi.org/10.1016/j.fcr.2017.05.005

Jiang, H., & Wang, S. (2016). Object Detection and Counting with Low Quality Videos. *Technical Report*.

Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment*, *198*, 105–114. https://doi.org/10.1016/j.rse.2017.06.007

Jin, X., Madec, S., Dutartre, D., de Solan, B., Comar, A., & Baret, F. (2019). High-Throughput Measurements of Stem Characteristics to Estimate Ear Density and Above-Ground Biomass. *Plant Phenomics*, *2019*, 1–10. https://doi.org/10.34133/2019/4820305

Karami, A., Crawford, M., & Delp, E. J. (2020). Automatic Plant Counting and Location Based on a Few-Shot Learning Technique. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5872–5886. https://doi.org/10.1109/JSTARS.2020.3025790

Khaki, S., Safaei, N., Pham, H., & Wang, L. (2021). WheatNet: A Lightweight Convolutional Neural Network for High-throughput Image-based Wheat Head Detection and Counting. *ArXiv Preprint*. Retrieved from http://arxiv.org/abs/2103.09408

Khanna, R., Rehder, J., Moeller, M., Galceran, E., & Siegwart, R. (2015). Studying Phenotypic Variability in Crops using a Hand-held Sensor Platform. *IROS Work, Agri-Food Robot*.

Kong, T., Yao, A., Chen, Y., & Sun, F. (2016). HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 845–853.

Le, T. H. N., Zheng, Y., Zhu, C., Luu, K., & Savvides, M. (2016). Multiple Scale Faster-RCNN Approach to Driver's Cell-Phone Usage and Hands on Steering Wheel Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 46–53. https://doi.org/10.1109/CVPRW.2016.13

Lempitsky, V., & Zisserman, A. (2010). Learning To Count Objects in Images. *Advances in Neural Information Processing Systems*, *23*, 1324--1332.

Li, C., Yang, T., Zhu, S., Chen, C., & Guan, S. (2020). Density Map Guided Object Detection in Aerial Images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 190–191. Retrieved from http://arxiv.org/abs/2004.05520

Li, W., Jiang, J., Weiss, M., Madec, S., Tison, F., Philippe, B., … Baret, F. (2021). Impact of the reproductive organs on crop BRDF as observed from a UAV. *Remote Sensing of Environment*, *259*, 112433. https://doi.org/10.1016/j.rse.2021.112433

Lian, D., Li, J., Zheng, J., Luo, W., & Gao, S. (2019). Density Map Regression Guided Detection Network for RGB-D Crowd Counting and Localization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1821–1830.

Liu, L., Lu, H., Li, Y., & Cao, Z. (2020). High-Throughput Rice Density Estimation from Transplantation to Tillering Stages Using Deep Networks. *Plant Phenomics*, *2020*, 1–14. https://doi.org/10.34133/2020/1375957

Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., … Comar, A. (2017). High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Frontiers in Plant Science*, *8*(November), 1–14. https://doi.org/10.3389/fpls.2017.02002

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., … Baret, F. (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

Mahlein, A.-K. (2010). Detection, identification, and quantification of fungal diseases of sugar beet leaves using imaging and non-imaging hyperspectral techniques (Rheinischen Friedrich-Wilhelms-Universität). Retrieved from http://hss.ulb.uni-bonn.de/2011/2428/2428a.pdf

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, *2020-July*, 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Quaglia, G., Visconte, C., Scimmi, L. S., Melchiorre, M., Cavallone, P., & Pastorelli, S. (2020). Design of a UGV Powered by Solar Energy for Precision Agriculture. *Robotics*, *9*(1), 13. https://doi.org/10.3390/robotics9010013

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Reynolds, M., Manes, Y., Izanloo, A., & Langridge, P. (2009). Phenotyping approaches for physiological breeding and gene discovery in wheat. *Annals of Applied Biology*, *155*(3), 309–320. https://doi.org/10.1111/j.1744-7348.2009.00351.x

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Slafer, G. A., Calderini, D. F., & Miralles, D. J. (1996). Yield components and compensation in wheat: opportunities for further increasing yield potential. In *Increasing yield potential in wheat: Breaking the Barriers* (pp. 101–133). Mexico, DF: CIMMYT.

Tong, K., Wu, Y., & Zhou, F. (2020). Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, *97*, 103910. https://doi.org/10.1016/j.imavis.2020.103910

Velumani, K., Lopez-Lozano, R., Madec, S., Guo, W., Gillet, J., Comar, A., & Baret, F. (2021). Estimates of maize plant density from UAV RGB images using Faster-RCNN detection model: impact of the spatial resolution. *Submitted to Plant Phenomics*.

Wang, Q., Han, T., Gao, J., & Yuan, Y. (2021). Neuron Linear Transformation: Modeling the Domain Shift for Crowd Counting. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13. https://doi.org/10.1109/TNNLS.2021.3051371

Wu, Y., Hu, Y., & Li, L. (2020). BTWD: Bag of Tricks for Wheat Detection. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: *Vol. 12540 LNCS* (pp. 450–460). https://doi.org/10.1007/978-3-030-65414-6_31

Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, *14*(6), 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., … He, Q. (2019). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, *109*(1), 43–76. Retrieved from http://arxiv.org/abs/1911.02685

# 4. An automatic method based on daily in situ images and deep learning to date wheat heading stage

The monitoring of crop phenology, which is essential for various stakeholders in agriculture, is a labor-intensive task requiring frequent revisits to the field. The rapid development in IoT (internet of things) sensors now permits to continuously monitor the field, providing image, weather and soil data in near real-time. This study evaluates the potential of such IoT sensors to facilitate phenology monitoring. We focus on the onset of wheat heading date, an important developmental stage to understand the genotype's reaction to seasonal changes which is also important to maximize the yield. Owing to the generalization ability of deep-learning algorithms, their suitability to identify wheat heads within the IoT time-series images covering a range of cultivars, crop maturity and diverse illumination conditions is tested. The heading dates estimated by the proposed methodology were then compared to reference heading dates from visual scoring by experts. Due to the lower resolution and more occlusions between heads induced by the 45° view orientation, an image classification method was applied over small patches of the image to decide whether they had at least one visible head. The dynamics of this simplified head counts was then exploited to estimate the heading date. The study was conducted on daily images collected over three growing campaigns in a total of 47 sites sown with winter wheat spread across France. This chapter corresponds to an article published in the July 2020 issue of Field Crop Research (https://doi.org/10.1016/j.fcr.2020.107793).

Field Crops Research 252 (2020) 107793

Contents lists available at ScienceDirect

Field Crops Research

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

# An automatic method based on daily *in situ* images and deep learning to date wheat heading stage

Kaaviya Velumani[a,b,*], Simon Madec[b], Benoit de Solan[c], Raul Lopez-Lozano[b], Jocelyn Gillet[a], Jeremy Labrosse[a], Stephane Jezequel[c], Alexis Comar[a], Frédéric Baret[b]

[a] Hiphen SAS, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9, France
[b] INRAE, Avignon Université, UMR EMMAH, UMT CAPTE, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9, France
[c] Arvalis, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9, France

ARTICLE INFO

*Keywords:*
Phenology
Internet of Things for Agriculture
Convolutional neural networks
Field sensors
Phenology modelling

ABSTRACT

Accurate and timely observations of wheat phenology and, particularly, of heading date are instrumental for many scientific and technical domains such as wheat ecophysiology, crop breeding, crop management or precision agriculture. Visual annotation of the heading date *in situ* is a labour-intensive task that may become prohibitive in scientific and technical activities where high-throughput is needed. This study presents an automatic method to estimate wheat heading date from a series of daily images acquired by a fixed RGB camera in the field. A convolutional neural network (CNN) is trained to identify the presence of spikes in small patches. The heading date is then estimated from the dynamics of the spike presence in the patches over time. The method is applied and validated over a large set of 47 experimental sites located in different regions in France, covering three years with nine wheat cultivars. Results show that our method provides good estimates of the heading dates with a root mean square error close to 2 days when compared to the visual scoring from experts. It outperforms the predictions of a phenological model based on the ARCWHEAT crop model calibrated for our local conditions. The potentials and limits of the proposed methodology towards a possible operational implementation in agronomic applications and decision support systems are finally further discussed.

## 1. Introduction

Phenological observations are essential in agronomy, as crop management strategies (irrigation, fertilizing or crop protection) are planned considering plant development (Brown et al., 2005; Chmielewski, 2013). In wheat, the heading stage is one of the critical developmental phases as the plant becomes highly sensitive to abiotic stress – heat stress, frost, water constraints – with a strong impact on yield components (Slafer and Rawson, 1994). Several studies have quantified the effect of post-heading abiotic stress on yield (Ferris et al., 1998; Gooding et al., 2003; Wheeler et al., 1996) due to a significant decrease in the grain weight, along with a fall in the number of grains per plant. More recently, Balla et al. (2019) analysed the possible impact of temperature on yield components at different development stages over a large wheat genotype panel, highlighting the high sensitivity of grain number to heat stress episodes around heading. Also, García et al. (2015) and Lobell and Ortiz-Monasterio (2007) have shown that the increase in minimum temperature during nights owing to changing climatic conditions accelerates the rate of crop development and reduces yield.

The timing and duration of stress through wheat developmental phases are thus essential to understand the impact of environmental factors on yield (Sadras and Slafer, 2012). Wheat phenology is driven by several eco-physiological mechanisms involving the response to temperature, photoperiod and vernalization (Gate, 1995) that are regulated by complex genetic pathways (Guedira et al., 2016; Whittal et al., 2018). Phenology has been traditionally one of the most important traits used to genetically improve wheat adaptation, matching crop development – particularly reproductive and grain-filling phases – to the optimal growing conditions of a target environment (Foulkes et al., 2011; Slafer, 2012). Indeed, phenology constitutes, in wheat breeding programs, one of the main levers enabling to optimize assimilates partitioning while reducing the impact of adverse weather events such as heat stress and frost during the grain-filling stage (Camargo et al., 2016; Chapman et al., 2012; Reynolds et al., 2009).

Accurate and timely observations of wheat phenology and, particularly, of heading date are, therefore, instrumental for many scientific and technical domains such as wheat ecophysiology, crop breeding,

* Corresponding author at: Hiphen SAS, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9, France.
*E-mail addresses:* kvelumani@hiphen-plant.com, kaaviya.velumani@inrae.fr (K. Velumani).

https://doi.org/10.1016/j.fcr.2020.107793
Received 18 December 2019; Received in revised form 25 March 2020; Accepted 27 March 2020

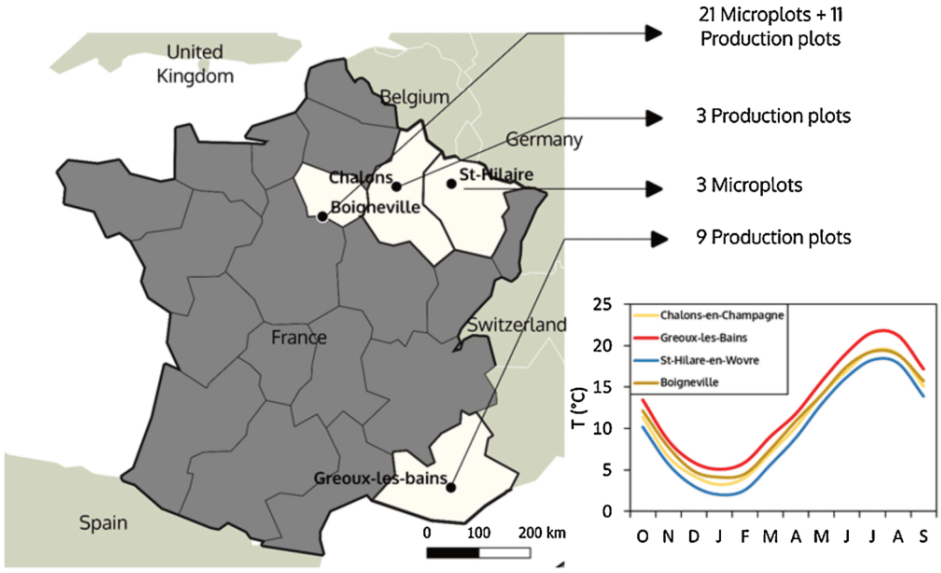


**Fig. 1.** Location of the 4 regions where field sensors installed during the 2017–2019 growing seasons. The graph indicates the average monthly temperatures in the period 1986–2018 observed in 4 near weather stations from INRA and Meteo france: St. Gilles (Gréoux-les-Bains), Gif-sur-Yvette (Boigneville, only since 1999), Erneville (St-Hilare-en-Wovre), and Fagnieres (Chalons-en-Champagne).

crop management or precision agriculture. Heading date is commonly scored visually *in situ* by operators that are frequently surveying the crops. It constitutes a labour-intensive task that requires skilled experts. For certain applications like high-throughput phenotyping in crop breeding programs on large genotype panels (Araus and Cairns, 2014; Cabrera-Bosquet et al., 2012) an accurate visual annotation of the heading date –with experts frequently visiting the field– can be difficult to achieve as the number of microplots increase geometrically. In other applications oriented towards crop management, phenological models may represent a valid alternative to *in situ* phenology observations (Bogard et al., 2014; White et al., 2008; Zheng et al., 2012), provided that those models are calibrated for each specific genotypes. However, such genotype-specific calibration necessitates extensive phenology observations from field experiments (Wallach et al., 2019).

The recent development of field sensors and unmanned platforms with imaging capabilities –including aerial (UAVs) and ground vehicles (UGVs)– have opened new avenues to monitor automatically crops in near-real time (Baret et al., 2018; Comar et al., 2012; Jay et al., 2017; Madec et al., 2017; White and Conley, 2013; Yang et al., 2017). In parallel, the advances in computational resources and data science achieved in the last years have fostered a significant breakthrough in computer vision and has paved the way to implement advanced algorithms to extract relevant information from high spatial resolution imagery. Deep learning algorithms including convolutional neural network (CNN), have shown excellent performances for object recognition (LeCun et al., 2015). These capabilities have favoured their progressive adoption in the fields of agronomy and phenomics (Kamilaris and Prenafeta-Boldú, 2018; Singh et al., 2018). For instance, they have been successfully used to detect and count individual cereal heads from RGB images (Hasan et al., 2018; Lu et al., 2017; Madec et al., 2019) and LiDAR data (Malambo et al., 2019). Nevertheless, the potential of such algorithms to provide accurate, automatic *in situ* estimations of crop phenology remains, up to now, underexploited. The use of well-known capabilities of CNNs to detect plant organs on individual images to derive crop phenology from image series needs to be further explored. At the time of writing this article, only very few studies have attempted to do that. The work of Yalcin (2018) proposing a CNN-based method for the discrimination of phenological stages for several crops, including wheat, seems promising, but no results on the absolute accuracy of the method have been provided. More recently Desai et al. (2019) have developed a deep learning approach to estimate the heading date on rice, but it was only tested over a small number of situations, preventing from drawing general conclusions about its performance under operational conditions.

The development of operational methods for automatic heading date detection using *in situ* images would enable to increase enormously the acquisition throughput at a reasonable cost. This would represent an important contribution to the afore-mentioned scientific and technical domains, where frequent phenology observations are critical.

In that context, the objective of this study is to present an automatic method to estimate wheat heading date from daily high-resolution images taken in the field from fixed cameras. The method developed rely on deep learning techniques: daily images are interpreted using a CNN classifier that detects the presence of wheat spikes in the image, and the dynamics of the presence of spikes along the season permits to determine the heading date. A major aspect that differentiates this study against the existing works on this subject is the use of an extensive dataset of observations used to validate the method proposed, which enables to discuss its possible implementation under operational conditions. The dataset comprises 47 field plots sown with several soft and durum wheat cultivars in different regions of France, where fixed cameras were installed, and actual heading dates were annotated by experts. The performance of our CNN-based method to estimate the actual heading date are also concurrently compared with those of a phenology model based on ARCWHEAT (Weir et al., 1984) and calibrated for local conditions. The robustness of our method and its potential operational implementation are discussed with emphasis on possible limitations–*e.g.* image quality issues, environmental conditions during image acquisition– that may impact the performances.

## 2. Materials and methods

### 2.1. Study sites

This study was conducted during the years 2017–2019 in different commercial and experimental fields belonging to four contrasted agro-climatic regions around the following cities: Gréoux-les-Bains (43.8 °N, 5.9 °E) in the south-east of France, Boigneville (48.3 °N 2.4 °E) in the center of France; Chalons-en-Champagne (49.0 °N, 4.4 °E) and Saint-Hilaire-en-Woëvre (49.1 °N, 5.7 °E) in the north-east of France (Fig. 1).

The climate in Greoux-les-Bains is Mediterranean (Kottek et al., 2006), with a maximum average temperature of 20 °C, 690 mm of rainfall. In Boigneville, the climate is temperate and humid, with a maximum average temperature of 15.3 °C over the year and rainfall of 677 mm (Meteo France). The climate in Chalons-en-Champagne is similar to that of Boigneville, whereas in Saint Hilaire-en-Wovre

**Table 1**
Description of the distribution of the 47 sites across years and regions. The number of sites available each year per region and year is indicated along with the corresponding number of cultivars.

| Locations | 2017 | | 2018 | | 2019 | |
|---|---|---|---|---|---|---|
| | Sites | Cultivars | Sites | Cultivars | Sites | Cultivars |
| Gréoux-les-Bains | – | – | 7 | 3 | 2 | 2 |
| Boigneville | 8 | 1 | 12 | 3 | 12 | 3 |
| Chalons en Champagne | – | – | – | – | 3 | 1 |
| Saint_Hilaire en Woëvre | – | – | – | – | 3 | 1 |

conditions are slightly colder, especially during winter (Fig. 1), and more humid (average precipitation close to 1000 mm/year).

Among these four sites, 24 field sensors equipped with RGB cameras (see 2.2) were installed in microplots with a size of 10 × 2 m belonging to larger experimental fields. The remaining 23 sensors were installed in production plots, with a size similar to a commercial field (around 800 × 200 m). Often, the production plots are subdivided in homogeneous units with different cultivars or agro management. In the sites at the north of France, production plots and microplots are sown with winter soft wheat (*Triticum aestivum*) cultivars Descartes, Oregrain, Fructidor, RGT Sacramento, Matheo and Rebelde. In *Greoux-les-Bains*, the winter durum wheat (*Triticum durum*) cultivars RGT Voilur, Anvergur and Toscadou were grown. A summary of the 47 sites considered is given in Table 1. Please refer to Table A1 in Appendix A for a detailed description of the sites, their location and average temperature over the growing season. The field sensors were installed in relatively homogenous areas of the fields which provided daily information over a footprint of about 10 $m^2$ (see 2.2). During the installation, special attention was paid to orientate the camera field-of-view towards the centre of the microplot to prevent possible border effects.

### 2.2. Acquisition of daily images of the canopy with IoTA systems

The field observation systems installed in the 47 sites were developed by Bosch and Hiphen (www.hiphen-plant.com/our-solutions/iot-field-sensor) and are named IoTA (Internet of Things for Agriculture). They consist of a telescopic pole placed vertically and equipped with an RGB camera as well as meteorological sensors (Fig. 2a). The RGB camera takes one image (Fig. 2b) each day at solar noon, and automatically uploads it to a cloud storage system through a GSM network. The image dimensions are 1024 × 768 pixels and is recorded in PNG format. The camera was set up at a height of approximately 1 m above the top of the crop (installed after emergence), with a 45° inclination angle oriented in a compass direction perpendicular to the row. In some of the fields of 2017 and 2018, the length of the pole was adjusted mid-growing campaign to ensure that the camera was always well above the plant canopy. In the 2019 campaign, the height of the pole was fixed at 1.5 m to avoid this mid-campaign intervention. It has a field of view (FOV) of 55° × 41° providing a footprint of 10.8 $m^2$. Because of the relatively large FOV, the ground resolution is non-uniform throughout the image, particularly in the vertical direction (Fig. 2b).

### 2.3. Heading date determined by experts

The heading stage corresponds to the emergence of the developing spike from the flag leaf sheath (Bonnett, 1936; Zadoks et al., 1974). In the field, the heading stage is identified according to the definition given by Zadoks et al. (1974): 50 % stems with spikes, at least, half-emerged, corresponding to the phase code 54 of the Zadoks scale. For the 27 sites monitored in 2017 and 2018 (Table 1), an online questionnaire was prepared with IoTA images covering a 9–13 consecutive days period approximately centered on the heading date. A panel of 14 experts was asked to identify the reference heading date for all the sites by applying the definition of Zadoks et al. (1974) for heading stage: 50 % stems with spikes at least half-emerged. The experts could view the images at their full resolution. Eight of the experts had more than 10 years of experience in working with wheat phenology and only four of them had less than two years of experience. The reference heading date for each of the 27 sites was considered as the average date of all the 14 experts. The standard deviation of the heading dates for each site was also calculated to quantify the variability of the expert replies.

In 2019, the actual heading dates were determined *in situ* by experts on all the 20 available sites (Table 1). They followed the definition by Zadoks et al. (1974) and visited the fields every two or three days. Note that this reference heading date might be different from the one derived from the 2017 to 2018 questionnaire since experts were scoring the heading date from images, not from direct observation of the crop in the field.

### 2.4. Heading date estimation from IoTA images and CNN

#### 2.4.1. Image preparation

To get a more uniform ground pixel size, each image was first

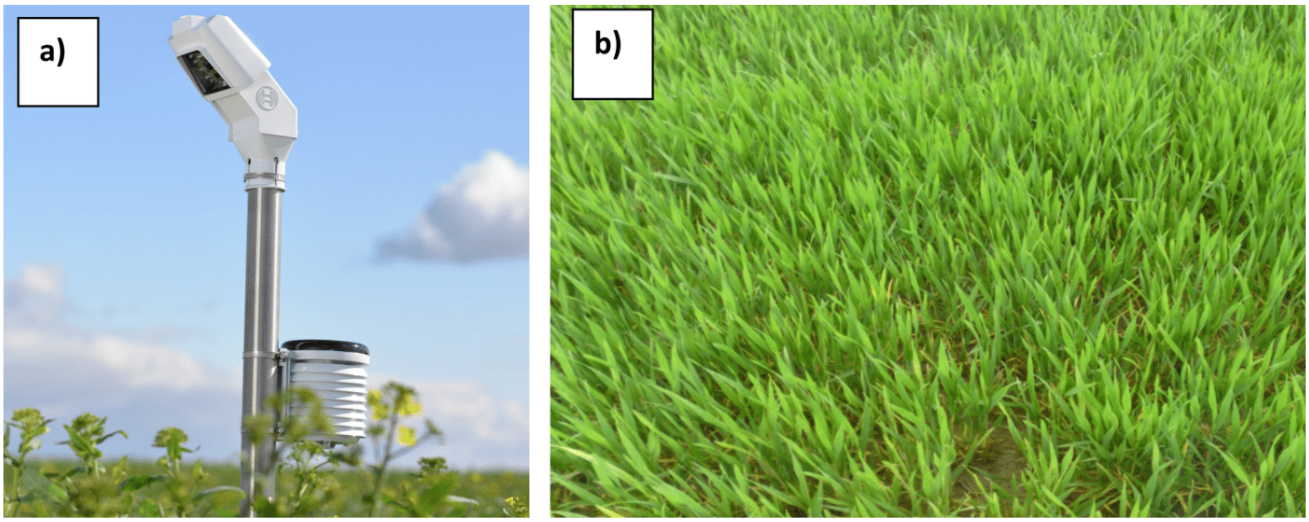


**Fig. 2.** a) The IoTA systems installed in a field. The top part inclined at 45° hosts the RGB camera. The cylinder attached to the vertical pole includes sensors to monitor the local air temperature and moisture. b) An example of a raw image.

cropped into 1024 × 384 pixels by removing the top region of the image where the crop-sensor distance is too large, resulting in blurred objects (Fig. 2b). Then 14 overlapping patches of size 256 × 256 pixels were extracted from the cropped image. The overlap between patches was 50 % in either vertical or horizontal directions to minimize possible border effects. Working with patches permits to benefit from the full resolution of the IoTA image while avoiding memory issues.

#### 2.4.2. Spikes labelling

The patches from the 27 sites observed during the 2017 and 2018 growing seasons were labelled into two classes: "spikes present" or "spikes absent". This represents a total of 40,500 patches out of which 17,000 were labelled as "spikes present" and 23,500 as "spikes absent". All the patches belonging to images acquired until five days before the actual heading date determined by experts were automatically assigned to the "spikes absent" class. Similarly, the patches from the images acquired from five days after the actual heading date onwards were assigned to the "spikes present" class. Therefore, only those patches within a window of ± 5 days around the actual heading date were visually attributed to their respective classes. Few patches with unclear assignation, such as emerging and sparse spikes were excluded from the training dataset. A few examples belonging to the two classes can be found in Fig. 3.

#### 2.4.3. Identifying the presence of wheat spikes with the ResNet50

Wheat spikes were identified in the images using the ResNet50 (He et al., 2016), which obtained the best results in object detection at the ImageNet Large Scale Visual Recognition Challenge 2015 (Russakovsky et al., 2015). This network has a depth of 50 layers and uses residual blocks with identity mappings. The ResNet50 pre-trained on the ImageNet dataset which is available in the Keras Python deep learning library (Chollet, 2015) was used.

We replaced the original top layers of the pre-trained ResNet50 by two fully connected layers with dimensions, respectively of 512 and 1 to build a binary classifier (spikes present/absent). The network was re-trained with the labelled image patches by fine-tuning the weights of the entire network with a low learning rate to identify only the high-level features which were relevant to detect spikes and classify the patches as "spikes present" or "spikes absent". This strategy, called transfer learning, performs generally better than training the full network from scratch (Lee et al., 2015; Tajbakhsh et al., 2016). The training dataset comprises the 27 sites available in 2017 and 2018. First-order data augmentation was applied to the patches: translation, rotation, zoom, flip and changes in brightness levels. This improves the generalization capacity of the neural network and increases the size of the training set at marginal cost. Two different training and validation schemes were followed:

- **Twofold-cross validation on 2017 and 2018 sites**: the 27 sites available in 2017 and 2018 were divided randomly into folds of 13 and 14 sites with approximately 19,500 and 21,000 labelled patches in each fold. ResNet50 was re-trained on each of the two folds independently and validated with the other one.
- **Independent validation on 2019 sites**: In this second scheme, a unique set of the 27 sites corresponding to 40,500 patches from the 2017 and 2018 sites is used for fine-tuning ResNet50. The purpose of this scheme is to mimic the operational conditions when CNNs are trained with data from the previous years: the CNN is trained on data acquired in years different from those used for the validation.

In both schemes, 20 % of the data were held back for performance testing at the end of each epoch of the ResNet50 re-training. The "binary cross entropy loss" available in the Keras library was used as the loss function. To avoid over-fitting, we reduced the learning rate by a factor of 0.5 when the validation mean absolute error did not improve after three consecutive epochs and stopped the training when the validation mean absolute error did not improve after five consecutive epochs. The pertinence of the ResNet50 network was further evaluated using Gradient-weighted Class Activation Maps (Grad-CAM, developed by Selvaraju et al., 2017). These maps highlight the regions that contribute to the output score using the gradient values input to the final convolutional layer (shown in Figs. 6 and 8).

#### 2.4.4. Heading date estimation

The heading date is determined from the dynamics of the presence of spikes in the images. For each day, $d$, the fraction of patches per image classified by the CNN as 'spikes present', $f_{head}(d)$, is calculated along the growing season (Fig. 4). It provides a fair approximation of the Zadoks definition followed by the experts: the proportion of patches with emerged spikes in an image is a reasonable proxy to the proportion of stems with spikes emerged in the image. Then, a three-parameter logistic function is fitted to the time-series of $f_{head}(d)$ for every site:

$$f_{head}(d) = \frac{L}{1 + e^{-k(d-d_0)}} \quad (1)$$

where $L$ is the maximum value of $f_{head}(d)$, fixed by construction to $L = 1.0$ (Fig. 4). The maximum growth rate, $k$, and $d_0$ are estimated using the Scipy Python package (Jones et al., 2001). Parameter $d_0$ represents the date when 50 % of the patches have spikes.

### 2.5. Heading date from ARCWHEAT model

A version of the phenology module of ARCWHEAT (Weir et al., 1984) adapted to the French local conditions by Gate (1995) was used. It is based on cumulated temperature with the effects of vernalization and photoperiod (Gouache et al., 2012). The model was run with actual

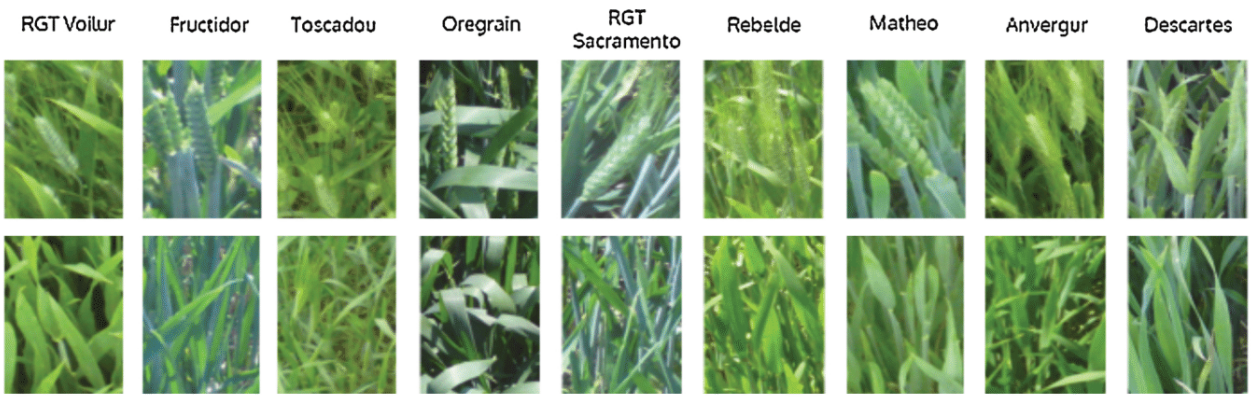


**Fig. 3.** Samples extracted from patches belonging to the 9 wheat varieties monitored in our study. The patches were classified as 'spikes present' (top row) and 'spikes absent' (bottom row).

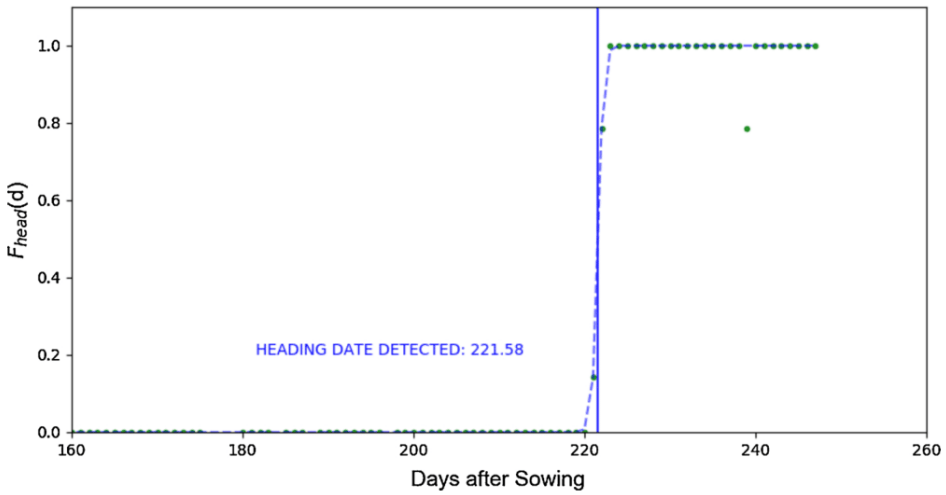


**Fig. 4.** Typical dynamics of the fraction of patches with spikes ($f_{head}$ shown as green dots). The logistic curve fitted to the dynamics of $f_{head}$ is shown using dashed blue line. The heading date estimated from adjusted parameter $d_0$ is represented by the vertical blue line (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

daily temperatures collected by the Meteo France weather stations that are the closest to the sites in the period 2016–2019. Cultivar-specific parameters for vernalization and photoperiod corresponding to the nine cultivars observed in this study were adjusted using optimization algorithms based on independent field experiments (Thépot, 2014).

### 2.6. Performance metrics

The accuracy of the CNN classification was evaluated based on the overall accuracy (OA):

$$OA = \frac{T_p + T_n}{N} \quad (2)$$

where $T_p$ and $T_n$ are, respectively, the number of patches correctly classified as "spikes present" (true positive) and "spikes absent" (true negative); and $N$ is the total number of patches in the test dataset.

The root mean squared error (RMSE) was computed to quantify the errors between estimated and observed heading dates:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (HD_r - HD_e)^2}{N}} \quad (3)$$

where $HD_r$ is the reference heading date obtained from experts, $HD_e$ is the heading date estimated by the indirect method (CNNs or phenology model) and $N$ is the number of sites used.

## 3. Results

### 3.1. Accuracy of heading date estimates from the CNN model

The automatic method proposed estimated the heading date with a RMSE close to 2.0 days (Fig. 5), as compared to the reference dates given by the experts. Moreover, the errors when the CNN is trained and validated with images from years 2017 and 2018 using a twofold cross-validation (Fig. 5a) are similar to those obtained when validating the method against a completely independent set of images from year 2019 (Fig. 5b). Moreover, the coefficient of determination $R^2$ is very close to 1 especially in 2019, where the variability in the time to heading among the plots monitored is large. In that year, heading was observed about 180 days after sowing for two durum wheat plots (varieties Anvergur and RGT Voilur) at Greoux-les-Bains, more than 20 days earlier than both varieties in 2018. This is explained by higher seasonal temperatures in the 2019 season compared to the 2018 season (seasonal average of 13.3 °C in 2019, against 10.1 °C in 2018, see details in the Appendix Table A1).

For two sites at Boigneville monitored in 2018 (Fig. 5a), the CNN estimations show discrepancies reaching up to 6 days with the reference dates from the experts. The dynamics of spike appearance for these two sites are shown in Fig. 6 along with the GradCAM (Gradient Class Activation Maps) which highlights the regions that influence the CNN prediction. In one of the sites (Fig. 6a) the misclassification was due to the poor quality of the images: leaf blades appear bluish because of a poor white balance camera setup that was fixed only after day 203 (Fig. 6a). These quality issues introduced substantial artifacts in the time-series that impacted the logistic curve fit. In the second site (Fig. 6b), the water droplets on leaves observed on days 195 and 196 after sowing were wrongly identified by the CNN as spikes. Although the CNN seems to slightly over-detect the presence of spikes in this site (Fig. 6b) even when no droplets were observed, the errors due to presence of droplets contributed to increase substantially the discrepancies against observed dates. Besides these two specific cases, the issues of CNN misclassification were marginal over the whole dataset. The cross-validation conducted with images from 2017 and 2018 revealed an overall accuracy of 98.45 % for classifying individual patches as spikes present/absent. Moreover, the classification errors observed did not exhibit any systematic bias.

### 3.2. Accuracy of heading date prediction by the ARCWHEAT phenology model

The ARCWHEAT crop model, adapted to French conditions had an RMSE of 4.1 days to predict the heading date for the 47 sites monitored from 2017 to 2019 (Fig. 7). This represents twice the error yielded by our CNN based method. A similar study conducted by Bogard et al. (2014), which evaluated the use of phenology models specifically calibrated for different cultivars to predict heading dates, reported errors comparable to those given by ARCWHEAT in this study. Other works where phenology models are applied at the regional or continental scale with no cultivar-specific calibration show even larger discrepancies reaching up to 20 days (McMaster and Smika, 1988; Ceglar et al., 2019).

Although the ARCWHEAT model is capable to simulate the variability of the heading date among sites and years ($R^2$ = 0.92), it clearly underperformed when compared to the automatic method based on CNN. Only 37 % of the sites were within the ± 1.5 days interval that represent the variability of the heading dates determined by the experts. Moreover the discrepancies between the model predictions and

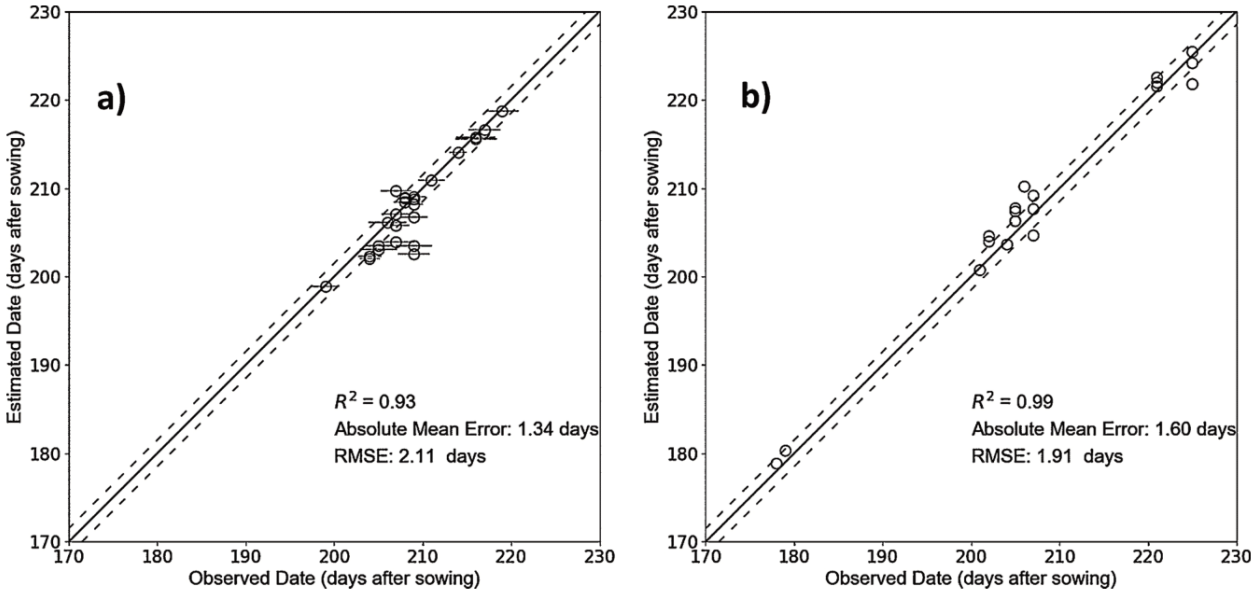


**Fig. 5.** Comparison of reference heading dates and those estimated by our CNN-based method. a) 2017 and 2018, using a twofold cross-validation for years 2017 and 2018; b) independent validation over year 2019. Horizontal error bars represent the standard deviation of the expert annotations for each site in 2017 and 2018. The dotted black lines represent the standard deviation (1.5 days) between experts when identifying the heading dates from photographs.

the observed heading dates were above 10 days for two sites (Fig. 7), which is not acceptable for precision agriculture or phenotyping applications.

## 4. Discussion

### *4.1. Suitability of the automatic CNN-based methodology to estimate wheat heading date in operational applications*

Our proposed CNN based method estimated heading dates with about 2 days uncertainties as evaluated over the 47 sites spanning across several regions, years and cultivars. The performances of the method proposed are considered satisfactory since the errors are close to the standard deviation of the expert panel replies provided for years 2017 and 2018. Moreover, this is close to the expected accuracy of an expert visiting the fields every two or three days to annotate crop development stage. The recent study of Desai et al. (2019) reported estimation errors between one and two days using also a CNN-based approach for paddy rice heading date, but the number of observations and sites were substantially smaller than in our study. Further, among the 20 sites monitored in the 2019 validation dataset, 6 of them were sown with cultivars that were not included in the training set of the CNN, indicating that the method is resistant to possible morphological differences among spikes between cultivars. The robustness of our method is essentially due to three factors: the reliability of the ResNet50 CNN to identify the presence of spikes in the patches; the statistics computed across patches within each image that smooth out individual errors; and the use of a logistic fit to determine the heading date from the daily statistics.

The ResNet50 CNN was used here to classify the patches with spikes. This approach was preferred to directly identifying spikes in the image and counting them (Madec et al., 2019). Using CNN for image

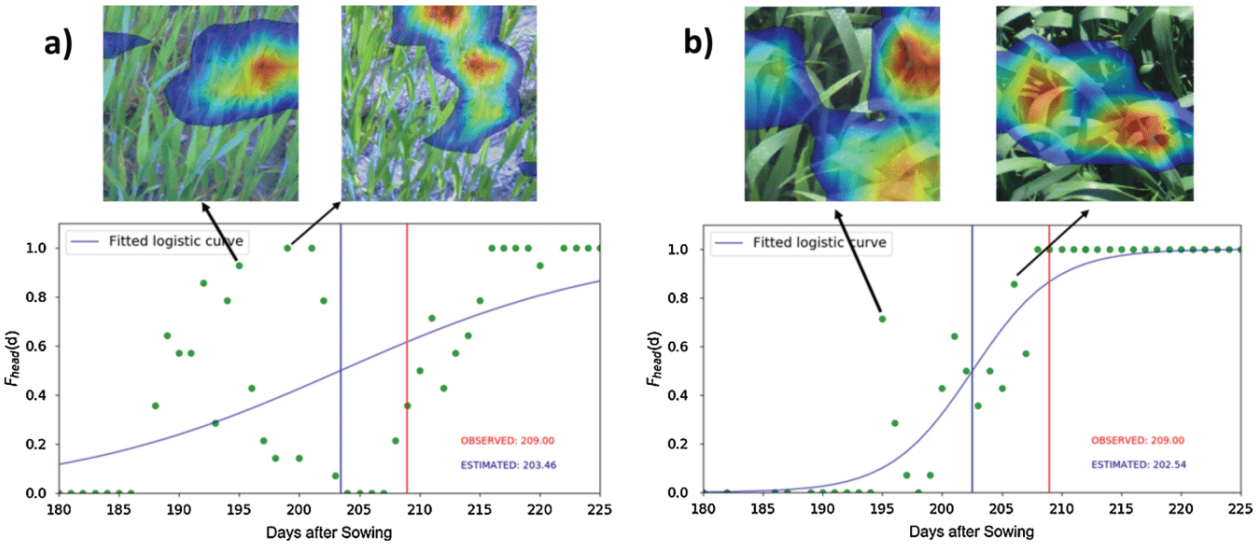


**Fig. 6.** Dynamics of $f_{head}$ for two sites in Boigneville in 2018 where large errors on heading date estimations by the CNN method were observed due to non-optimal image quality (a); image misclassification due to the presence of water droplets (b).

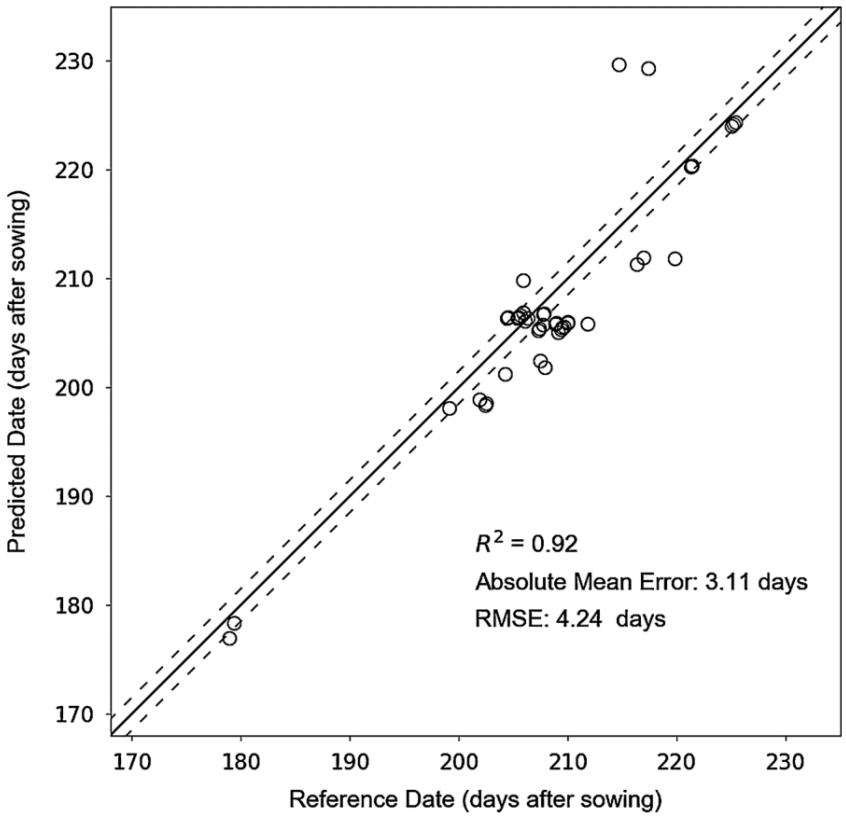


**Fig. 7.** A comparison between the heading date predictions from the crop phenology model with the reference heading date for all the 47 sites in our study from the 3 growing seasons. The dashed lines around the 1:1 line corresponds to the 1.5 days interval. To ensure that all the 47 points are visible on the graph despite the overlap, a negligible random noise was added while plotting the points.

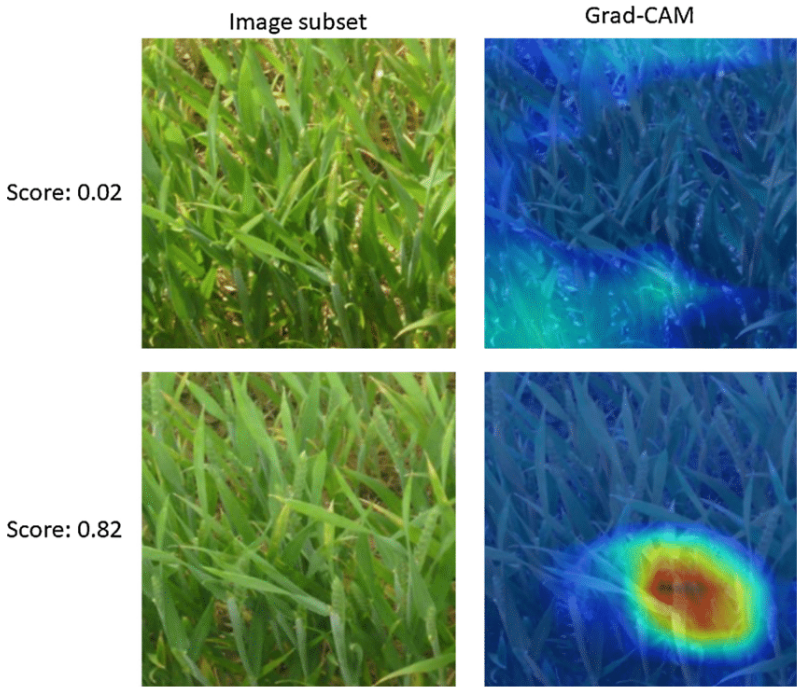


**Fig. 8.** Patches extracted from two images of the same IoTA system and acquired on the reference heading date (top) and one day after (bottom) with their corresponding class activation maps. The confidence score provided by the CNN towards the presence of spikes is shown along with the Grad-CAM heat map which shows the areas which influence the CNN output.

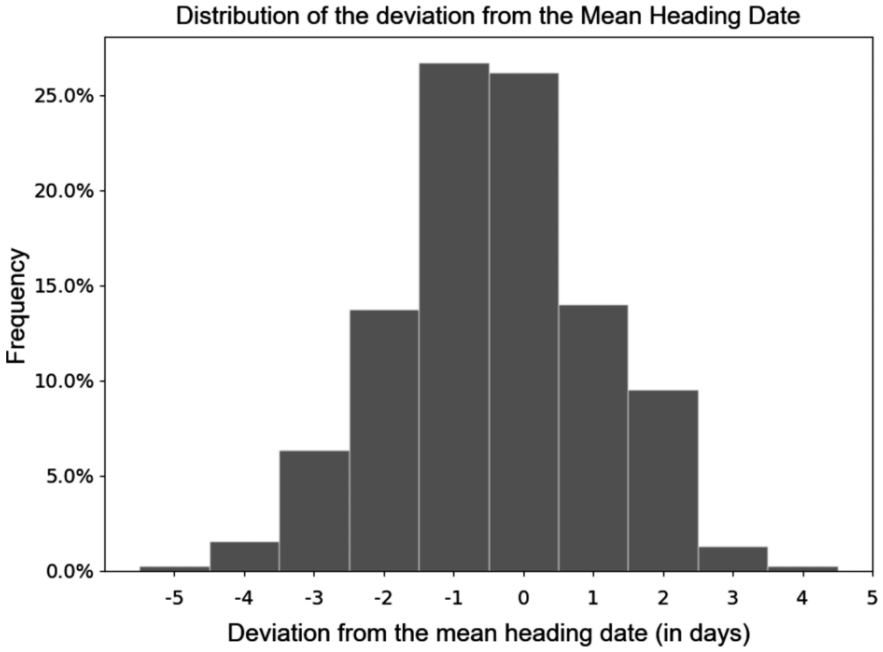


**Fig. 9.** Distribution of the difference in days between the individual heading date identified by the experts and the mean heading date value.

classification increases the efficiency with which the training dataset is generated. Indeed, assigning individual images to the spikes present/absent classes is relatively straightforward: it permitted to generate a training dataset of more than 40,000 patches that contributed to improve the robustness of ResNet50. By contrast, annotating images for spikes identification or spikes counting is time-consuming. Although CNN-based classification provides much less information than object counting/detection algorithms, the results indicate that this approach is sufficient for heading date estimation. In Fig. 8, the first image (top row) spikes are not yet emerged and the Grad-CAM heat map shows low confidence, whereas in the second one (bottom row) the presence of emerged spikes is obvious, and the probability increases up to 0.82. ResNet50 may only detect the presence of a proportion of the spikes present in the patch, but this appears sufficient to correctly classify it. This makes the classification approach more robust for phenology identification as compared to the approaches based on object detection or counting by regression, where the variability of spike size and shape among cultivars or environmental conditions during image acquisition may severely affect the performances (Park et al., 2010). The presence of water droplets induced only a moderate bias in heading date estimation in two of the sites (Fig. 6b). Images with droplets represent less than 1% of the training dataset cases. That was probably not sufficient to teach the CNN to distinguish between droplets and spikes. These issues can only be solved by increasing the variability by including more images taken under diverse environmental conditions.

The logistic function smooths the daily values of $f_{head}(d)$, and reduces the influence of possible unsystematic classification errors on the heading date estimation. Furthermore, to minimize the possible impact of classification errors, in operational conditions it would be possible to prevent unrealistic estimations by imposing some constraints on the logistic model based on prior knowledge of the heading date:

$$f_{head}(d) = 0.0 \; if \; d < d_{prior} - \delta$$
$$f_{head}(d) = 1.0 \; if \; d > d_{prior} + \delta \quad (4)$$

With $d_{prior}$ being the prior value of the heading date derived from previous years observations or from phenological models, and $\delta$ being the associated maximum error. Anyhow, the use of a logistic function fit makes difficult to estimate the heading date in real time. The correct function fit is only possible *a posteriori*, *i.e.* when some images where $f_{head}(d) = 1.0$ have been already observed, which may only happen 5–10 days after the actual heading date. This may limit the use of the method proposed on operational applications of crop management requiring a rapid assessment of the heading date.

Compared to a phenological model calibrated under local conditions, our CNN-based method provides better accuracy while not requiring ancillary information (*e.g.* sowing dates, variety-specific model parameters) or daily weather data. This is an important advantage, since the accuracy achieved by ARCWHEAT over the 47 sites monitored in this study (RMSE = 4 days) was mainly due to the cultivar-specific calibration of model parameters. However, such a cultivar-specific calibration is time-consuming and requires frequent observations throughout the crop cycle, repeated over several locations and years (Cabelguenne et al., 1990; Jégo et al., 2010). Moreover, phenological models largely rely on the quality of meteorological variables, which sometimes are interpolated from weather stations that are far from the sites to be monitored without accounting for possible microclimatic effects (Joly et al., 2011; Monestiez et al., 2001). In any case, mechanistic modelling of plant phenology remains always necessary on prospective studies and to predict cultivar performance under a range of climatic scenarios. It also constitutes an attractive alternative for heading date estimation for some applications where the acquisition of high-throughput canopy images is not feasible. Phenological models can also largely benefit from the method developed in this paper: the use of IoTA field systems and deep learning approaches would substantially reduce the cost of calibration experiments and would permit also to increase the environmental variability of the field trials giving access to frequent observations from remote locations.

Our CNN based method constitutes a robust and cost-efficient approach for heading date estimation for operational applications when daily images of the canopy are available, as it is the case for some high-

throughput phenotyping platforms. In those platforms, vectors such as unmanned ground vehicles and hand-pushed carts are often used to frequently monitor the plant development and characterize biophysical traits of the different cultivars using optical images (Deery et al., 2014; Mueller-Sim et al., 2017; White and Conley, 2013). In such applications, the proposed automatic method showcased in this study could be directly integrated into other data processing pipelines at almost no cost to estimate heading date from the existing RGB images.

In agronomic or breeding applications where the phenology on distant fields need to be monitored –*e.g.* regional or national networks of agricultural fields– the whole system presented, including the fixed camera and the CNN-based method to estimate heading dates, may present important advantages in terms of costs compared to *in situ* visual annotation. The initial investment in each camera (including the pole, batteries and the hardware for data transmission) raises up to, approximately, 650 €; and the yearly costs of system maintenance and real-time data transmission by GSM is about 150 € per camera. The cost-efficiency –and the benefits in terms of environmental impact– of a system based on network of cameras compared to expert visits has to be determined case by case, and will largely depend on the distance between the fields to be monitored: the larger the network is, the more efficient remote observations are compared to field visits.

### *4.2. Subjectivity of visual annotations of the heading date from experts using RGB images*

The visual determination of heading date is a task that includes some degree of subjectivity. For the 27 sites monitored in 2017 and 2018, the panel of 14 experts provided different estimates of the heading dates after inspecting visually the IoTA images. In most of the cases, the panel members selected 5–7 different dates per site out of those proposed in the questionnaire, although they were asked to follow the same definition of the heading date (Zadoks et al., 1974). The distribution of the deviations between individual replies and the average date for each site was roughly Gaussian with a standard deviation of 1.5 days (Fig. 9), very close to the RMSE of our CNN based method proposed in this paper.

The subjectivity when determining visually heading date from images is obviously higher compared to *in situ* scoring. Issues regarding the image quality exacerbated the discrepancies among experts. According to the feedback provided by the panel, image saturation and blur due to a suboptimal camera setup made difficult to see the emergence of spikes in some of the images. Further, spikes were harder to identify in images taken under direct illumination conditions due to low image contrast, as well as for cultivars with awns.

The variability of expert replies evaluated in the 27 sites constitutes a good benchmark for our CNN based method, as image quality issues may also affect the identification of spikes by ResNet50 when classifying patches. The similarity between the 2 days RMSE associated to our CNN based method with the 1.5 days confidence interval of the expert date demonstrates that the performances of our CNN based method can be considered comparable to the expert reply when observing the same images.

## 5. Conclusion

In this study, we propose a CNN based method to estimate the heading date from daily images acquired over wheat crops using an RGB camera fixed in the field. Images are processed per patch based on a binary CNN-classifier to construct the dynamics of spikes appearance. Our method is easy to implement since the labelling of patches is not time-consuming as compared to individual object annotation required for other CNN models used for object identification or counting. The reliability of the CNN-classifier to identify the presence of emerged spikes was marginally affected by the illumination conditions and cultivar diversity, since the training dataset included images acquired under diverse environmental conditions. Our method achieved satisfactory performances with RMSE ≈ 2.0 days, which is close to the uncertainties of expert annotations, and substantially better than phenological models specifically calibrated for the cultivars monitored.

The proven robustness of the proposed method suggests a strong potential for cost-efficient operational applications in the field of phenotyping and agronomic applications. However, our method is limited by the image footprint close to 5 $m^2$ if only the bottom half of the image is used, as done in our study. However, the good consistency with expert observations taken over a larger sampling area demonstrates that our restricted sampling was sufficient. Nevertheless, the representativeness of such small footprint estimations to characterize phenology over large and heterogenous fields remains an open question for future works.

This method has been developed using a camera looking to the crop in a fixed position, but a similar approach could be transposed to time series of images from other vectors used in phenotyping experiments, such as unmanned ground and aerial vehicles, providing that the revisit time and resolution are sufficient. Further, the method could be adapted to identify other crop development stages associated with the identification of certain organs, such as the appearance of anthers for wheat to date flowering, or the appearance of tassels for the male flowering in maize.

## CRediT authorship contribution statement

**Kaaviya Velumani:** Methodology, Software, Writing - original draft, Visualization. **Simon Madec:** Methodology, Writing - review & editing. **Benoit de Solan:** Data curation, Resources. **Raul Lopez-Lozano:** Methodology, Supervision, Writing - review & editing. **Jocelyn Gillet:** Supervision, Writing - review & editing. **Jeremy Labrosse:** Data curation, Resources. **Stephane Jezequel:** Data curation, Resources. **Alexis Comar:** Writing - review & editing, Supervision, Funding acquisition. **Frédéric Baret:** Conceptualization, Supervision, Writing - review & editing, 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.

## Acknowledgements

The authors would like to thank the Association Nationale de Recherche Technologique (ANRT) for their grant CIFRE. The authors acknowledge Guy Deshayes (Arvalis) for his kind support with consolidating the reference data from the experts in the field.

## Appendix A

**Table A1**
Description of the 47 sites considered in our study over 3 growing seasons and 9 different varieties of wheat.

| Field Id | Coordinates | Year | Mean Seasonal Temperature Oct to July (°C) | Growing Degree Days °C (sowing to heading) | Heading date | Variety | Sowing date | Harvest date | Comments |
|---|---|---|---|---|---|---|---|---|---|
| D-D1 | 48.322452, 2.383738 | 2017 | 11.43 | 1518.15 | 17-05-2017 | Descartes | 20-10-2016 | – | Experimental plots |
| D-D2 | 48.322483, 2.383812 | 2017 | 11.43 | 1518.15 | 16-05-2017 | Descartes | 20-10-2016 | – | |
| D-D3 | 48.321520, 2.382034 | 2017 | 11.43 | 1518.15 | 17-05-2017 | Descartes | 20-10-2016 | – | |
| D-I1 | 48.321570, 2.382026 | 2017 | 11.43 | 1518.15 | 17-05-2017 | Descartes | 20-10-2016 | – | |
| D-I2 | 48.321530, 2.383126 | 2017 | 11.43 | 1518.15 | 16-05-2017 | Descartes | 20-10-2016 | – | |
| D-I3 | 48.322499, 2.383748 | 2017 | 11.43 | 1518.15 | 17-05-2017 | Descartes | 20-10-2016 | – | |
| D-S1 | 48.320248, 2.379639 | 2017 | 11.43 | 1518.15 | 18-05-2017 | Descartes | 20-10-2016 | – | |
| D-S2 | 48.320221, 2.379579 | 2017 | 11.43 | 1518.15 | 19-05-2017 | Descartes | 20-10-2016 | – | |
| E-A | 44.238959, 5.928383 | 2018 | 7.45 | 1612.904 | 11-05-2018 | RGT Voilur | 25-10-2017 | – | Production plots |
| E-ME | 44.261415, 5.870530 | 2018 | 8.61 | 1841.848 | 06-05-2018 | Anvergur | 04-10-2017 | – | |
| E-MO | 44.262088, 5.869253 | 2018 | 8.61 | 1894.37 | 09-05-2018 | Anvergur | 04-10-2017 | – | |
| E-MN | 43.766323, 6.098798 | 2018 | 10.08 | 1780.29 | 17-05-2018 | Toscadou | 10-10-2017 | – | |
| E-MS | 43.765653, 6.100184 | 2018 | 10.15 | 1754.06 | 14-05-2018 | Toscadou | 10-10-2017 | – | |
| E-S | 43.812474, 5.772345 | 2018 | 11.22 | 1620.32 | 09-05-2018 | RGT Voilur | 22-10-2017 | – | |
| E-V | 43.791706, 6.120393 | 2018 | 10.08 | 1878.21 | 14-05-2018 | Toscadou | 10-10-2017 | – | |
| A-201 | 48.323864, 2.378989 | 2018 | 10.06 | 1586.40 | 11-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | Experimental plots with different nitrogen |
| A-207 | 48.348859, 2.432189 | 2018 | 10.06 | 1575.05 | 10-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | management |
| A-210 | 48.345231, 2.432945 | 2018 | 10.06 | 1613.05 | 13-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | |
| A-301 | 48.349030, 2.432504 | 2018 | 10.06 | 1586.40 | 11-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | |
| A-305 | 48.323589, 2.378951 | 2018 | 10.06 | 1601.60 | 12-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | |
| A-308 | 48.347419, 2.43389 | 2018 | 10.06 | 1586.40 | 11-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | |
| A-310 | 48.344959, 2.432586 | 2018 | 10.06 | 1613.05 | 13-05-2018 | Oregrain | 16-10-2017 | 10-07-2018 | |
| 22-DS | 48.325583, 2.386405 | 2018 | 10.06 | 1624.20 | 14-05-2018 | Fructidor | 20-10-2017 | 09-07-2018 | Production plots |
| 22-SS | 48.324231, 2.385779 | 2018 | 10.06 | 1639.85 | 15-05-2018 | Fructidor | 20-10-2017 | 09-07-2018 | |
| 331-MS | 48.321029, 2.384333 | 2018 | 10.06 | 1613.05 | 13-05-2018 | RGT Sacramento | 20-10-2017 | 09-07-2018 | |
| 331-SS1 | 48.319074, 2.380753 | 2018 | 10.06 | 1601.60 | 12-05-2018 | RGT Sacramento | 20-10-2017 | 09-07-2018 | |
| 331-SS2 | 48.319046, 2.380693 | 2018 | 10.06 | 1601.60 | 12-05-2018 | RGT Sacramento | 20-10-2017 | 09-07-2018 | |
| SI-1 | 49.073911, 5.704641 | 2019 | 8.80 | 1582.76 | 27-05-2019 | Fructidor | 14-10-2018 | – | |
| SI-2 | 49.073735, 5.704535 | 2019 | 8.80 | 1582.76 | 27-05-2019 | Fructidor | 14-10-2018 | – | |
| SI-3 | 49.074086, 5.704128 | 2019 | 8.80 | 1582.76 | 27-05-2019 | Fructidor | 14-10-2018 | – | |
| C-B1 | 48.958538, 4.253838 | 2019 | 9.48 | 1671.19 | 23-05-2019 | Matheo | 14-10-2018 | – | |
| C-B2 | 48.958576, 4.253507 | 2019 | 9.48 | 1671.19 | 23-05-2019 | Matheo | 14-10-2018 | – | |
| C-B3 | 48.95826, 4.253713 | 2019 | 9.48 | 1671.19 | 23-05-2019 | Matheo | 14-10-2018 | – | |
| T-217 | 48.323273, 2.379553 | 2019 | 9.62 | 1529.25 | 16-05-2019 | Oregrain | 23-10-2018 | – | Experimental plots with different nitrogen |
| T-205 | 48.323624, 2.379038 | 2019 | 9.62 | 1529.25 | 16-05-2019 | Oregrain | 23-10-2018 | – | management |
| T-202 | 48.323761, 2.378927 | 2019 | 9.62 | 1543.0 | 17-05-2019 | Oregrain | 23-10-2018 | – | |
| T-305 | 48.323589, 2.378951 | 2019 | 9.62 | 1529.25 | 16-05-2019 | Oregrain | 23-10-2018 | – | |
| T-201 | 48.323864, 2.378989 | 2019 | 9.62 | 1557.70 | 18-05-2019 | Oregrain | 23-10-2018 | – | |
| T-110 | 48.323586, 2.379289 | 2019 | 9.62 | 1529.25 | 16-05-2019 | Oregrain | 23-10-2018 | – | |
| 332-DS | 48.32201, 2.383564 | 2019 | 9.62 | 1518.65 | 16-05-2019 | Rebelde | 24-10-2018 | – | Production plots |
| 332-SS | 48.319752, 2.37894 | 2019 | 9.62 | 1558.70 | 19-05-2019 | Rebelde | 24-10-2018 | – | |
| 332-MS | 48.320652, 2.381057 | 2019 | 9.62 | 1558.70 | 19-05-2019 | Rebelde | 24-10-2018 | – | |
| 411-SS | 48.317192, 2.387024 | 2019 | 9.62 | 1570.45 | 27-05-2019 | Fructidor | 06-11-2018 | – | |
| 411-DS | 48.318287, 2.388875 | 2019 | 9.62 | 1553.5 | 26-05-2019 | Fructidor | 06-11-2018 | – | |
| 411-MS | 48.316154, 2.384733 | 2019 | 9.62 | 1570.45 | 27-05-2019 | Fructidor | 06-11-2018 | – | |
| E-F1 | 43.710176, 4.534401 | 2019 | 13.31 | 1854.88 | 21-04-2019 | RGT Voilur | 25-10-2018 | – | |
| E-M1 | 43.843360, 4.442453 | 2019 | 13.12 | 1874.5 | 22-04-2019 | Anvergur | 25-10-2018 | – | |

## References


Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 19, 52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

Balla, K., Karsai, I., Bónis, P., Kiss, T., Berki, Z., Horváth, Á., Mayer, M., Bencze, S., Veisz, O., 2019. Heat stress responses in a large set of winter wheat cultivars (Triticum aestivum L.) depend on the timing and duration of stress. PLoS One 14, e0222639. https://doi.org/10.1371/journal.pone.0222639.

Baret, F., Madec, S., Irfan, K., Lopez, J., Comar, A., Hemmerlé, M., Dutartre, D., Praud, S., Tixier, M.H., 2018. Leaf-rolling in maize crops: from leaf scoring to canopy-level measurements for phenotyping. J. Exp. Bot. 69, 2705–2716. https://doi.org/10.1093/jxb/ery071.

Bogard, M., Ravel, C., Paux, E., Bordes, J., Balfourier, F., Chapman, S.C., Le Gouis, J., Allard, V., 2014. Predictions of heading date in bread wheat (Triticum aestivum L.) using QTL-based parameters of an ecophysiological model. J. Exp. Bot. 65, 5849–5865. https://doi.org/10.1093/jxb/eru328.

Bonnett, O.T., 1936. The development of the wheat spike. J. Agric. Res. 53, 445–451.

Brown, B., Westcott, M., Christensen, N., Pan, B., Stark, J., 2005. Nitrogen management for hard wheat protien enhancement. Pac. Northw. Ext. Publ. 578, 1–14.

Cabelguenne, M., Jones, C.A., Marty, J.R., Dyke, P.T., Williams, J.R., 1990. Calibration and validation of EPIC for crop rotations in southern France. Agric. Syst. 33, 153–171. https://doi.org/10.1016/0308-521X(90)90078-5.

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M.D., Luis Araus, J., 2012. High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J. Integr. Plant Biol. 54, 312–320. https://doi.org/10.1111/j.1744-7909.2012.01116.x.

Camargo, A.V., Mott, R., Gardner, K.A., Mackay, I.J., Corke, F., Doonan, J.H., Kim, J.T., Bentley, A.R., 2016. Determining phenological patterns associated with the onset of senescence in a wheat MAGIC mapping population. Front. Plant Sci. 7, 1540. https://doi.org/10.3389/fpls.2016.01540.

Ceglar, A., van der Wijngaart, R., de Wit, A., Lecerf, R., Boogaard, H., Seguini, L., van den Berg, M., Toreti, A., Zampieri, M., Fumagalli, D., Baruth, B., 2019. Improving WOFOST model to simulate winter wheat phenology in Europe: evaluation and effects on yield. Agric. Syst. 168, 168–180. https://doi.org/10.1016/j.agsy.2018.05.002.

Chapman, S.C., Chakraborty, S., Dreccer, M.F., Howden, S.M., 2012. Plant adaptation to climate change—opportunities and priorities in breeding. Crop Pasture Sci. 63, 251. https://doi.org/10.1071/CP11303.

Chmielewski, F.-M., 2013. Phenology in agriculture and horticulture. In: Schwartz, M.D. (Ed.), Phenology: An Integrative Environmental Science. Springer Netherlands, Dordrecht, pp. 539–561. https://doi.org/10.1007/978-94-007-6925-0_29.

Chollet, F., 2015. Keras. https://keras.io.

Comar, A., Burger, P., de Solan, B., Baret, F., Daumard, F., Hanocq, J.-F., 2012. A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. Funct. Plant Biol. 39, 914. https://doi.org/10.1071/FP12065.

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., Furbank, R., 2014. Proximal remote sensing buggies and potential applications for field-based phenotyping. Agronomy 4, 349–379. https://doi.org/10.3390/agronomy4030349.

Desai, S.V., Balasubramanian, V.N., Fukatsu, T., Ninomiya, S., Guo, W., 2019. Automatic estimation of heading date of paddy rice using deep learning. Plant Methods 15, 76. https://doi.org/10.1186/s13007-019-0457-1.

Ferris, R., Ellis, R.H., Wheeler, T.R., Hadley, P., 1998. Effect of high temperature stress at anthesis on grain yield and biomass of field-grown crops of wheat. Ann. Bot. 82, 631–639. https://doi.org/10.1006/anbo.1998.0740.

Foulkes, M.J., Slafer, G.A., Davies, W.J., Berry, P.M., Sylvester-Bradley, R., Martre, P., Calderini, D.F., Griffiths, S., Reynolds, M.P., 2011. Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. J. Exp. Bot. 62, 469–486. https://doi.org/10.1093/jxb/erq300.

García, G.A., Dreccer, M.F., Miralles, D.J., Serrago, R.A., 2015. High night temperatures during grain number determination reduce wheat and barley grain yield: a field study. Glob. Change Biol. 21, 4153–4164. https://doi.org/10.1111/gcb.13009.

Gate, P., 1995. Ecophysiologie du blé. Tec & Doc-Lavoisier, France.

Gooding, M.J., Ellis, R.H., Shewry, P.R., Schofield, J.D., 2003. Effects of restricted water availability and increased temperature on the grain filling, drying and quality of winter wheat. J. Cereal Sci. 37, 295–309. https://doi.org/10.1006/jcrs.2002.0501.

Gouache, D., Le Bris, X., Bogard, M., Deudon, O., Pagé, C., Gate, P., 2012. Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France. Eur. J. Agron. 39, 62–70. https://doi.org/10.1016/j.eja.2012.01.009.

Guedira, M., Xiong, M., Hao, Y.F., Johnson, J., Harrison, S., Marshall, D., Brown-Guedira, G., 2016. Heading date QTL in winter wheat (Triticum aestivum L.) coincide with major developmental genes VERNALIZATION1 and PHOTOPERIOD1. PLoS One 11, e0154242. https://doi.org/10.1371/journal.pone.0154242.

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, 100. https://doi.org/10.1186/s13007-018-0366-8.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778.

Jay, S., Maupas, F., Bendoula, R., Gorretta, N., 2017. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: comparison of vegetation indices and PROSAIL inversion for field phenotyping. Field Crops Res. 210, 33–46. https://doi.org/10.1016/j.fcr.2017.05.005.

Jégo, G., Pattey, E., Bourgeois, G., Morrison, M.J., Drury, C.F., Tremblay, N., Tremblay, G., 2010. Calibration and performance evaluation of soybean and spring wheat cultivars using the STICS crop model in Eastern Canada. Field Crops Res. 117, 183–196. https://doi.org/10.1016/j.fcr.2010.03.008.

Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., Wavresky, P., 2011. Temperature interpolation based on local information: the example of France. Int. J. Climatol. 31, 2141–2153. https://doi.org/10.1002/joc.2220.

Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open Source Scientific Tools for Python. https://www.scipy.org.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: plant identification with convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE. pp. 452–456. https://doi.org/10.1109/ICIP.2015.7350839.

Lobell, D.B., Ortiz-Monasterio, J.I., 2007. Impacts of day versus night temperatures on spring wheat yields: a comparison of empirical and CERES model predictions in three locations. Agron. J. 469–477. https://doi.org/10.2134/agronj2006.0209. American Society of Agronomy.

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C., 2017. TasselNet: counting maize tassels in the wild via local counts regression network. Plant Methods 13, 79. https://doi.org/10.1186/s13007-017-0224-0.

Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., Hemmerlé, M., Colombeau, G., Comar, A., 2017. High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates. Front. Plant Sci. 8, 1–14. https://doi.org/10.3389/fpls.2017.02002.

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.

Malambo, L., Popescu, S.C., Horne, D.W., Pugh, N.A., Rooney, W.L., 2019. Automated detection and measurement of individual sorghum panicles using density-based clustering of terrestrial lidar data. ISPRS J. Photogramm. Remote Sens. 149, 1–13. https://doi.org/10.1016/j.isprsjprs.2018.12.015.

McMaster, G.S., Smika, D.E., 1988. Estimation and evaluation of winter wheat phenology in the central Great Plains. Agric. For. Meteorol. 43, 1–18. https://doi.org/10.1016/0168-1923(88)90002-0.

Monestiez, P., Courault, D., Allard, D., Ruget, F., 2001. Spatial interpolation of air temperature using environmental context: application to a crop model. Environ. Ecol. Stat. 8, 297–309. https://doi.org/10.1023/A:1012726317935.

Mueller-Sim, T., Jenkins, M., Abel, J., Kantor, G., 2017. The Robotanist: a ground-based agricultural robot for high-throughput crop phenotyping. In: Proceedings - IEEE International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers Inc. pp. 3634–3639. https://doi.org/10.1109/ICRA.2017.7989418.

Park, D., Ramanan, D., Fowlkes, C., 2010. Multiresolution models for object detection. In: Daniilidis, K., Maragos, P., Paragios, N. (Eds.), Computer Vision – ECCV 2010. Springer, Berlin, Heidelberg, pp. 241–254. https://doi.org/10.1007/978-3-642-15561-1_18.

Reynolds, M., Foulkes, M.J., Slafer, G.A., Berry, P., Parry, M.A.J., Snape, J.W., Angus, W.J., 2009. Raising yield potential in wheat. J. Exp. Bot. 60, 1899–1918. https://doi.org/10.1093/jxb/erp016.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sadras, V.O., Slafer, G.A., 2012. Environmental modulation of yield components in cereals: heritabilities reveal a hierarchy of phenotypic plasticities. Field Crops Res. 127, 215–224. https://doi.org/10.1016/j.fcr.2011.11.014.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. The IEEE International Conference on Computer Vision (ICCV) pp. 618–626.

Singh, A.K., Ganapathysubramanian, B., Sarkar, S., Singh, A., 2018. Deep learning for plant stress phenotyping: trends and future perspectives. Trends Plant Sci. 23, 883–898. https://doi.org/10.1016/j.tplants.2018.07.004.

Slafer, G., 2012. Wheat development: its role in phenotyping and improving crop adaptation. In: Reynolds, M., Pask, A., Mullan, D. (Eds.), WHEAT Physiological Breeding I: Interdisciplinary Approaches to Improve Crop Adaptation. CIMMYT, Mexico, pp. 107–121.

Slafer, G., Rawson, H., 1994. Sensitivity of wheat phasic development to major environmental factors: a re-examination of some assumptions made by physiologists and modellers. Funct. Plant Biol. 21, 393. https://doi.org/10.1071/PP9940393.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging 35, 1299–1312. https://doi.org/10.1109/TMI.2016.2535302.

Thépot, S., 2014. Utilisation d'une population multi-parentale et hautement recombinante de blé tendre pour l'étude de l'architecture génétique de la précocité de floraison. Util. d'une Popul. multi-parentale hautement Recomb. blé Tendr. pour l'étude l'architecture génétique la précocité floraison. Univ. Paris Sud (Paris 11) (2014). Université Paris Sud 11.

Wallach, D., Palosuo, T., Thorburn, P., Seidel, S.J., Gourdain, E., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hochman, Z., Hoek, S., Horan, H., Hoogenboom, G., Huang, M., Jabloun, M., Jing, Q., Justes, E., Kersebaum, K.C., Klosterhalfen, A., Launay, M., Luo, Q., Maestrini, B., Moriondo, M., Zadeh, H.N., Olesen, J.E., Poyda, A., Priesack, E., Pullens, J.W.M., Qian, B., Schütze, N., Shelia, V., Souissi, A., Specka, X., Srivastava,

A.K., Stella, T., Streck, T., Trombi, G., Wallor, E., Wang, J., Weber, T.K.D., Weihermüller, L., de Wit, A., Wöhling, T., Xiao, L., Zhao, C., Zhu, Y., 2019. How well do crop models predict phenology, with emphasis on the effect of calibration? bioRxiv. https://doi.org/10.1101/708578. 708578.

Weir, A.H., Bragg, P.L., Porter, J.R., Rayner, J.H., 1984. A winter wheat crop simulation model without water or nutrient limitations. J. Agric. Sci. 102, 371–382. https://doi.org/10.1017/S0021859600042702.

Wheeler, T.R., Hong, T.D., Ellis, R.H., Batts, G.R., Morison, J.I.L., Hadley, P., 1996. The duration and rate of grain growth, and harvest index, of wheat (Triticum aestivum L.) in response to temperature and CO 2. J. Exp. Bot. 47, 623–630. https://doi.org/10.1093/jxb/47.5.623.

White, J.W., Conley, M.M., 2013. A flexible, low-cost cart for proximal sensing. Crop Sci. 53, 1646. https://doi.org/10.2135/cropsci2013.01.0054.

White, J.W., Herndl, M., Hunt, L.A., Payne, T.S., Hoogenboom, G., 2008. Simulation-based analysis of effects of and loci on flowering in wheat. Crop Sci. 48, 678. https://doi.org/10.2135/cropsci2007.06.0318.

Whittal, A., Kaviani, M., Graf, R., Humphreys, G., Navabi, A., 2018. Allelic variation of vernalization and photoperiod response genes in a diverse set of North American high latitude winter wheat genotypes. PLoS One 13, e0203068. https://doi.org/10.1371/journal.pone.0203068.

Yalcin, H., 2018. Phenology recognition using deep learning: DeepPheno. In: 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE. pp. 1–4. https://doi.org/10.1109/SIU.2018.8404165.

Yang, G., Liu, J., Zhao, C., Li, Zhenhong, Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Zhenhai, Li, H., Yang, H., 2017. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. Front. Plant Sci. 8, 1111. https://doi.org/10.3389/fpls.2017.01111.

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x.

Zheng, B., Chenu, K., Fernanda Dreccer, M., Chapman, S.C., 2012. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivium) varieties? Glob. Change Biol. 18, 2899–2914. https://doi.org/10.1111/j.1365-2486.2012.02724.x.

# 5. Conclusion

This PhD, structured in three scientific papers, investigated the use of deep learning methods applied on RGB images to estimate three phenotypic traits – early-stage plant counting, head density, and heading date – that, traditionally, were determined *in situ* from visual notation. This PhD was developed thanks to a CIFRE convention with the private company HIPHEN that specializes in the development of high-throughput phenotyping services. Therefore, the three studies presented in this dissertation were also designed to produce operational and transferable solutions that can be of direct benefit of HIPHEN to produce reliable phenotypic traits estimations, with accuracy comparable or better than the visual notations. In this context, special attention was paid to solutions maximizing the observation throughput, and particularly to UAV images for the estimation of plant and head density. Compared to ground vectors, the use of UAV introduces scaling issues that need to be addressed in the application and evaluation of deep learning methods. The main conclusions of the work conducted in this PhD are detailed hereafter.

## 5.1 Can deep learning methods replace in-situ measurements for estimating maize plant density, wheat head density and heading date?

The observed accuracy in the estimation of the three phenotypic traits studied in this dissertation from deep learning methods can be considered close to manual in situ determinations. For early-stage maize plant counting (Chapter 2) the two-stage object detection network, Faster-RCNN, applied to UAV images provided a relative error of 6% compared to the ground truth, which is considered highly satisfactory. Therefore, the deep learning algorithms could very well replace manual counting in the field. In this study, the ground-truth was generated by manual plant counting on the UAV images. Given the minimum overlap observed between maize plants at the early stages, manual annotations on high-resolution images can be safely used as a reference ground truth equivalent to in situ counting, since all instances of the plants are clearly distinguishable on the images.

For wheat head density estimation (Chapter 3) the regression-based object counting method based on SFC$^2$Net (L. Liu et al. 2020) yielded an average error of 14% against annotations, which can be considered satisfactory. Nevertheless, when comparing against in-situ head density observations, both image annotations and estimations from deep learning presented a moderate underestimation. This systematic bias is inherent of the use of RGB images in crowded canopy with a high overlap between objects: not all wheat heads were visible within the images. Theoretically, this moderate bias could be a priori mitigated in regression-based methods as SFC$^2$Net by training the model with in-situ observations of the actual plant density for the same

training images. Nevertheless, the broad-sense heritability of the model-estimated wheat head density was higher compared to that of the in-situ measurements. The lower heritability of the latter could be partly explained by the limited representativeness of the sampling area and human errors. The deep learning estimations are, therefore, more repeatable than human measurements and could be useful to understand the genotype interaction within a target environment. However, it should be kept in mind that deep learning algorithms could possibly exhibit bias over cultivars not well-characterized within the training dataset. This learning bias is a recurrent problem actively investigated under the ethics of artificial intelligence (Müller 2020) for a range of applications including facial recognition (Raji et al. 2020) or gender classification (Buolamwini and Gebru 2018). Under the field phenotyping context, the broad-sense heritability could be artificially boosted when the models exhibit a bias towards certain cultivars. Further investigations should be directed to quantify this possible problem.

The results obtained from the studies were compared, in all cases, to reference measurements, in each case performing well with minimal errors. Depending on the study, the reference measurements used were collected in-situ by manual intervention or from image annotations or a combination of the two. The trueness of the reference measurements was also analyzed and questioned. For phenological traits like wheat heading date estimation the proposed method based on ResNet50 to detect wheat heads in daily RGB images, yielded an average error of 2 days when compared to visual scoring from experts. The uncertainties of this method are lower compared to those introduced by human subjectivity, as discovered during the experiments conducted with a panel of wheat experts. In fact, the standard deviation of the heading dates provided by several experts that were observing the same canopy was above the error associated with the proposed method, indicating that this automatic method could replace in situ annotations of the heading date.

## 5.2 Data augmentation and data preparation strategies are efficient to minimize scaling issues

Two recurrent issues were faced while estimating wheat head and maize plant density from UAV using deep learning methods: the hyper-specialization of the deep learning algorithms to the object scale and the degradation of performances for detecting small objects.

As shown in Chapter 2, a systematic overfit of Faster-RCNN to the training dataset was observed, leading to a drastic decrease in performance when the GSD varied significantly in the inference dataset. Applying a CNN trained with low-resolution images (GSD≈0.63cm) to detect plants on high-resolution ones (GSD≈0.3cm) led to a systematic over-detection: the CNN proposed multiple small boxes to cover the image area occupied by single plants. Conversely, when detecting maize plants on low-resolution images with a CNN exclusively trained on high resolution, an important

under-detection was observed. The most efficient strategy to overcome this problem was data augmentation in the scale domain, down-sampling the training dataset at different resolutions to enable Faster-RCNN to recognize plants with different sizes. This data augmentation strategy provided even better results than CNN predictions on the same native resolutions. It is therefore recommended to use it for any UAV-based object detection and counting application.

Further, it should be noted that object detection networks, like the Faster-RCNN used in this study, were designed and built to achieve the best performances on datasets with large-scale objects such as COCO (Lin et al. 2014) or ImageNet (Russakovsky et al. 2015). Hence, when these networks are used to detect objects smaller than 20 x 20 pixels, the object information is reduced to a few bytes during the convolutional operations, thus failing on small objects. If objects are below this size, they have to be up-sampled to prevent a systematic under-detection.

The algorithm used to down-sample images as part of a data augmentation or data preparation strategy is also an important factor determining the accuracy of CNN detections. In Chapter 3, the use of images acquired from ground sensors were used to train deep learning methods applied to detect and count wheat heads on UAV images at 4-7 times lower resolution. Down-sampling images of the training dataset was thus necessary and two different methods were used: bicubic down-sampling and Gaussian filter followed by motion blur. The results indicated important differences in the performance which resulted from the down-sampling method, particularly for the SFC$^2$Net regression-based algorithm: 14% error on total number of heads per scene with Gaussian blur against 35% error with bicubic down-sampling. For the same GSD, the down-sampling method determines the amount of textural information contained in the image. Therefore, when re-sampling the images in the training dataset as part of a data augmentation or data preparation strategy, the realism of the method used is essential to avoid possible biases. Gaussian filter was therefore found to be a more realistic method when resampling UAV images. In fact, the relative error of 14% achieved by down-scaling the training dataset using the Gaussian filter was very similar to the uncertainties observed when training and applying SFC$^2$Net with images from ground sensors at their full resolution, which demonstrates the suitability of the down-sampling method.

Lastly, inadequate UAV camera settings may degrade image quality and impact negatively the performances of the object detection algorithm. In Chapter 2, the UAV was flying at two different altitudes with the same camera settings. A visual inspection of the images revealed that those settings were not optimal for the flight taken at a higher altitude, and the large uncertainties in plant detection observed from faster-RCNN confirmed it (rRMSE=0.48). This problem was partially mitigated using an advanced CNN-based super-resolution method (Z. Han et al. 2019; Xintao Wang et al. 2019) to up-sample the UAV images taken at a higher altitude, as part of a data preparation strategy. This super-resolution model was trained jointly with a CycleGAN module (Zhu et al. 2017) that allows the generation of a paired high-and-low resolution

training dataset in an unsupervised manner, thanks to a cyclic adversarial loss. While this strategy has the drawback of manipulating several losses, which complicates the model convergence and sometimes adds artifacts to the images, a significant improvement in plant counting was observed (rRMSE=0.22).

## 5.3 Object detection versus object counting algorithms for plant/organ density estimations

In Chapter 3, two types of deep learning algorithms were compared to estimate wheat head density: Faster-RCNN object detection algorithm; and the regression-based SFC$^2$Net algorithm. SFC$^2$Net provided systematically better results than Faster-RCNN. Whilst Faster-RCNN exhibited a slight negative bias against visual notations –taken from both, ground and from UAV– SFC$^2$Net estimations showed no systematic bias. The advantage of regression-based methods against object detection are summarized in Table 1. The most important factor in favor of regression-based methods is that the algorithm output is the object count or object density. When the training dataset contains a realistic object count, the algorithm can implicitly handle possible biases between the actual number of objects present in the scene and the number of visible objects. This is an important advantage for traits like wheat head counting, as canopies are dense and substantial overlap exist among wheat heads, but also between heads and other plant organs.

Object detection algorithms, by contrast, do not directly provide an object count, but rather a proposal of regions that may contain an object with a confidence score associated. The number of objects will result from the choice of a confidence score threshold to select actual objects –which is not straightforward in networks that are not calibrated– and therefore, in case of overlapping objects, the actual object density will be underestimated. The main advantage of object detection methods is that they provide additional information on the objects such as size and location, which may be necessary in some phenotyping applications. Faster-RCNN performed satisfactorily for maize plant counting at early stages (Chapter 2) as the scenes were relatively simple and overlap between small maize plants was not frequent. In these types of applications, the performance of object detection methods is expected to be comparable to the regression-based algorithms.

*Table 4 Summary of advantages of regression-based methods and object detection methods for object counting.*

| **CNN approach** | **Required object scale** | **Advantages** | **Throughput/Processing speed** |
|---|---|---|---|
| Object detection | > 20 x 20 pixels | - Exact localization<br>- Object size | - 2-stage networks are relatively slow.<br>- Possibility to use one stage networks with a |

| | | | |
|---|---|---|---|
| | | | trade-off between accuracy and speed. |
| Regression-based Object Counting | > a few pixels | - The algorithm output is object count<br>- Approximate localization of objects<br>- Can handle densely packed, overlapping scenes | - Relatively faster, may support real-time applications |

Regarding computing resources, regression-based methods are more computationally efficient, enabling them to be used for real-time applications. Object detection algorithms, by contrast, often requires training 2-stage networks (region proposal and classification) that increases computation time. Furthermore, the need of a minimum size for object detection algorithms may require an image pre-processing of large datasets, whereas regression methods are more versatile.

## 5.4 Temporal resolution requirements for the estimation of phenological dates

The satisfactory results obtained in the automatic estimation of the heading date relied strongly on the availability of daily images from fixed sensors (IoTA). As opposed to other approaches where the CNN outputs directly the percentage of crop maturity (X. Wang et al. 2019), the methodology implemented in Chapter 4 to estimate the heading date focused on the dynamics rather than on the number of objects in the image. The estimation of the heading date was based on the temporal dynamics of the presence of heads in small patches of the scene, which allowed to simplify the problem into a binary classification (ears present/absent). This allowed to build a large and diverse set of training images, only using the reference heading dates, without requiring image labels from experts. On each plot observed, fitting the temporal dynamics in the presence of wheat heads with a logistic curve permitted to minimize the possible effects of uncertainties in the classification of specific daily images, increasing the robustness of the estimation.

Therefore, the temporal resolution of the observations is the main factor determining the accuracy of phenological dates estimation. Observations every day or every two days are necessary to retrieve specific phenological appearances. This temporal resolution is difficult to achieve, for instance with some unmanned ground vehicles, as the costs of running a Phenomobile-like vehicle daily would be very high, and probably a human operator can be more efficient. The use of UAV could be better suited, but the image resolution may not be sufficient for some developmental stages like anthesis, that requires very high resolution. Therefore, fixed cameras are the most suitable vectors for phenological traits, but the cost of installing such cameras on large

phenotyping experiments with thousands of microplots is still high. Fixed sensors and cameras are better tailored to create regional/national networks of sensors with frequent observations over distant fields, thanks to their connectivity. The impact of temporal resolution in the accuracy of the wheat heading date was not evaluated in Chapter 4, and further works will be needed to investigate the best strategies to use other phenotyping vectors (UAV, ground vehicles, portable sensors) to retrieve reliable phenological dates.

## 5.5 Perspectives

### 5.5.1 Deep learning approaches

Based on the results obtained during this study, the following avenues could be explored to encourage the adoption of deep learning algorithms under operational field conditions for plant phenotyping:

**A) Domain Shift** may be defined as the shift in distribution between the source domain (of the training dataset) and the target domain. It negatively impacts the model performances over the target domains. This was sometimes encountered in our studies, especially in the scenarios where the phenological stages of the crops and the image quality differed across the training and target dataset. This is a recurring problem in applying deep learning models to a new dataset since the models are trained under the assumption that the training and the target datasets are independently and identically distributed, which is always not the case (Zhu 2005). Domain adaptation, an actively studied branch of transfer learning, thus attempts to improve the model performance over the target domain using different strategies (Farahani et al. 2020). For example, by minimizing the distance between the target/source distributions computed in terms of Wasserstein metric, KL divergence etc. (Shen et al. 2017; Sun, Feng, and Saenko 2016), taking into consideration the intra-class and inter-class divergence in the source domain (Kang et al. 2019), generating pseudo labels on the target domain either from multiple learners (Lee et al. 2018; Mendel et al. 2020) or denoising them while model training using per class representative prototypes (Han, Luo, and Wang 2019; Zhang et al. 2021). Other techniques such as lifelong learning attempt to improve model generalization over new targets without forgetting the learnings from the source domain (Wu et al. 2019; Yoon et al. 2017). This is often addressed by adding a new loss using a domain classifier and a gradient reversal layer (Ganin and Lempitsky 2015) which forces the top layers of the network (feature extractor) to be domain invariant. These techniques could be of great utility to improve model performances covering the diversity due to genotypes, sowing densities, and soil conditions (Ayalew, Ubbens, and Stavness 2020).

**B) Dataset bottleneck**: It has been well established that the generalization capability of deep learning algorithms improves when increasing the size and

diversity of the training dataset (Alom et al. 2019). This has accelerated the publication of open datasets with collaboration between institutes such as the GWHD dataset (David et al. 2020) used in this study. However, the annotation of such a diverse dataset is a time-consuming and expensive task. In such situations where only limited labeled data is available, the use of semi-supervised approaches could be explored: the top-layers of the models are first trained on a large sample of the unlabeled dataset to learn the image representations in an unsupervised manner. Following this, the full model is finetuned in a task-specific manner using the limited labels available (Chen et al. 2020; Y. Liu et al. 2020; Siddharth et al. 2017). The generation of synthetic datasets provides another promising alternative to overcome limited labels. This strategy is already widely applied to urban scene segmentation for autonomous car driving (Richter et al. 2016; Ros et al. 2016; Wrenninge and Unger 2018), crowd counting (Q. Wang et al. 2019) or crop phenotyping under controlled conditions (Toda et al. 2020; Ward, Moghadam, and Hudson 2018). The approaches presented by Barth et al. (2018) and Liu et al. (2019) could be good starting points to capture the complexity of canopy structure under field conditions. Once again, the use of domain adaptation strategies (Zhuang et al. 2019) would be required to improve the generalization of the models trained on the synthetic dataset to the real-world datasets. Another alternative, outside deep learning, would be the use of agent-based methods which do not require labelled datasets, such as proposed by Jacopin et al. (2021) for plant counting. These methods exploit the spatial organization of the objects within the image and hence their suitability for complex scenes such as wheat canopy may be a limitation and needs to be tested.

C) **Scale sensitivity**: In this thesis, the object scale was treated as a domain problem and we proposed solutions to overcome this through data augmentation and appropriate image rescaling strategies. But it is also possible to address this at the network level using scale aware networks that employ the recent attention mechanisms (Dong et al. 2020; Hossain et al. 2019; Jiang et al. 2020). Attention mechanisms, first popularized for applications involving natural language processing, attempt to imitate the human cognitive system by focusing on the most important parts of the sentences or phrases to process the information (Galassi, Lippi, and Torroni 2020). Similarly, in the case of image processing, this approach focuses on the most important features output by the convolutional neural network to successfully accomplish the task and has been proved to provide state of the art results for small object detection and counting, where context of the objects is extremely important (Lim et al. 2021; Zhou et al. 2021). In this way, we would not need to resample the target dataset, which requires a priori knowledge on the object scale, which is not always correlated to the GSD owing to variations among genotypes and crop maturity.

### 5.5.2 In terms of methodologies developed

The methodologies developed in this study could be improved or extended in the following ways:

- All the three methodologies could be transposed to other crops, provided the occurrence of the phenological stages or plant/organ density can be visually discerned. In addition, unsupervised domain adaptation approaches based on open-set transfer learning to exploit the existing annotations to perform cross species detection/counting could be experimented to minimize labelling effort (Ayalew et al. 2020; Gebru, Hoffman, and Fei-Fei 2017).
- The wheat heading date estimation method, demonstrated over the daily RGB images from fixed sensors, could be transposed to other higher throughput data acquisition vectors like unmanned aerial or ground vehicles. In this case, the appropriate revisit period should be defined to ensure that the occurrence of the phenological stage can be detected without significant delay.
- The regression-based methodologies for organ counting may be combined with an object detection approach to facilitate a density-map guided detection (Li et al. 2020; Lian et al. 2019). This would allow accurate localization, access to the object size and region-specific upscaling to improve detection of small objects.

## References

Alom, Md Zahangir, Tarek M. Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C. Van Essen, Abdul A. S. Awwal, and Vijayan K. Asari. 2019. "A State-of-the-Art Survey on Deep Learning Theory and Architectures." Electronics 8(3):292. doi: 10.3390/electronics8030292.

Ayalew, Tewodros W., Jordan R. Ubbens, and Ian Stavness. 2020. "Unsupervised Domain Adaptation for Plant Organ Counting." Pp. 330–46 in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 12540 LNCS. Springer Science and Business Media Deutschland GmbH.

Barth, R., J. IJsselmuiden, J. Hemming, and E. J. Van Henten. 2018. "Data Synthesis Methods for Semantic Segmentation in Agriculture: A Capsicum Annuum Dataset." Computers and Electronics in Agriculture 144:284–96. doi: 10.1016/j.compag.2017.12.001.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Pp. 77–91 in Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Vol. 81, Proceedings of Machine Learning Research, edited by S. A. Friedler and C. Wilson. New York, NY, USA: PMLR.

Chen, Ting, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. 2020. "Big Self-Supervised Models Are Strong Semi-Supervised Learners." ArXiv Preprint.

David, Etienne, Simon Madec, Pouria Sadeghi-Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchgessner, Goro Ishikawa, Koichi Nagasawa, Minhajul A. Badhon, Curtis Pozniak, Benoit de Solan, Andreas Hund, Scott C. Chapman, Frédéric Baret, Ian Stavness, and Wei Guo. 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 2020:1–12. doi: 10.34133/2020/3521852.

Dong, Li, Haijun Zhang, Yuzhu Ji, and Yuxin Ding. 2020. "Crowd Counting by Using Multi-Level Density-Based Spatial Information: A Multi-Scale CNN Framework." Information Sciences 528:79–91. doi:

10.1016/j.ins.2020.04.001.
Farahani, Abolfazl, Sahar Voghoei, Khaled Rasheed, and Hamid R. Arabnia. 2020. "A Brief Review of Domain Adaptation." ArXiv Preprint ArXiv:2010.03978.
Galassi, Andrea, Marco Lippi, and Paolo Torroni. 2020. "Attention in Natural Language Processing." IEEE Transactions on Neural Networks and Learning Systems 1–18. doi: 10.1109/TNNLS.2020.3019893.
Ganin, Yaroslav, and Victor Lempitsky. 2015. "Unsupervised Domain Adaptation by Backpropagation." Pp. 1180–89 in International conference on machine learning. PMLR.
Gebru, Timnit, Judy Hoffman, and Li Fei-Fei. 2017. "Fine-Grained Recognition in the Wild: A Multi-Task Domain Adaptation Approach." Pp. 1358–67 in 2017 IEEE International Conference on Computer Vision (ICCV). IEEE.
Han, Jiangfan, Ping Luo, and Xiaogang Wang. 2019. "Deep Self-Learning From Noisy Labels." Proceedings of the IEEE International Conference on Computer Vision 2019-Octob:5137–46.
Hossain, Mohammad, Mehrdad Hosseinzadeh, Omit Chanda, and Yang Wang. 2019. "Crowd Counting Using Scale-Aware Attention Networks." Pp. 1280–88 in 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE.
Jacopin, Eliott, Naomie Berda, Léa Courteille, William Grison, Lucas Mathieu, Antoine Cornuéjols, and Christine Martin. 2021. "Using Agents and Unsupervised Learning for Counting Objects in Images with Spatial Organization." Pp. 688–97 in Proceedings of the 13th International Conference on Agents and Artificial Intelligence. Vol. 2. SCITEPRESS - Science and Technology Publications.
Jiang, Xiaoheng, Li Zhang, Mingliang Xu, Tianzhu Zhang, Pei Lv, Bing Zhou, Xin Yang, and Yanwei Pang. 2020. "Attention Scaling for Crowd Counting." Pp. 4706–15 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Kang, Guoliang, Lu Jiang, Yi Yang, and Alexander G. Hauptmann. 2019. "Contrastive Adaptation Network for Unsupervised Domain Adaptation." Pp. 4893–4902 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
Lee, Kuang-Huei, Xiaodong He, Lei Zhang, Linjun Yang, and JD AI Research. 2018. "CleanNet: Transfer Learning for Scalable Image Classifier Training with Label Noise." Pp. 5447–56 in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Li, Changlin, Taojiannan Yang, Sijie Zhu, Chen Chen, and Shanyue Guan. 2020. "Density Map Guided Object Detection in Aerial Images." Pp. 190–91 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.
Lian, Dongze, Jing Li, Jia Zheng, Weixin Luo, and Shenghua Gao. 2019. "Density Map Regression Guided Detection Network for RGB-D Crowd Counting and Localization." Pp. 1821–30 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Lim, Jeong-Seon, Marcella Astrid, Hyun-Jin Yoon, and Seung-Ik Lee. 2021. "Small Object Detection Using Context and Attention." Pp. 181–86 in 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE.
Lin, Tsung Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. "Microsoft COCO: Common Objects in Context." Pp. 740–55 in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 8693 LNCS. Springer Verlag.
Liu, Liang, Hao Lu, Yanan Li, and Zhiguo Cao. 2020. "High-Throughput Rice Density Estimation from Transplantation to Tillering Stages Using Deep Networks." Plant Phenomics 2020:1–14. doi: 10.34133/2020/1375957.
Liu, Shouyang, Pierre Martre, Samuel Buis, Mariem Abichou, Bruno Andrieu, and Frédéric Baret. 2019. "Estimation of Plant and Canopy Architectural Traits Using the Digital Plant Phenotyping Platform." Plant Physiology 181(3):881–90. doi: 10.1104/pp.19.00554.
Liu, Yan, Lingqiao Liu, Peng Wang, Pingping Zhang, and Yinjie Lei. 2020. "Semi-Supervised Crowd Counting via Self-Training on Surrogate Tasks." Pp. 242–59 in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 12360 LNCS. Springer Science and Business Media Deutschland GmbH.

Mendel, Robert, Luis Antonio de Souza, David Rauber, João Paulo Papa, and Christoph Palm. 2020. "Semi-Supervised Segmentation Based on Error-Correcting Supervision." Pp. 141–57 in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 12374 LNCS. Springer Science and Business Media Deutschland GmbH.
Müller, Vincent C. 2020. "Ethics of Artificial Intelligence and Robotics." in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University.
Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. "Saving Face." Pp. 145–51 in Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. Vol. 7. New York, NY, USA: ACM.
Richter, Stephan R., Vibhav Vineet, Stefan Roth, and Vladlen Koltun. 2016. "Playing for Data: Ground Truth from Computer Games." Pp. 102–18 in European conference on computer vision. Cham: Springer Verlag.
Ros, German, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. 2016. "The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes." Pp. 3234–43 in Proceedings of the IEEE conference on computer vision and pattern recognition.
Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision 115(3):211–52. doi: 10.1007/s11263-015-0816-y.
Shen, Jian, Yanru Qu, Weinan Zhang, and Yong Yu. 2017. "Wasserstein Distance Guided Representation Learning for Domain Adaptation." Proceedings of the AAAI Conference on Artificial Intelligence 32(1).
Siddharth, N., Brooks Paige, Jan-Willem van de Meent, Alban Desmaison, Noah D. Goodman, Pushmeet Kohli, Frank Wood, and Philip H. S. Torr. 2017. "Learning Disentangled Representations with Semi-Supervised Deep Generative Models." Advances in Neural Information Processing Systems 2017-Decem:5926–36.
Sun, Baochen, Jiashi Feng, and Kate Saenko. 2016. "Correlation Alignment for Unsupervised Domain Adaptation." Pp. 153–71 in Advances in Computer Vision and Pattern Recognition. Cham: Springer London.
Toda, Yosuke, Fumio Okura, Jun Ito, Satoshi Okada, Toshinori Kinoshita, Hiroyuki Tsuji, and Daisuke Saisho. 2020. "Training Instance Segmentation Neural Network with Synthetic Datasets for Crop Seed Phenotyping." Communications Biology 3(1):1–12. doi: 10.1038/s42003-020-0905-5.
Wang, Qi, Junyu Gao, Wei Lin, and Yuan Yuan. 2019. "Learning from Synthetic Data for Crowd Counting in the Wild." ArXiv Preprint.
Wang, Xu, Hong Xuan, Byron Evers, Sandesh Shrestha, Robert Pless, and Jesse Poland. 2019. "High-Throughput Phenotyping with Deep Learning Gives Insight into the Genetic Architecture of Flowering Time in Wheat." GigaScience 8(11):1–11. doi: 10.1093/gigascience/giz120.
Ward, Daniel, Peyman Moghadam, and Nicolas Hudson. 2018. "Deep Leaf Segmentation Using Synthetic Data." ArXiv.
Wrenninge, Magnus, and Jonas Unger. 2018. "Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing." ArXiv Preprint.
Wu, Zuxuan, Xin Wang, Joseph E. Gonzalez, Tom Goldstein, and Larry S. Davis. 2019. "ACE: Adapting to Changing Environments for Semantic Segmentation." Pp. 2121–30 in Proceedings of the IEEE/CVF International Conference on Computer Vision.
Yoon, Jaehong, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. 2017. "Lifelong Learning with Dynamically Expandable Networks." ArXiv Preprint.
Zhang, Pan, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. 2021. "Prototypical Pseudo Label Denoising and Target Structure Learning for Domain Adaptive Semantic Segmentation." ArXiv.
Zhou, Joey Tianyi, Le Zhang, Du Jiawei, Xi Peng, Zhiwen Fang, Zhe Xiao, and Hongyuan Zhu. 2021. "Locality-Aware Crowd Counting." IEEE Transactions on Pattern Analysis and Machine Intelligence 1–1. doi: 10.1109/TPAMI.2021.3056518.
Zhu, Xiaojin (Jerry). 2005. Semi-Supervised Learning Literature Survey. University of Wisconsin-Madison

Department of Computer Sciences.
Zhuang, Fuzhen, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. 2019. "A Comprehensive Survey on Transfer Learning." Proceedings of the IEEE 109(1):43–76.