Deep learning applied to multi-component imagery for variety testing problems

Hadhami Garbouge

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

université angers

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ D'ANGERS
COMUE UNIVERSITÉ BRETAGNE LOIRE

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *(voir liste des spécialités)*

Par

« Hadhami GARBOUGE »

« Deep learning applied to multi-component imagery for variety testing problems »

Thèse présentée et soutenue à « INRAe - Angers », le « 29 November 2022»
Unité de recherche : LARIS - Laboratoire Angevin de Recherche en Ingénierie des Systémes
Thèse N° :

Rapporteurs avant soutenance :

Pr Christian GERMAIN	IMS, Bordeaux, France
MCF-HDR Frédéric COINTAULT	Agrosup, Dijon, France

Composition du Jury :

Président :	Pr. Julia BUITINK	INRAe, Angers, France
Dir. de thèse :	Pr. David ROUSSEAU	LARIS-INRAe, Angers, France
Co-encadrant :	Dr. Pejman RASTI	LARIS-CERADE, Angers, France
Co-encadrant :	Dr. Natalia SAPOUKHINA	INRAe, Angers, France

Invité(s) :
Philippe VERMEULEN,CRA-W, Gembloux, Belgique
Pierre ROUMET, INRAe, Montpellier, France

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my supervisor, Pr. David ROUSSEAU for his patience, motivation, immense knowledge and his continuous support during three years of my Ph.D study and related research,. His guidance helped me in all the time of research, from my first internship for my master's degree to the writing of this thesis. Thank you again for all the opportunities you have given me during the five years we have worked together.

I greatly appreciate my co-supervisor, Dr. Pejman RASTI, for his excellent feedbacks, encouragement, and guidance. It was such a pleasure working with you on several research works. Also, I could never thank my second co-supervisor, Dr. Natalia SAPOUKHINA, for all your support and advice, especially during the writing of this manuscript. I will never forget your special hug during my difficult moments.

I am deeply grateful to my previous colleagues from GEVES and my partner in the INVITE project, Nicolas MASCHER, for helping me install and manage the data acquisition system that I used in my work and Didier DEMILLY for his insightful comments and suggestions during our monthly meetings.

I would like to extend my sincere thanks to Valérie CADOT from GEVES for the great year, I enjoyed working under your supervision at GEVES and making great trips to collect data, I was very pleased to continue collaborating with you on my thesis.

Thanks should go to all my friends and colleagues in INRAe Angers, Polytech Angers, and especially the ImHorPhen team, including Mouad ZINE EL ABDINE, who shared the office with me for three years.

I would like to thank the H2020 European project INVITE for financing this PhD. Also, our partners in the INVITE project, with special thanks to Philippe VERMEULEN from CRA-W Belgium, for your help in collecting database.

I would like to express my sincere gratitude to my mother institute ISET'Com in Tunisia, for the high level of academic skills they provided to me, especially Dr. Amin ZRIBI, who offered me the opportunity to start my first steps in scientific research here in France, also Dr. Belgacem AOUDI for all his recommendations and believing on me to be able to become Dr. Hadhami.

My father, my all, you left us so early, but you are always present with us in our hearts. I wish you could be with me on two particular occasions in my life, my wedding ceremony, and my thesis defense, but 'Alhamdulillah,' there is a secret I should reveal today. I hope you are proud of me where you are.

My mother, the best mother in the world, without you, I would never have gotten here, a huge thanks to you, may God bless you and protect you forever.

My partner, my half, Ismail 'Pa', I cannot find the right words to say thank you for leaving ten years of your career in Tunisia, joining me and starting a new life from scratch, for your patience, and for being with me during difficult and fun moments, and for the sacrifices you have made to me to pursue, I'm grateful to be Hadhami GARBOUGE OUSSAIFI. I am very blessed to have you in my life.

My two brothers, Mohamed Helmi, and my sister Hafidha, you are the best gift in my life, thank you for always being with me and encouraging me to never give up. I can not forget my loves Aziz, Ayoub, Adem, Haroun and my new sisters Emna Maamouri, Karima OUSSAIFI, my brothers Mohamed OUSSAIFI and Mohamed KOCHTAN.

Special thanks to my new family, the OUSSAFI family, I have the pleasure of becoming a member of your family. Especially, my second father Ali and mother Baya for your hospitality and given love which has encourage me this last period.

I would like to express my sincere gratitude to my cousins, Dr. Malek GARBOUGE and Mariam GARBOUGE, for their support and help since I decided to continue my studies here in France.

Many thanks go to all my friends who are considered part of my family: Fatma,

Marwa, Maha, Nawel, Angélina, Audrey, Therese, Aladdin, Ibrahim, and Khalil.

Last and foremost, I would like to thank "Allah" for blessing me and helping me to choose the right path.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

To commercialize a new variety of agricultural or vegetable species in the European Union, a plant breeder has to follow a process managed by a national authority and delegated to an examination office (EO) that will describe and evaluate the variety for its registration on the national catalog (Figure 1.1). The national catalogs of all the EU Member States (MS) are compiled by the European Commission to form the Common Catalog allowing the variety to be marketed throughout the EU. Evaluation results, including variety descriptions, also grant Plant Variety Rights (PVR) both at the national and European level and for specific crops for the certification of seed lots. Depending on the MS, the legal mandate of a national examination office (EO) covers part or all of these missions. According to this framework, the EOs run field tests either under the supervision of their competent national authorities or upon request of the Community Plant Variety Office (CPVO)[1] in charge of granting PVR on the territory of the EU.

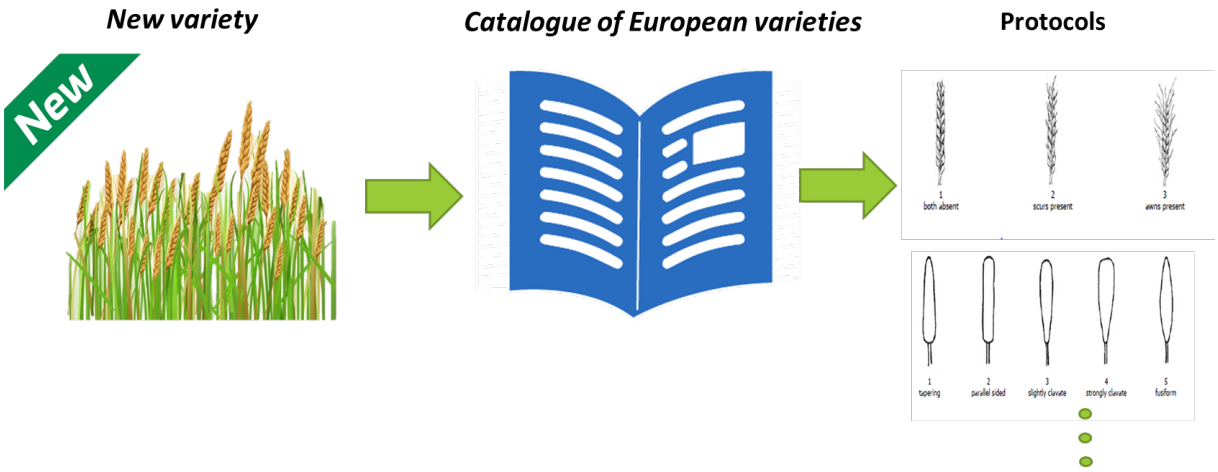


Figure 1.1 – The process of registration of a new variety in the European catalog.

Most of these tests are based on manual measurements performed from visual inspection. This method has consequences in terms of efficiency because of the time-consuming

nature of these tests. It is also an issue for the tests' reproducibility when some characteristics are based on qualitative features, suffering from subjectivity in their assessment. Improving the efficiency and reproducibility of these observations would be extremely useful for EOs continuously seeking optimized testing methods implemented in testing protocols. It could also provide means to assess new characteristics developed in response to new agricultural constraints, particularly from the perspective of climate change. In addition, more efficient measurement methods would assist in addressing the challenge of the constant increase in the number of varieties that must be tested. More reproducible measurements would also contribute to harmonizing practices between European EOs (supporting, for example, the use of historical data to predict the expected behavior of varieties toward different climatic scenarios). The described challenges encourage us to head toward using sensors and numerical practices to progressively replace classical manual methods of examination whenever there is a need to speed up a measurement or increase their reproducibility and objectiveness. The trend of using more and more imaging for plant science started some decades ago and has been extensively reviewed (see [2, 3] for the most recent ones), including cost-effective strategies [4]. While imaging modalities used in plant science and variety testing may be similar, the types of measures in plant science and variety testing differ by their nature or technical aspects. So far, little attention from the academic imaging community has focused on these specific aspects of variety testing. Variety testing is performed among networks of offices and has to be accessible to breeders. Consequently, measurements should rely on cost-effective technologies that can easily be replicated. The introduction chapter is organized as follows. After explaining the variety of testing specificities, we propose a rationale for selecting characteristics that may benefit the most from using low-cost imaging systems. This rationale is illustrated in crops of significant interest in the food industry, such as wheat, maize, sunflowers, and tomato. We then propose possible technologies for the measurement of these characteristics. We conclude by pointing toward the needs, challenges, and opportunities for deploying the low-cost imaging system in various testing protocols. Finally, we end up with a list of our methodological contributions and explain the structure of the PhD manuscript.

1.1 Variety testing specificities

1.1.1 DUS: Distinctness, Uniformity and Stability

Two types of evaluation are mainly performed for a variety testing. **DUS** [5] tests (for **D**istinctness, **U**niformity, **S**tability) are conducted to ensure that a new variety is distinct from existing varieties, that it is sufficiently uniform in its characteristics, and that the variety is stable with consistent phenotypic characteristics from one generation to the next. For most species, these tests are harmonized worldwide by UPOV (Union Pour la Protection des Obtentions Végétales) members. They are carried out according to standardized technical protocols (CPVO TPs), based on UPOV guidelines, and using reference plant material provided by the breeders. For example, morphological features and color are mostly used for agricultural crops and phenological features such as flowering and ripening phases. Some species are also tested for disease resistance. This produces a « variety description » (VD) which forms the identity card of the tested variables. The VDs are also used -as one tool amongst others- to enforce the PVR to which they are associated. Thanks to the harmonization of the guidelines, the members of the **UPOV**[6] convention may (if they wish to) accept **DUS** reports established by another UPOV member (meaning that another UPOV member can use a given **DUS** report established in one UPOV member as a basis for a decision to grant a PVR, without the breeder having to pay again for the same field tests but only an administrative fee of Swiss Fr 350). In terms of data processing, **DUS** measurements correspond to a classification problem. Deciding to classify is by nature a non-linear problem. Consequently, it can be done with non-linear sensors and may not need fully linear and calibrated sensors. What needs to be calibrated is the performance of the classification, but this classification can be done on possibly distorted data, provided distortion does not degrade classification performance. This means that **DUS** can, by nature, benefit from low-cost imaging systems.

1.1.2 VCU: Value for Cultivation Use

The second type of evaluation is **VCU** [5] tests (for **V**alue for **C**ultivation, **U**se which are performed for many agricultural crops. These tests aim to evaluate the variety's suitability for growing in local agro-climatic conditions and the technical value of the harvest e.g., protein, oil content,.... To qualify for registration, the new variety must have an « added value » in the country where it is evaluated. This is established by comparing

it to a set of existing reference varieties over two testing cycles of 5 to 20 trials per year. Unlike **DUS**, **VCU** measurements are not harmonized among the countries. Also, in terms of data processing, **VCU**, corresponds to a regression. It is more demanding in terms of precision and less likely to benefit from low-cost imaging systems and will not be addressed in this PhD. This choice here does not mean that **VCU** would be less important than **DUS**, but rather that **DUS** is more straightforward to address with low-cost systems than **VCU**. Also, **VCU** characteristics have received relatively more attention than **DUS** from the imaging community for their applications in yield assessments or their value as input data in crop models. For all these reasons, we focus more on **DUS** characteristics in this PhD.

1.2 A rationale to identify most promising characteristics in DUS protocols

Assessment of **DUS** characteristics for each crop is explained in the UPOV Test Guidelines. This constitutes thousands of traits. Switching current manual practices to numerical practices will require a lot of time and effort. In this section, we propose a rationale to select the most promising characteristics to start the work. We first give the different types of measurement which are performed in **DUS**.

For the registration of new varieties, two modes of observation are currently performed. The first is visual observations (**V**) which rely on the expert's judgment. It includes observations where the expert uses reference points (e.g., diagrams, example varieties, side-by-side comparison) or non-linear charts (e.g., color charts). Visual observations can also include sensory observations of the experts (smell, taste, and touch). The second type is the measurement (**M**), which corresponds to objective observations relative to calibrated linear scales, e.g., using a ruler, colorimeter, dates, counts, etc. These two types of observations can be recorded as a single record for a group of plants or parts of plants (G) or may be recorded as records for many single, individual plants or parts of plants (S). Therefore four possible combinations are found in **DUS** protocols: **VG**: Visual assessment by a single record per group of plants or plant parts for the evaluation of distinctness; **VS**: Visual inspection by individual records for each plant or plant parts; **MS**: Assessment by measurements and separate records for each plant or plant parts for the assessment of distinctness; **MG**: Assessment by measurement and one record per group of plants or plant parts of the evaluation of distinctness.

Based on the four different types of measurement in **DUS**, we can consider that quantitative characteristics are the ones that suffer less from subjectivity in a classical human visual inspection. They are therefore suitable for translation in automated, sensor-based protocols, which can be compared with standard protocols. Among these, the most difficult objective characteristics are those attached to the assessment of dynamical processes (emergence, time of flowering, ...). The difficulty comes from the fact that evaluation can currently be carried out only for a fixed and limited number of time points. Having continuous recording would possibly improve the accuracy of such monitoring. Second, one can focus on characteristics common to different crops so that the development of a sensor can serve several usages. Third, characteristics that are laborious to access in proxy detection, such as plant height, or ear size (especially in the field and for large crops at the mature stage such as maize), could be assessed much faster with remote sensing technologies, such as UAVs and high-resolution cameras. At last, the quantification of characteristics, which could be measured simultaneously with a single snapshot acquisition, such as diameter, length, number of grains, and shape, would also be accelerated with imaging systems.

Following the rationale described above, a list of characteristics to be chosen in priority can be extracted from the UPOV Test Guidelines. For illustration, we applied this rationale to four crops of significant importance to the food industry and came up with the shortlist in Table 1.1.

	Organ	**characteristic**	**Description of the characteristic**	**Scale of observation**	**Visual(V)/Measure(M)**
Wheat	Plant	Length	short ->long	MG	M
	Ear	Length	short ->long	MS	M
Maize	Plant	Length	short ->long	MS	M
	Plant	Length	short ->long	MS	M
	Ear	Length	short ->long	MS	M
	Ear	Diameter	small ->large	MS	M
Sunflower	Time	Time of flowering	very early ->very late	MS / MG	M
	Plant	Natural height	very short ->tall	MS / MG	M
Tomato	Plant	Height	short ->long	VG / MS	V/M
	Fruit	Time of flowering	early ->late	MS	M

Table 1.1 – Most promising characteristics proposed for the four crops taken for illustration.
MS: Assessment by measurements and individual records for each plant or plant parts for the assessment of distinctness; MG: Assessment by measurement and one record per group of plants or plant parts for the assessment of distinctness.

1.3 Challenges for affordable imaging systems dedicated to most promising characteristics

Imaging devices are nowadays largely available at low-cost, and are embedded in connected objects such as smartphones, tablets, or mini-computers which have been largely reviewed in the recent literature [7, 8, 9, 10, 11, 4, 12, 13, 14, 15, 16]. These imaging systems and connected objects can be fixed on various devices such as Unmanned aerial vehicles (UAV), unmanned ground vehicles or connected sticks (see Fig. 1.2). To translate the current variety testing protocols into sensor-driven protocols, it would be more strategic to provide ergonomic systems directed carried by the variety testers. In this PhD, we will mostly deal with handy light cameras.

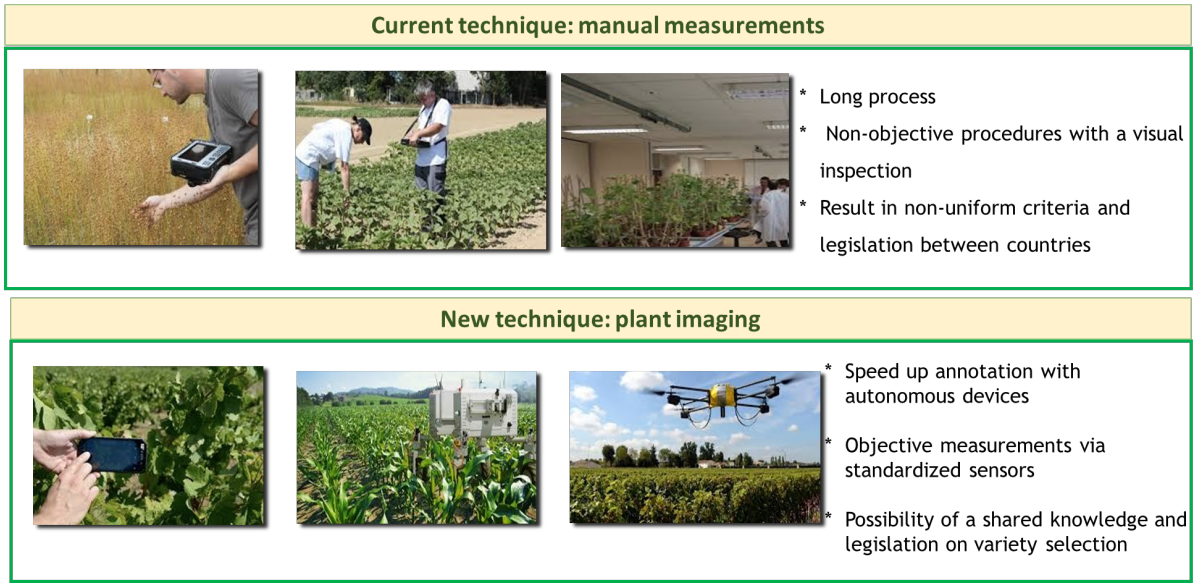


Figure 1.2 – The current practices in variety testing and possible outcome of more numerical practices.

Some affordable sensors are already available for the most-promising characteristics to be measured in the field. For repeated event measurements (e.g., monitoring of dynamic traits), time-lapse (TL) camera systems may help as they can acquire images over larger periods without user interaction. Such cameras are available off-the-shelf like Wild-Vision cameras [17], originally designed as animal photography traps but also capable of delivering TL image series forr nature monitoring [18]. Modern DSLR cameras are equipped with internal TL mode or may be triggered with commercial external intervalometers.

Finally, mini-computers like Raspberry Pi or micro-controllers like Arduino may also be used as intervalometers. For length annotation and measurements, there exists a bunch of applications for smartphones. The application scenarios range from annotation like ImageMeter Pro [19] to measurements of length and areas like in Smart Measure [20], Smart Measure Tool Kit [21], partly using augmented reality (AR) methods for measurement and display, e.g. Measure Tools AR ruler [22] and EasyMeasure [23].

The limitation of all these available technologies for variety testing is primarily due to image processing. Although a wide range of image processing software has been developed, for an overview, see [24, 25], only a minimal selection of these softwares is exactly following the protocols of variety testing [26, 27]. Moreover, the available software particularly dedicated to variety testing [26, 27] only focuses on post-harvest assessments in controlled environments.

1.4 Contributions

We have highlighted the interest in developing accessible imaging acquisition systems and image processing algorithms to accelerate and increase the objectivity of assessed characteristics in variety testing. Considering the massive amount of traits to be measured in variety testing, we proposed a rationale for selecting the automatable traits for low-cost sensors. While several low-cost sensors and efficient machine learning algorithms are available, the remaining challenge is to design ergonomic imaging systems assisted by a processing software. This Ph.D. was funded by the European project INVITE H2020(https://www.h2020-invite.eu/). One of this project's objectives was to propose ergonomic vectors and sensors with associated software to address numerically some of the selected characteristics as identified in this introductory chapter. While achieving this task, our work contributed methodologically to machine learning and instrumentation, as described in this section.

Most of the recent literature in image processing now relies on artificial intelligence (AI) approaches like Convolutional Neural Networks (CNNs) deep learning [28]. With this machine learning technique, both features and decision-making are learned simultaneously. This approach, which has been successfully applied in all domains of computer vision, including plant imaging [29, 30, 31], has produced state-of-the-art performances for all image processing tasks. Standard deep neural networks are now accessible to address many types of problems like for image classification [32, 33] , for object recognition [34,

35], for segmentation [36, 37].

Deep learning is now used worldwide in almost all domains of image analysis as an alternative to traditional purely handcrafted tools. Nevertheless, as recently declared by Yann Lecun, one of the field pioneers, "Deep learning is not suitable for all applications." A requirement for good deep learning applications is the possibility to produce an extensive annotated database (typically at least thousands). This is not the case in all domains. For instance, assembling thousands of patients in medical imaging is a heavy task. Also another requirement is to have images including complex structures in which the depth of the neural network will integrate some context that would not be easily modeled with simple geometrical shapes. Again, this is not the case in all domains. For instance, in the industrial vision where manufactured shapes are to be controlled, the variety of shapes might be minimal and does not systematically require to resort to deep learning, while some 3D CAD models of the object to be controlled exist and may be helpful for classical handcrafted tools.

Plant science, in this context, is one of the especially well-adapted applications for deep learning. Firstly, it involves many biological variables, such as growth, response to biotic and abiotic stress, and physiology. Secondly, plants display huge variability (e.g., size, shape, color), and the consideration of all these variables surpasses the human capacity for software development and response to the needs of plant scientists. Thirdly, thanks to phenotyping systems or the use of robots in the field, the throughput of image acquisition is relatively high, so the observed large population of plants can meet the needs of big data required for effective deep learning.

Deep learning promises to offer universal algorithms for definite informational tasks. Mainly three informational tasks can be found: classification, object detection, and segmentation. This builds a corpus of ready to be used tools to address variety testing tasks such as the ones addressed in this PhD: phenological stages determination, which can be seen as a classification of images in a time series, flowering which can be seen as an object detection task, and disease rating which can be seen as a segmentation task. Although codes to address these tasks are now publicly available, some challenges are still rising when one targets real world implementation.

First, most of the literature on deep learning has been derived from RGB images corresponding to standard resolution. In plant imaging, one may get interested in adding more components such as Depth which is very contrasted for plant [9]. Depth comes in sensors at a low cost and offers interesting LIDAR contrast about the spatial structure of

the plant. How to fuse the information of such a low-cost Depth map with RGB images in neural networks for plant imaging is a challenge we address in this PhD. For this purpose, an entire setup of field and in-door RGB-Depth imaging system has been constructed from scratch in the PhD. A specificity of plants is their continuous growth. Following this growth process can be done with time-lapse from fixed cameras. Several families of neural network architectures have been designed for the process of time-lapse. We compare and discuss them in the Phd on a variety testing use case. This includes some methods (transformers [38]) from neural language processing tested for the first time in plant imaging.

No public data sets were available for a variety testing to benefit from the opportunity opened by deep learning. While we produced some original annotated data set in this PhD we also investigated ways to benefit from prior trained models on other annotated data. Indoor experiments similar to those carried out in the field by variety testing have been carried out for decades in plant phenotyping centers. We investigated, for the first time to the best of our knowledge, the possibility of performing transfer learning from data acquired in a controlled environment to similar data acquired in a non-controlled environment. When already existing closely-related real data are not available, another approach can be to use synthetic data set automatically annotated. Some approaches have been designed with several modeling, or generative models in our laboratory [39]. However, the generation of the synthetic model can be time-consuming, and we investigated the possibility of taking benefit of virtual environments from video gaming to perform pretraining.

Finally, the contrast in plant imaging can show subtle spectral details, not optimally captured by standard RGB images. This is especially the case for plant diseases. From the literature [40, 41], it is not yet clear if the gain of contrast observed in indoor controlled conditions is kept in field conditions. We focus on this matter and design a multispectral imaging and associated machine learning system in an end-to-end fashion for a plant disease use case in variety testing.

1.5 Structure of the document

The document follows the description of the resolution of the challenges listed in the previous section. In chapter 2, we investigate the value of RGB-Depth fusion for two variety testing characteristics: on individual seedlings along time-lapse and small

parcels of overlapping plants in snapshot images. In Chapter 3, we focus on the question of transfer learning from indoor data to outdoor conditions illustrated in the seedling emergence problem of chapter 2. We also investigate a transfer learning from synthetic to real data on another variety testing characteristic. In chapter 4, we move to other multi-component images with multispectral images. We present another way of transfer, which is the transfer from a hyper-spectral camera used indoor to a multispectral camera used in the field. We propose a global pipeline from the most effective wavelengths to detect wheat diseases and build a multispectral camera. We point toward perspectives in the conclusion chapter detailed in the annex A, on the design of automatic acquisition protocol in variety testing. As a disclaimer, the document includes contributions articulated around various testing characteristics and addressing the specific challenges identified for the progress toward more numerical practices. State-of-the-art and most related works are to be found in each chapter and are not centralized in a bibliographic chapter.

Chapter 2

Low cost sensor and RGB-Depth fusion for variety testing

In this chapter, we focus on two characteristics prioritized in the introduction chapter. First, we tackle plant emergence in controlled environment. Second, we tackle the development stage of wheat heading in the field. We propose to perform measurements using a RGB-depth sensors, the Intel® RealSense D435 [42]. The D435 stereo camera is part of the new D400 series of depth cameras featuring the Intel® RealSense™ D4 vision processor. In a very compact and lightweight and rather low-cost format, Intel® RealSense combined a depth sensor with an RGB sensor. We introduce the implementation and original algorithms associated with the fusion of the RGB and depth components. The detail of the sensor and how it was assemble in a network together with the soft developed to pilot and pre-process the data are described in Annex B; The aim is to investigate on both characteristics the added value of the depth component by comparison with a sole RGB image. The study was published in [43].

2.1 Seedling growth

The detection of the seedling growth stages is a fundamental problem in plant science. This covers the emergence of seedling from the soil, the opening of cotyledons and appearance of the first leave which correspond to the earliest stages of development of plant. The success or failure of these developmental stages and their kinetics have a huge impact on the evolution of the future plant. Recently, seedling growth monitoring has received attention from the computer vision community [44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59]. Among these works, the state-of-the-art approach based on deep learning proposed in [59] has shown the possibility to automatically classify the stages of development of seedling with RGB sequences of images from top view with an accuracy higher than 90%.

One of the limitations of the work proposed in [59] is that the monitoring was done only during daylight with RGB images. Consequently, any events happening during the night would be missed and/or possibly estimated with a temporal bias. In this work, we propose an extension of the work of [59] and investigate the possibility to push forward the monitoring of the seedling growth during the day and the night. To this purpose, RGB-Depth camera were used. These technologies have been demonstrated of wide value in plant phenotyping [9, 60, 7, 61, 62, 63, 64, 65]. The depth images are computed by an active LIDAR camera operating in infrared (IR). This camera can be activated during day and night without impact on the development of the plants. As in [59] we selected low-cost versions of these RGB-Depth cameras. These low-cost constraints are specially important in plant phenotyping [4] when moving the plants or the camera is not an option and that replication of cohorts of cameras is to be chosen to monitor large populations of plants. Low-cost RGB-Depth cameras are also coming with artifacts and noise. Such artifacts and metrological limitations of low-cost RGB-Depth cameras have been extensively studied (see [66] for a recent survey). In our case, we rather work at an informational level. We focus on a classification task, i.e. a nonlinear decision, which is by nature more robust to noise since it does not have to provide a high-fidelity, metrological, linear estimation. The hypothesis investigated in this study is that these low-cost RGB-Depth sensors despite their limited spatial resolution and the presence of artifacts may be of enough value to enhance the tracking of seedling growth during day and night.

We demonstrate, for the first time, to the best of our knowledge the value of these RGB-Depth images to monitor the early stages of seedling growth. We investigate fusion strategies between RGB and depth with several neural network architectures. The underlying motivation to use multimodal data is that complementary information give a richer representation that may be utilized to create better results than a single modality. The multimodal fusion research community has made significant progress in the past decade [67]. Different fusion strategies have been reviewed [68, 69]. Specifically for RGB and Depth with deep learning architectures, fusion has been extensively studied in the literature [70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81]. Mainly two types of fusion can be distinguished. First, images can be stacked at the input: this is the early fusion [70, 71, 72, 73, 74], that we call image fusion. Second, deep features can be independently extracted and then fused before a classification stage: this is the feature fusion [75, 76, 77, 78]. In this work, we investigate these fusions scenarios that we applied to the important problem of seedling growth stage monitoring. Since we process sequences of images we

considered time-dependent neural network architectures. As in [59], we included a base line convolutional neural network (CNN) and LSTM [82]. We also added TD-CNN GRU [83] and transformer [38] which were not included in [59].

2.1.1 Materials and methods

- **Imaging system and data set**

We have conducted similar experiments as the ones described in detail in [59] and shortly recalled here. A set of minicomputers, connected to RGB-Depth cameras [84], was used to image seedlings from the top view as illustrated in Fig.2.1. We used, instead of the RGB cameras of [59], Intel real sense cameras [42] (model D435) which natively produces registered RGB-Depth pairs of images and calibrated Depth maps. We installed eight of these RGB-Depth cameras in a growth chamber where cameras followed the growth of seedlings from top view. During experiment, soil pots were hydrated to saturation for 24h after which excess water was removed. After 24h, seeds were sown at a depth of 2 cm, and trays were placed in a growth chamber at 20°C/16°C, with 16h for photoperiod at $200\mu Mm^{-2}s^{-2}$. The soil was kept wet throughout the experiments. Each experiment took one week with a frame rate of 15 minutes. The time lapse program (made in Python) was implemented on a central minicomputer controlling, via ethernet wires, the eight minicomputers connected to the RGB-Depth cameras.

Concerning the biological material, seedling growth was recorded for two experiments using seed lots from different accessions of beans such as Flavert, Red Hawk, Linex, Caprice, Deezer and Vanilla. Each experiment consisted of 3 trays with 40 pots in which 120 seeds of accessions were sown. There is a similarity between the species in this experiment and the two species which were used in [59] as all of them consist in dicotyledon species. The main difference between them comes from the number of varieties in this experiment which is three times higher than the one in [59].

In total, the database consists of 72 temporal sequences of RGB and depth images of size 66×66 pixels where each temporal sequence consists of 616 individual images. Example of images from the database is shown in Fig. 2.1. RGB-Depth temporal sequences acquired during daylight were annotated by expert in biology while looking at RGB images. This ground-truth annotation consisted of four classes: soil, first appearance of the cotyledon (FA), opening of the cotyledon (OC), and appearance of the first leaf (FL). The algorithms presented in this work for seedling emergence identification following these

Table 2.1 – Description of the RGB-Depth dataset used in this study.

	Species	**No.of temporal sequences**	**Totale No. of images during days**	**Totale No.of images during nights**	**Totale No.of all images**
Training dataset	Flavert	10	4240	1920	36960
	Red Hawk	10	4240	1920	
	Linex	10	4240	1920	
	Caprice	10	4240	1920	
	Deezer	10	4240	1920	
	Vanilla	10	4240	1920	
Validation dataset	Flavert	1	424	192	3696
	Red Hawk	1	424	192	
	Linex	1	424	192	
	Caprice	1	424	192	
	Deezer	1	424	192	
	Vanilla	1	424	192	
Testing dataset	Flavert	1	424	192	3696
	Red Hawk	1	424	192	
	Linex	1	424	192	
	Caprice	1	424	192	
	Deezer	1	424	192	
	Vanilla	1	424	192	

four phases of growth were trained, validated, and tested against this human-annotated ground-truth. In order to train robust models, we used the cross-validation approach by considering image sequences of bean varieties in three split of train, validation, and test dataset. Table 2.1 provides a synthetic view of the data set used for training and testing of the models. For the training dataset, we applied data augmentation using a simple horizontal flip on each temporal sequence.

Depth images can contain artifacts with missing values. This can happen on part of the scene where not enough light is reflected or for objects that are too close or too far from the camera. While neural networks should be able to cope with such noise, it is better to correct them to use the capability of these networks on clean data. In order to correct these artifacts, we applied a classical inpainting technique [85] of depth images to reduce the noise.

- **RGB-Depth Deep learning fusion strategies**

We describe here the different neural network architectures tested in this study to fuse the RGB and Depth for the classification of seedling growth stages as depicted in Fig. 2.2

CNN-based image early fusion learning structure

We first integrated, as in [86], RGB and Depth data stacked in a four-channel as input to a CNN (see Fig. 2.3.a). The feature extraction block from four-channel input images is followed by the classification block (shown in Fig. 2.3a). The CNN architecture is the one of [59, 83] that we shortly recall. The feature extraction block of a CNN model is

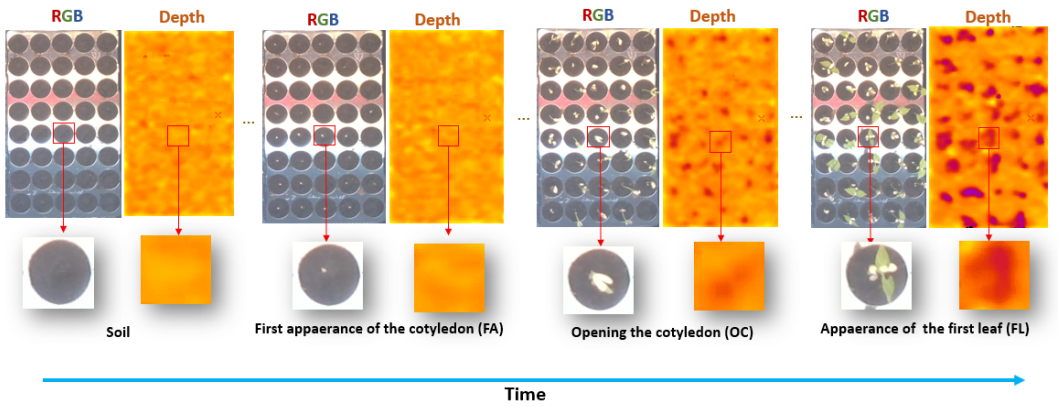


Figure 2.1 – Overview of the time-lapse collected for this work. Upper row, view of a full tray with 72 pots from top view. Lower row, a zoom on a single pot at each stage of development to be detected from left to right: soil, first appearance of the cotyledon (FA), opening the cotyledons (OC) and appearance of the first leaf (FL).

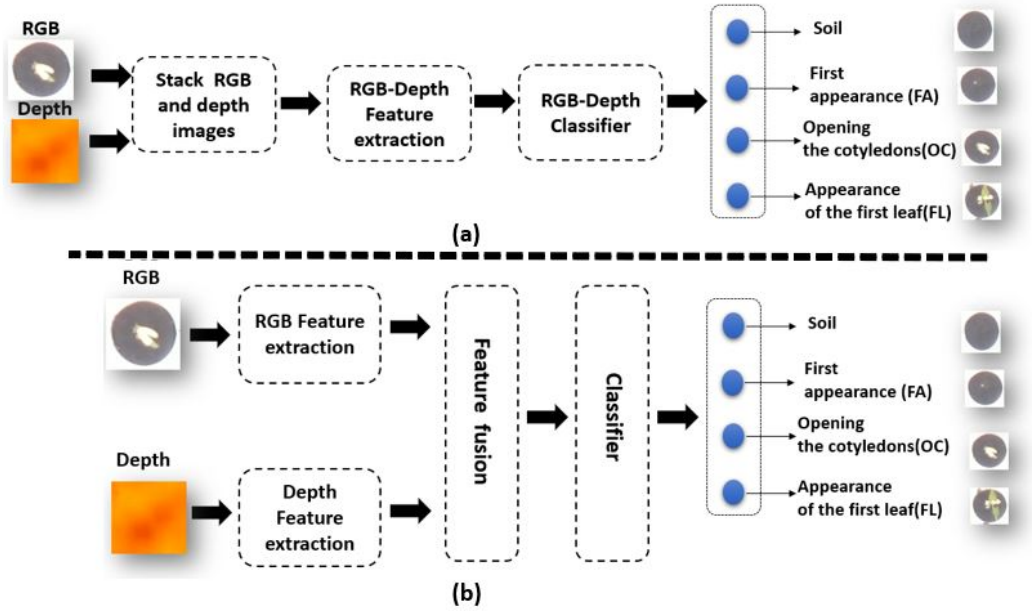


Figure 2.2 – Different types of RGB-Depth fusion architectures tested in this work for image classification. (a) Image-based RGB-Depth fusion. (b) Feature-based RGB-Depth fusion.

responsible for extracting features from input images using convolutional layers, whereas the classification block determines classes. To keep the amount of train parameters low, we used an AlexNet [32] like CNN structure. This architecture reads as follows: four convolutional layers with filters of size 3×3 and respective numbers of filters 64, 128, 256,

and 256 each followed by rectified linear unit (RelU) activations and 2×2 max-pooling; a fully connected layer with 512 units, ReLU activation and dropout (p=0.5) and a fully connected output layer for four classes corresponding to each stage with a softmax activation. This proposed CNN architecture has been optimized on a hold-out set and was demonstrated in [59] to be optimal by comparison with other standard classical architectures (VGG16, ResNet, DenseNet). The network was trained from scratch since the size of the input tensor (4 channels and small spatial resolution) was different from existing pre-trained networks on large RGB data sets.

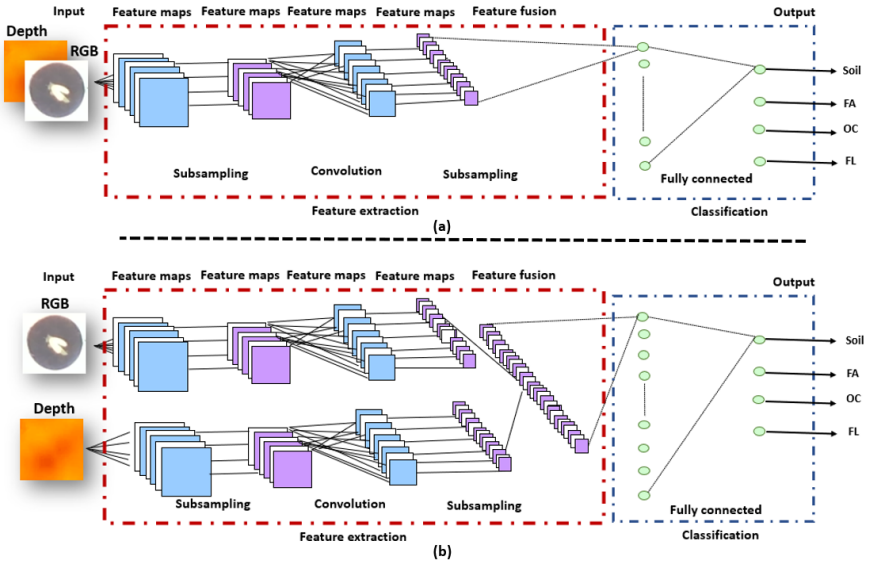


Figure 2.3 – (a) CNN architecture of image fusion for RGB-Depth. (b) CNN architecture of features fusion for RGB-Depth.

CNN-based feature fusion learning structure

Our architecture, shown in Fig.2.3.b, is made up of two convolutional network streams that operate on RGB and Depth data, respectively. The same structure of image fusion CNN has been developed for each stream of the feature fusion CNN. The feature extractor part of the CNN architectures of RGB and Depth images consists of four convolutional layers which have 64, 128, 256, and 256 filters, respectively (similar to the AlexNet like structure of the previous subsection). The ReLU activation function is considered for each convolutional layer followed by a max-pooling layer. On the classification part of the CNN architectures, a fully connected layer with 512 units, and an output layer with four

neurons corresponding to each event with a softmax activation function.

TD-CNN-GRU-based image and feature fusion learning structure

We demonstrated in [83, 59] the possible added value to embed in controlled environment a memory in the process of the sequence of images. We demonstrated in [83], the superiority of Time dependent CNN with gated recurrent units (TD-CNN-GRU) by comparison with other memory based methods such as long short term memory (LSTM) and CNN-LSTM architectures. GRU uses two gates: the update gate and the reset gate while there are three gates in LSTM. This difference makes GRU faster to train and with better performance than LSTMs on less training data [87]. The same CNN architecture of our model in [59] was embedded in our TD-CNN-GRU model where the optimal duration of the memory was found to be 4 images in [83, 59] corresponding to 1 hour of recording. Fig.2.4 shows a schematic view of the proposed TD-CNN-GRU for images and feature fusion respectively.

Transformers-based image and feature fusion learning structure

A last class of neural network dedicated to time series are the transformers. Since their introduction in [38] they have been shown to outperform recurrent neural networks such as LSTM and GRU specially in the field of natural language processing as they do not require that the sequential data be processed in order. Transformers have been shown suitable to process temporal information carried by single pixels in satellite images time series [88, 89, 90]. Transformers have recently been extended to the process of images [91] where images were analysed as a mosaic of subparts of the original images creating artificial time series. In our case, we directly have meaningful original images which corresponds to the field of view of the pots. We, therefore, provide the transformer of [91] with time series of consecutive images of the same pot (we used the same time slot as in the other spatio-temporal methods). We used 32 transformer layers with batch size 64, feed forward layer as classification head layer and the size of our patch size was equal to 66×66 pixels for both architectures of Fig. 2.5.

For all our training, we used the NVDIA DGX station. This station is composed of 4 GPUs and each one of them have a RAM memory of 32 Gb. We used Python version 3.7.8, Tensor-flow version 2.7.0 and Keras library version 2.3.1.

- **Accuracy**

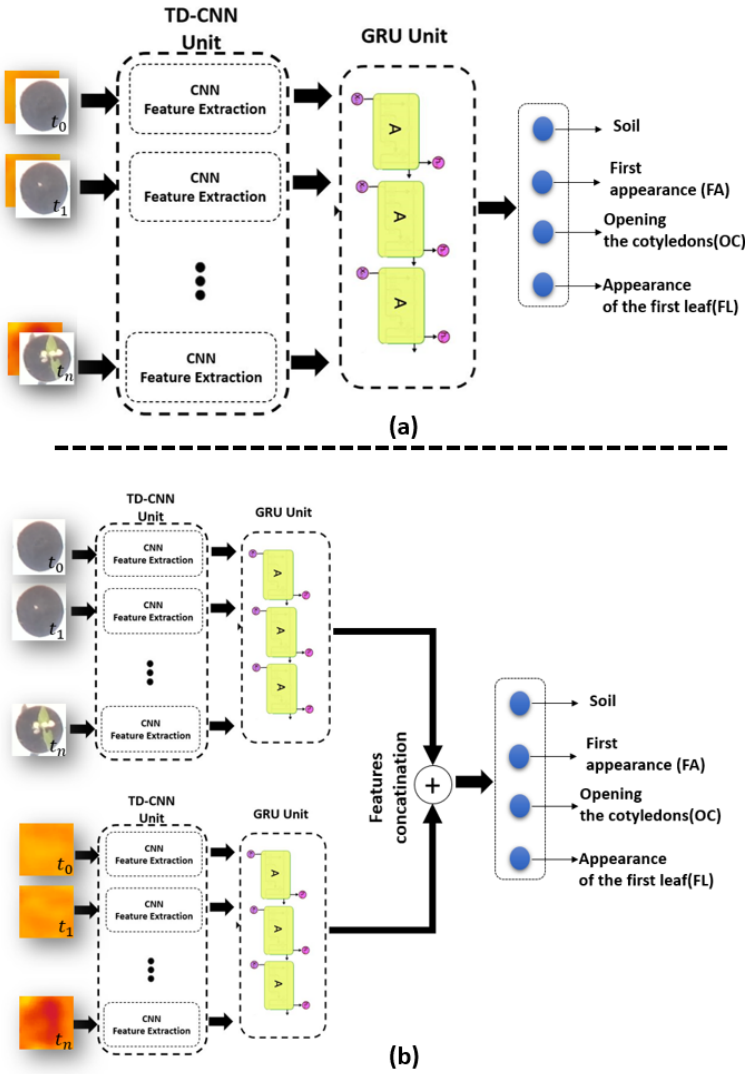


Figure 2.4 – (a) TD-CNN-GRU architecture of image fusion for RGB-Depth. (b) TD-CNN-GRU architecture of features fusion for RGB-Depth.

The performances of the different fusion strategies tested on our dataset were classically assessed with Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2.1}$$

where TP, TN, FP, and FN stands for true positive, true negative, false positive, and false

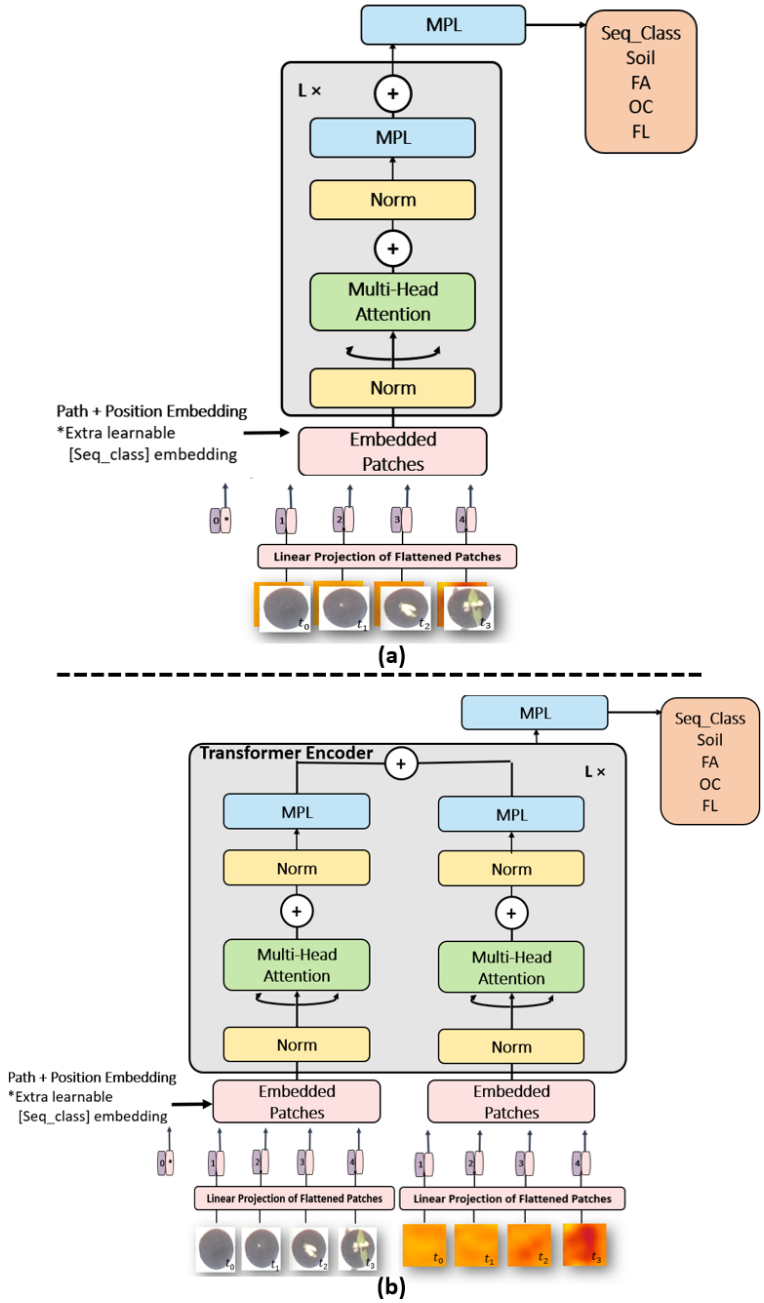


Figure 2.5 – (a) Transformer architecture of image fusion for RGB-Depth. (b) Transformer architecture of features fusion for RGB-Depth.

negative).

2.1.2 Results

- **Fusion strategies**

The proposed deep learning methods CNN, TD-CNN-GRU, and Transformers with image or feature RGB-Depth fusion were applied to the produced dataset as described in the section 2.1.1. The performances are provided in Tables 2-4 and Fig.2.6.

Table 2.2 – Seedling growth stage classification average accuracy and standard deviation when performed over 10 repetitions of CNN model.

	Training	**Validation**	**Test**
RGB	0.95 ± 0.02	0.91 ± 0.03	0.88 ± 0.05
Image fusion RGB-Depth	0.97 ± 0.02	0.95 ± 0.02	0.94 ± 0.04
Features fusion RGB-Depth	0.97 ± 0.01	0.96 ± 0.01	0.94 ± 0.01

Table 2.3 – Seedling growth stage classification average accuracy and standard deviation when performed over 10 repetitions of TD-CNN-GRU model.

	Training	**Validation**	**Test**
RGB	0.87 ± 0.02	0.85 ± 0.01	0.80 ± 0.01
Image fusion RGB-Depth	0.91 ± 0.01	0.87 ± 0.02	0.82 ± 0.01
Features fusion RGB-Depth	0.90 ± 0.01	0.86 ± 0.02	0.81 ± 0.01

Table 2.4 – Seedling growth stage classification average accuracy and standard deviation when performed over 10 repetitions of transformer model.

	Training	**Validation**	**Test**
RGB	0.90 ± 0.02	0.86 ± 0.01	0.82 ± 0.01
Image fusion RGB-Depth	0.96 ± 0.02	0.91 ± 0.01	0.88 ± 0.03
Features fusion RGB-Depth	0.92 ± 0.03	0.89 ± 0.02	0.84 ± 0.01

Tables 2-4 show that all methods performed better when RGB and Depth data are fused by comparison with the sole use of RGB data. This improvement is obtained both with image fusion and with feature fusion. This demonstrate the value of RGB-Depth fusion with a gain of 5% (on average) compared to the use of the sole RGB images. This is obtained at a reasonable training time of around 1 to 3 hours as detailed in Table 5. The best results are obtained with the CNN method, i.e. the spatial method by comparison with the spatio-temporal method. This CNN is showing the best absolute performance, the smallest training time and also minimum decrease of performance between training,

Table 2.5 – Training time of the different deep learning architectures.

	Model	Training time
RGB	CNN	1h00min
	Transformer	1h30min
	TD-CNN-GRU	3h00min
Image fusion RGB-Depth	CNN	1h15min
	Transformer	1h35min
	TD-CNN-GRU	3h30min
Features fusion RGB-Depth	CNN	1h20min
	Transformer	1h30min
	TD-CNN-GRU	3h20min

validation and test. This is in agreement with our previous results found in [83, 59], where spatio-temporal methods outperformed memoryless spatial ones only when the kinetic of growth were homogeneous among the dataset. This was not the case in this study.

The confusion matrix of the CNN method is displayed in Fig.2.6 for RGB images and RGB-Depth images. Interestingly errors with both RGB and RGB-Depth only occur on adjacent classes along the developmental order. These are situations where even the human eye can have uncertainty to decide the exact time of switching from one class to the next one. Remaining errors can thus be considered as reasonable errors. The confusion matrices also clearly demonstrate that the main gain brought by the Depth channel is on the stage of opening the cotyledons for which the error are divided by a factor two. First appearance out of the soil, or the appearance of the first leave produce very limited variations on the depth. By contrast, the opening of the cotyledons produces an abrupt variation of the Depth. Therefore, the impact of Depth on the improvement of the performance of classification on this developmental stage is consistent with this rationale. Following also this rationale, one can notice that the errors on opening the cotyledon slightly increase when Depth is added but the overall impact of Depth is on average beneficial to the global accuracy.

- **Detection of event changes at night using depth information**

The advantage of using the depth is not limited to enhance the performance during the day as shown in the previous subsection. Depth is also expected to be specifically useful during the night since the RGB cameras are then non operating while the Depth images can still be acquired. If the growth stage switches during the night the RGB imaging devices detect the switch only on the first frame of the next day time as illustrated in

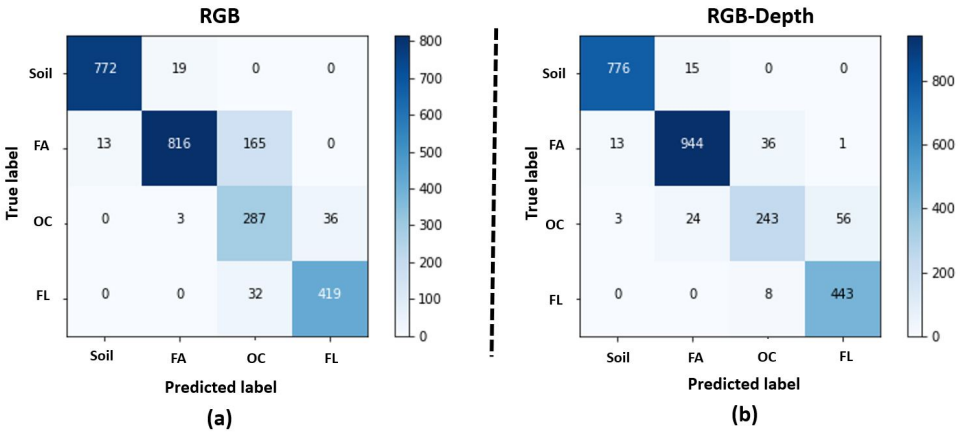


Figure 2.6 – Confusion matrix for the best method found in Table 2.5, i.e. CNN. Left for the RGB images and right for the RGB-Depth images.

Fig.2.8. It is possible to screen for Depth alone during these nights and observe the start of a growth pattern actually occurring before the beginning of the day. We demonstrate in this subsection how to take benefit quantitatively of the sole Depth channel during these nights.

We analyzed the number of switches from one growth stage to another happening on the first image acquired during the day in the data set of [59] and found out that it represented 35 percent of the events (see Fig. 2.7). This is similar to what we found with the dataset of in this work where we had 100 sets of pots from different varieties. In these frames, we have 115 switches of growth stages with 43 happening during night time. While some could be triggered by the action of light others could also happen earlier during the night. To detect a possible change during the night, we quantitatively used Depth. We designed Algorithm 1 which acts as follows. We first detects nights where a switch between a growth stage to another growth stage is found in RGB images. During these nights, the algorithm then detects the depth frame on which the switch is the most likely to occur. In short, this is obtained by choosing the time where the average spatial depth is permanently (computed over a sliding window of 4 images=1hour) closer to the average spatial depth of the next growth stage.

To validate Algorithm 1, we could not establish ground truth during the night. As a workaround, we used daylight events and applied the depth channel only to the Algorithm 1. Then, we used the annotated ground truth obtained from the RGB images to compute the performance of Algorithm 1. We found 80% of these 115 switches with a shift of less than 4 frames on average (standard deviation of 2 frames) by comparison with the

Algorithm 1: Detection of night events using depth information.

Input:
S^{night} = Sequences of depth images of a night during which a switch a growth stage is observed in RGB images.
S^a = Sequences of depth images from the last day before the switch of growth stage A to B.
S^b = Sequences of depth images from the first day after the switch of growth stage A to B.
Output: P_t = Precise time of switch of growth stage.

$\overline{\text{DA}} \leftarrow$ mean(S^a); Spatial average of S^a
$\overline{\text{DB}} \leftarrow$ mean(S^b); {Spatial average of S^b}
$\overline{\text{DN}_\text{k}} \leftarrow$ mean(S^{night});{Spatial average of S^{night}}
$< M_{DA} > \leftarrow$ mean($\overline{\text{DA}}$);{Temporal average of $\overline{\text{DA}}$}
$< M_{DB} > \leftarrow$ mean($\overline{\text{DB}}$);{Temporal average of $\overline{\text{DB}}$}
$GA \leftarrow \overline{\text{DN}}$ - $< M_{DA} >$;{Difference between $\overline{\text{DN}}$ and $< M_{DA} >$}
$GB \leftarrow \overline{\text{DN}}$ - $< M_{DB} >$; {Difference between $\overline{\text{DN}}$ and $< M_{DB} >$}
$bin \leftarrow$ sign (GA - GB); {Binary vector of the sign for the difference between GA and GB }
$Idx \leftarrow$ find(bin==1111);{Get the index of first pattern (1111) in the binary vector. }
$P_t \leftarrow Length(S^a) + Idx$; {Add the length of S^a to the index of the first pattern (1111) to get the precise time }

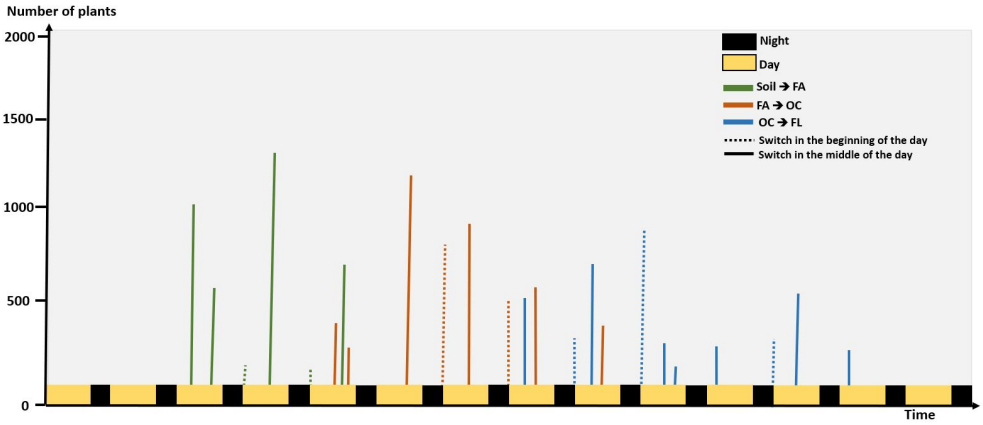


Figure 2.7 – Histogram of detection of growth stage change during day and night from 4000 plants.

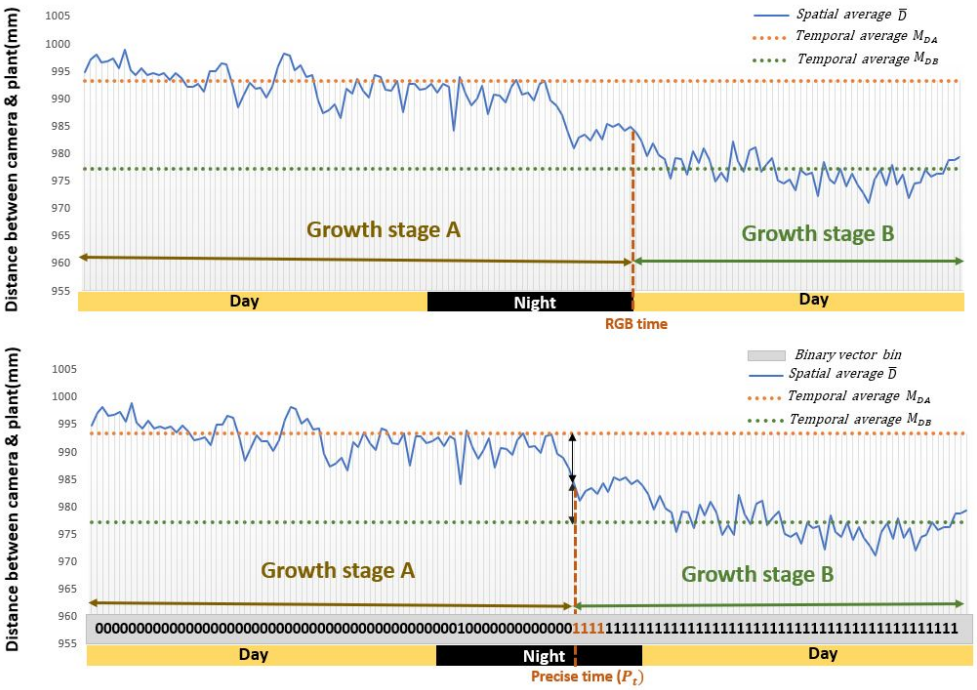


Figure 2.8 – First row: the detection of switch from growth stage A to growth stage B using only daytime RGB images. Second row: the more precise detection of switch from growth stage A to growth stage B using the Depth pattern during the night time as proposed by Algorithm 1.

manually annotated ground truth. This corresponds to an uncertainty (bias here) of 1 hour which is very reasonable and much lower than the error duration of the night itself

(8 hours) if no Depth were used.

2.1.3 Discussion

We analyzed the remaining errors of the proposed algorithms and discuss them in this section together with some open perspectives of the work.

Two main sources of errors can be attached to the acquisition protocol and instrumentation itself. These are illustrated in Fig. 2.9. First, some seedlings growth so fast that their leaves or cotyledons go out of the observation window (Fig. 2.9a). This causes drop in depth and change in the RGB pattern. With our current approach, we do focus on individual pots. For such seedlings growing at early stages outside of their pot, we would need to either use larger pots or develop tracking algorithms. This falls outside of the scope of this study which focused on the added value of Depth when fused to RGB for the detection of early growth stages of seedlings. Another source of errors happens due to noise on the Depth channel (Fig. 2.9b). Such noises were observed when too much or too low amount of IR light was reflected on pots. This happens for instance when the plastic material of the pots has a high reflectance or when some remaining water(absorbing IR) is present. These noises can be reduced by carefully choosing the material used for the pot and the watering process. Another type of error comes from the inherent large heterogeneity of shapes and sizes of the bean varieties considered in this study and illustrated in Fig. 2.10. This affects specially the detection of growth stage which shows the tiniest changes, i.e. the opening of the cotyledons. To solve these errors, one could simply add more data or use more advanced data augmentation techniques such as zoom, stretch, color jitter, ... We wanted to provide basic results here which already happen to be of rather high quality without the use of such approach to robustify the model since the main goal was the fusion of the RGB and Depth for seedling growth monitoring.

One may wonder about the robustness of the model proposed given the relatively small size of the plant population considered. First, the overfit measured with the best method was found to be limited together with the difference of performance between cultivars. It is important to recall here that the point of the work is to quantify the added value of RGB-Depth images by comparison with sole RGB. This is what we do on the same data sets. Interestingly, the performance with RGB images obtained with only 72 samples are similar to the larger data set used in [59] (90% against 88% here). However, we cannot ensure a perfect robustness to large change of phenotypic shapes. If such variability in scale was expected, larger data sets would have to be constituted. The comparison

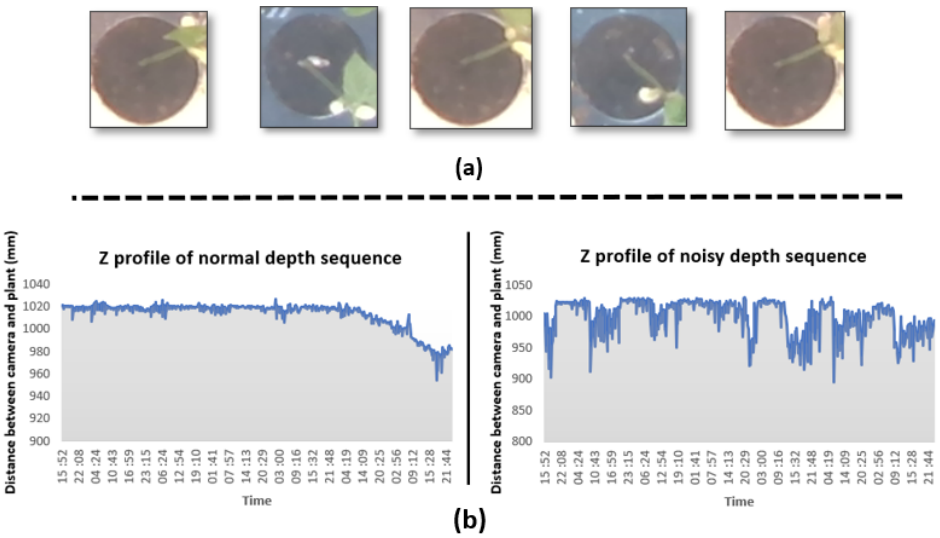


Figure 2.9 – Sources of errors due to the acquisition protocol (a) and instrumentation (b).

Varieties	Flavert	Red Hawk	Linex	Caprice	Deezer	Vanilla
Cotyledons shape						
Cotyledons size (pixels)	576	710	132	165	221	256
First leaf shape						
First leaf size (pixels)	1482	2280	743	853	736	1764

Figure 2.10 – Heterogeneity of shape and size in the two events OC and FL for the different bean varieties used in the training.

between RGB and RGB-Depth would remain unchanged.

In this work, we focused on early fusion and feature fusion of RGB and Depth. One may also consider decision fusion where the classification from the RGB image and the Depth image would be made. We performed this analysis and found a pure random decision when the classification was made on Depth alone. Therefore, at the decision level, no added value of Depth was to be expected on average. Fusion between RGB and Depth for such small images and low-cost sensors as the one considered in this study is found to be beneficial on average at earlier stages of processing (image or features).

However, after analysing the confusion matrix in detail, one could imagine to selectively using the added value of Depth at the stages of growth where it is expected to be the most significant. This was found to be between the FA and OC in our case and more generally when large contrast in Depth happens. On the contrary, one could discard the use of Depth when the growth process is estimated to lay at stages where no contrast in Depth is expected (between Soil and FA in our case).

This study could be developed in several other future directions. First, we could revisit this study with higher resolution Depth sensors [66] to investigate how the reduction of noise and improvement of resolution in Depth could help to further improve the classification results. More advanced stages of development yet still accessible from the top view, could be investigated without targeting 3D reconstruction [92]. An issue comes with the possible overlapping between plants. One solution would be to decrease the density of plants but this would come with a lower throughput for the experiments. Another solution would be to investigate the possibility to track leaves during their growth in order to decipher partial occlusions. Here again, RGB depth sensors coupled with advanced machine learning approaches could be tested to further extend the capability to monitor seedling growth [93]. Last but not least, we can now directly apply the developed algorithms to analyze biologically in detail the statistical distribution of seedling growth events at night on large datasets. This may unravel new knowledge on the physiological impact of light on these growth kinetics in addition to their links with circadian rhythms [94]. We selected another option for the investigation in the following of the chapter. We investigate the value of the RGB-Depth sensor of this section for another variety testing experiment.

2.2 Wheat heading stage

In previous section 2.1, the monitoring of seedling development stages, we focused on individual plant growth. In more advanced stages, the plants overlap and the individual tracking becomes complicated. A workaround approach consists of considering an ensemble of touching plants. This group of plants appears as texture to the camera. As stressed in the introduction chapter, this corresponds to actual observation scale in variety testing where a parcel scale is rated: height, flowering time, etc. In the following section, we revisit the question of RGB-Depth fusion of the previous section at this observation scale. For illustration, we focus on the automatic detection of wheat heading stage.

The hundreds of millions of tons of wheat produced each year are of different varieties. The biologists manually evaluate a parcel according to BBCH reference. The BBCH code (Bundesanstalt, Bundessortenamt und CHemische Industrie) [95] is a scale to identify the stages of phenological development of a plant. BBCH scale splits the wheat growth stage into ten principal stages, from germination (stage 0) to Senescence (stage 9), as we can see in Fig. 2.11. Each principal stage is divided into ten sub-stages. We obtain a two-digit code composed of the principal stage and the sub-stage (see Table 2.12). A parcel is considered to have achieved a sub-stage when 50% of the plants have attained that sub-stage. This part focuses on the wheat heading stage (stage 6 in BBCH code). The heading is the process whereby the seed head emerges from the sheath of the flag leaf.

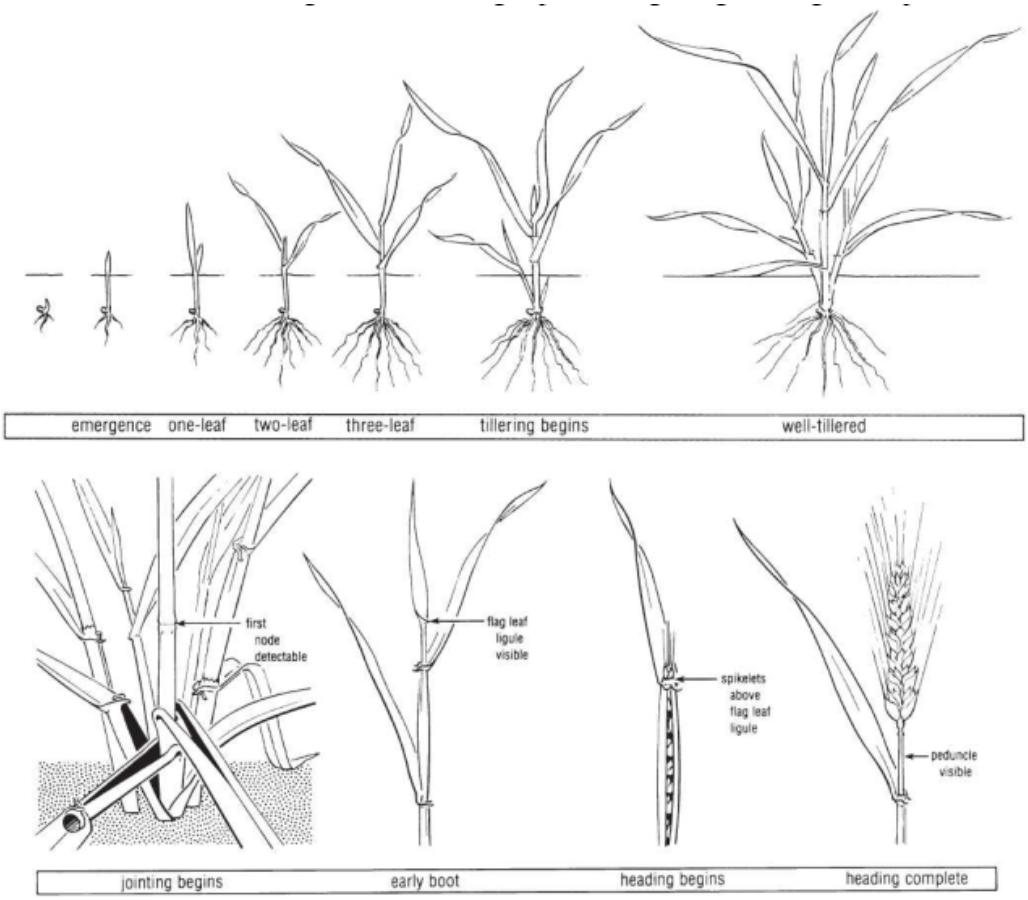


Figure 2.11 – Illustration representing the general growth pattern of wheat plant from emergence to heading.

0. Sprouting/Germination	
00	Dry seed (caryopsis)
01	Beginning of seed imbibition
03	Seed imbibition complete
05	Radicle emerged from caryopsis
06	Radicle elongated, root hairs/side roots visible
07	Coleoptile emerged from caryopsis
09	Coleoptile penetrates soil
1. Leaf Development	
10	First leaf through coleoptile
11	First leaf unfolded
12	2 leaves unfolded
13	3 leaves unfolded
1...	Stages continuous till ...
19	9 or more leaves unfolded
2.Tillering	
20	No tillers
21	First tiller detectable
22	2 tillers detectable
23	3 tillers detectable
2...	Stages continuous till ....
29	Max no. of tillers detectable
3. Stem Elongation	
30	Pseudostem & tillers erect, first internode elongating, top of inflorescence at least 1 cm above tillering node
31	First node at least 1 cm above tillering node
32	Node 2 at least 2 cm above node 1
33	Node 3 at least 2 cm above node 2
3...	Stages continuous till ...
37	Flag leaf just visible, rolled (last leaf)
39	Flag leaf unrolled, ligule just visible
4. Booting	
41	Early boot: flag leaf sheath extending
43	Mid boot: flag leaf sheath just visibly swollen
45	Late boot: flag leaf sheath swollen
47	Flag leaf sheath opening
49	First awns visible (in awned forms only)
5. Inflorescence Emergence, Heading	
51	Tip of inflorescence emerged from sheath, first spikelet just visible
52-54	20% to 40% of inflorescence emerged
55	Half inflorescence emerged
56-58	60% to 80% inflorescence emerged
59	Inflorescence fully emerged
6. Flowering, Anthesis	
61	First anthers visible
65	Full flowering: 50% of anthers mature
69	End of flowering: all spikelets flowered some dry anthers may remain
7: Development of Fruit	
71	Watery ripe: first grains half final size
73	Early milk
75	Medium milk: grain content milky, Grains final size, still green
77	Late milk
8. Ripening	
83	Early dough
85	Soft dough: grain content soft but dry. Fingernail impression not held
87	Hard dough: grain content solid Fingernail impression held
89	Fully ripe: grain hard difficult to divide with thumbnail
9. Senescence	
92	Over-ripe: grain very hard, cannot be dented by thumbnail
93	Grains loosening in day-time
97	Plant dead & collapsing
99	Harvested product

Figure 2.12 – BBCH growth stages for wheat.

2.2.1 Materials and methods

We acquired videos using the Intel real-sense sensor on the field at the same time as manual evaluation. Two different angles of view have been used, 90° and 45°. The informational task corresponds to a classification task with RGB or RGB-Depth images as input and phenological stages as output. Three phenological sub-stages were observable and corresponded to the sub-stages 51, 55 and 59 according to the BBCH code. In the

following section we will define them as 3 classes: 1, 5 and 9.

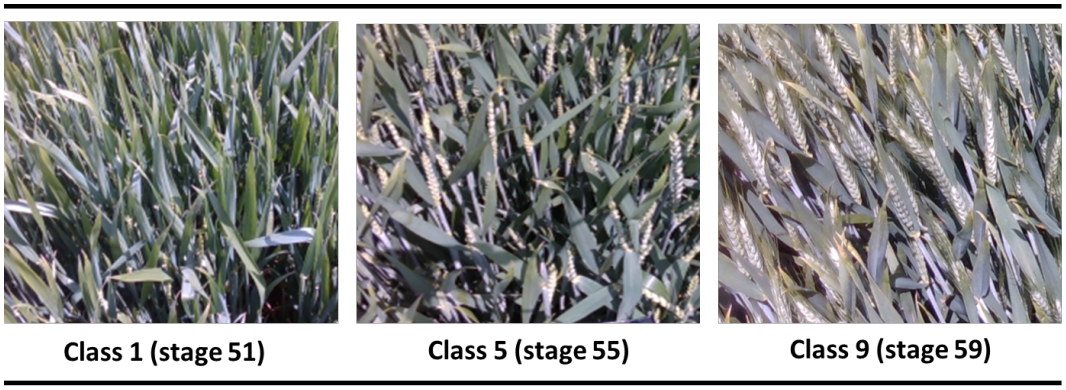


Figure 2.13 – The three heading classes in the field.

- **Dataset**

The dataset includes 31 videos RGB-Depth acquired at 5 different dates and with two various view angles: at zenith (90°) and 45° using RGB and depth sensors. The resolution of the images is 1920x1080 pixels for the RGB and 1280x720 pixels for the depth maps. From each video, we extract 100 RGB images and 100 depth maps of the same 720x1280 pixels resolution an alignment function function (see Fig. 2.13).

The angles of view provide complementary information. On the one hand, the images taken at 45° give a closer view of stems with curvatures that contain extensive information. They allow other parcels to appear in the frame and complicate the analysis of the depth maps. On the other hand, the proximity of the camera to the plants in the zenith shot offers a better resolution of the ears.

- **Support vector machine for multi-class problems.**

Because we operated with a limited database here, we use machine learning with the classical support vector machines (**SVM**) for classification. Developed in the 1990s by Vladimir Vapnik [96, 97], the SVM model, in its simplest version, cannot natively perform multi-class classification. It provides binary classification and the split of data points into two classes. The same principle is applied for multi-class classification after splitting the multi-classification problem into several binary classification problems. The data points are mapped to high dimensional space to gain mutual linear separation between every two classes. This approach is called One-to-One. In One-to-One classification, for the N-class instances dataset, we have to generate the $N * (N - 1)/2$ binary classifier models.

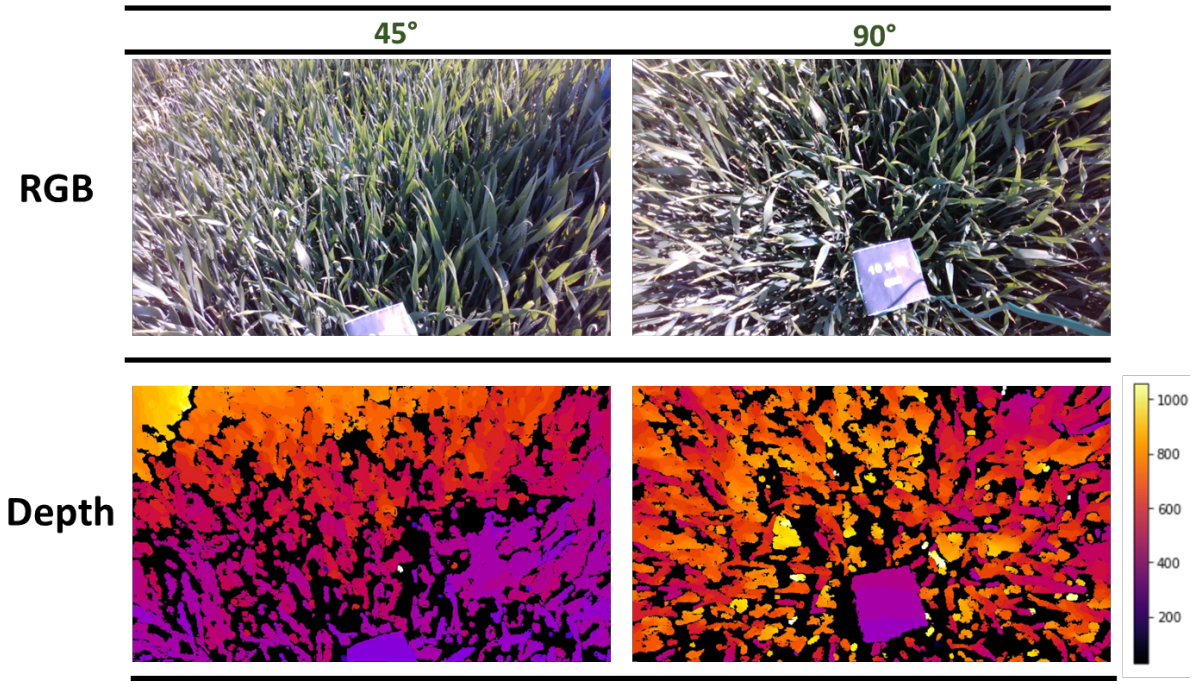


Figure 2.14 – RGB images (top) and depth maps (bottom) for Chevignon variety at stage 5 with viewing angles of 90° (right) and 45° (left).

- **Textural descriptor selection**

The percentage of visible spikes defines the wheat heading stages in variety testing. However, with the used sensor, the resolution is limited. Furthermore, due to the overlapping and limited resolution of our sensors, individual counting is a challenging task. Therefore, we propose investigating the problem as a pattern or texture recognition. Texture description is a classical problem in image processing [98]. Repeated shapes in the image characterize textures. The grains are organized in the spike creating a regular and repeated pattern.

Several groups of descriptors were implemented to investigate the added value of Depth on the texture classification task: Haralick features, local binary patterns **LBP**, and the scattering transform. However, the color information is not considered in the heading stages. For this reason, we applied some descriptors to the gray level of RGB images.

Haralicks descriptors called also GLCM [99] are based on grey level co-occurrence matrices (**GLCM**). These matrices encode the repetition of greyscale over a specific distance and direction in the whole image. There are four matrices according to the angles 0°, 45°, 90°, 135°, and 14 statistical indicators(Mean, entropy, variance) are computed for each one. The average of these statistics in the various directions constitutes the 14

GLCM descriptors.

LBP was introduced in [100]. In the LBP method, each pixel in the image is associated with a value to these neighbors, a 0 if they are lower than the central pixel and one if they are higher [101]. These values are concatenated clockwise or counterclockwise to get the smallest binary number to ensure invariance by the rotation of the descriptor. The value is assigned only to the central pixel to create an image based on the relative variations of the gray levels around this pixel. Therefore, the descriptors are robust to lighting variations in the image.

The scattering transform generates an invariant representation of a signal as a function of rotation and scale change. The first application of this descriptor in plant science was made in [102]. In this study, they presented a convolution network that performs a decomposition into wavelets using the complex module. The wavelet decomposition is done at various scales and orientations to build larger image blocks. After several layers, the images are reduced to their average value to create the descriptor vector.

- **Fusion strategies**

Similarly to what was done in the previous section on individual seedlings, we present the results from three fusion strategies: Image, feature, and late. The two first strategies are the same as the previous work, section 2.1. The image fusion strategy merges RGB and Depth images before extracting descriptors. As trivially indicated in its name, the second fusion merges at the features level. Finally, the last strategy (late fusion) is composed of two independent classifiers, each trained on a different type of data (one in the RGB image and another in Depth maps). Both of them are used for prediction. If the two classifiers are in agreement, there is no ambiguity. Otherwise, the highest prediction score is retained. These fusion strategies were compared with the three types of textural descriptors independently.

2.2.2 Results

To increase the number of images, the images were divided into two by three, i.e., a resolution of 240x640 pixels and the different angles of view were mixed. The database contained 1584 images in total, with 528 images per class. The database was split into 80% for the training and 20% for the test. The descriptors have different dimensions: 14 for the GLCM descriptors, 59 for the local binary patterns, and 417 for the scattering transform. They were reduced to the lowest dimension (14) using the PCA method

(Principal Component Analysis).

All the SVM models were trained ten times in a cross-validation mode. For each training, we calculate the accuracy (Eq.2.1). Then, we computed the average and standard deviation over the ten folds of the cross-validation. The results are presented in Table 2.6.

Table 2.6 – Wheat heading stage classification average accuracy and standard deviation when performed over ten repetitions.

		GLCM	**LBP**	**Scatter transform**
Without fusion	**Gray scale**	31.6 ± 1.0	72.6 ± 1.2	78.6 ± 1.3
	Depth maps	56.8 ± 1.9	55.1 ± 1.0	57.4 ± 1.8
With fusion	**Image fusion**	30.8 ± 0.9	70.0 ± 1.7	70.1 ± 1.5
	Features fusion	56.0 ± 2.4	72.6 ± 1.3	68.2 ± 1.2
	Late fusion	40.1 ± 3.7	71.1 ± 1.3	76.6 ± 1.3

In the result presented in Table 2.6, we first observe the classification performances on single components. The performance on the gray level component reaches overpasses 70%. This demonstrates that the textural approach, although challenging, at first sight, provides already interesting results. Based on the results of three descriptors based on gray level images, the scattering transforms descriptor provides the best results with an accuracy of around 79%. For the results obtained using the depth maps, as opposed to the previous results, the three descriptors have almost the same results (around 55%), but they still are insufficient. According to the fusion of gray-scale information and depth information performance, the best fusion strategy is the late one. Nevertheless, the fusion results are lower than those using single-channel information.

- **Typical errors of the late fusion.**

Fig. 2.15 a shows the confusion matrices obtained with the scattering transform descriptor based on gray-scale images. The most exciting point in confusion matrices is that errors are mainly produced in adjacent classes. The visual annotation can also have uncertainties in deciding between the adjacent classes. The obtained errors can therefore be considered acceptable errors.

In order to provide guidelines for future research in fusion strategies, we studied the typical errors provided by the late fusion models. We used in this test 316 images, and there are 67 images misclassified (see Fig. 2.15.b). The database comprises bearded and

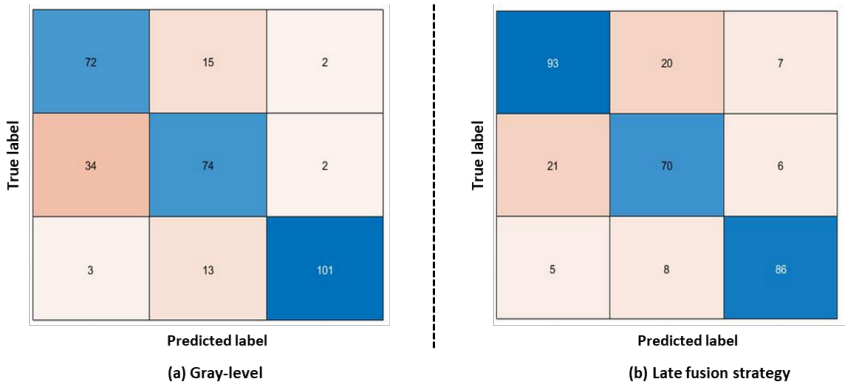


Figure 2.15 – Confusion matrix for the best method found in Table 2.16, i.e., scattering transform. Left for the gray-scales images and right for the gray-scales-Depth late fusion.

non-bearded wheat and images taken from different angles. We have identified two error sources. Around 56.5% of the errors originate in zenith images, and 44.0% come from bearded varieties with a standard deviation of around 6 for both of them. Based on this analysis, for the subsequent acquisition, we recommend to chose the 45° angle.

2.2.3 Conclusion

In this chapter, we have demonstrated the added value of Depth when fused with RGB images for the important problem of detecting seedling growth stage development. During the daytime, Depth was shown to improve by 5% the classification performances on average. Also, Depth was shown value to refine the estimation of the switch of growth stage during the night period. These results were established on different fusion strategies, including CNN, TD-CNN-GRU, and transformers. These methods were compared to incorporate the prior information on the order in which the different stages of development occur. The best classification performance on these types of images was found with our optimized CNN, which achieved 94% accuracy of detection. In our experiments, all models and fusion strategies were trained and tested on several genotypes of beans.

We extended the investigation of our RGB-Depth sensor with the possibility of using a machine learning method to classify the wheat heading stages in the field. We used the support vector machine (SVM) as a classification method. We tested three different texture descriptors the scattering transform, Haralick, and local binary. The best results obtained is around 79 % accuracy using a scatter transform descriptor. This demonstrates

the feasibility of performing global rating with textural features for this variety testing traits instead of individual detection. The performance was nevertheless limited to envisioning a direct application, and a larger database would be necessary to further push the application toward usability. Additionally, we investigated three fusion strategies between the RGB images and the depth information. We found the best fusion method to be late fusion. However, this fusion did not improve the performance of the best component. For this second characteristic of variety testing, contrary to the results shown for the individual seedling stage in the first section, depth information is not helping the classification.

Chapter 3

Transfer knowledge from controlled environment to noisy environment

3.1 Introduction

The implementation of deep learning models needs a large amount of annotated data. Nevertheless, having such large datasets is generally challenging in several fields. One of these domains is variety testing. In the field, the collecting data process is related to the season of each crop as opposed to what is accessible in controlled environment. For this reason, there are more large databases acquired in controlled environments than in the field. In this chapter, we will be interested in the possibility of using such existing extensive databases to predict limited databases. For illustration, we remain on one of two use case of the previous chapter. We focus on the classification of the four early development stages of plants by exploiting our database[59] and the flowering time detection using a synthetic database. In the first part, we present the validity of the proposed method of transfer from controlled environments to greenhouses. The second part describes the same approach, with this time the transfer from indoor to the field. These two studies were published in [103][104]. In the last part of the chapter, we present another approach of transfer,with the transfer from synthetic environment to real environment.

3.2 Indoor to greenhouse transfer on seedling growth

Several work around approaches have been proposed to address the bottleneck of annotation in applied computer vision including the development of ergonomic tools to speed up annotation, data augmentation, transfer learning, generation of simulated images

or the use of generative neural networks. These approaches have been applied to the domain of plant imaging and the communication here is in this trend [105, 39, 106, 107]. We recently developed a spatio-temporal deep learning algorithm to monitor the growth of seedlings in controlled environment from top-view in RGB images [59]. Here, we propose an extension of this work by investigating the possibility to transfer this knowledge to greenhouse environment where the lighting conditions are not controlled and shadow may occur due to the position of the sun or the presence of clouds passing by. This is to the best of our knowledge the first trial of this type in plant imaging.

As most related works to our proposal, one can point that the computer vision community has in recent years addressed the automatic detection and removal of shadows in RGB images with deep learning [108, 109, 110]. As often encountered when considering the translation of such literature to other application domains some basic practical issues may appear. In the current work, the spatial content and resolution from [108, 109, 110] are clearly different from the one considered in seedling growth. As a consequence direct transfer learning would very likely fail and would require additional annotated images. Our proposal here is rather to investigate the possible transfer of knowledge from plant observed in indoor conditions to greenhouse conditions.

3.2.1 Datasets

- **Real indoor and out door data**

Two distinct datasets have been produced. The first dataset consists of 449286 images (600 different pots) from red clover (Trifolium pratense) and alfalfa (Medicago sativa) which were captured in a fully controlled environment [59]. This dataset or a pre-processed version of it will serve as training dataset in this study. The second dataset includes 22212 images (36 different pots) captured from sunflower seedling in a non controlled environment (greenhouse). This second datasets serves as testing dataset in this study. Both datasets have been recorded with the frame rate of one image every 15 minutes. Figure 3.1 shows an example of each dataset. Both datasets record the first developmental stages of growth of seedlings. This includes four stages with the soil, the first appearance of the cotyledon (FA), the opening of the cotyledons (OC), the appearance of the first leave (FL).

The objective of the work is to transfer knowledge from a model trained on the first dataset to the second dataset as illustrated in Figure 3.2. While the species of both

datasets are different they are both dicotyledons so that they share similar shapes at early stages of development. Moreover the two cameras share the same spatial resolution. As visible in Figure 3.1, the color of the crop observed indoor and greenhouse are not exactly the same. This color discrepancy happened to be none critical to transfer knowledge from indoor to greenhouse. As done in [59], the plant is filtered from the soil with a standard thresholding approach to avoid any impact on the difference of soil and surrounding background. The challenge in the proposed experiment therefore lay in the presence of shadows which occurs in greenhouse environment only.

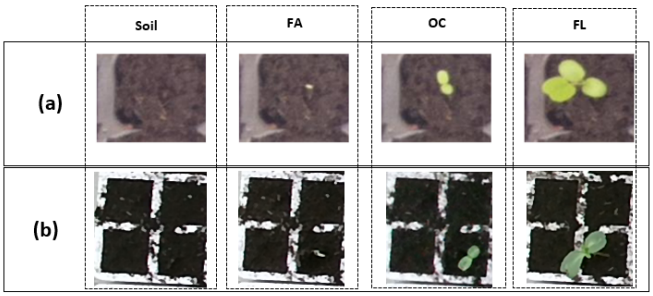


Figure 3.1 – (a) Images from controlled environment on which seedling development is trained. (b) Images from greenhouse environment on which we want to test the trained model. The four developmental stages to be detected are the soil, the first appearance of the cotyledon (FA), the opening of the cotyledons (OC), the appearance of the first leave (FL).

- **Simulated greenhouse data**

To simulate images acquired in the greenhouse environment from indoor images, we propose an automatic shadow generator as detailed in Algo. 2. The shadows are randomly positioned by using a thresholded speckle generator [111, 107]. All sizes of shadow can be present greenhouse. However, only shadows larger than the typical size of seedling organs and smaller than a single plant are expected to impact the detection of seedling development. We adjusted the number of phasors in the speckle generator in order to fit with this prior knowledge and produce shadows corresponding to the maximum area of the seedling (40% of the size of the pot in our training dataset). Modulation of maximum intensity during the day was recorded in the validation dataset. This information was used to adjust the value of the threshold in the algorithm (found to $threshold = 0.5$ in our validation dataset). Each images in the indoor database is then spatially modulated by the generated shadow with a simple multiplication.

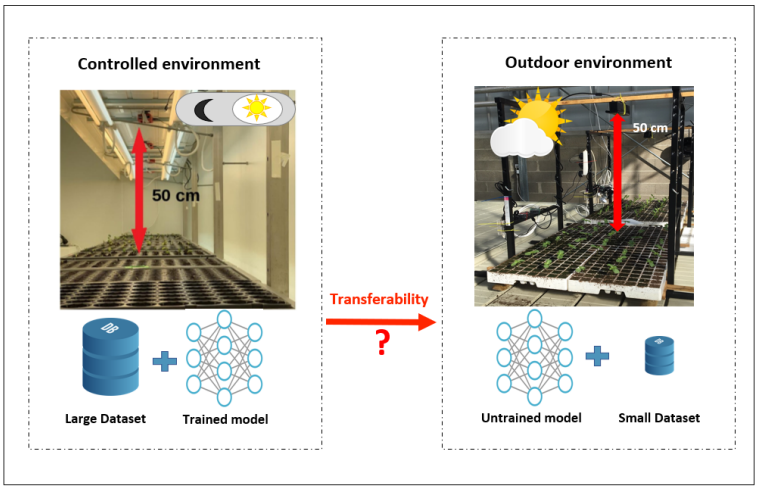


Figure 3.2 – Left panel illustrates the imaging system in controlled environment associated with the large database of [59]. Right panel illustrates the imaging system in an greenhouse environment with a smaller database. We investigate the possibility of transfer of knowledge from left to right panels.

- **Proposed Methods**

We shortly recall the deep neural networks used in [59] and tested here on the transfer of knowledge from indoor to greenhouse environmental conditions. We then extend to other methods, not included in [59] and tested for the first time in plant imaging.

First we included in [59] a basic CNN architecture performing a 4 classes classification to discriminate between the images of Figure 3.4.(a). The architecture of CNN is composed of five convolutional layers with filters of size 3×3 and respective numbers of filters 64, 128,128, 512 and 512 each followed by rectified linear unit (RelU) activations and 2×2 max-pooling; a fully connected layer with 512 units and ReLU activation, a fully connected output layer with 4 classes corresponding to each event and a softmax activation. We use cross entropy as loss function and adam as optimizer. The architecture optimized for this 4 classes task is visible in Figure 3.4 and served as baseline in [59] since it does not embed any memory about the growth process. We demonstrated in [59] the added value to embed in controlled environment such a memory and demonstrated the superiority of a CNN-LSTM (see Figure 3.4.b) by comparison with a sole LSTM architecture (see [59]). The optimal duration of the memory was found to 4 images in [59] corresponding to 1 hour of recording.

To further enrich the investigation on memory, we added other neural network archi-

Algorithm 2: Pseudo-code to simulate random shadows

Data: I: Original image, n:number of phases, s:threshold $(0, 1)$.
Result: I_{aug}: image with shadow

```
l ← height of original image
c ← width of original image
shadow ← zeros (l,c)
Phases ← exp (2 * π * Rand(n, n) * i)
shadow (1:n,1:n) ← Phases
shadow ← FFTshift(IFFT(shadow))
shadow ← shadow / (Max(shadow)
for i ← 1 to l do
    for j ← 1 to c do
        if shadow(i, j) < threshold then
        shadow(i,j) ← threshold
I_shadow=I * shadow
```

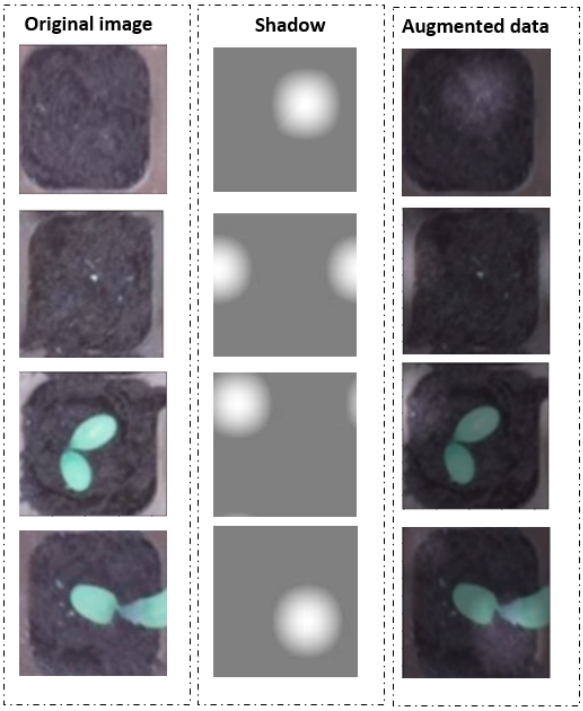


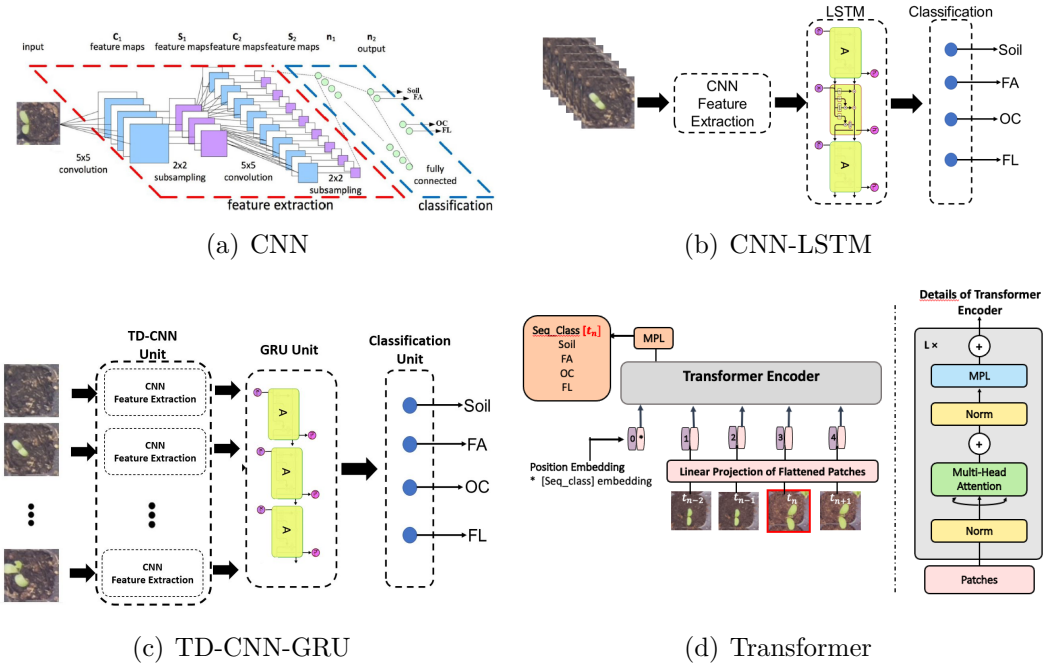
Figure 3.3 – Example of original indoor images (left), shadows generated with Alg. 2 (middle) and, indoor images with simulated shadows (right).

tectures. We tested gated recurrent unit (GRU) networks [112], an alternative to LSTM, which has been demonstrated empirically to converge faster. GRU uses two gates: the update gate and the reset gate while there are three gates in LSTM. This difference makes GRU faster to train and with better performance than LSTMs on less training data [87]. A last class of neural network dedicated to time series are the transformers. Since their introduction in [38] they have been shown to outperform recurrent neural networks such as LSTM and GRU specially in the field of natural language processing as they do not require that the sequential data be processed in order. Transformers have been shown suitable to process temporal information carried by single pixels in satellite images time series [88, 89, 90]. Transformers have recently been extended to the process of images [91] where images were analysed as a mosaic of subparts of the original images creating artificial time series. In our case, we directly have meaningful subparts of the original images which corresponds to the field of view of the pots. We therefore provide the transformer of [91] with time series of consecutive images of the same pot (we used the same time slot as in the other spatio-temporal methods). We used 32 transformer layers with batch size 64, feed forward layer as classification head layer and the size of our patch size was equal to 89×89 pixels.

The performance of the models proposed in [59] for controlled conditions are recalled in table 3.1 in addition to the three new methods added in this communication CNN-GRU, TD-CNN-GRU, Transformer. The performance of the TD-CNN-GRU model and Transformer are found to outperform the other methods in controlled environment. A possible interpretation is that, in the TD-CNN-GRU and Transformer models, time and space are stacked and processed at the same time while CNN-LSTM first processes space and then time in a sequential way. In the following, we investigate how the performances of the methods shown in table 3.1 evolve when the models are applied in greenhouse environment. For this experiment we selected the memoryless CNN model and the best time-dependent neural network models: TD-CNN-GRU and Transformer.

3.2.2 Results

Several transfer of knowledge have been tested from indoor conditions to greenhouse conditions. First, as baseline we have applied a brute transfer where the models trained indoor have directly been applied to predict the greenhouse images. The performance with the CNN model, visible in Tab. 3.2, shows a clear drop although it does not vanishes to pure randomness. Then, we have used data augmentation based on the simulation of



(a) CNN (b) CNN-LSTM

(c) TD-CNN-GRU (d) Transformer

Figure 3.4 – Neural networks architecture tested. (a) Optimized CNN proposed in [59]. (b) Optimized CNN-LSTM model proposed in [59]. (c) Optimized TD-CNN-GRU proposed here. (d) Transformer adapted from [91].

Table 3.1 – Tested models in the fully controlled environment. Mean and standard deviation of the accuracy from 5 different trials for each model.

Models	Accuracy
CNN	0.80 ±0.08
CNN-LSTM	0.90 ±0.08
CNN-GRU	0.91 ±0.06
TD-CNN-GRU	0.96 ± 0.01
Transformer	0.92 ± 0.01

shadows applied on indoor images as presented in section 12 As visible in Table 3.2, this simple simulation brings a significant increase of 10% to the overall accuracy on the CNN model. Fine tuning the model trained on these simulated greenhouse data with a small amount of real greenhouse data improved the performance up to 91% while the model trained on the same amount of real data produced 70% accuracy on the CNN model. Interestingly, as demonstrated in Figure 3.5, fine tuning training on data augmented indoor data with shadow converges to a high plateau of performance with a very small

number of input plants. This plateau of performance reached with 7 plants produces a confusion matrix shown in Figure 3.6. Remaining errors are limited to adjacent classes of seedling development and therefore constitute reasonable errors.

Table 3.2 – Performance of CNN in greenhouse conditions.

Models	**Train**	**Validation**	**Test**	**Accuracy**
Brut transfer	400	200	4	0.53 ±0.02
Data augmentation	800	400	4	0.64 ±0.10
Greenhouse training	26	6	4	0.81± 0.02
Greenhouse training	7	6	4	0.70± 0.03
Fine tuning training	7	6	4	0.91± 0.02

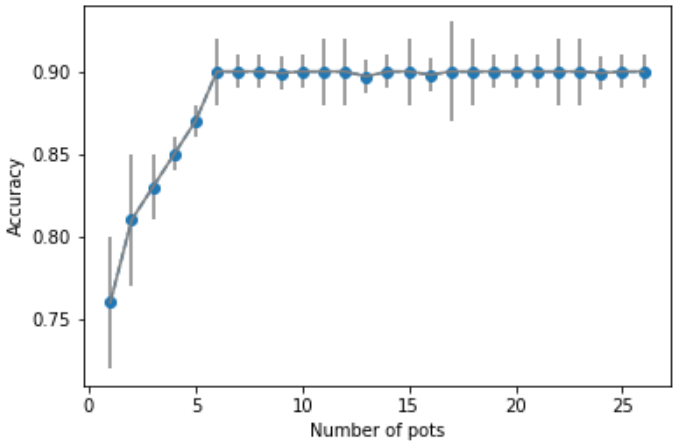


Figure 3.5 – Classification accuracy as a function of number of pots used in train database after data augmentation and fine tuning.

Similar experiments have been carried with the TD-CNN-GRU model as provided in Tab. 3.3 and with the Transformer in Tab. 3.4. Indoor classification performances with these spatio-temporal methods were better than the spatial CNN. However, they appear to drop when applied to greenhouse data and become less interesting than the pure spatial CNN approach. The data augmentation approach with the proposed shadow generator is improving the performance of the TD-CNN GRU and the Transformer by comparison with a direct brut transfer. Yet, they perform in the end with this data augmentation at the same level as if they had been trained fully greenhouse.

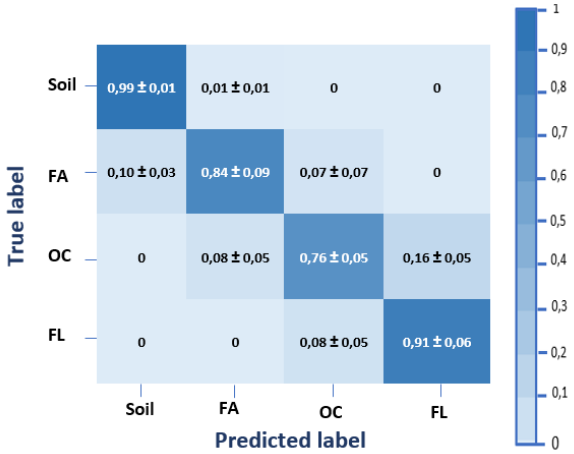


Figure 3.6 – Confusion Matrix of CNN after data augmentation and fine tuning training model using seven pots.

Several parameters could influence the temporal information from indoor to greenhouse. Despite similar speed of the seedling development (approximately 72 hours for the whole process on average) for indoor and greenhouse conditions, the difference of growing conditions may have influenced the kinetics to pass from one developmental stage to another. Therefore a systematic analysis of the statistics to pass from developmental stage to another could be interesting to carry out. However, data augmentation with shadow systematically improved all tested methods. This demonstrates that the presence of these shadows is a critical limitation when moving from indoor to greenhouse.

Table 3.3 – Performance of TD-CNN GRU in greenhouse conditions.

Models	**Train**	**Validation**	**Test**	**Accuracy**
Brut transfer	400	200	4	0.32 ±0.04
Data augmentation	800	400	4	0.59± 0.04
greenhouse training	26	6	4	0.72 ± 0.04
Fine tuning training	26	6	4	0.74 ± 0.02

Table 3.4 – Performance of Transformer in greenhouse conditions.

Models	**Train**	**Validation**	**Test**	**Accuracy**
Brut transfer	400	200	4	0.23 ±0.03
Data augmentation	800	400	4	0.56± 0.04
greenhouse training	26	6	4	0.74 ± 0.03
Fine tuning training	26	6	4	0.76 ± 0.02

In this part of the study, we have investigated the possibility of transferring knowledge from indoor to greenhouse conditions in a plant science application. We have considered the automatic detection of early stages of seedling development to this purpose. While in controlled conditions, time dependence was found to bring additional information, we found that the presence of shadows in greenhouse conditions are destroying this information. However, we have demonstrated that the transfer of knowledge from indoor was possible via the simulation of shadows to be applied to indoor images. We have demonstrated an interest in training on such simulated data and fine tune on a limited amount of real data. The proposed approach is of interest in plant science since greenhouse conditions are important for agricultural practice, while indoor conditions have received considerable attention via the development of phenotyping platforms.

3.3 Indoor to field transfer on seedling growth

In section 3.2, we have investigated the transfer from fully controlled conditions to the greenhouse environment [103] on the question of seedling emergence [59]. We propose to push forward again and extend this study to an entirely uncontrolled environment, i.e., the field. Connected cameras settled on sticks have been positioned to monitor the emergence of various cultures in the field. We explore the value of transferring the knowledge gained in a controlled environment via transfer learning approaches.

3.3.1 Materials and methods

Here, we consider the problem of automated classification of early stages of seedling development of mono- and dicotyledons plants under field conditions. Figure 3.7 illustrates our approach to overcome the time-consuming annotation of the dataset acquired under the field conditions. Hereinafter, we use "indoor dataset" to designate seedlings images acquired under controlled conditions and "outdoor dataset" - the seedling images acquired under field conditions. We applied the model trained on the indoor dataset to classify the seedling development stages on the outdoor images (Fig.3.7). Although such transfer learning is common in image processing, this is the first transfer from indoor to outdoor datasets to the best of our knowledge. The previous study [6] investigated the transfer from fully controlled conditions to the greenhouse environment. In this communication, we tested three transfer strategies of the model trained on the indoor dataset to

two outdoor datasets and derived the most appropriate one. First, we introduced three datasets used for the CNN model training and inference. Second, we performed transfer strategies of the model trained on two indoor datasets to outdoor datasets of maize and rapeseed seedlings. Finally, we discussed what constitutes an effective transfer and gave some recommendations for the image acquisitions that could improve the performance of automated stage classification of seedling development.

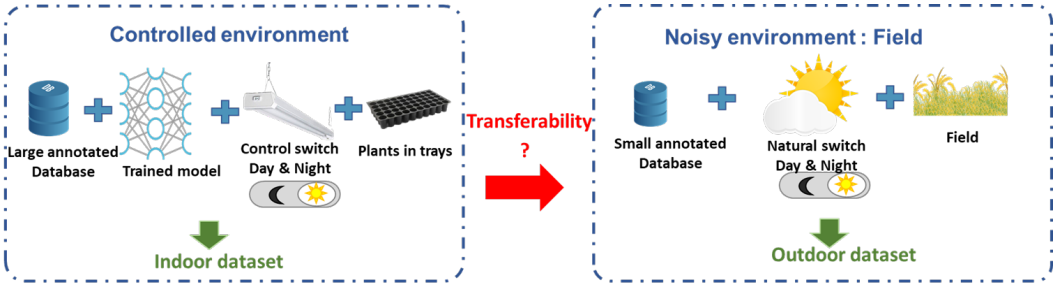


Figure 3.7 – Studied problem. Left panel illustrates the imaging system in a controlled environment associated with the large database. Right panel illustrates the imaging system under field conditions with a smaller database.

The method proposed for the classification of seedling growth stages consists of three main elements: (1) the imaging system designed to generate the datasets; (2) preprocessing images to separate plants from the soil; (3) testing transfer strategies of the knowledge gained by the model on the indoor dataset to the outdoor dataset.

- **Imaging system**

A group of 21 RGB top-view cameras was implemented in the field for three weeks in June 2021 to follow the emergence of maize and in September 2021 to monitor the rapeseed. Cameras were connected to minicomputers to manage the acquisition and storage of images and the power bank (Fig. 3.8). We configured our cameras with an interface to acquire images every 30 minutes during the daytime. The distance between sensors and soil was 1m. It was chosen to track plants in 2 or 3 rows and get images with a spatial resolution of 3264×2448.

- **Datasets**

We produced three distinct datasets. The first dataset consisted of 600 temporal sequences of RGB images from red clover (Trifolium pratense L.) and alfalfa (Medicago sativa) which were captured in a fully controlled environment and used before in study

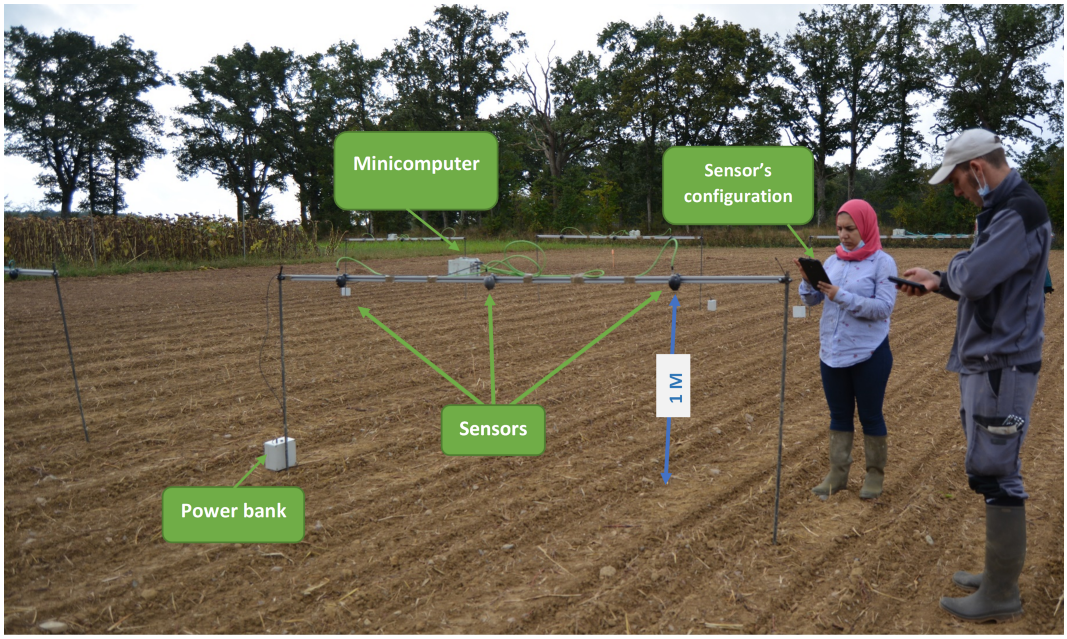


Figure 3.8 – The imaging system in an outdoor environment (filed).

[59]. Here, this dataset or its pre-processed version served as the training dataset. The second and third datasets included 57 time-lapse sequences images of size 89×89×3 pixels captured from rapeseed and maize seedlings in the field. The indoor dataset was recorded with the frame rate of one image every 15 minutes in the daytime, and outdoor datasets were taken with a frame rate of one image every 30 minutes. Figure 3.9 shows an example of each dataset. Table 3.5 summarizes the details of the datasets. The spatial resolution of the indoor data set was similar to the outdoor dataset. This is essential since convolutional neural networks are not scale invariant by design. We took images of the first developmental stages of the growth of seedlings. All resulting image datasets were manually annotated by plant experts. The ground truth included four stages with the soil, the first appearance of the cotyledon (FA), the opening of the cotyledons (OC), and the appearance of the first leaf (FL) for dicotyledons plants (alfalfa and rapeseed) and the soil, the appearance of the first leaf (FL), the appearance of the second leaf (SL) and the appearance of the third leaf (TL) for monocotyledons plants (maize). Before applying the deep learning method for the classification, the raw sequences of images were treated with the algorithm described in the 3.5.3 part to remove the soil background as in the previous study [59].

- **Individual plant extraction**

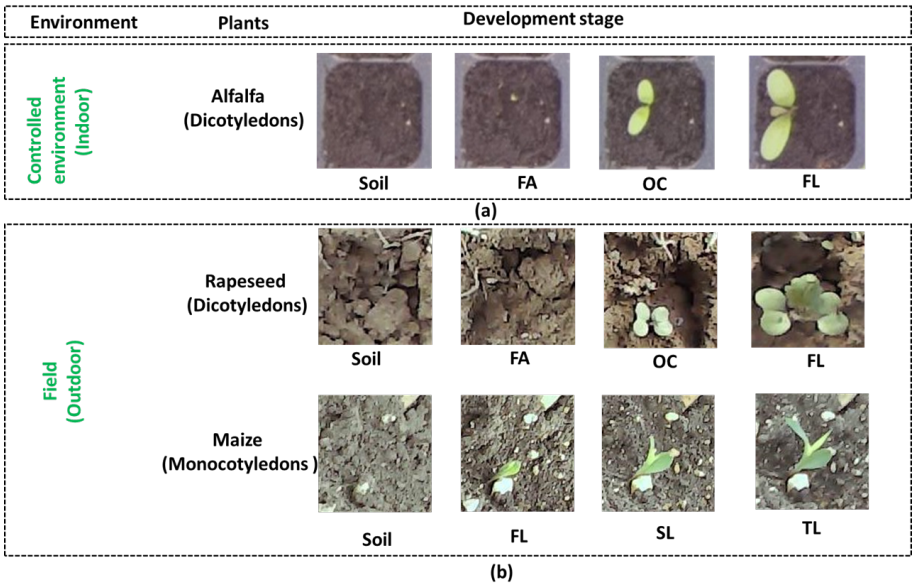


Figure 3.9 – The four developmental stages to classify are the soil, the first appearance of the cotyledon (FA) or First leave (FL), the opening of the cotyledons (OC) or Second leaf (SL), the appearance of the first leave (FL) or Third leaf (TL). (a) Images from the indoor environment. (b) Images from the outdoor environment.

Table 3.5 – Datasets used in the study for model training and inference.

Dataset	Plant Species	N° Images	N° Plants	Environment	Train N° Plants	Validation N° Plants	Test N° Plants
Indoor	Alfaalfa	449 286	600	controlled	400	200	-
Synthetic	Alfaalfa	449 286	600	controlled+simulated shadow	400	200	-
Outdoor	Rapeseed	14 022	57	filed	26	6	25
Outdoor	Maize	14 592	57	field	26	6	25

We used a color-based object detection method to perform plant/background segmentation in temporal sequences of RGB images (Fig. 3.10a). First, we converted our RGB images into HSV color space that decomposes the colors into their hue and saturation components plus the value component. Then, we applied the following lower and upper boundaries defined empirically: (36, 25, 25) and (86, 255, 255) - to filter the green color

of vegetation on HSV images and to get binary masks. After, we cleaned the noise on binary masks, applying mathematical morphological operations opening and closing (Fig. 3.10b). Finally, we applied the resulting binary masks to original RGB images to separate plants from background pixels. The next step was to extract the individual plants to build the database. First, we selected the most contrasted image in the last emergence stage, FL for rapeseed and TL for maize (Fig. 3.10c). Afterward, we designed bounding boxes around each plant (Fig. 3.10d). Finally, we used the coordinates of bounding boxes for cropping the area of each plant on temporal sequences of images (Fig. 3.10e).

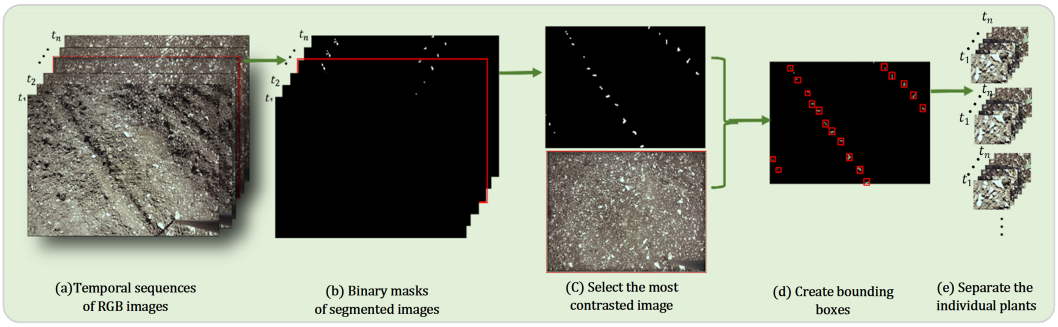


Figure 3.10 – Pipeline of individual plant extraction.

- **Deep learning method**

We used a basic CNN architecture from previous studies [59, 103] to classify images from three datasets presented in Fig. 3.9 into four classes of seedling development stages. The architecture of CNN was composed of five convolutional layers with filters of size 3×3 and respective numbers of filters 64, 128, 128, 512, and 512 each, followed by rectified linear unit (RelU) activations and 2×2 max-pooling, a fully connected layer with 512 units and ReLU activation, a fully connected output layer with four classes corresponding to the development stage and a softmax activation. In addition, we used cross-entropy as a loss function and Adam as an optimizer.Figure 3.4 presents the model architecture optimized for the image classification in four classes. We tested different transfer strategies of knowledge learned from a source, indoor images, to target outdoor images:

1. **Direct transfer:** training the CNN model from scratch on the indoor dataset and testing on the outdoor dataset.
2. **Data augmentation:** we simulated outdoor images from indoor images, adding automatically generated shadows, as proposed earlier in [113]. After that, we trained the model from scratch using the indoor dataset and the obtained synthetic dataset.

3. **Fine-tuning:** the training does not start from scratch. We took the model's weights estimated on synthetic data (2) as initial weights and re-train the model using outdoor images. Then, the model was tested on the outdoor images.

The results of transfer strategies were compared with the model trained and tested on the outdoor dataset, named "Reference" in Tables 3.6 and 3.7. The split ratio of datasets into train, validation and test parts is presented in Table 3.5.

- **Evaluation metrics**

To evaluate model performance, we used false positive FP, false negative FN, true positive TP, and true negative TN that constituted the following metrics: recall, precision, F1-score (Tables 3.8 and 3.9), and accuracy (Tables 2, 3):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \ , \tag{3.1}$$

$$Precision = \frac{TP}{TP+FP} \ , \tag{3.2}$$

$$Recall = \frac{TP}{TP+FN} \ , \tag{3.3}$$

$$F1 - Score = 2 \times \frac{Precision*recall}{Precision+recall} = \frac{2*TP}{2*TP+FP+FN} \ , \tag{3.4}$$

Table 3.6 – Performance of CNN model in outdoor datasets of rapeseed.

	Direct transfer	**Data augmentation**	**Fine-tuning training**	**Reference**
Without background	0.35 ±0.82	0.45 ±0.01	**0.83 ±0.02**	0.73±0.03
With background	0.35 ±0.03	0.46±0.04	**0.85 ±0.01**	0.76 ±0.01

3.3.2 Results

First of all, we compared the models trained on images without soil background in the sequence of images with the models trained on images with soil background. Unlike previous results [59], removing the background in outdoor images did not improve the accuracy results Tables (3.6 and 3.7). Background suppression while keeping the plant

Table 3.7 – Performance of CNN model in outdoor datasets of maize.

	Direct transfer	**Data augmentation**	**Fine-tuning training**	**Reference**
Without background	0.23 ±0.02	0.32 ±0.02	**0.79 ±0.02**	0.70±0.01
With background	0.30 ±0.01	0.35±0.02	**0.81 ±0.02**	0.72 ±0.02

Table 3.8 – Confusion matrix for the best results of CNN method for rapeseed.

True classes	**Predicted**							
Soil	**FA**	**OC**	**FL**		**Precision**	**Recall**	**F1-Soore**	
Soil	1775	195	0	0	**Soil**	0.95	0.90	0.92
FA	102	293	54	0	**FA**	0.51	0.65	0.57
OC	0	91	1674	354	**OC**	0.97	0.79	0.87
FL	0	0	2	1610	**FL**	0.82	1.00	0.90

is a complex task in the outdoor images because the soil is non-uniform. Segmentation errors might explain this result. In the same tables, we can see that the simple simulation of shadow in the indoor dataset has a significant increase of around 9% in both outdoor databases. However, performance values were low compared to the model trained and tested on outdoor images. Then, we fine-tuned the model with simulated outdoor images with small real outdoor datasets. The fine-tuning strategy resulted in significant accuracy improvement, from 72% to 81% for the maize dataset, and from 76% to 85% for the rapeseed dataset.

Tables 3.8 and 3.9, present the confusion matrix of the best performance of the CNN

Table 3.9 – Confusion matrix for the best results of CNN method for maize.

True classes	**Predicted**							
Soil	**FA**	**OC**	**FL**		**Precision**	**Recall**	**F1-Soore**	
Soil	1302	197	2	0	**Soil**	0.88	0.87	0.87
FL	175	1241	59	0	**FL**	0.66	0.84	0.74
SL	0	425	2115	87	**SL**	0.92	0.81	0.86
TL	0	23	129	596	**TL**	0.87	0.80	0.83

model, obtained with fine-tuning strategy on outdoor images. Also, we computed recall, precision, and F1-score for every class to interpret the results and errors types.

For rapeseed, all classes had good precision and F1-score except the first appearance class (FA), 30% of images were misclassified in soil class. This error is logical because detecting the emergence of plants in the field is very difficult.

The comparison of Tables 3.8 and 3.8 revealed that the performance of the classification of the second class (FA, FL) was better for maize than for rapeseed. It can be explained by the fact that seeing the first leaf is simpler than seeing a plant's first appearance. Table 5 shows that around 20% of images in the SL class are incorrectly classified as the FL class, and we observe the same between classes TL and SL. These errors are due to the variable lighting during the day, resulting in exposed or underexposed images that may hide the leaves. This error is not visible for the rapeseed dataset (Table 3.8) since the dataset was not acquired in the same season (autumn for rapeseed and summer for maize). Finally, we demonstrated that our approach works for both types of plants: dicotyledon and monocotyledon.

3.4 Synthetic to real transfer on sunflower flowering

In the two previous sections (section 3.2 and 3.3), we presented two works of transfer from a real environment (indoor) to another real environment (greenhouse or field). In these two approaches we were relying on existing annotated dataset from similar crops acquired in environments different from the ones met in variety testing. These dataset may not always be accessible. Another approach to still benefit from transfer learning can be to generate synthetic dataset automatically annotated [114]. This approach is widely used already in plant phenotyping based on deep learning (see for some recent proofs of feasibility [115, 107, 39, 116]). In most of these successful attempts, a specific pipeline of image generation is specially designed for a given purpose. This pipeline is not generic and therefore needs to be redone for each use-case. This somehow limits the interest as the generation of the synthetic and automatically annotated images requires a significant time of software development. We wanted to test another framework and investigate how virtual reality gaming environment could offer a framework for the generation of synthetic annotated data of plants. We selected Unity the 3D engine gaming environment. Unity includes libraries with hundreds of realistic models. Some libraries are dedicated to plants. This includes the main crops. We selected one of these models and focus on flowering sunflower detection for this attempt as it was one of our identified important variety testing trait to be automated. We describe this pilot investigation with the generation of the synthetic environment and then the transfer learning experiment.

3.4.1 Material and methods

- **Synthetic data**

In the following, we first describe how to prepare Unity's 3D engine (and their Perception package) to generate automatically annotated synthetic images, then explore how to design the virtual sunflower field and use the synthetically generated data to bootstrap the project and get a prototype running.

- **Preparation the Unity's 3D engine**

Unity is a cross-platform game engine developing 2D and 3D multiplatform games and interactive experiences. We need a Unity account and license to start working with Unity first. After signing up, we need to install Unity Hub, a management tool that will allow us to switch between versions of Unity. We installed Unity 2019.4.18f1 (LTS) – the current

"Long Term Support" version of Unity which support the Perception package that we used in this project to generate perfectly annotated synthetic images. The chosen template for this project is Universal Render Pipeline (URP). We also installed the Perception Package through the Unity package manager module and imported it into our project.

- **Methodology to design a virtual environment.**

— **Initial Analysis and Preparation:**

Initial analysis and preparation is the first step of building the virtual scene in Unity 3D. This stage includes every preparation task required for completing the virtual scene. When creating a virtual environment, one of the most crucial requirements will often be its similarity with the real environment. Therefore, the user must analyze every detail of the environment, such as the plants, lighting, camera's positioning, and environment layout. The best practice is to collect as many references as possible. This includes photos and videos of the environment, object dimensions, and 2D layout. Appropriate references will make creating 3D assets and the virtual environment much more straightforward. The virtual background for our scene includes the terrain, the sky, and the trees in the background shown in Figure 3.11.

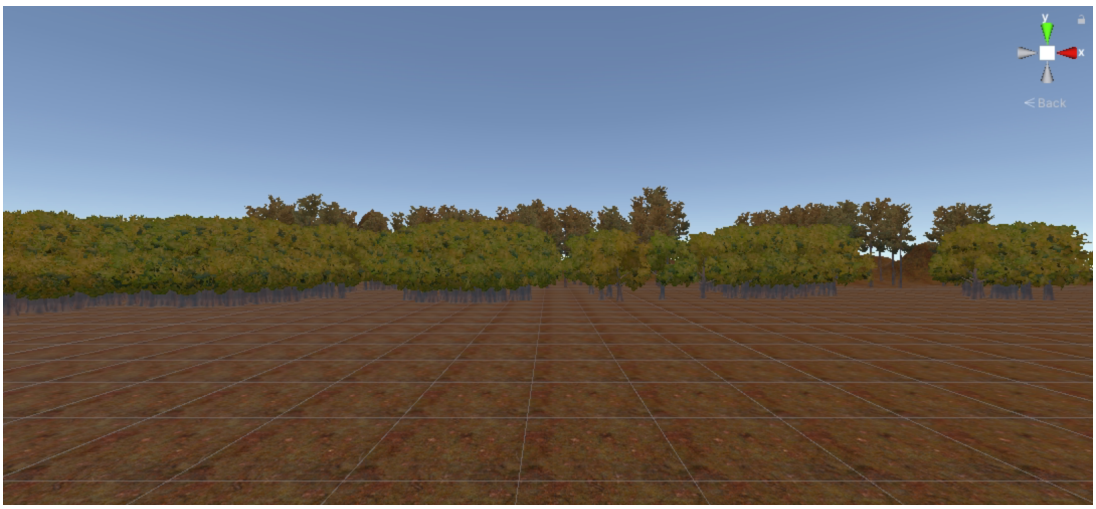


Figure 3.11 – Background of the scene.

As mentioned in [117], the direction and elevation angle of mature sunflower heads are varied based on the time of the day and the age of the flower. In this project, we also consider different angles for the head of mature sunflowers positioned randomly. In each scene, we also consider the portion of the sunflower immature. Automatically by running

a scrip, we have the evolution from 10 percent to 80 percent of the maturity on sunflowers. The sun's position in the sky also follows the daylight pattern, making it possible for us to simulate the day's shadow and light. We consider the weather a clear sunny day in all the scenes.

— **Design prefabs of sunflowers:**

The main game object in this virtual scene is the 3D sunflower. There are many free samples of 3D sunflowers in Unity asset stores. We provide a realistic 3D high-poly sunflower model from the SpeedTree library for this project. However, the structure of this 3D model does not let us split it into different components. So we create two main game objects, the body and the head. For the body part, we consider the shading model of the petals to the "transparent" surface and the rendering face to the "back" by this trick, we transform the sunflowers into immature.

The second game object is the head; in this one, we consider the transparent shading for all the materials except the head. We add the "Flower" tag to this object to make it easier to access in coding. We save both game objects as a prefab.

Figure 3.12, shows the head and body prefabs of sunflower.

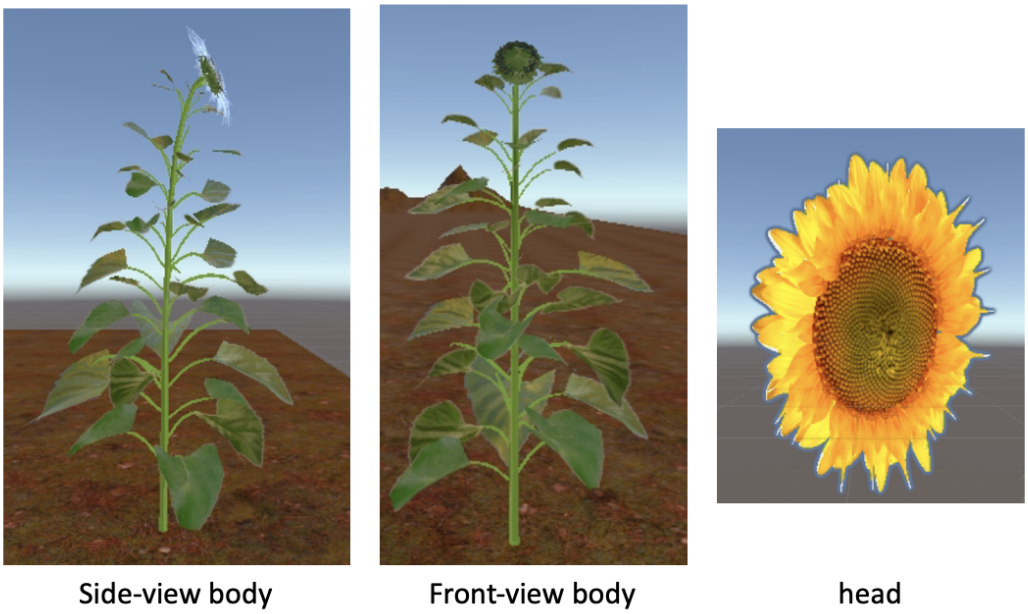


Figure 3.12 – Different components of sunflower.

As a reference field, we add 100 body prefabs to the scene that consists of the field and the trees as a background. This scene is considered as a field of zero percent maturity of

sunflower. After this stage, we need to write two scripts, one for adding the head on the top of the sunflower's body in a random position and slightly different angles and also for simulating day lighting. At this stage, each time we run the game, we see the sunflower field with 10 to 80 percent of mature plants randomly positioned.

Figure 3.13 illustrates one example of a virtual environment designed by 3D unity which simulated the real scene. We consider several images captured in the field with different backgrounds for creating this virtual environment.

Figure 3.13 – Simulated sunflower field designed by 3D unity.

To use the perception package first, we must add "Ground Truth Renderer Feature" from the ForwardRenderer.asset to the project. This step prepares the project to render tailor-made images that will be later used for labeling the generated synthetic data.

We then add the necessary components to the camera to equip it for the Perception workflow. To do this, we need to add a Perception Camera component to the main camera through add component feature and then define which types of ground-truth we wish to generate using this camera. There are seven common labelers for object-detection and human keypoint labeling tasks such as keypoint labeling, 3D bounding boxes, 2D bounding boxes, object counts, object information (pixel counts and ids), instance segmentation, and semantic segmentation. We used 2D bounding boxes, object counts, object information, and semantic segmentation in this project.

The labeler added to the Perception Camera should know which objects it should label in the generated dataset. To do that, we should first create label configurations. This

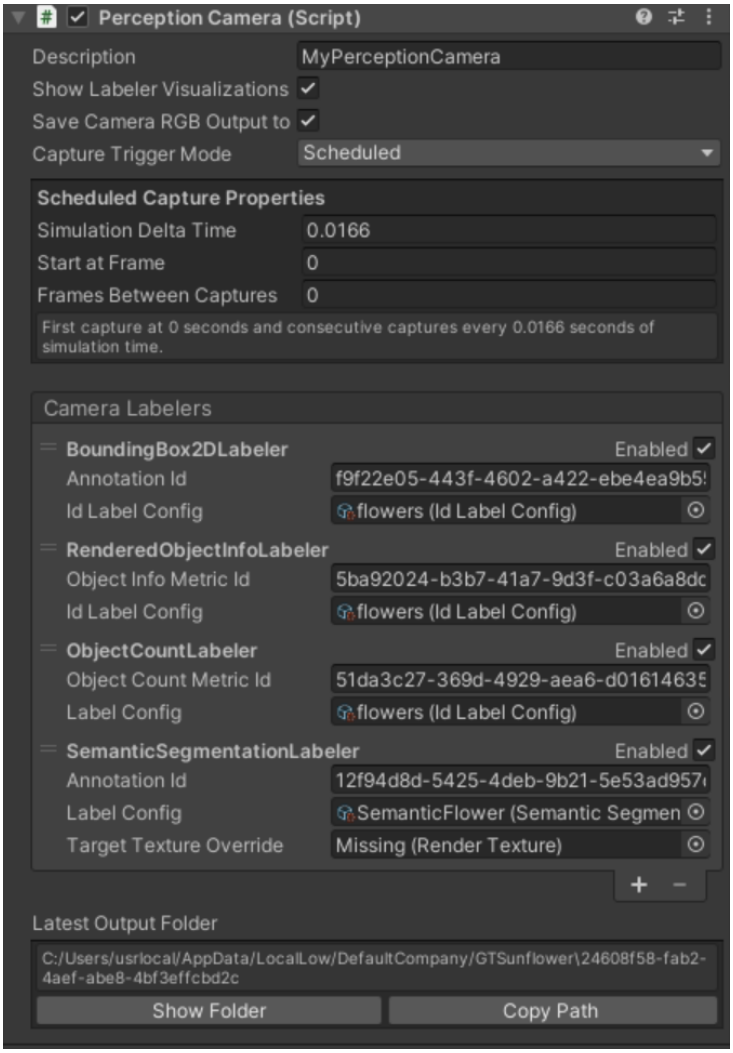


Figure 3.14 – Properties of Perception Camera component.

way, the labeler looks for specific labels within the scene and ignores the rest.

After adding the labelers and defining the IDLableConfig, the Inspector view of the Perception Camera component will look like Figure 3.14. The next step is to assign labels to the sunflower's head prefab that are supposed to be detected by an eventual object-detection model and add those labels to the label configurations we have created.

The prefab has a component, namely Labeling. This component is specific to the Perception package and is used to assign object labels. Each object can have multiple labels assigned and thus appear as different objects to Labelers with different label configurations. At this level, since we enable the visualizations option on our Perception Camera, we can see a bounding box drawn around the sunflower's head in the scene and the object

itself being colored according to its label's color, as illustrated in Figure 3.15.

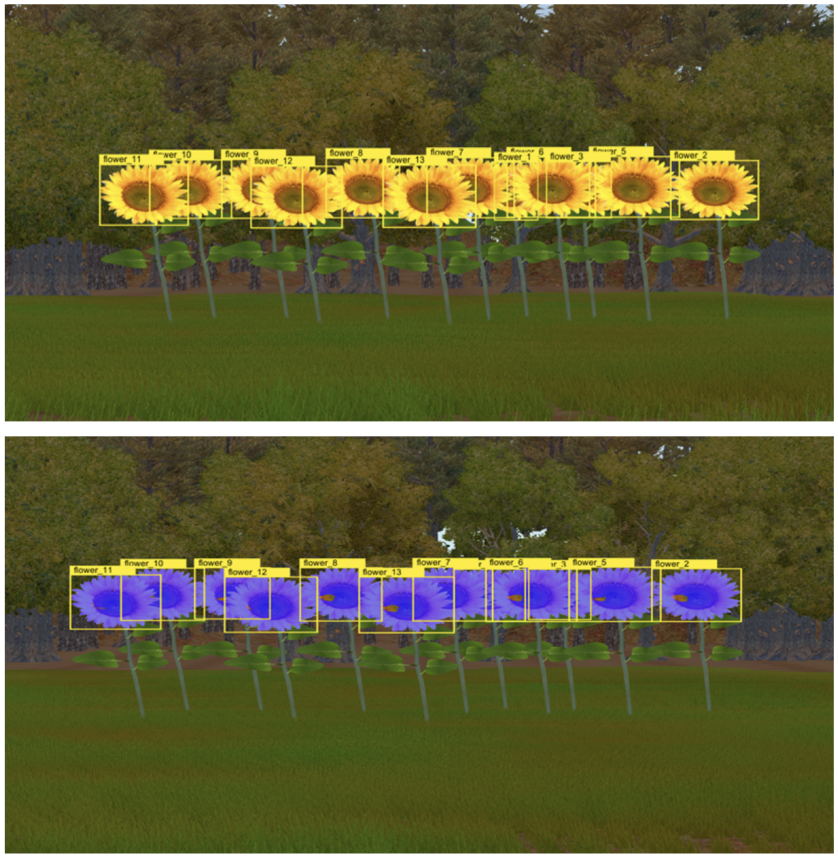


Figure 3.15 – Annotated sunflowers.

In this project, we defined a fixed-length scenario to capture 1550 RGB images and corresponding ground truth by randomly adding sunflowers heads to the scene. The generated dataset is in the Perception format and contains four types of data, Logs files, JSON data, RGB images (raw camera output), and semantic segmentation images. All the bounding box position information is presented in the JSON file format.

- **Real data**

We have 193 RGB images acquired in the field during the flowering time from a different parcel. Figure 3.16 shows an example of dataset. The size of image is 900×550 pixels. All images are annotated using a bounding boxes around each flower in images.

- **Deep learning method**

YOLO (You Only Look Once) is a deep learning method for object detection [118]. This model is based on a neural network that takes an image as input and predicts the

Figure 3.16 – Real images of sunflowers in the field.

bounding boxes and the class labels for each bounding box. The model works by first dividing the input image into a grid of cells, where each cell is responsible for predicting a bounding box. A class prediction is also based on each cell. YOLO is composed of a total of 24 convolutional layers followed by two fully connected layers. The layers are separated by their functionality as follows:

1) The first twenty convolutional layers are pre-trained on the ImageNet 1000-class classification dataset. The layers include 1×1 reduction layers and 3×3 convolutional layers.

2) The final four convolutional layers followed by two fully connected layers are added to train the network to detect objects with our database. The final layer predicts class probabilities and bounding boxes.

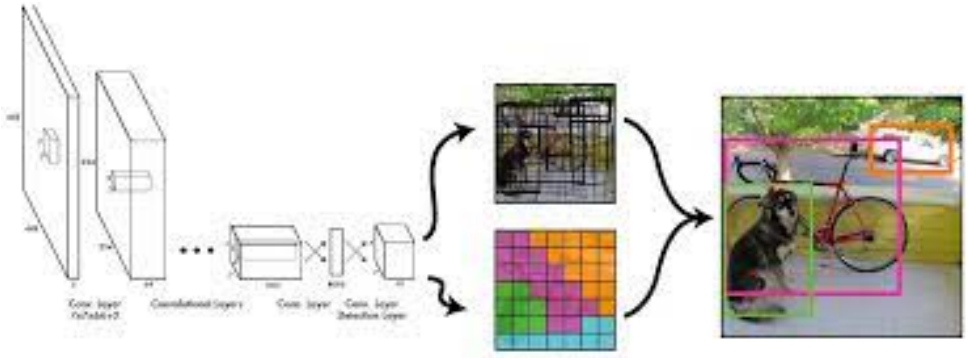


Figure 3.17 – YOLO architecture.

We tested two different transfer strategies. In the first strategy, we trained the YOLO model using synthetic data. Then, we split the data using 965 images in training, 340 images for validation, and 245 for the test. Furthermore, we test the model directly in real data. This approach is called direct transfer. In the second strategy, we used the weights estimated on synthetic data as initial weights and re-train the model using different numbers of images in the training dataset from real data: 30, 50, 70, 90, 110, and 37 images in validation and 50 images in the test. This approach is called fine-tuning. Finally, to evaluate the added value of this second approach, we train the same models from scratch using real images. Table 3.10 present all the dataset used in a different approach.

- **Evaluation metrics**

The mean Average Precision (mAP) is an evaluation metric for Object Detection. The mAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in detecting. The mAP is calculated by finding the Average Precision(AP) for each class and then average over a number of classes:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_I \tag{3.5}$$

3.4.2 Results

We first tested the performance of the model trained and tested in synthetic data. As visible in Table 3.10, we get good results with 82% of mAP. Then, we tested the direct transfer approach. We get 26% as an mAP. As visible in Figure 3.15 and Figure 3.16, there are some differences in the flower size, and we see more depth in the real image compared to the synthetic one. Based on the preliminary results in Table 3.10, we can see an increase in performance by some percent when using the fine-tuning approach. This is however not systematically the case depending on the amount of real data used for the fine-tuning. Figure 3.18 shows a representative result of flower detection using our best model. We can see some error detection mostly related to false detection of small objects. These would certainly be removable easily by post-processing. Despite not perfect, the described experiment shows that the Unity environment can be of value for the generation of automatically annotated data set to boost deep learning training stage.

Table 3.10 – Description of the datasets and the performance of each approach.

		Train	Validation	Test	mAP
Synthetic data		965	340	245	82%
Real Data	**Without fine tuning**	110	37	50	71%
		90	37	50	68%
		70	37	50	65%
		50	37	50	60%
		30	37	50	50%
	With fine tuning	110	37	50	73%
		90	37	50	72%
		70	37	50	64%
		50	37	50	55%
		30	37	50	45%

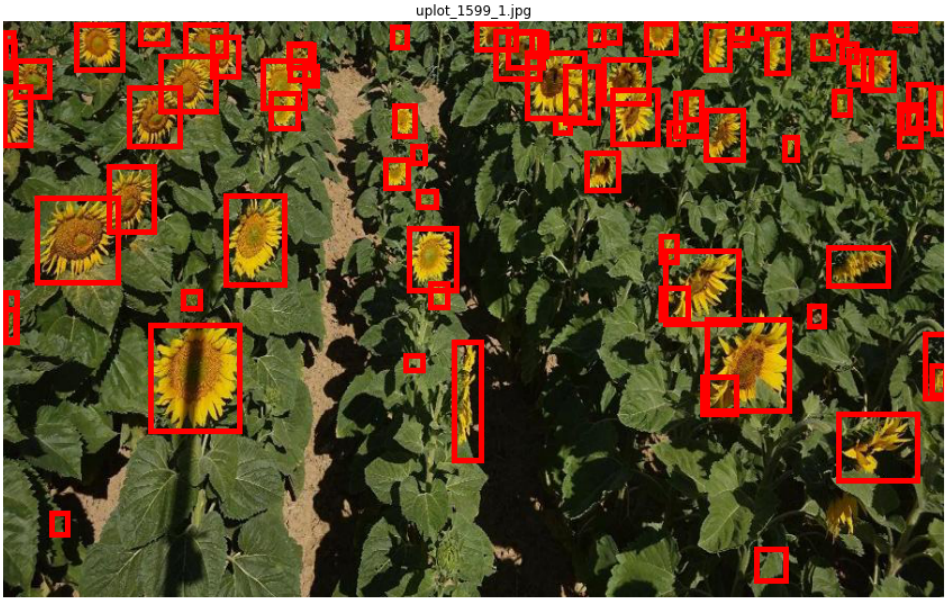


Figure 3.18 – Example of result of flowering detection by the best performance (training on synthetic and transfer learning with fine tuning).

3.5 Conclusion

In this work, we have demonstrated the possibility of using indoor images to transfer knowledge to deep learning algorithms operating on greenhouse and outdoor images. This was illustrated quantitatively on a task of seedling emergence for crops in variety testing trials. To pragmatically quantify the gain brought by our transfer learning approach, we estimate two weeks of work to annotate the 600 sequences of images acquired in indoor

conditions (at a speed of around 60 sequences per day). Thanks to transfer learning from indoor to outdoor environment, the amount of data requested to reach a plateau of performance was found to be 30 pots. This can be achieved in one day of annotation work. This study can be used to reduce the time-consuming annotation task. It would be interesting to extend these results to other informational tasks and a variety of plant developmental stages. In our approach, the outdoor noise considered was limited to shadow. However, other sources of noise could also be included to extend the result of this study. This includes for instance the presence of wind causing motion blur which could also easily be simulated with data augmentation following the approach presented in this study. In this chapter, because of chronological constraints during the progress of the Phd we used RGB images. We have shown in the previous chapter that the depth was boosting the discriminative value of the images for seedling emergence. We would of course recommend to use both Depth and transfer learning since the Depth sensors used in the previous chapter can operate outdoor.

A last attempt of transfer learning was carried out on the detection of sunflower during flowering. We tested on a preliminary investigation mode, a virtual gaming environment to generate automatically annotated images. We have described the protocol to generate easily such virtual environment and hack it for machine learning purposes. We have shown that some model designed outside any consideration of scientific agronomical purposes could be used to very efficiently design virtual fields. Also, we have shown that a small benefit of some percent of performance could be obtained via the pre-training on such virtual environment. While preliminary this approach is promising as it can adapt to any situation to mimic realistic conditions including variable lighting conditions or image acquisition setup. Nevertheless, the fixed aspect of the virtual plant available in the environment seems limiting here and having different stages of development would certainly help. A way to address this issue could be to connect the structure-function plant models [119] which incorporate variability and more anatomically relevant features with 3D engine from gaming. This could constitute an interesting perspective to improve the synthetic to real transfer learning approach tested here.

Chapter 4

Optimized multispectral imaging and machine learning for wheat disease quantification.

4.1 Introduction

In this chapter, we propose the design of a new multi-spectral camera to estimate the *Fusarium* area on wheat spikes more objectively and faster in the framework of plant variety testing. First, we explain the transferring process from a hyperspectral camera to the new multi-spectral camera. Then, we validate the functionality of the multi-spectral camera under controlled conditions. Finally, we demonstrate its performance under field conditions.

In this study, we focus on *Fusarium* head blight(FHB) infecting wheat spikes. FHB can cause significant economic losses for a producer, especially since the fungicide treatment, under optimal application conditions, has only 50 to 75% effectiveness. Moreover, there is a critical sanitary issue since *Fusarium* produces mycotoxin deoxynivalenol (DON) in the grains, threatening both human and animal health [120, 121]. Thus, since the first of July 2006, the cereal industry has been subject to the European regulation 1881/2006, which sets maximum levels of deoxynivalenol (DON). Respect for the regulatory limits has become a new reality for the cereal market. The economic repercussions are heavy in case of downgrading of non-compliant batches. As a result, selecting wheat varieties resistant to FHB has become a priority.

The permanent technical committee for plant breeding (CTPS: le Comité Technique Permanent de la Séléction) encourages the development of resistant varieties by facilitating their registration in the Official Catalogue of Species and Varieties of Cultivated Crops in France and penalizing the susceptible varieties. The current methods of testing varietal resistance are based on human observation or chemical analysis of wheat spikes.

However, visual disease scoring is a time-consuming task, giving subjective results since it depends on the qualification of experts. In parallel to visual disease scoring, chemical analysis is usually performed, including liquid chromatography coupled to mass spectrometry (HPLC/MS-MS). However, this approach is costly, challenging to implement, and inadequate for studying many grain samples. Thus, there is still a need to develop more efficient high-throughput automated phenotyping tools for FHB detection.

Currently, RGB [122, 123, 124, 125] and hyperspectral [**agriculture4010032**, 126, 127, 128, 129, 130, 131, 132, 133, 134, 133] imaging systems are widely used for FHB disease detection. It was shown that hyperspectral imaging is more performant than RGB for FHB detection [41]. Thanks to its high sensitivities, hyperspectral imaging can be used for disease detection before the emergence of visible symptoms [135]. However, despite their numerous advantages, the hyperspectral systems are inconvenient for field conditions. The acquisition protocols are unsuitable for practical implementation. In [**agriculture4010032**, 133, 134], they used a big box for the acquisition which the implementation process is very time consuming (see an example in Figure 4.1). In addition, huge generated data volumes are complicated to process in real time. Thus, our study aims to design an original light multispectral imaging system prototype adapted to field conditions, overcoming the weaknesses of hyperspectral systems.

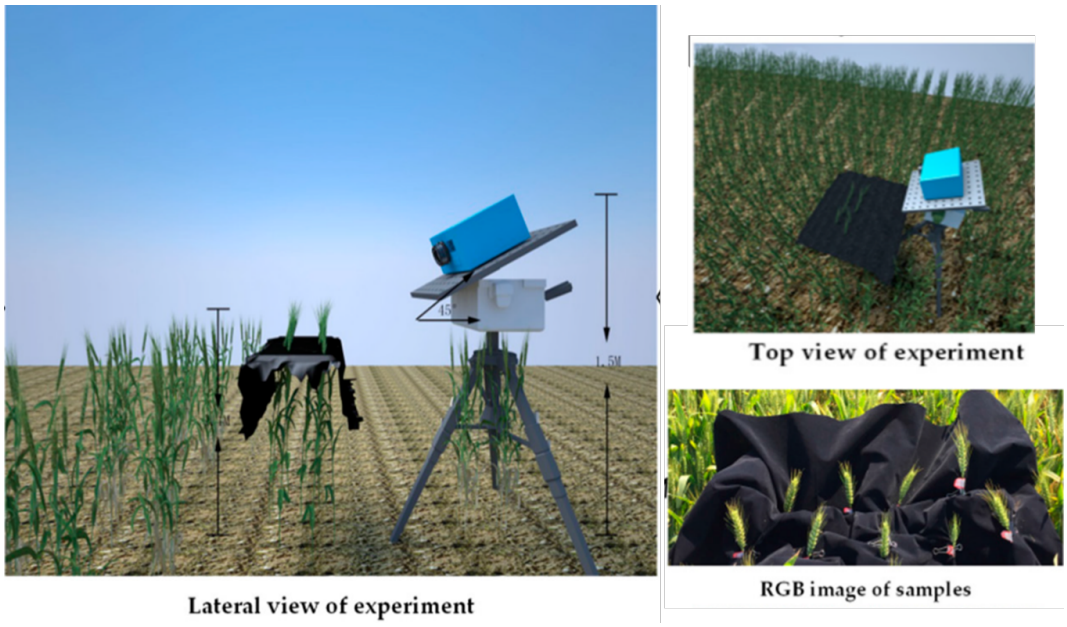


Figure 4.1 – The acquisition protocol uses a hyperspectral imaging system designed for field conditions [134].

One of the methods to reduce the amount of data is to select the most discriminating wavelengths within hyperspectral images [130, 127, 136, 133]. A small dataset is used in other studies [126, 127, 128]. For example, in [126], they used a dataset containing 86 samples. Where in our study, we selected wavelengths using 1300 images of infected wheat spikes collected from two experimental sites with distinct weather conditions and acquired under indoor conditions for four years. To our knowledge, this is the first time that the optimal wavelength shows stability over four years. Moreover, the selected wavelengths are used to build a new lightweight multispectral camera suitable to field conditions, simplifying the acquisition protocol by comparison with other studies [**agriculture4010032**, 133, 134]. We develop models for estimating the percentage of *Fusarium* detection in the control and field environments. Furthermore, we proposed a segmentation model of the first row of spikes associated with our acquisition protocol without physically isolating the row of interests, as seen in [122].

4.2 Materials and Methods

In this study, we focus on the estimation of the percentage of *Fusarium* disease. The building process of the new multispectral camera followed the four steps illustrated in Figure 4.2. First, the optimal wavelength to detect the *Fusarium* disease was selected in a controlled environment from a hyperspectral camera Figure 4.2.A. Then, a multispectral camera was designed using the selected wavelength and tested 4.2.B in indoor conditions. Third, segmentation of spikes of wheat was performed in the field Figure 4.2.C. Finally, *Fusarium* severity was detected on wheat spikes in the field 4.2.D. In the following sections, we will describe the steps from part A to part D of the global pipeline. For the rest of this chapter, we define the percentage of detected *Fusarium* as the severity :

$$severity = \frac{\text{disease spike area}}{\text{whole spike area}} \times 100(\%). \tag{4.1}$$

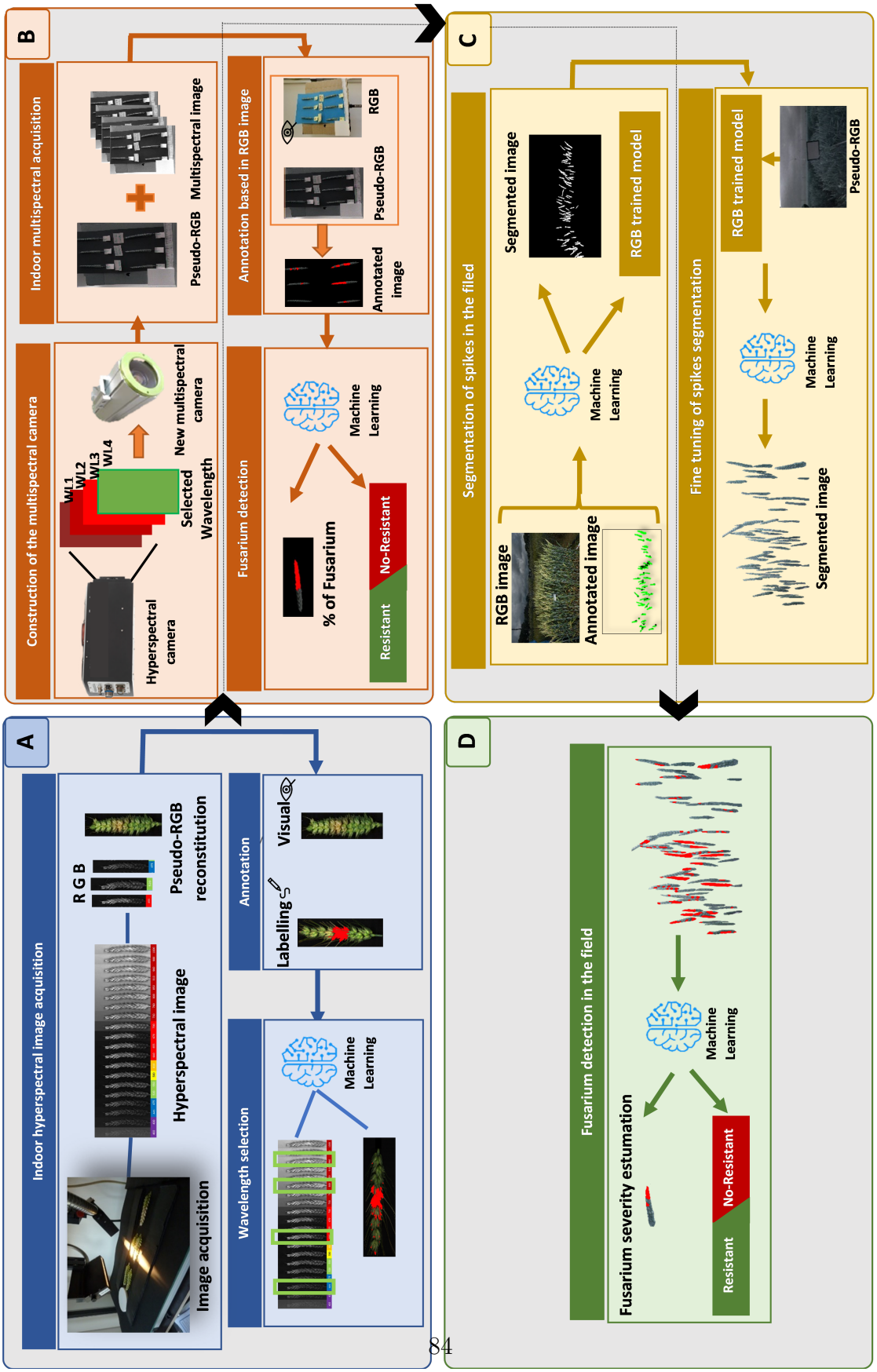


Figure 4.2 – Global pipeline of building and testing the multispectral camera. A: optimal wavelengths selection in a controlled environment from a hyperspectral camera. B: Designing and testing of multi-specral camera. C: Wheat spikes segmentation in the field. D: *Fusarium* severity estimation in the field.

4.2.1 The building of optimized multi-spectral camera

- **Optimal wavelengths selection**

In this section, we explain the different phases of the optimal wavelengths selection (Figure 4.3). A hyperspectral camera (NEO Hyspex VNIR-1800), including 216 wavelengths over a spectral range from 400 nm to 1000 nm, was used. The hyperspectral image acquisitions were realized in a controlled environment and repeated over four years: from 2016 to 2019. We collected spikes from three wheat species, durum wheat, soft wheat, and triticale, in two different sites in France, Angers and Clermont Ferrand, with distinct weather conditions. The spikes development stage is between 250°C/d and 550°C/d after inoculation. We used ten varieties from each wheat species. For each variety, site and year, five spikes were harvested in the field and placed on a dark background under the hyperspectral camera at a distance of 30 cm. The obtained database of hyperspectral images included an overall of 1500 spikes.

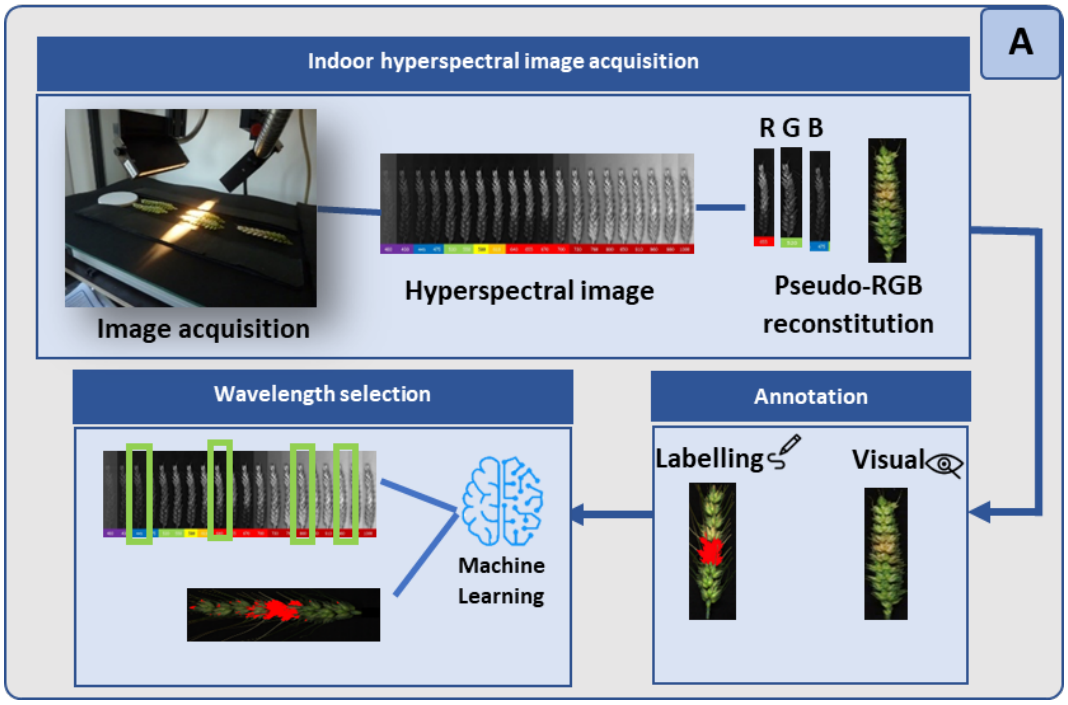


Figure 4.3 – Optimized wavelength selection from the hyperspectral camera in a controlled environment.

The annotation of *Fusarium* severity in the hyperspectral images was estimated by two independent approaches. First, pseudo-RGB images have been reconstructed from

hyperspectral images using three wavelengths associated with the Red, Green, and Blue channels. The spikes included in each image were segmented using a simple color thresholding algorithm. Then, these reconstructed images were analyzed by three different experts providing the *Fusarium* severity in every spike. The *Fusarium* area was manually annotated by the expert based on the results of chemical analysis [137] to be sure that each annotated pixel corresponds to an FHB pixel.

After creating the annotated database, we applied several classical machine learning methods to select optimal wavelengths. These methods included linear Discriminant analysis sequential step by step (**DASS-Seq**), Covariance Selection (**CovSel**) [138], and non-sequential linear Discriminant analysis sequential step by step for 2λ or 3λ (**DASS** 2λ and **DASS** 3λ).

DASS is ascending discrimination by computing the Mahalanobis distance [139] on the total covariance. A measure of the distance between the values of observation and the average of all observations on the independent variables. A large Mahalanobis distance identifies an observation with extreme values for independent variables. **CovSel** method is adapted to the multi-response calibration of spectrometers and can apply to the problem of discrimination considering indicator variables as responses. The **CovSel** technique has been specially designed for spectral bands selected for the treatment of two common problems, first, the huge number of spectral bands that yield a huge solution space, and second the strong correlation between them. For each machine learning method, the training database contained 500.000 healthy pixels and 500.000 infected pixels selected from 60 images.

After the training, **CovSel** method and **DASS-Seq** classified twenty wavelengths from the most discriminating to the least discriminating one. Then, we compute the accuracy metric of the classification using a test database starting with the first wavelength and adding one more wavelength at each iteration until we reach all twenty wavelengths. Then, using a test database, we apply several classification tests. Again, we start the classification with the images of the first classified wavelength and add one more wavelength at each test until we reach the final test based on all twenty wavelengths. For each test, we compute the accuracy metric. Then, we plot the accuracy curve as a function of the wavelength number used (see Figure 4.4). Finally, we choose the number of wavelengths when the curve reaches a horizontal asymptote. Figure 4.4 presents an illustrative example of an accuracy curve for the **DASS-Seq** method. In this example, the curve becomes asymptotic with the first five wavelengths. In consequence, we keep those five

wavelengths as optimal ones.

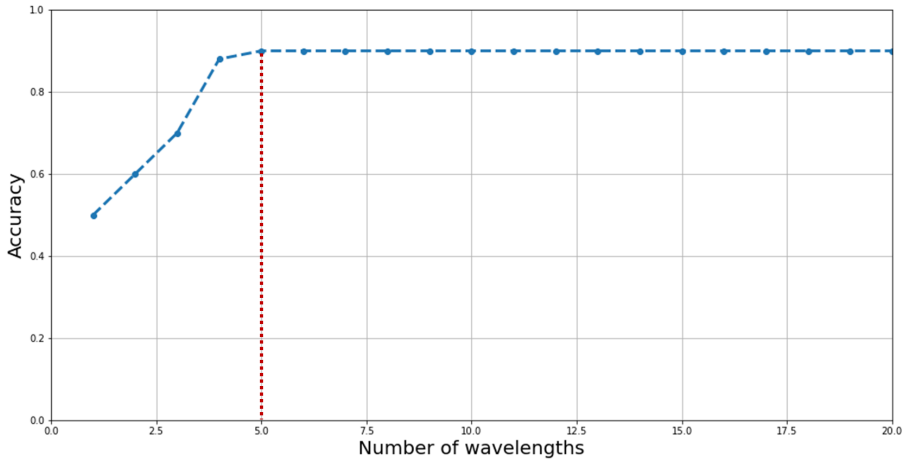


Figure 4.4 – An illustrative example of the choice of optimal wavelength number based on the DASS-Seq method: the accuracy of disease detection as a function of the number of wavelengths used; the curve reaches a horizontal asymptotic with five wavelengths.

To evaluate the performance of selected wavelengths, we apply two different methods. In the first evaluation, we compute the accuracy metric between the pixels predicted by the machine learning models and the pixels annotated by experts. The second one is the R^2, the determination coefficient between the *Fusarium* severity estimated by experts and the severity predicted by the models:

$$R^2 = 1 - \frac{\sum(Y_i - \widehat{Y}_i)^2}{\sum(Y_i - \overline{Y}_i)^2}. \tag{4.2}$$

The R^2 is the proportion of the variance of a dependent variable explained by one or more independent variables in the regression model. The R^2 is expressed as a value between 0 and 1.

- **Test of the new multi-spectral camera in controlled conditions**

After selecting the optimal wavelengths, we continue in our global pipeline. Now, we move to the building of the multispectral camera (Figure 4.5). In the commercial cameras, SILIOS company [**silios**] have a multispectral camera called CMS4 used on field applications. The CMS4 cameras are mainly designed for high integration of multispectral VIS/NIR systems. These lightweights (less than 170g) and compact (52x62x40mm) cameras split the image into eight spectral bands plus one B/W channel. These cameras are designed by hybridizing a Bayer-like mosaic filter on a commercial 4.2 MPixel CMOS

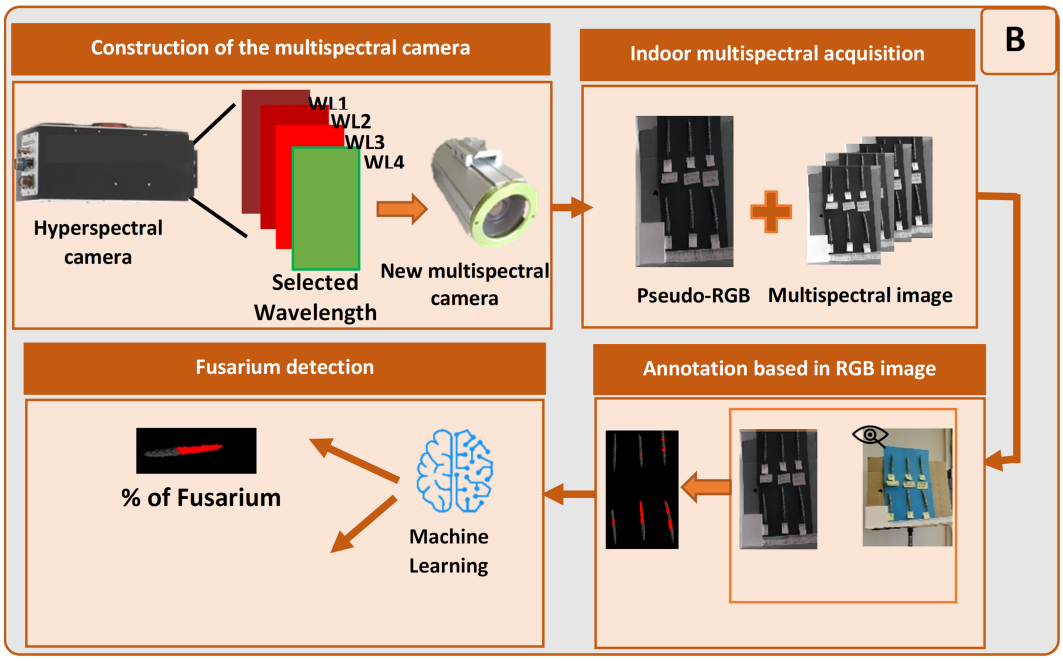


Figure 4.5 – The building of the multispectral camera CMS4 and its experiment controlled conditions.

Sensor. The existing version of CMS4 does not include our optimal selected wavelengths. For that reason, we design a new version of the CSM4 (Figure 4.6) in collaboration with SILIOS company. The camera is assisted by software to acquire and save images. The CSM4 camera is covered by a box to be protected in the field.

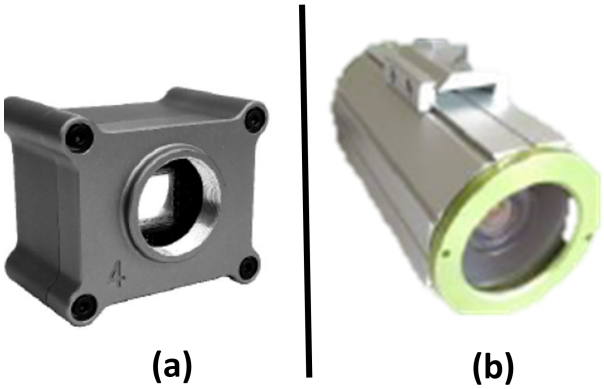


Figure 4.6 – (a) CMS4 camera without box. (b) CMS4 camera with outdoor box.

- **Preliminary test of the CMS4 camera in a controlled environment**

We are certainly losing precision from a hyperspectral camera to an optimized mul-

tispectral camera. For this reason, a preliminary test of our new camera is done under controlled conditions. We acquired 176 images of wheat spikes under controlled conditions using the CMS4 camera and a high-resolution RGB camera. Similarly to creating hyperspectral databases, we collected six spikes from ten varieties of three wheat species on two sites in France on two different dates. Each image has four different layers and a pseudo-RGB image (see Figure 4.7). After the acquisition, chemical analyses were performed on spikes to confirm the presence of FHB disease. Then, based on the results of chemical analyses and the help of RGB images, the experts annotated the *Fusarium* area and estimated the severity in pseudo-RGB images of the multispectral camera.

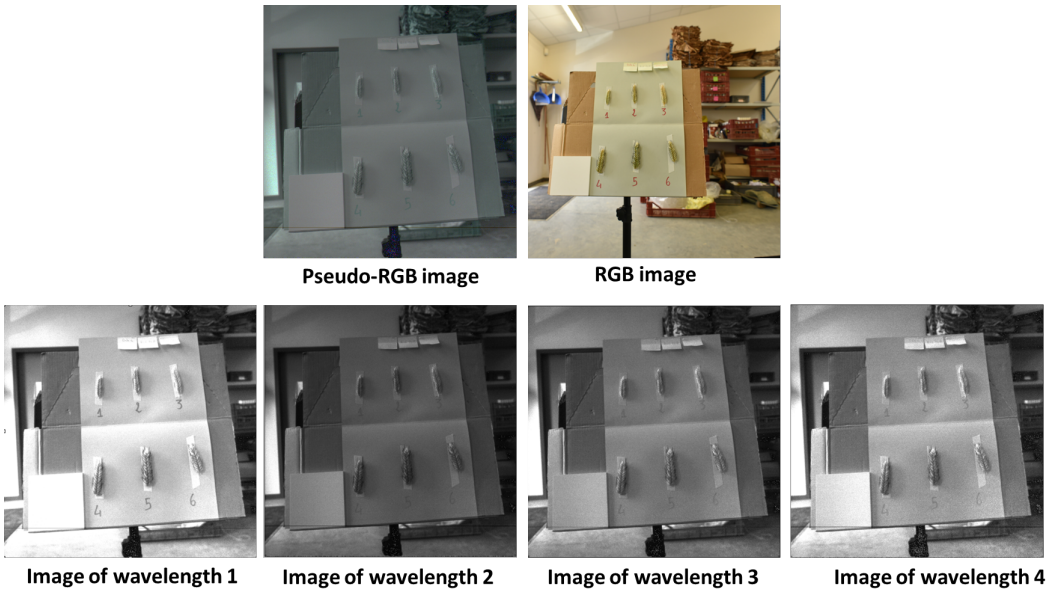


Figure 4.7 – Example of images acquired in controlled condition with RGB and CMS4 camera.

A supervised binary classification to estimate the *Fusarium* severity was implemented using four different machine learning methods: Bagged Trees, Cubic k-nearest neighbors (**GKNN**) [140], Weighted k-nearest neighbors (**WKNN**) [140], and fine Gaussian Support Vector Machine (**FGSVM**) [141]. We use 70% of our annotated data in training and 30% in the test for each method. The performance of all methods was evaluated using recall, precision, and accuracy metrics [142] along the following equations:

$$Precision = \frac{TP}{TP + FP}, \tag{4.3}$$

$$Recall = \frac{TP}{TP + FN}, \tag{4.4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4.5}$$

where TP corresponds to the number of disease pixels correctly detected, TN represents the number of disease pixels incorrectly identified, FN corresponds to the number of healthy pixels correctly detected, and FP represents the number of healthy pixels detected as disease pixels.

4.2.2 In the field: proposed models for segmentation of spikes and FHB detection

Based on the promising results obtained from the CMS4 camera in a controlled environment, we switched to testing the new camera in the field (part C of the pipeline, Figure 4.8).

In the field conditions, each row represents a wheat variety. In order to annotate the variety, the experts compute the average of the severity from twenty spikes. Thanks to the lightweight CMS4 camera, we can easily install it on the tripod. Then, we placed this tripod between the rows of wheat spikes and in front of the interested row. Thus, we can acquire our images with a simple and suitable protocol in the field, unlike other works that used specific imaging boxes [**agriculture4010032**, 129, 130, 131, 132, 134, 143].

Before estimating the *Fusarium* severity in the field, we need to segment the wheat spikes. Some previous works on spikes detection and segmentation have been proposed in the literature [122, 123, 134], but they do not focus separately on each row or use a physical separation of the interested row. Instead, we aim to automatically segment the first row of spikes in the image using an image processing algorithm, and then we estimate the *Fusarium* severity. First, we proposed a segmentation model of the wheat spikes of the first row. Moreover, we present the benefits of transfer learning from a model trained with RGB images to a new model for multispectral images. Then, we applied several classification methods using segmented images to estimate the *Fusarium* severity.

- **Segmentation of wheat spikes in the field**

While constructing the new multispectral camera, we acquired high-resolution RGB images with the same acquisition protocol. We used this data to build and test a first segmentation model of the spikes. In this part, we present the segmentation of the first row

of spikes using RGB images. Then the transfer learning process from the segmentation model trained with RGB images to a model for multi-spectral images.

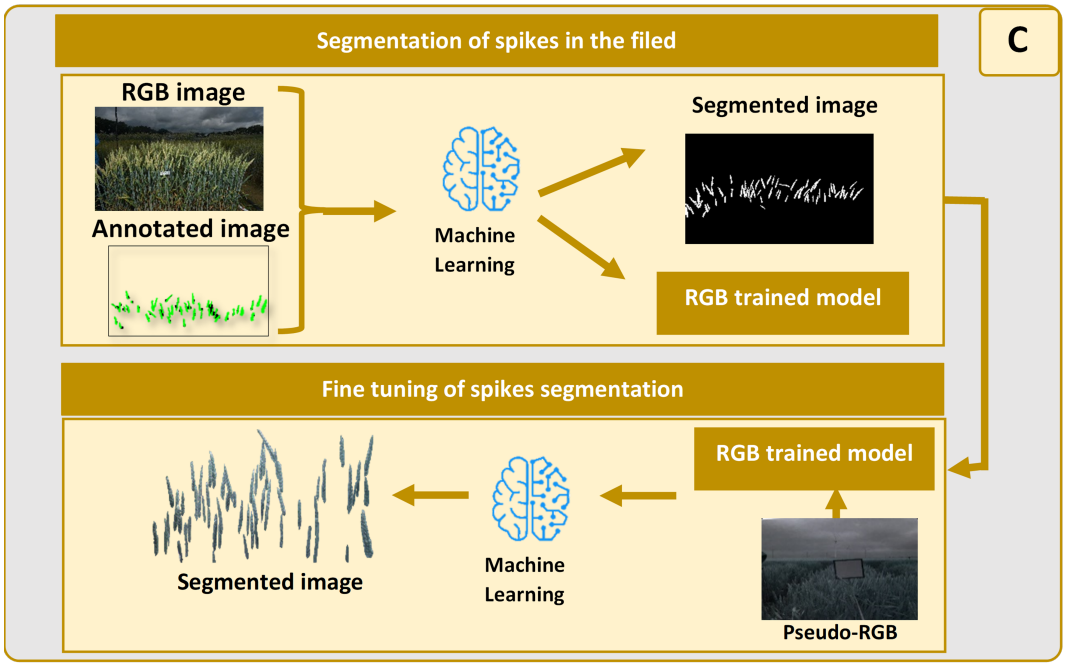


Figure 4.8 – The segmentation of the first row of wheat spikes using RGB and multi-spectral images acquired in the fields environment.

— **Segmentation of wheat spikes on RGB images**

The database includes 220 RGB images acquired in the field environment at two sites. This database contained several varieties: durum wheat, soft wheat, and triticale. Experts manually segment all spikes on the first row in the images. Then, we train a standard U-Net [36] model to be able to segment the spikes automatically. The database is split in the following way: 120 images in the training dataset, 20 images in validation, and 80 images in the test. Evaluation of the results was computed with the Sørensen-Dice coefficient [144]:

$$D = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{4.6}$$

where X is the predicted segmentation and Y is the ground truth.

— **Transfer knowledge from RGB model to multispectral model**

To take benefit of the annotated RGB database, we use the weights of the model trained with RGB images as initial weights for training a segmentation model for multi-

spectral images. This process is called fine-tuning. In the field condition, we have 160 annotated images acquired with a CMS4 camera (see Figure 4.9). We trained the model several times using a different number of images in the training database; we used 0, 40, 60, 80, 100, and 120 images. And we used 20 images for validation and 20 images in the test. As a comparison, we trained from scratch the U-net model using training, validation, and testing from CMS4.

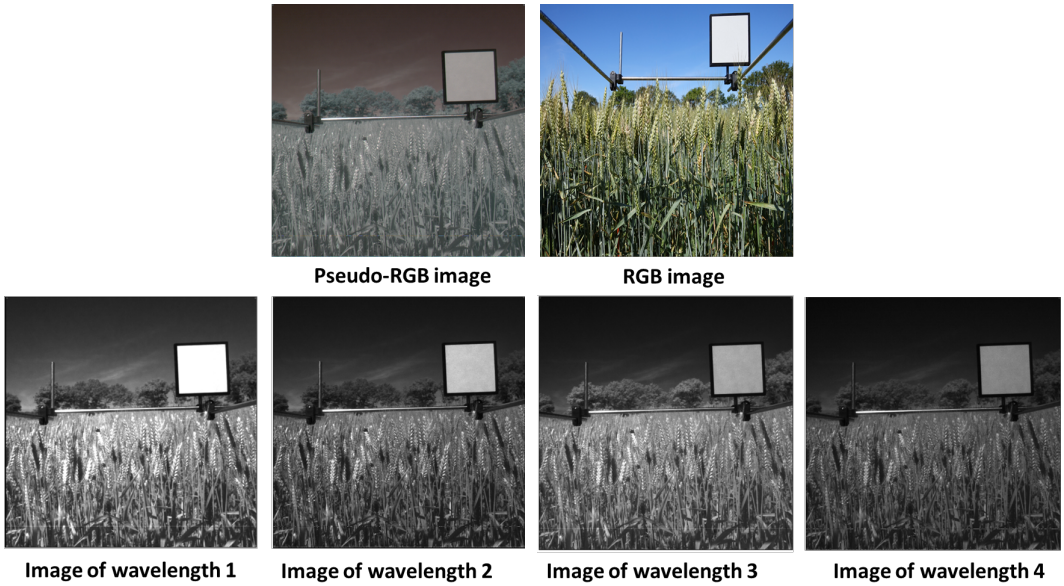


Figure 4.9 – Example of images acquired in the field environment with RGB and CMS4 camera.

- ***Fusarium* detection in the field using multispectal images**

Following the segmentation of spikes, we continue to the last part of the global pipeline, which is the validation of the multispectral camera in the field condition (see figure 4.10)

The manual annotation of the *Fusarium* area on segmented images is done. Next, we apply the same machine learning methods for the estimation of *Fusarium* severity used in the previous section of the test of the multispectral camera in controlled conditions (Part B in the global pipeline). Then, we evaluate the results of *Fusarium* severity estimation in two different methods. The first evaluation is based on the pixel annotation of *Fusarium* computing the accuracy, recall, and precision. The second evaluation is the correlation between the severity of *Fusarium* provided by the expert on image compared to the severity predicted by our classification models.

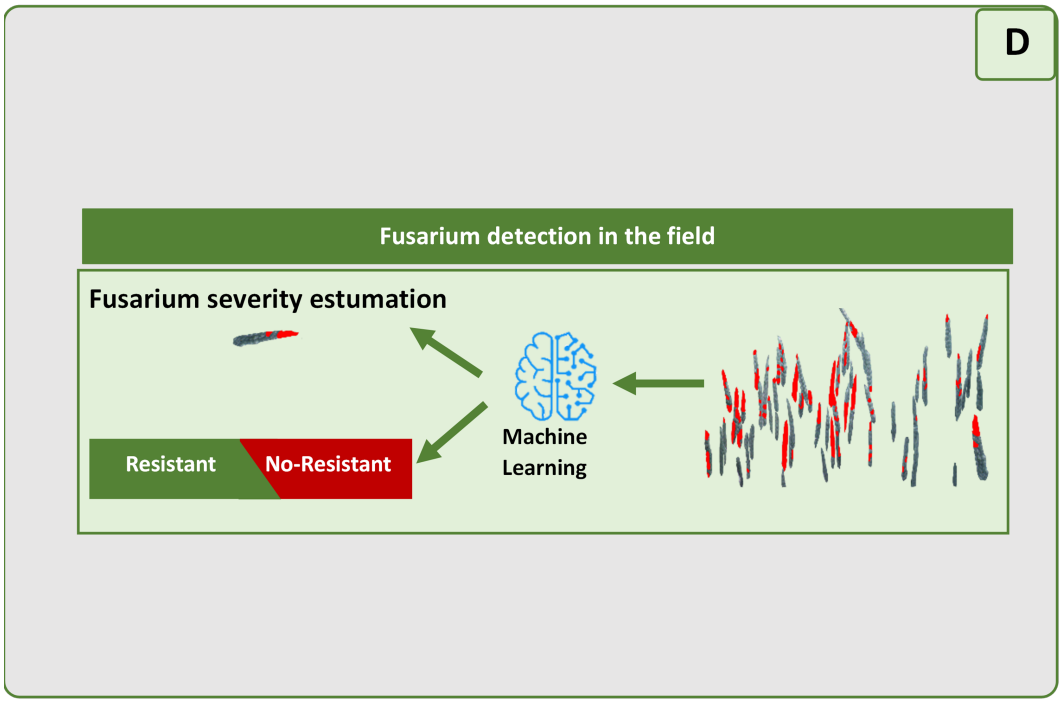


Figure 4.10 – *Fusarium* detection by machine learning methods on segmented images acquired in the field environment using the CMS4 camera.

A global view of the produced data set for this study is given in Table 4.2.2. These data have been used to obtain the following results.

Task	Type of images	Year	N° of images
Wavelengths selection	Hyperspectral images	2016	189
		2017	389
		2018	189
		2019	533
Segmentation of wheat spikes in the field	RGB images	2020	220
	Multispectralimages	2021	160
Fusarium detection in controlled environment	RGB images	2020	179
	Multispectral images	2020	179
Fusarium detection in field environment	RGB images	2021	160
	Multispectral images	2021	160

4.3 Results

4.3.1 Optimized wavelengths selection

Following all methods of selecting the optimized wavelengths to discriminate between the healthy part and the parts contaminated with *Fusarium* on wheat spikes presented in the previous section 4.2.1, the results are provided in Figure 4.11 using four different databases from four years. We obtain between two to five different wavelengths depending on the method used. As shown in Figure 4.11, the selected wavelengths are located in the visible and near-infrared. Moreover, the selected wavelengths are almost similar over the four years, thus showing the stability of selected wavelengths.

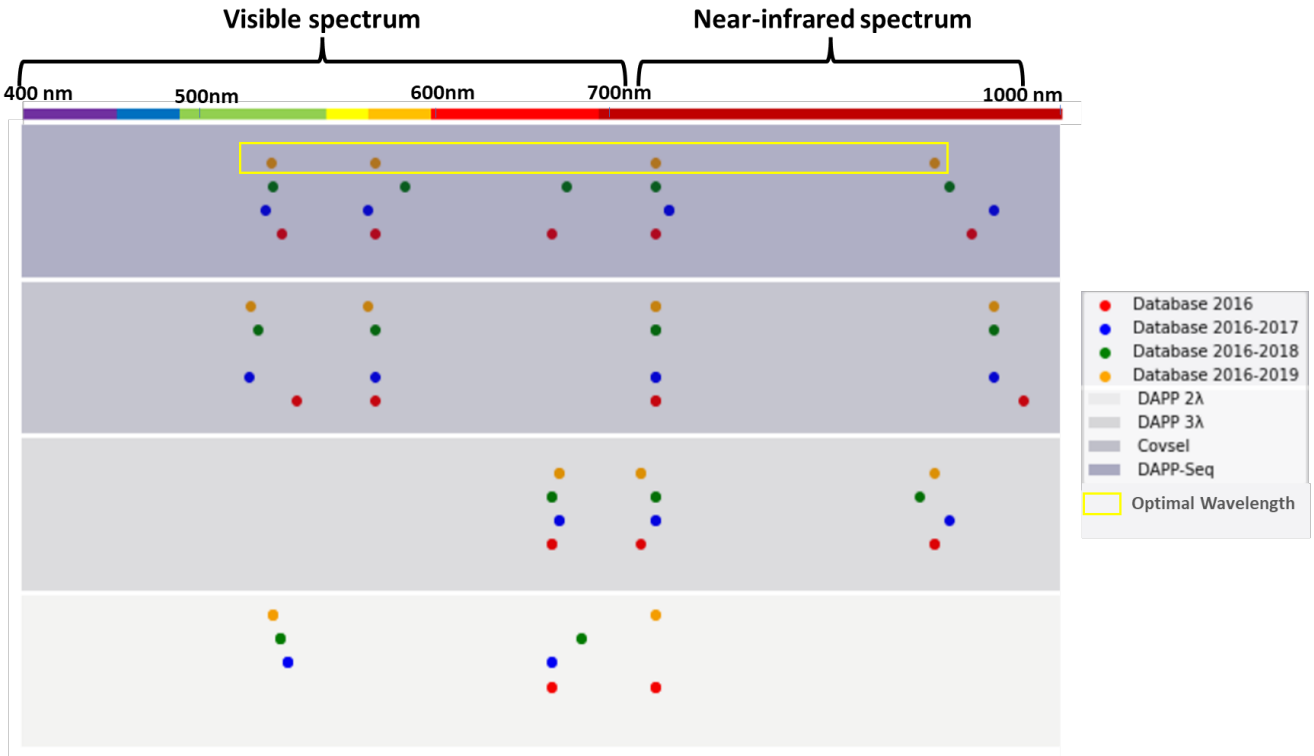


Figure 4.11 – Optimal selected wavelengths for *Fusarium* detection over four years.

Table 4.1 shown the accuracy results of DASS seq, CovSel, DASS 3λ and DASS 3λ methods using test database over four years. Looking at the results in Table 4.1, we see almost the same accuracy value on each database. In addition, the results of the four methods are very close. Moreover, we compute the R^2 coefficient between the *Fusarium* severity precited and *Fusarium* severity annotated by three different experts

Table 4.1 – The accuracy results of all discrimination methods using a test database over four years.

	DASS Seq	**CovSel**	**DASS 3 λ**	**DASS 2 λ**
Database 2016	0.88 ± 0.01	0.87 ± 0.01	0.85 ± 0.01	0.85 ± 0.01
Database 2016 - 2017	0.87 ± 0.01	0.86 ± 0.01	0.86 ± 0.01	0.85 ± 0.01
Database 2016 - 2018	0.88 ± 0.01	0.86 ± 0.01	0.85 ± 0.02	0.84 ± 0.01
Database 2016 - 2019	0.88 ± 0.02	0.87 ± 0.01	0.85 ± 0.01	0.83 ± 0.02

using the resulting wavelength selected from the database of four years (database 2016-2019). The R^2 results are present in Table 4.2. As a result, we can find the best correlation with all experts with the DASS-Seq method. Consequently, we have retained these four wavelengths resulting from the last database of years marked in yellow in Figure 4.1.

Table 4.2 – The R^2 coefficient between the *Fusarium* severity annotated by experts and the predicted one using wavelength from the database of four years(database 2016-2019).

	DASS Seq	**CovSel**	**DASS 3 λ**	**DASS 2 λ**
Expert 1	0.89 ± 0.01	0.88 ± 0.01	0.86 ± 0.01	0.84 ± 0.01
Expert 2	0.90 ± 0.01	0.88 ± 0.01	0.85 ± 0.01	0.86 ± 0.01
Expert 3	0.91 ± 0.01	0.89 ± 0.01	0.87 ± 0.01	0.82 ± 0.01

- **Preliminary test of the CMS4 camera in a controlled environment**

Following the selection of wavelengths, we build the four wavelengths on an optimized multispectral camera called CMS4. Now, we move to test the new multispectral camera CMS4 in the controlled condition. First, we apply pixels classification to each wheat spike in two classes: *Fusarium* and healthy. In Table 4.3, we present the precision, recall, and accuracy for each classification method.

Based on Table 4.3, the weighted KNN method and bagged Trees method provide the best performance of *Fusarium* detection. Both methods have an accuracy of more than 78% with a recall of around 90%. Furthermore, the results prove the ability to keep a significant performance of *Fusarium* severity estimation with an accuracy of 80% by using the optimized wavelengths selected from the hyperspectral camera and built into a multispectral camera.

Table 4.3 – Results of different classification models for *Fusarium* disease detection on wheat.

	Precision	**Recall**	**Accuracy**
Bagged Trees	0.72 ± 0.01	0.90 ± 0.02	0.78 ± 0.02
Cubic KNN	0.68 ± 0.01	0.87 ± 0.02	0.72 ± 0.02
Fine Gaussian SVM	0.69 ± 0.03	0.77 ± 0.02	0.72 ± 0.03
Weighted KNN	0.74 ± 0.03	0.91 ± 0.02	0.80 ± 0.02

4.3.2 In the field: proposed models for segmentation of spikes and FHB detection

- **Wheat spikes segmentation in field conditions**

The image acquired in the field include several rows of wheat spikes behind the row of interest. Our goal is the segmentation of the first row of spikes. We started the segmentation on RGB images acquired before building the multispectral camera. The results of the U-Net segmentation model are present in Table 4.4. As visible in the first line of Table 4.4, we get a promising result. Therefore, we can prove the possibility of automatic segmenting of spikes in the first row without the need to put an extensive background behind this row.

Since in the future, we will use the new CMS4 sensors, we need to build a segmentation model for the multispectral images. To get the benefit of the first segmentation model based on RGB images, we transfer the final weights of this model as input weights for the new model for the pseudo-RGB images of the multispectral camera. Ultimately, we train another model from scratch using the pseudo-RGB images. Table 4.4 illustrates the performance of these models. As we see in this Table, using fine-tuning, we can improve segmentation results by 10%. These results are highly suitable to validate our simple acquisition protocol in the field.

Table 4.4 – Dice coefficient of different segmentation models of wheat spikes on the images acquired in field environment.

	Training	**Validation**	**Test**
RGB	0.83 ± 0.03	0.79 ± 0.02	0.75 ± 0.02
Pseudo-RGB	0.78 ± 0.01	0.75 ± 0.02	0.71 ± 0.02
Fine tuning	0.88 ± 0.03	0.84 ± 0.02	0.79 ± 0.03

In addition, we want to see the impact of the number of images used on transfer learning. So, we train several models using a different number of images in training, and we plot the Dice coefficient as a function of the images used number. Looking at Figure 4.12, we can admit that we obtained a gain of 10% using only 100 multispectral images in training.

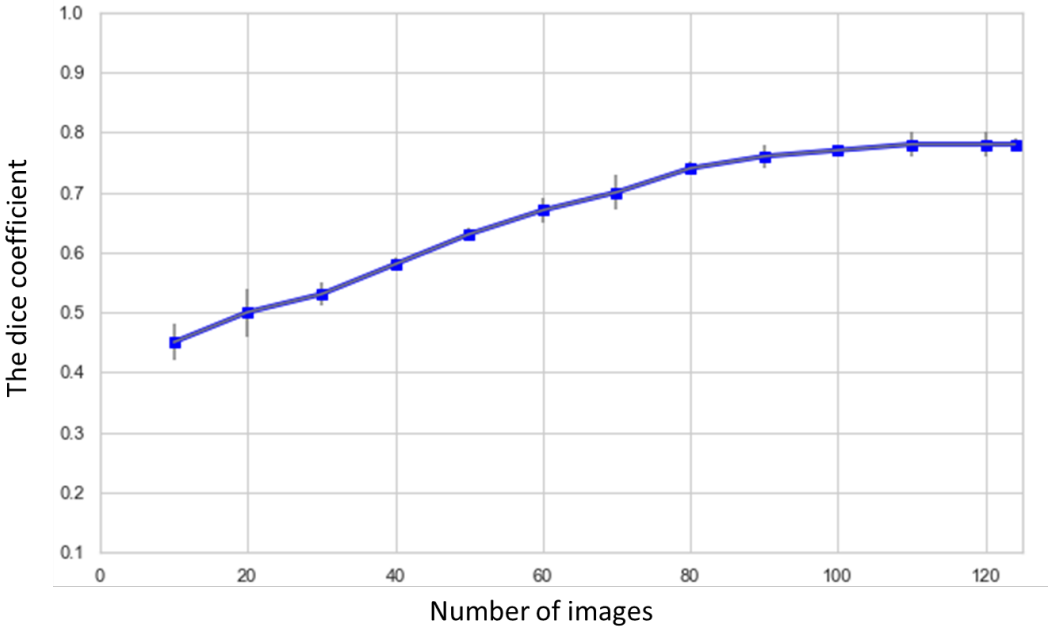


Figure 4.12 – The Dice coefficient as a function of number of images in train database for fine tuning.

- ***Fusarium* severity estimation in the field using multispectral images**

After segmenting spikes of the first row on the multispectral images, we apply a binary classification method to detect the *Fusarium* disease in the field environment. Then, we evaluate our results using two methods. In the first one, we calculate the precision, recall, and accuracy between the pixels predicted and annotated pixels, as we can see their values in Table 4.5. In the second method, we compute the R^2 correlation coefficient between the severity estimated by the best machine learning model with the severity estimated by the expert based on the images. The Figure 4.13 and Figure 4.14 illustrate the R^2 coefficient for winter wheat (from V_1 to V_n) and durum wheat.

Based on R^2 coefficient results are shown in Figure 4.13 and Figure 4.14. We observe a high correlation from 0.86 to 0.93 between the evaluation of expert and the prediction

Table 4.5 – Results of different models for *Fusarium* detection on wheat spikes based on multispectral images acquired in the field environment.

	Precision	**Recall**	**Accuracy**
Bagged Trees	0.74 ± 0.01	0.90 ± 0.02	0.78 ± 0.02
Cubic KNN	0.68 ± 0.01	0.87 ± 0.02	0.72 ± 0.02
Fine Gaussian SVM	0.69 ± 0.03	0.77 ± 0.02	0.72 ± 0.03
Weighted KNN	0.74 ± 0.03	0.91 ± 0.02	0.79 ± 0.03

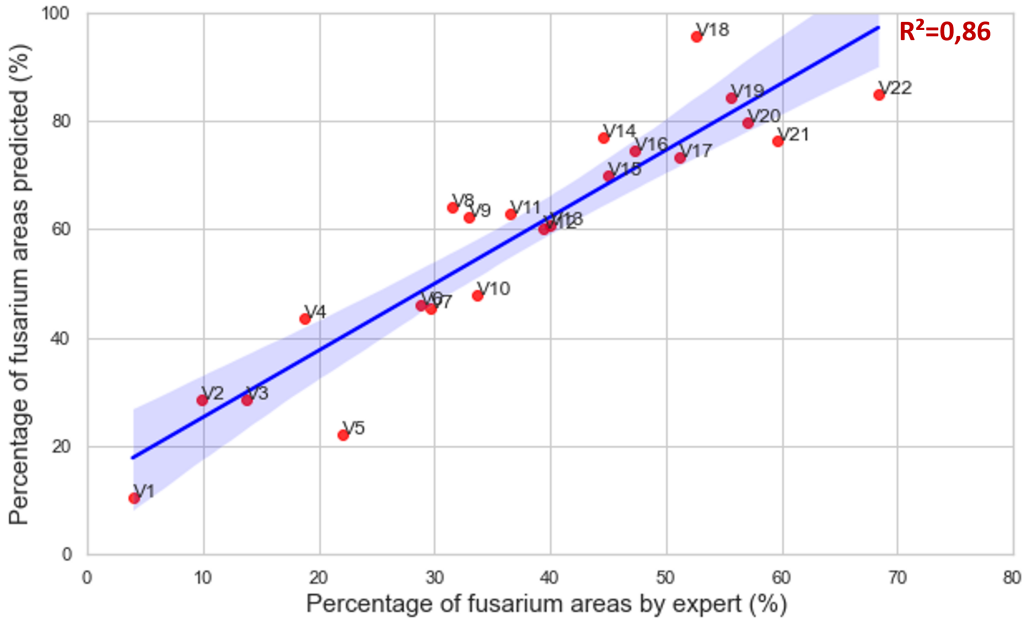


Figure 4.13 – Correlation between severity estimated by the expert based on image and severity predicted by the KNN model for winter wheat.

of our model. These results are promising for replacing the manual annotation in the field with our multispectral camera.

4.4 Discussion and Conclusions

In this chapter, we proposed a global pipeline of building an optimal multispectral camera for estimating the severity of *Fusarium* on wheat spikes in the field environment using a suitable protocol acquisition. First, we acquired hyperspectral images of wheat spikes from two different places for four years in a controlled environment with the NEO Hyspex VNIR-1800 camera to select the discriminated wavelengths. Then, we applied

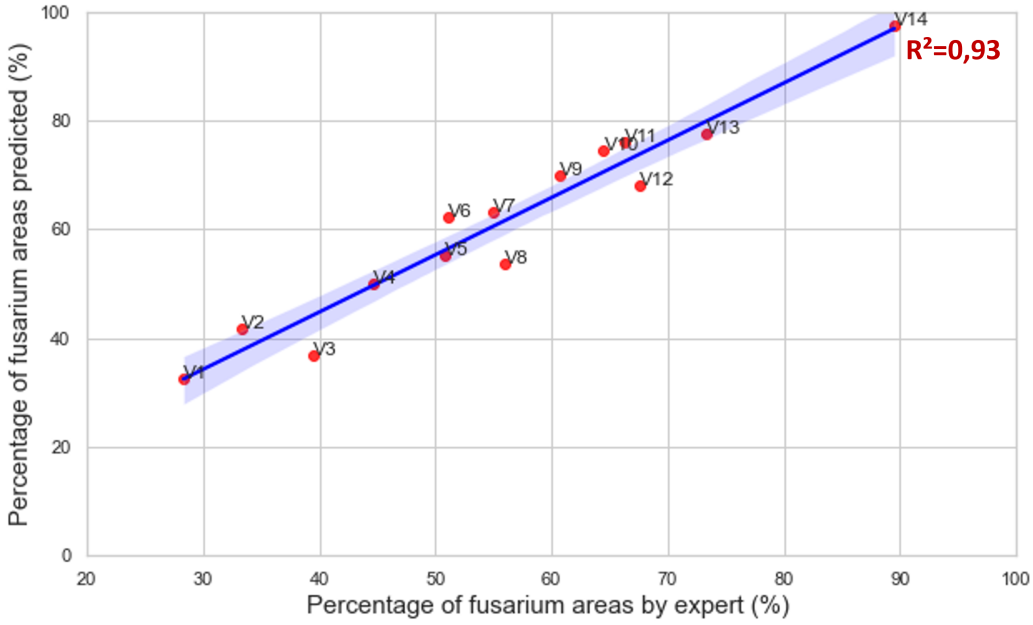


Figure 4.14 – Correlation between severity estimated by the expert based on image and severity predicted by the KNN model for durum wheat.

on four different linear discriminant methods: Discriminant analysis sequential step by step (**DASS-Seq**), Covariance Selection **CovSel**, and non-sequential linear Discriminant analysis sequential step by step for 2λ or 3λ (**DASS** 3λ and **DASS** 3λ) in order to discriminate *Fusarium* area. The resulting wavelengths are showing the stability over four years. We demonstrated the possibility of estimating the *Fusarium* severity of diverse wheat varieties, even those with distinct species and sizes, using only four wavelengths. Next, we build a new multispectral sensor called CMS4 using these optimal wavelengths. We test in the beginning our new camera in controlled conditions using supervised machine learning methods such as Bagged Trees, Cubic KNN, Fine Gaussian SVM, and Weighted KNN. The results achieved 80% accuracy between an expert's manual annotation of the image and the prediction of machine learning models. Based on these promising results in controlled conditions, we moved to test the new camera in the field condition. In the field, the *Fusarium* severity is based on visual estimating each row of spikes. We acquired images in the field using a simple acquisition protocol. Next, we started segmenting wheat spikes of the first row using U-net deep learning method. We showed we could segment the first row of spikes with a Dice coefficient of more than 0.75. Also, we demonstrated the gain of 10% of Dice using the transfer learning with fine-tuning method between a

model trained with RGB images to a segmentation model for multispectral images. We showed that we could segment the first row of spikes using a simple acquisition protocol of images contrarily to other works that previously resorted to specific imaging box [**agriculture4010032**, 129, 130, 131, 132, 134, 143].

Afterward, we returned to the primary goal, *Fusarium* severity estimation on the field condition. So, we apply Bagged Trees, Cubic KNN, Fine Gaussian SVM, and Weighted KNN as a binary classification method on the segmented images. We get the best results with the KNN model with an accuracy of 0.79. This means with almost no loss of performance from what was obtained in indoor conditions. Moreover, we get a good correlation between the visual annotation based on the image and the prediction of the KNN model with a R^2 coefficient equal to 0.86 on winter wheat and 0.93 on durum wheat.

Therefore, we can conclude that the new multispectral camera could be very useful for the quantification of *Fusarium* in the field. In 2022, we plan to test the CMS4 camera in nine different places in France to validate our approach on a larger scale. The data is collected during the 2022 season, and the images are currently being processed.

Chapter 5

Conclusion and Perspectives

5.1 Conclusion

In this PhD, we proposed a methodological approach to automate selected traits in variety testing based on deep learning and computer vision using multi-component images.

There are several hundreds of traits that could be automated in a variety testing. In the introduction chapter, we have proposed a rationale to select the ones that would benefit the most from a shift from manual to numerical practices. Based on this selection, we have developed two multicomponent imaging systems: a network of low-cost RGB-Depth cameras to monitor seedling emergence or wheat heading and an optimized multispectral camera to detect wheat diseases.

While designing these elaborated engineering solutions for plant science applications, we have addressed methodological challenges specific to multicomponent imaging systems. In chapter 2, we have explored fusion strategies of RGB and Depth information within deep learning models. The depth was shown to improve the performance of seedling emergence detection thanks to additional size information and acquisition capabilities during the night. However, depth was not helpful for the detection of wheat heading stages.

In chapter 3, we investigated transfer learning approaches from indoor to outdoor conditions encountered in a variety testing. The transfer learning approach was boosted by adding simulated shadows to account for the nonuniform lighting that can occur in the outdoor environment. This was illustrated by detecting development plant stages using various spatio-temporal deep learning methods. We applied for the first time the transformer models for plant imaging processing. While these methods are successful for natural language or image segmentation, they do not outperform the other classical deep learning methods (LSTM, GRU) when processing a series of developmental images. Although this would have to be confirmed in more use cases, one can raise that there might be fundamental reasons for this. In natural language processing, several patterns

can be found, with words occurring in different orders and having similar meanings. In developmental biology, an arrow of time imposes a phenomenological order in the stages of development. This prior knowledge is not natively included in the transformer approaches. We have finally proposed a last possibility of transfer from synthetic data to real ones. This was shown on the sunflower during the flowering process. Video gaming development environments are modified, oriented, and used for transfer learning. Despite promising preliminary results, more variability in the synthetic objects provided by these environments are to be provided for higher efficiency of the transfer.

In chapter 4, we proposed a global pipeline to design an optimized multispectral imaging system to detect wheat disease. In addition, we specially investigated the possibility of transferring the value of the optimized wavelength from indoor to outdoor conditions.

In addition to these methodological contributions, we have provided tools in Annex B and Annex C. This includes software to process RGB-Depth sequences of images and original annotated data sets.

5.2 Perspectives

Specific perspectives on each chapter were provided in its conclusion. Consequently, we provide more generic perspectives here. In this PhD, we developed image processing pipelines using data produced by low-cost sensors to address several characteristics of high interest in variety testing (emergence, flowering, disease quantification). These results are promising but constitute, at this stage, proof of feasibility.

One must keep in mind that the current observation time for one DUS characteristic by an expert is often shorter than the image acquisition and processing of this characteristic. Accordingly, the efficiency of the variety examiner should be improved when several DUS characteristics can be assessed from one image of a pot, plant, or organ. An interesting perspective would be to develop models capable of extracting several characteristics from one image. Another one would be to analyze the need for sensors in VCU testing. As mentioned in the introduction, the current situation is that VCU testing protocol for species is not normalized between European countries due to the specific conditions and needs (climate, soil, diseases...) of each country, which drives local evaluation that differs from one country to another. One way to support VCU assessment would be to select phenotypic characteristics which can constitute the input for agronomical models [145]. Such models have been designed for phenotyping purposes with some high-resolution sensors. From

this perspective, it would be interesting to analyze the effect of lowering the resolution while keeping the predictive value of the agronomical models of the literature. Another way is using sensors and data fusion to identify DUS and VCU characteristics. Those perspectives open up analytical approaches to be investigated. Last, VCU characteristics assessment, such as biotic and abiotic stresses on plants and quality of fruits, also need other types of more expensive imaging systems (fluorescence, multispectral and hyperspectral near-infrared, thermal, LIDAR) [146],[147]. Lowering the cost of these imaging systems could, as we did in this PhD, significantly increase the potential impact of sensor-based DUS and VCU characteristics assessment and help the seed sector in general. For vector, we focused on a handy camera in this PhD. More ergonomic alternatives may be wearable glasses positioned on the head of the testers and leaving both hands free for manipulation of the plants (as recently done in our laboratory [105]).

The image processing algorithms developed in this PhD have been trained with the help of powerful GPU-equipped computers. All data have been trained at rest. It would be interesting to head toward instrumentation that could process the data with possibly re-training stages in the field. However, most of the deep learning architectures used in the PhD were very demanding in computation during training. For these reasons, specific light versions have been designed to run in embedded mode. For variety testing, it would be imperative to consider such light architectures for smartphone field applications, as recently stressed in [148]. Currently, the developed models run in jupyter notebooks, which are not directly usable by non-experts, as is mostly the case in variety testing.

Also, unfortunately, available solutions are currently not accessible within applications dedicated explicitly to a manual rating in the field for various testing, such as [149]. A simple and helpful development would thus be to use the existing literature of algorithms (including our original contributions), which is suited for variety testing, and implement it in the Internet of Things (IoT) platform [150] to record measurements and meta-data associated with variety testing. During this PhD, we initially expected to benefit from data from our parterner around Europe in the framework of the INVITE project. Because of the COVID-19 pandemic, we mostly had to generate the data ourselves. As a more collaborative gathering of data starts again, we took time to envision possible difficulties arising in multi-centric acquisition trials. In such trials, the acquisition protocol may vary from site to site. This will result in a variety of quality images. A challenging step is, therefore, to determine how to converge toward a standard acquisition protocol. We propose a first pilot study in this direction in Annex A.

5.3 Valorization of the work

Journal Articles

- Hadhami Garbouge, Pejman Rasti and David Rousseau,"Enhancing the Tracking of Seedling Growth Using RGB-Depth Fusion and Deep Learning", Sensors 21.24 (2021), p.8425.
- Hadhami Garbouge, Valérie Cadot, Fred Serre, Sylvie Roche and David Rousseau "Optimized multispectral imaging and machine learning for wheat disease quantification in variety testing; A lab-to-field perspective.", Sensors(2023). **(in progress)**
- Hadhami Garbouge, Geoffroy Couasnet and David Rousseau," Detection of seedling development Software", SoftwareX (2022). **(Under review)**

International conferences

- Mathis Cordier, Hadhami Garbouge, Salma Samiei, Pejman Rasti, and David Rousseau, "Growth-data a new tool to characterize spatio-spectral patterns of plant growth", North American Plant Phenotyping Network(2020).
- Hadhami Garbouge, Salma Samiei, Pejman Rasti and David Rousseau,"Machine-learning assisted determination of best acquisition protocols in variety testing ", AI for Agriculture and Food Systems (2021).
- Hadhami Garbouge, Pejman Rasti and David Rousseau,"Deep Learning-Based Detection of Seedling Development from Indoor to Outdoor", International Conference on Systems, Signals and Image Processing (2022), pp. 121–131, Springer
- Hadhami Garbouge, Natalia Sapoukhina, Pejman Rasti and David Rousseau,"Deep learning-based detection of seedling development from controlled environment to field", In 31st International Horticultural Congress (IHC), 2022.
- Hadhami Garbouge, Salma Samiei and David Rousseau,"A SIM2REAL transfer approach based on video gaming environment for sunflower detection", In International Conference on Computer Vision (ICCV), (2023) (in progress).

Chapter 6

ANNEX A: Machine-learning assisted determination of best acquisition protocols in variety testing

---

As a consequence of climate change, there is an urgent need to develop new varieties capable of facing new climatic scenarios. However, the process of variety selection is rather long (10 years). To commercialize a new variety of an agricultural or vegetable species, a plant breeder has to follow a process managed by a national authority and delegated to an examination office (EO) that will describe and evaluate the variety for its registration on the national list. Evaluation results including variety descriptions may also serve for the granting of Plant Variety Rights (PVR). Currently a large majority of these tests are based on manual measurements performed from visual inspection. This method has consequences in terms of efficiency due to the time consuming nature of these tests. It is also an issue for the reproducibility of these tests when some characteristics are based on qualitative characteristics which may suffer from subjectivity in their assessment. Improving efficiency and reproducibility of these observations would be extremely useful for EOs that are continuously seeking for optimized testing methods implemented in testing protocols. It could also provide means to assess new characteristics developed in response to new agricultural constraints, particularly in the perspective of climate change. In addition, more efficient measurement methods would assist in addressing the challenge of the constant increase in the number of varieties that have to be tested. The described challenges encourage to head toward the use of sensors and numerical practices to progressively replace classical manual methods of examination whenever there is a need to speed up measurement or increase their reproducibility and objectiveness [151]. The trend of using more and more imaging for plant science has started some decades

ago and has been extensively reviewed [2, 3] for most recent ones, including with cost-effective strategies [4]. While imaging modalities used in plant science and variety testing may be similar, the types of measures in plant science and variety testing differ either by their nature and technical aspects. So far, few attention from the academic imaging community focus on these specific aspects of variety testing. This ongoing numerical transition is currently encouraged at the European level via collaborative networking projects (including https://www.h2020-invite.eu/).

There are several challenges to address in order to reach common numerical practices in variety testing. One of them lays right at the level of image acquisition. How to define optimal protocols of acquisition which would be shared and strictly followed by several countries? A Top-down approach would consists in letting engineers propose a strict protocol including the brand and set up of a camera, lighting mode, vector on which to fix the camera, position of the imaging setup toward the targeted crops, ... Such a rigid approach would by sure normalize the practices, but would run the risk to face non-compliant behaviors among the local experts in charge of image acquisition since it may not systematically be applicable due to local environmental constraints not envisioned before-hand. Another bottom-up approach would consists in letting the local experts of all interested nations discuss before-hand with engineers to define a common protocol. A risk here is to have a low convergence of these discussions. We believe that another option is possible to help this process of selection of best acquisition protocol.

We propose in this communication to consider the situation where existing datasets gathered in several places for the same purposes are fed to an algorithm capable of identifying automatically the best images for a final measurement. This methodology is illustrated on three datasets. We finally discuss the perspectives opened by this first pilot trial which could be extended and enriched in many ways.

6.1 Method

We assume a dataset constituted of raw images is acquired with various acquisition protocols for the same purpose and the associated ground truth (binary masks for segmentation for instance (e.g. binary masks for segmentation)) exists. We propose the following method to automatically detect the best imaging conditions for acquisition protocols inside this dataset (See Figure 6.1). At the first step, we split the dataset to the train and the test. These datasets are composed of balanced (uniform) images from the

different acquisition protocols. In the second step, handcrafted features corresponding to the expected optical quality of the acquired images are computed and a clustering method is applied on these features. The clustering method includes two classes for the expected good and bad quality of images. A statistical test is then made to decide if the distribution of the quality metric inside each cluster can be considered distinct or not. Finally in step 3, based on the results of the statistical test, a recommendation setting of the optical parameters are generated for the users.

We did not identified clear most related work from the computer vision community on this problem. Ideally, we would like to come up with a caption associated to an image were the expected quality of the image would be directly indicated to the technician in the field if acquisition parameters (focal, focus, angle, light, ...) are not in agreement with the reference dataset.

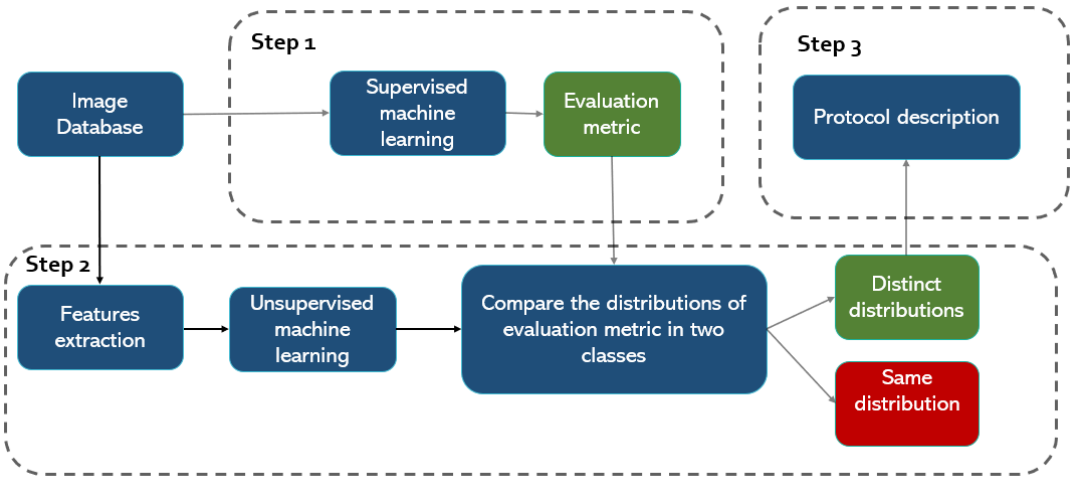


Figure 6.1 – Proposed generic pipeline proposed to select best acquisition protocol in variety testing.

6.1.1 Dataset

We implemented the generic pipeline of Figure 6.1 and tested it on three datasets shown in Figure 6.2. The first dataset includes 213 images (150 in training and 63 in test) of a sugar beets acquired under various illuminations including overexposed (i.e. where the sensor is saturating) conditions. The purpose of this dataset is the segmentation of the leaves from the soil. The percentage of coverage of the soil at a given date is an important

trait in variety testing. The second dataset includes 190 images(160 in training and 30 in test) of wheat observed for side view. The task is the segmentation of the spikes from the first raw of the micro-parcel. The last dataset is taken from the global wheat data challenge [152] with a subset of 3422 images (2758 images in training and 664 in test) of wheat ears observed from top view in the field. The task is to segment the ears. Here again the angle of view may vary from top view (90 degrees) to 45 degrees from top view. Images from these three datasets have been manually annotated to produce binary masks of the objects to be segmented.

6.1.2 Algorithms

We now provide more details about the specific algorithms used in the generic pipeline of Figure 6.1. The three considered datasets being dedicated to segmentation, we used a standard U-Net neural network architecture [36] for the image processing algorithm of step 1. The evaluation metric was chosen as the Sørensen-Dice coefficient D of the segmentation

$$D = \frac{2|X \cap Y|}{|X| + |Y|} \tag{6.1}$$

where X is the predicted segmentation and Y the ground truth.

The features extracted were selected to test the impact of variations of acquisition conditions on the final result. The sugar beet dataset were acquired under various spatial illumination including risks of image saturation and low exposure. We proposed for this dataset to simply count the percentage of pixels having low values, arbitrarily chosen from 0-30 after RGB to gray conversion, and the pixels close to saturation level, arbitrarily chosen from 227-255. An image with correct exposition is expected to have low percentage of pixels in these saturation part of its input-output characteristic. Wheat from side view were acquired under various angles of the camera toward the ground. To probe this optical parameter, we included an estimation of the depth from RGB monocular view (arbitrarily chosen from [153] among many deep learning variants from the literature) and simply computed the standard deviation of the estimated depth map. An image with low standard deviation in this depth map is expected to be acquired with an angle of 90 degree from the main vertical axis of the wheat heads. Last, to also probe the angle of view, the percentage of vegetation was computed from a standard semantic segmentation such as the one used in [154]. A high percentage of vegetation indicates a side or top view with low part due to the sky or additional non plant items (humans, tractors, ...). These

three simple features were applied on each images to feed the clustering method.

Image quality control by binary clustering (K-means with K=2) is applied to test the hypothesis of the quality of images based on the defined acquisition protocol. All features were normalized to 1 to avoid distortion effects when using Euclidean distance in the K-Means algorithm. A Wilcoxon rank-sum test [155] was applied on the distribution of the Dice coefficient inside each cluster. The null-hypothesis was chosen as the equality of the medians. This null hypothesis is validated at the default 5% on the P-Value. A recommendation of specific care about the tested optical parameter is finally recorded based on the result of this test.

6.2 Results

The distribution of the Dice coefficient in each cluster produced by the K-Means algorithm are displayed in Figure 6.3 for the three tested data sets. The P-value indicates in all these cases that hypothesis H_0 can be rejected. This indicates that the optical parameters tested (Illumination for dataset 1 and 3, Orientation for dataset 2) have an impact on the quality of the segmentation performance. Interestingly, when gazing at the image in each cluster (see Figure 6.4) the clustering indeed corresponds to uniform optical conditions, i.e. saturated or well exposed images in dataset 1 and 3 and uniform angle of view is dataset 2. On could use the result of such an experiment to identify the most important optical parameters and define in a data driven way the best practices. Here the experiment indicates to avoid saturation and favor side view or 45 degree view rather than top view. One can also notice that the distribution of the Dice coefficients are overlapping in the three conditions. This means that despite a statistically grounded difference in the performance in each cluster the difference is limited and could probably be reduced again by extending significantly the size of the training data sets with optical parameters in the range of what was included in the first. With both analysis our pipeline of Figure 6.1 provides fruitful feedback and strategy to define the best acquisition protocol depending on the size of the dataset and the associated effort of image annotation.

6.3 Conclusion and perspective

In this communication, we have introduced the problem of normalization of acquisition protocol in variety testing. We believe that machine learning can help to define the

best protocol in a reverse engineering mode. In this pilot study, first we proposed a supervised approach where handcrafted features correlated to optical parameters were used to cluster images. The approach was then successfully illustrated on datasets dedicated to segmentation tasks.

The work could be extended in many ways. While the problem appears to us original and challenging for computer vision some clear limitations can be underlined on the way we tackled it so far. Because we have chosen a supervised approach, we have to deliver a similar amount of data for all the tested variants of the protocol. This may seem problematic since we especially do not completely specify the protocol itself but rather propose to dive into the dataset to select the best practices. Also, annotation of the images has to be done on the whole dataset while we suspect that some of these data has insufficient quality. This appears as a loss of time. We can expect that expert that will do the annotation, will, by common sense, be able to identify the quality of the images by themselves and may not in the end have to wait for the answer of our algorithm to sort out the good from the bad quality images. One could envision heading toward a fully unsupervised and end-to-end approach. Variational auto-encoders (VAE) [156] could be used to produce a latent space where the clustering would operate. A possible limitation is that this latent space would still depend on the composition of the initial dataset. What would happen if among all the protocols, the best one was represented with few images only. This last remark rely on the fact that in the implementation presented in this communication the datasets were limited. A direction would be to bet on unsupervised algorithms trained on huge dataset purposely acquired in diverse conditions in order to ensure from the data rather than from the protocol itself sufficient robustness.

Another direction would be to investigate the possible use of synthetic plants positioned in virtual environment such as the one used for video gaming conception. There are models of virtual plants for almost all crops of interest and the libraries are continuously growing. The production of these models benefit from extensive use of L-System grammars [157, 158, 159] to simply but very realistically produce in-silico plant models. Optical parameters such as lighting, angle, optics, depth of field, exposure, resolution of the cameras can automatically be simulated in virtual environment. Annotation of the plants themselves can also be automated since the ground truth is created by the computer directly. The selection of the optimal acquisition protocols would in this case be more direct since the optical parameters would directly be known and not only correlated with handcrafted features. Our group has expertise in this field of digital twin [39, 106]

and we are working in this direction to overcome some of the mentioned limitations of our proposed approach.

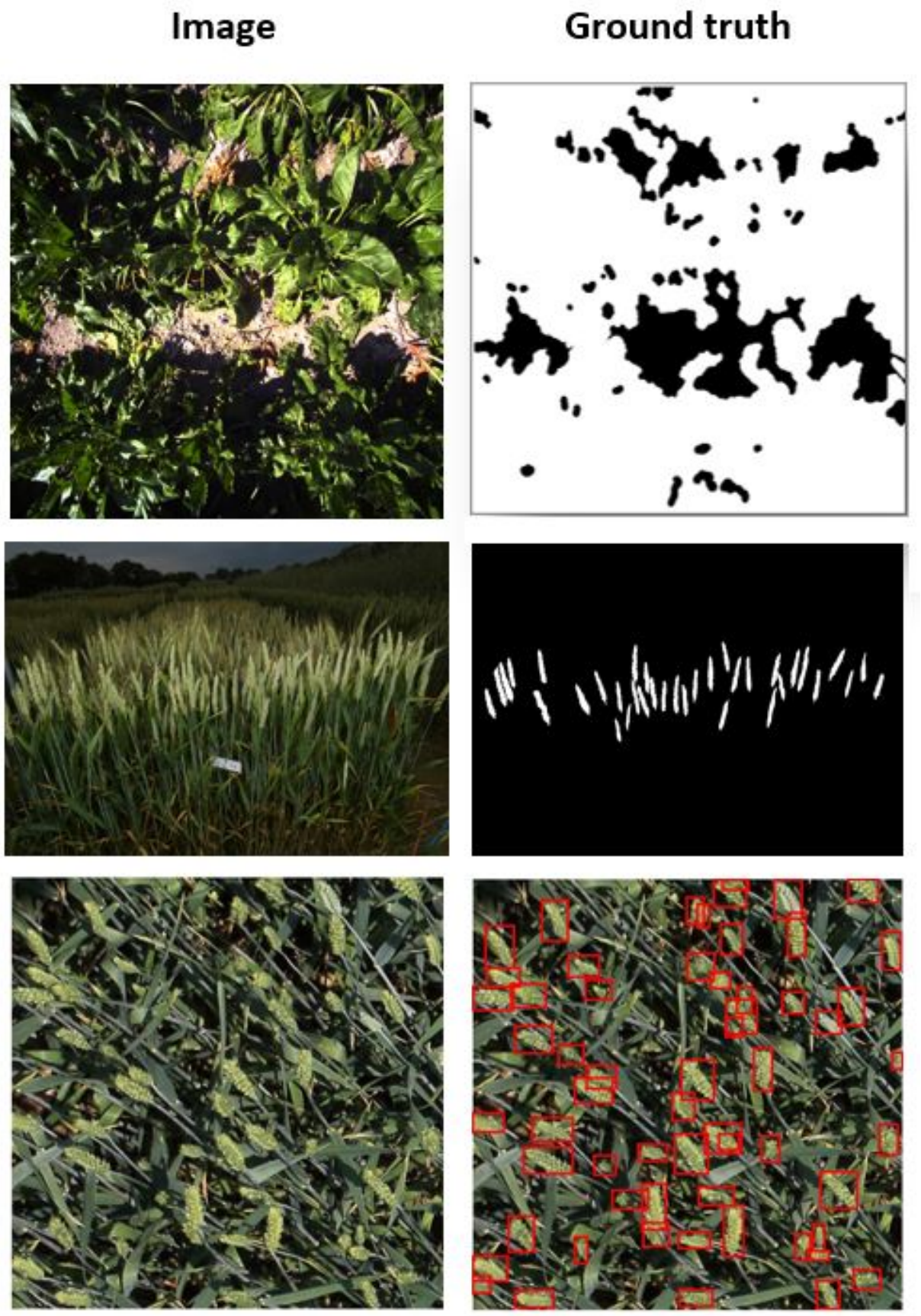


Figure 6.2 – Datasets and ground truth used to test the pipeline of Figure 6.1. Top raw: sugar beets observed from top view with various illuminations; Middle raw: wheat observed from the side view with various angles of the cameras; Bottom raw: hear observed from top view with various angles of the various angles of the cameras.

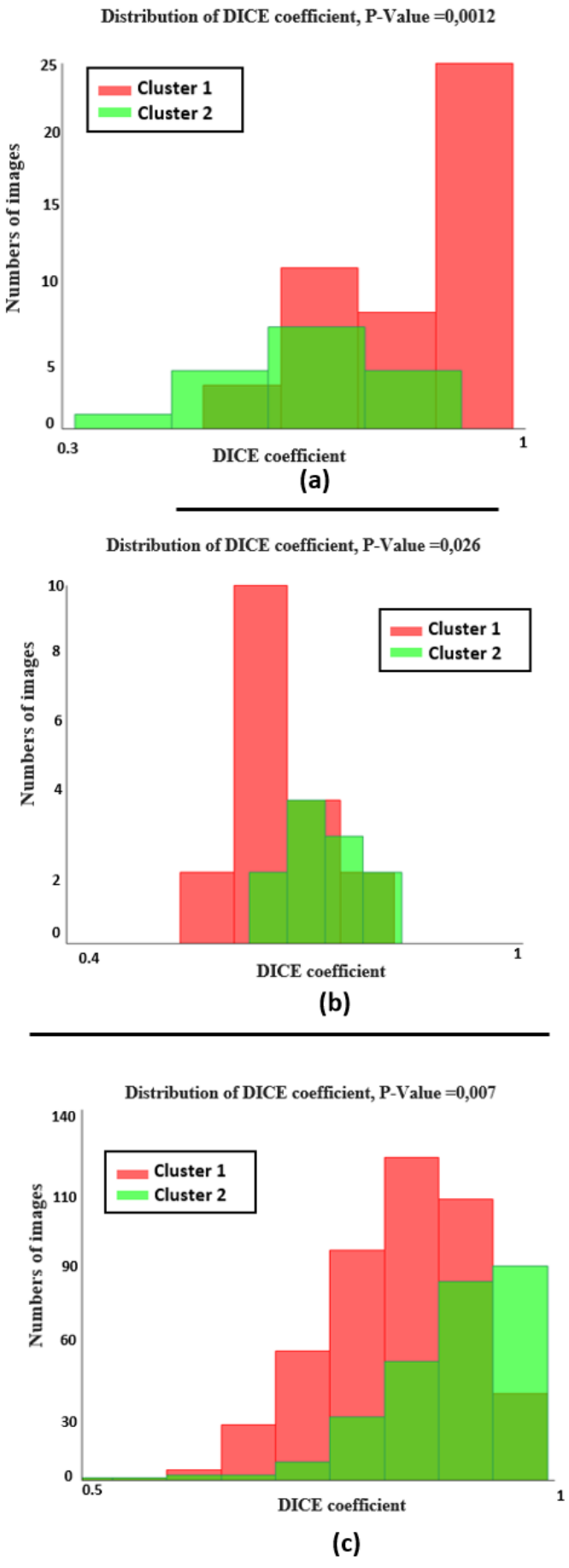


Figure 6.3 – Distribution of Dice coefficient in each cluster for the three datasets processed

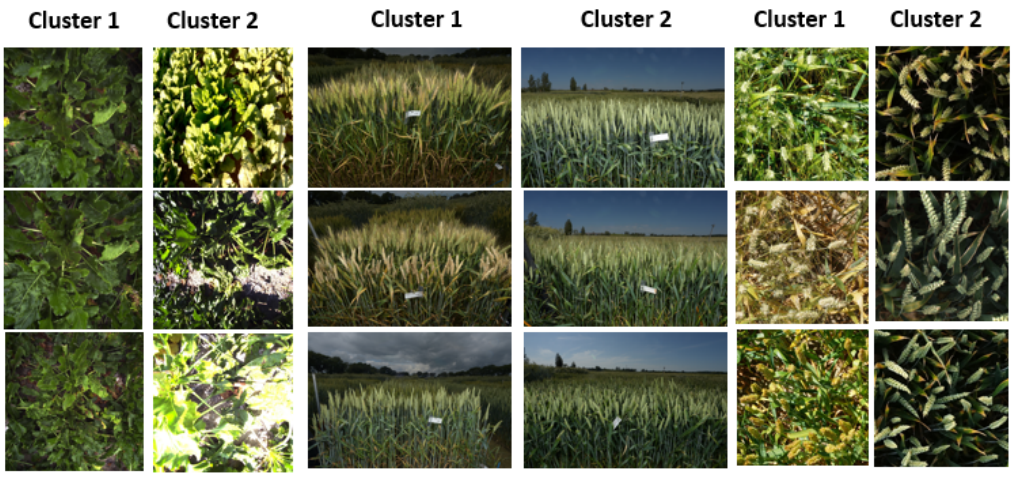


Figure 6.4 – Instances of each cluster in each of three datasets processed in this study.

Chapter 7

ANNEX B: User interface for image processing and analysis

7.1 Imaging system

7.1.1 Sensor choice

In this PhD, we have designed a network of affordable multi-component cameras to be deployed in growth chamber. This network has been used in Chapter 2 and we described its design in this annex. We started with a selection of a camera, as visible in Table 7.1. After testing some of the RGB-Depth solutions, we chose to work with Intel RealSense D435 (Figure 7.2). The D435 stereo camera is part of the new D400 series of depth cameras featuring the Intel® RealSense™ D4 vision processor. In a very compact and lightweight and rather low-cost format, Intel® RealSense combined a depth sensor with an RGB sensor and IR sensor. This camera was also used for outdoor investigation in Chapter 3.

Caméra	Depth FOV	Video resolution	Depth Range	RGB Sensor	Power	IOS
ZED 1 Stereo Camera	90° (H) x 60° (V) x 100° (D) max	Side by Side 2x (2208x1242) @15fps 2x (1920x1080) @30fps 2x (1280x720) @60fps	0.5 m to 25 m	1/3" 4MP CMOS	USB 5V / 380mA	Windows 10, 8, 7, Ubuntu 18, 16, Debian, CentOS (via Docker), Jetson L4T
ZED 2 Stereo Camera	110°(H) x 70°(V) x 120°(D)	Side by Side 2x (2208x1242) @15fps 2x (1920x1080) @30fps 2x (1280x720) @60fps 2x (672x376) @100fps	0.3 m to 20 m	1/3" 4MP CMOS	USB 5V / 380mA	Windows 10 - 64 bit, Ubuntu 16.04/18.04 - 64 bit, Debian, CentOS (via Docker), Jetson L4T
Intel® RealSense™ Camera D415	65°±2° x 40°±1° x 72°±2°	Up to 1280 x 720 active stereo depth resolution. Up to 90 fps.	0.1 m to 10 m	1920 x 1080 at 30 fps		Windows Ubuntu
Intel® RealSense™ Camera D435	87°±3° x 58°±1° x 95°±3°	Up to 1280 x 720 active stereo depth resolution. Up to 90 fps.	0.1 m to 10 m	1920 x 1080 at 30 fps		Windows Ubuntu
MYNT EYE	D: 146o H: 122o V: 76o	752 x 480 @ 60 FPS	0,5 m to 18 m	752 x 480 @ 60 FPS	2.7 W @ 5V DC	Ubuntu 14.04/16.04/18.04, Windows 10, ROS, Android 7.0 +
DUO 3D	165° Wide Angle Lens	45 FPS @ 752x480 49 FPS @ 640x480 98 FPS @ 640x240 192 FPS @ 640x120 86 FPS @ 320x480 168 FPS @ 320x240 320 FPS @ 320x120	min 0,3 m		~2.5 Watt @ +5V DC	Windows 7/8/10 Ubuntu 14 or later
IFM O3X100	60 H x 45 V	224 x 172	0,5m to 30 m		20.4, 28.8 DC	
Carnegie Robotics® MultiSense™ S7	80° x 49°(2MP sensor) 80° x 80°(4MP sensor)	7.5 FPS @ 2048 x 1088s 15 FPS @ 2048 x 544 30 FPS @ 1024 x 544		2048 x 1088 or 2048 x 2048	24V DC nominal	

Table 7.1 – Possible camera candidate for RGB-Depth imaging.

7.1.2 Network Description

We installed eight cameras to follow the experiment table's total surface in the individual room. Each camera is managed by a mini-computer, Raspberry Pi Model B (Figure 7.3). The Raspberry was equipped with PoE (Power on Ethernet) HAT. All the mini-computer were related to a local network, as shown in Figure 7.1. The local network is installed using a TP-Link router. Thanks to the local network, each camera can upload the acquired images to the server. The server is equipped with a Raspberry and a Hard Disk for data storage.

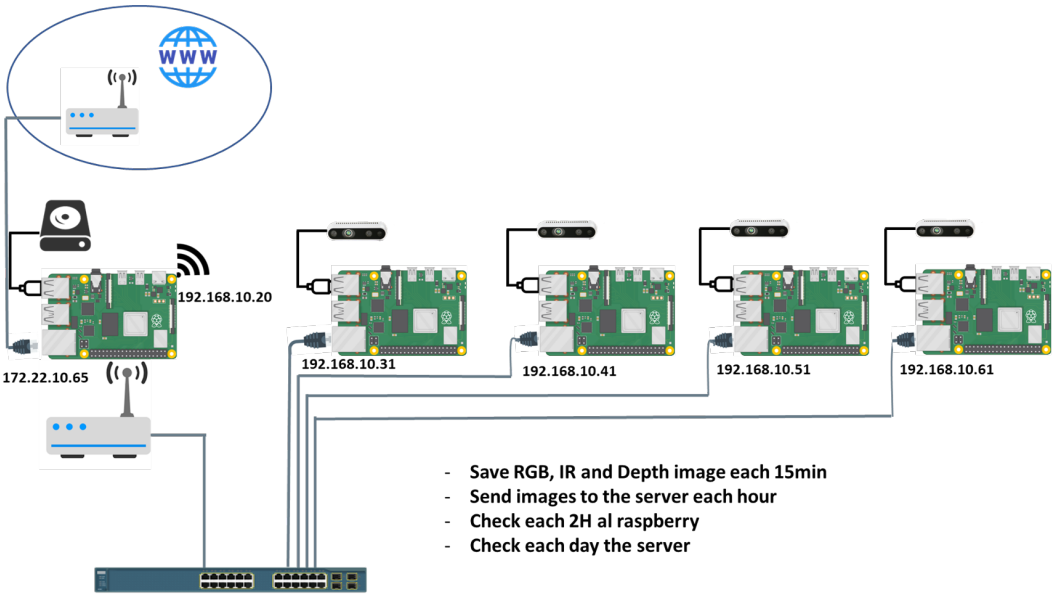


Figure 7.1 – Demonstration of the camera network installed in the growth chamber.

The server was then connected to our INRAe network. This allowed us to transfer the data to the biologists and to access to all cameras remotely. In addition, we also had a Raspberry Pi 4 with a second 4TB hard drive (identical to the one listed above) that took care of backing up the images daily. We performed these backups via Python scripts executed every night at the same scheduled time through the Raspberry Crontab. Plus, we have two notifications. In the first one, the server checked the file of each camera every two hours. If it didn't receive new images, it sent an mail. The second one, the server should send an email every day containing the situation of the cameras to be sure it didn't encountered any problem.

Such a room is thus composed of the following material for an approximate price of few keuros:

— **Server :** Raspberry Pi 4 Model B 2GB RAM ×1

— **Clients :** Raspberry Pi 4 Model B 2GB RAM ×8

— **Raspberry power supply :** PoE (Power on Ethernet) HAT ×8

— **Ethernet switch :** Netgear PoE+ Gigabit Ethernet Switch Model GS116LP ×1

— **Raspberry case :** The PiHut PoE+ HAT Case for Raspberry Pi 4 v2.0 ×9

— **Battery backup :** APC Back-UPS CS 650VA, 230V ×1

— **SD cards :** OKdo Micro SDHC cards 32Go class 10 pour Raspberry Pi (pre-installed operating system) ×9

— **Cameras :** Intel Realsense D435 ×8

— **Rooter :** TP-Link Archer C9 ×1

— **Data storage :** External hard drive LaCie Rugged USB-C 4TB ×1

7.2 Technical specifications

In this section, we will give some non-exhaustive technical specifications on the hardware that may be important and directly impact the data processed and analyzed by the software. We will not detail all the possible configurations of the hardware used but rather the parameters we used.

7.2.1 Intel RealSense D435

Global camera settings :

— **Dimensions length × depth × height) :** 90mm × 25mm × 25mm

— **Acquisition frequency :** 4 images per hour (1 image taken every 15 minutes)

— **Output :** USB-C 3.1 Gen 1 port

— **Effective distance :** from 0.3 to 3 meters (recommended min 0.5)

Settings used for RGB image capture :

— **Resolution :** 1920 × 1080 pixels

— **Image format :** PNG RGB images

— **Field of Vue (H × V) :** 69.4° × 42.5°

Settings used for depth image capture :

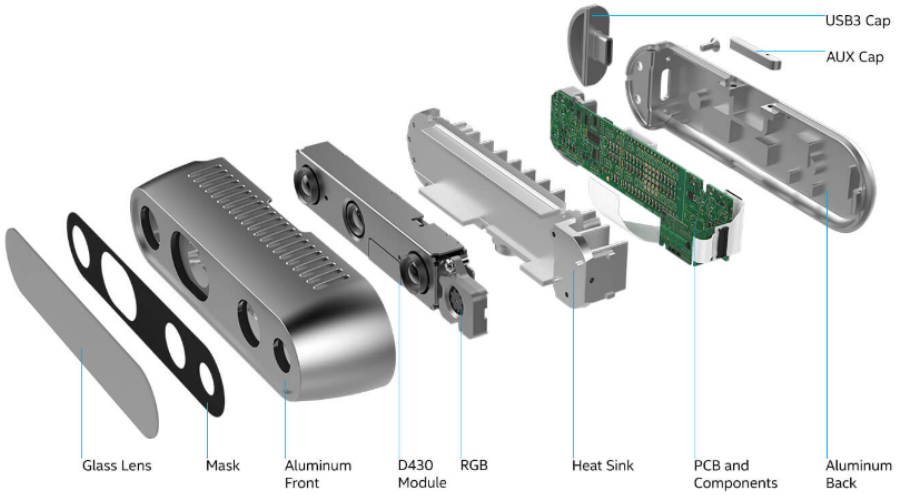


Figure 7.2 – Intel RealSense D435 camera.

— **Resolution :** 1920 × 1080 pixels

— **Image format :** PNG 16-bits

— **Field of Vue (H × V) :** 87° × 58°

Settings used for infra-red image capture :

— **Resolution :** 1280 × 720 pixels

— **Image format :** PNG 8-bits

— **Field of Vue (H × V) :** 90° × 63°

7.2.2 Raspberry Pi 4 Model B

The Raspberry we used in our systems is Raspberry Pi 4 Model B, which comes in several versions depending on the amount of RAM required. In our case, we initially chose the model with 2GB of RAM, but this seems limited for this use. That's why, during the following installations, we chose the version with 4GB of RAM. Here are the technical specifications of these models:

— **CPU :** Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz

— **Mémoire vive :** 2GB, 4GB or 8GB LPDDR4-3200 SDRAM (4GB minimum is recommended)

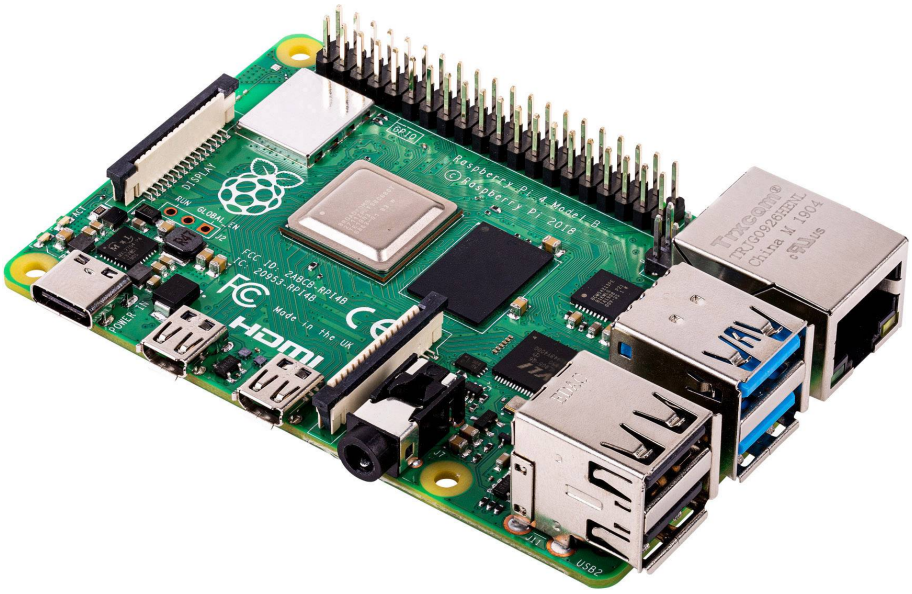


Figure 7.3 – Raspberry Pi model B.

— **Carte réseau :** 2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0, BLE Gigabit Ethernet

— **USB ports :** USB 3.0 ports ×2, USB 2.0 ports ×2

— **GPIO ports :** Raspberry Pi standard 40 pin GPIO header (fully backwards compatible with previous boards)

— **HDMI ports :** micro-HDMI ports ×2 (up to 4kp60 supported)

7.3 Description of the program

In this part, we present the different functionalities of the program we developed for our experiments and how it works. Consequently, we will start by presenting the program's structure, which can be divided into three main steps using an illustration diagram.

7.3.1 Program structure

As shown in the diagram below (figure 7.4), the program works in 3 steps: data import (image sets and associated Excel files), image processing, and finally, data analysis (feature extraction and/or predictions). We detail these 3 points in the following subsections and the nature of the data processed by the program.

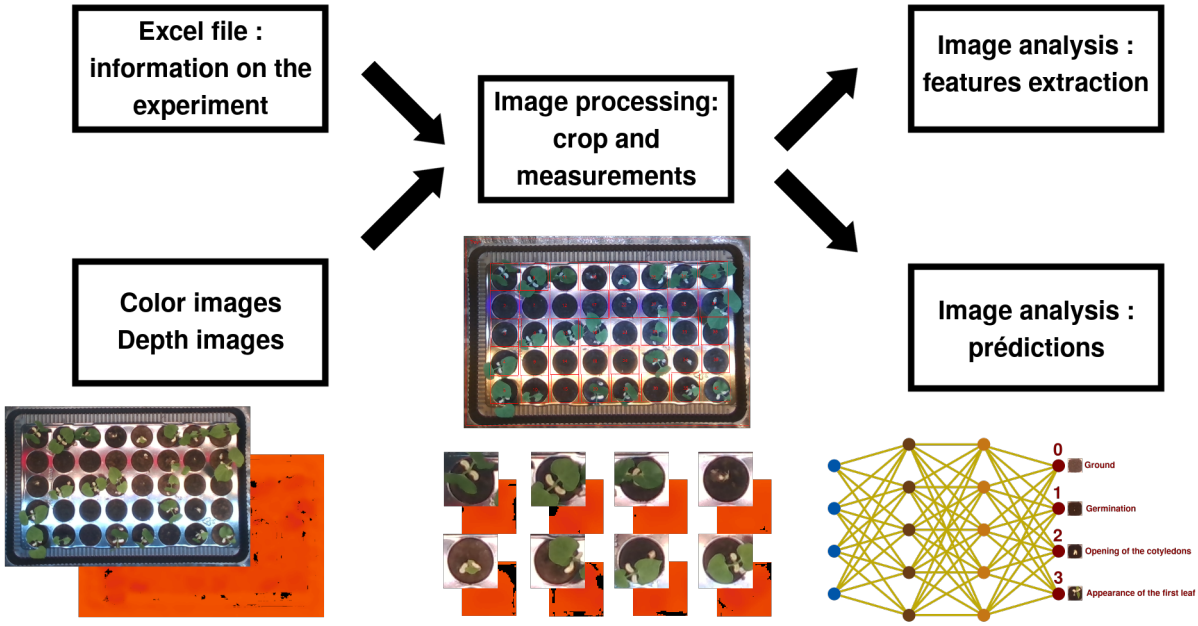


Figure 7.4 – Diagram illustrating the structure of the program, broken down into 3 main steps.

- **Import Data**

At the beginning of the experiment, the plants are installed in rooms whose environment can be controlled, with Raspberry equipped with cameras allowing the acquisition of color and depth images every 15 minutes. In parallel to these images, an Excel file is provided for each tray containing plants, in which the tray configuration is represented (same number of rows and columns) and each cell is a value associated with the corresponding plant of the tray. This file has as many sheets we want to associate variables to the plants (for example, variety, the quantity of water given, etc.). Each sheet name will be the name of the associated variable, and on each sheet is represented the same tray containing the values of the variable associated with the corresponding plants (see figure 7.5).

Before importing the Excel file containing the variables associated with the experiment's plants, the color and depth images of the plants taken every 15 minutes must be imported. To do this, we simply place them in 2 sub-folders named "Color" and "Depth" in any folder on the computer and then indicate where the latter is located after clicking on the button "*Image directory*". Once this is done, it remains to indicate in the same

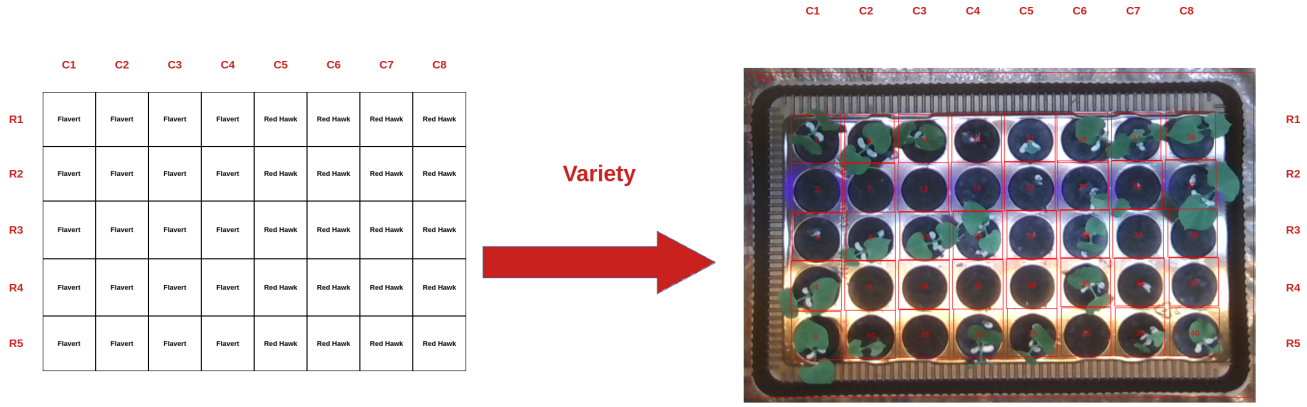


Figure 7.5 – Example of association of varieties to plants in an experiment from an Excel file.

way the location of the Excel file(s) with a button "*Import Excel File*". A dedicated area of the interface will display the images loaded by the software, and navigation buttons allow to navigate between the images and between the data sets "color" and "depth". The import of the Excel file allows the software to know the number of trays present on the images (number of Excel files imported), the number of rows and columns present on each of the trays, and the variables to be associated with each of the plants.

- **Image processing**

Once the images and Excel files are imported, the program should determine the location of each tray and plant it on the image. For this operation, there are two methods: an automatic detection method based on "*template matching*" and a manual method in case of failure of the first method. Just click on the button "*Search trays*" to use the automatic tray detection. Otherwise, we have to manually select the tray on the image and click on the button "*Add label*". If the software has never registered the tray type, the "*template matching*" algorithm will not detect it, but it is possible to select it manually and save it as a future *template* by simply clicking on "*Add tray template*". Once the tray(s) are selected, the manipulation for the selection of the plants is the same, except that the corresponding buttons are the "*Search plants*" for automatic detection, and the button "*Add label*" remains the same to add the selection (once one of the trays is finished), and to save a selection of plant as a template it is necessary this time to use the button "*Add plant template*". Once the trays and plants have been selected, the program associates with each of them a unique label visible on the displayed image. It is also possible to save

this image to have a trace of the notation. The user then has to start the measurement by clicking on the button "*Start measures*". For each plant selection zone, the program will then calculate the average measurement of this zone from the depth images. The program has different functionalities at this stage, so it is possible to:

— **Noise reduction :** application of an algorithm called "*inpainting*" to fill the missing pixels on the depth images related to noise.

— **Export of the coordinates of the selection areas :** It is possible to export the list of coordinates of the selection zones of each tray and each plant in the format *csv* (File → Export coordinates).

— **Export of the displayed image :** It is possible to save the image displayed by the software, allowing the user to have a visual plan of the identifiers of each plant and each tray (File → Export displayed image).

— **Export of measurements :** It is possible to export the measurements made on the depth images in the format *csv* (File → Export measures).

— **Saving image stacks :** The program cuts the images on each plant selection zone during the measurements. It is possible to save for each plant a set of RGB and Depth images containing the temporal sequence of images cut on the selection area of the plant. These image stacks will be helpful in making predictions using neural networks that take this type of data as input.

- **Image analysis**

As can be seen in the diagram illustrating the overall structure of the program (figure 7.4), the data analysis can be separated into two parts: feature extraction and stage prediction. These two functionalities are independent and do not use the same types of input data. The feature extraction is based exclusively on the measurements made during the image processing, while the predictions are made via convolution neural networks using color and depth image stacks as input. We will therefore present these two types of analysis separately in this section.

Feature extraction As said before, the features are calculated from the measurements made on the depth images. These measurements give us the heights (between the plateau and the highest point of the plant) throughout the experiment in pixel value (average value calculated on each selection area). We thus obtain a growth curve from which we can extract the desired features, which are :

— The final height of the plant

— Daily slope (average daily growth rate)

— Daily harmonic distortion rate

— Daily minimum amplitude of the circadian cycles

— Daily maximum amplitude of the circadian cycles

To export these different features, simply go to the menu "*File* → *Export features*" after having made the measurements on the images, then indicate the destination path.

Predictions of the stages of evolution When processing the images, image stacks are created for each plant, one with the color images, one with the depth images, and a third composed of the depth images retouched by the *inpainting* algorithm. By importing the color image stacks only or the depth image stacks retouched to reduce noise. Once loaded, these image stacks are sent to a convolution neural network which will, at each image of the stack, predict the stage of evolution in which the plant. There are four stages of evolution, each associated with a label :

0. Soil
1. Germination
2. Opening of the cotyledons
3. Appearance of the first leaf

Once the predictions are calculated, they are exported in the format *csv* accompanied by the plates and identifiers of plants as well as a variable indicating for each image of the stacks if the image was taken by day or night.

7.3.2 Saving and loading projects

We have previously presented the different functionalities of the program, except for one: saving and loading an existing project. Indeed, the processes carried out by the interface can take more or less time depending on the number of images contained in an experiment. If the user has to restart the processes from scratch for the same experiment as soon as he needs new data or because he forgot to export the measurements, it can quickly become a waste of time. We, therefore, decided to add the possibility of saving the project's current state so that the user can resume it later without starting over. This system works as in most programs, the user can save via the menu "*File* → *Save project*"

or "*File* → *Save project as...*" or by using the shortcuts "*Ctrl + S*" and "*Ctrl + Shift + S*". To load a project, simply use the menu "*File* → *Load project*".

Chapter 8

ANNEX C: ANNOTATED DATASETS, MACHINE AND DEEP LEARNING MODELS

8.1 Plants emergence in greenhouse : sunflower

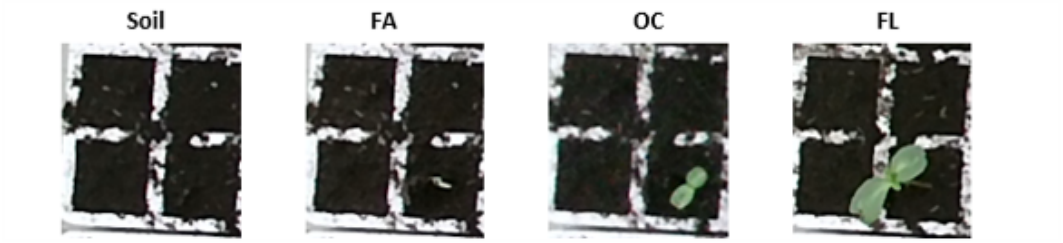


Figure 8.1 – The four developmental stages to be detected are the soil, the first appearance of the cotyledon (FA), the opening of the cotyledons (OC), the appearance of the first leave (FL).

- **Date of acquisition:** 13/02/2020 to 01/03/2020
- **Location:** l'Anjouère, GEVES, France
- **Image number:** 36 plants / 1398 RGB images and 1398 depth images for each plant
- **Sensor description :** Microsoft Kinect V2
- **Vector :** Top view with a fixed vector
- **standard reference measurements:** Manuel annotation based in images
- **Potential use of the image set :** The dataset was obtained to count the plants after emergence as a preparatory work to detect and count young plants after emergence in the field.

- **Machine and deep learning models:** CNN model

8.2 Plants emergence in the field

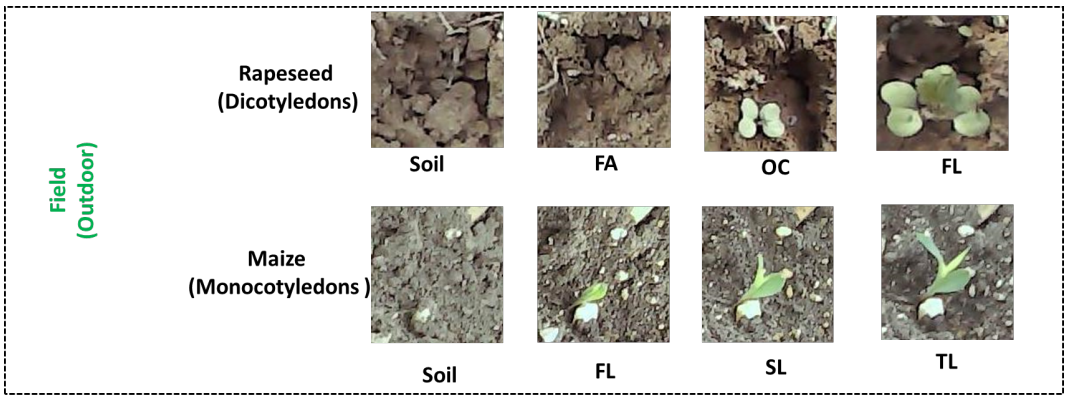


Figure 8.2 – The four developmental stages to classify are the soil, the first appearance of the cotyledon (FA) or First leave (FL), the opening of the cotyledons (OC) or Second leaf (SL), the appearance of the first leave (FL) or Third leaf (TL).

8.2.1 Rapeseed

- **Date of acquisition:** 09/09/2020 to 30/03/2020
- **Location:** l'Anjouère, GEVES, France
- **Image number:** 57 plants / 14022 RGB images
- **Sensor description :**
- **Vector :** Top view with a fixed vector
- **standard reference measurements:** Manuel annotation based in images
- **Potential use of the image set :** The dataset was obtained to count the plants after emergence as a preparatory work to detect and count young plants after emergence in the field.
- **Machine and deep learning models:** CNN model

8.2.2 Maize

- **Date of acquisition:** 5/06/2021 to 26/06/2020

- **Location:** l'Anjouère, GEVES, France
- **Image number:** 57 plants / 14592 RGB images
- **Sensor description :**
- **Vector :** Top view with a fixed vector
- **standard reference measurements:** Manuel annotation based in images
- **Potential use of the image set :** The dataset was obtained to count the plants after emergence as a preparatory work to detect and count young plants after emergence in the field.
- **Machine and deep learning models:** CNN model

8.3 Wheat height

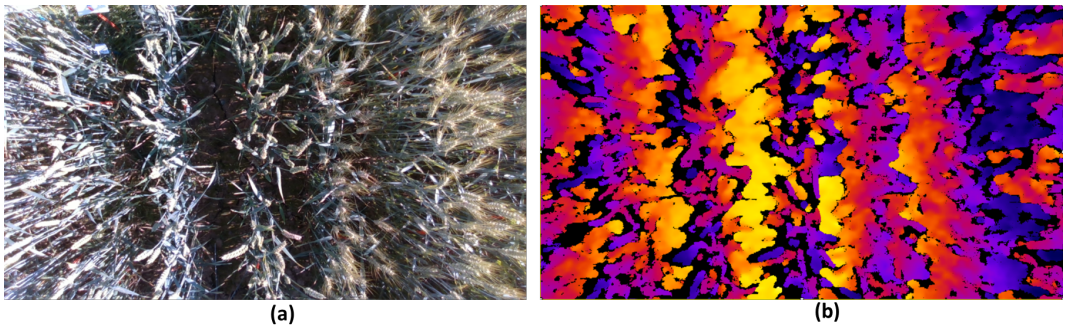


Figure 8.3 – Images of wheat in the field in order to measure the height. (a) RGB image. (b) Depth image.

- **Date of acquisition:** 15/06/2021
- **Location:** l'Anjouère, GEVES, France
- **Videos number:** 13 RGB and 13 Depth for each plant
- **Sensor description :** Intel RealSense D435
- **Vector :** Camera mounted on a Stick
- **Standard reference measurements:** measurements of the height on several plants of the plot with a metre stick and then we take the average of measurements
- **Potential use of the image set :** Plant length determination according UPOV and CPVO n°13 ; classification as "very short, short, medium, long, very long"

8.4 Sunflower : flowering detection

8.4.1 Real data

Figure 8.4 – (a) Sunflower plant in the field. (b) Bounding boxes around the flower.

- **Date of acquisition:** 5/07/2021
- **Location:** l'Anjouère, GEVES, France
- **Images number:** 197 RGB
- **Sensor description :** Nikon camera
- **Vector :** Hand
- **Standard reference measurements:** coordinates of bounding boxes
- **Potential use of the image set :** Detection of flowering time
- **Machine and deep learning models:** YoLo model

8.4.2 Synthetic data

- **Images number:** 1550 RGB
- **Standard reference measurements:** coordinates of bounding boxes of flower and segmentation of flower.
- **Machine and deep learning models:** YoLo model

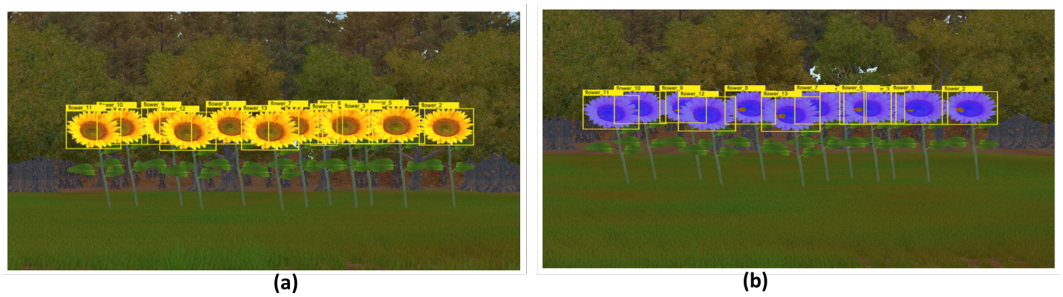


Figure 8.5 – (a) Bounding boxes of flower detection. (b) Segmentation of flower.

BIBLIOGRAPHY

[1] *Community Plant Variety Office*, `https://cpvo.europa.eu/en`, Accessed July 09,2022.

[2] Lei Li, Qin Zhang, and Danfeng Huang, « A review of imaging techniques for plant phenotyping », *in*: *Multidisciplinary Digital Publishing Institute* 14.*11* (2014), pp. 20078–20111.

[3] Ruicheng Qiu et al., « Sensors for measuring plant phenotyping: A review », *in*: *International Journal of Agricultural and Biological Engineering* 11.*2* (2018), pp. 1–17.

[4] Daniel Reynolds et al., « What is cost-efficient phenotyping? Optimizing costs for different scenarios », *in*: *Plant Science* 282 (2019), pp. 14–22.

[5] *DUS and VCUS: the core of variety testing*, `https://www.geves.fr/about-us/variety-study-department/dus-vcus-testing`, Accessed July 07, 2022.

[6] *International Union for the Protection of New Varieties of Plants*, `https://www.upov.int/overview/en/upov.html`, Accessed July 07,2022.

[7] Stefan Paulus et al., « Low-cost 3D systems: suitable tools for plant phenotyping », *in*: *Sensors* 14.*2* (2014), pp. 3001–3018.

[8] Sotirios A Tsaftaris and Christos Noutsos, « Plant phenotyping with low cost digital cameras and image analytics », *in*: *Information Technologies in Environmental Engineering*, Springer, 2009, pp. 238–251.

[9] Yann Chéné et al., « On the use of depth camera for 3D phenotyping of entire plants », *in*: *Computers and Electronics in Agriculture* 82 (2012), pp. 122–127.

[10] Robert T Furbank and Mark Tester, « Phenomics–technologies to relieve the phenotyping bottleneck », *in*: *Trends in plant science* 16.*12* (2011), pp. 635–644.

[11] Gustavo A Pereyra-Irujo et al., « GlyPh: a low-cost platform for phenotyping plant growth and water use », *in*: *Functional Plant Biology* 39.*11* (2012), pp. 905–913.

[12] Thomas Roitsch et al., « Review: New sensors and data-driven approaches—A path to next generation phenomics », *in*: *Plant Science* 282 (2019), The 4th International Plant Phenotyping Symposium, pp. 2–10.

[13] Daniel Reynolds et al., « CropSight: a scalable and open-source information management system for distributed plant phenotyping and IoT-based crop management », *in*: *GigaScience* 8.*3* (2019).

[14] Joaquim Miguel Costa et al., « Opportunities and Limitations of Crop Phenotyping in Southern European Countries », *in*: *Frontiers in Plant Science* 10.*1125* (2019).

[15] Alan Bauer et al., « Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production », *in*: *Horticulture research* 6.*1* (2019), p. 70.

[16] Aude Coupel-Ledru et al., « Multi-scale high-throughput phenotyping of apple architectural and functional traits in orchard reveals genotypic variability under contrasted watering regimes », *in*: *Horticulture research* 6.*1* (2019), p. 52.

[17] *Wildvision TL*, `https://www.wildkamera.net/wildkamera`, Accessed July 07, 2022.

[18] *Brinno*, `https://www.brinno.com/construction-camera/BCC100`, Accessed July 07, 2022].

[19] *ImageMeter Pro: Android application*, `https://play.google.com/store/apps/details?id=de.dirkfarin.imagemeterpro&hl=en_US`, Accessed July 07, 2022.

[20] *Smart Measure: Android application*, `https://play.google.com/store/apps/details?id=kr.sira.measure&hl=en_US`, Accessed July 07, 2022.

[21] *Smart Measure Tool Kit: Android application*, `https://play.google.com/store/apps/details?id=com.pcmehanik.measuretools&hl=en_US`, Accessed July 07, 2022.

[22] *Measure tools: Android application*, `https://play.google.com/store/apps/details?id=com.craftars.measuretools&hl=en_US`, Accessed July 07,2022.

[23] *EasyMeasure :Ios application*, `https://apps.apple.com/fr/app/easymeasure-original/id349530105`, Accessed July 07,2022.

[24] Guillaume Lobet, Xavier Draye, and Claire Périlleux, « An online database for plant image analysis software tools », *in*: *Plant methods* 9.*1* (2013), pp. 1–8.

[25] Lobet and Guillaume, « Image analysis in plant sciences: publish then perish », *in*: *Trends in plant science* 22.*7* (2017), pp. 559–566.

[26] Marin Talbot Brewer et al., « Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species », *in*: *Plant physiology* 141.*1* (2006), pp. 15–25.

[27] G Polder, G Blokker, and GWAM van der Heijden, « An ImageJ plugin for plant variety testing », *in*: *Proceedings of the ImageJ User and Developer Conference* (2012), pp. 168–173.

[28] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.

[29] Sharada P Mohanty, David P Hughes, and Marcel Salathé, « Using deep learning for image-based plant disease detection », *in*: *Frontiers in plant science* 7 (2016), p. 1419.

[30] Andreas Kamilaris and Francesc X Prenafeta-Boldú, « Deep learning in agriculture: A survey », *in*: *Computers and electronics in agriculture* 147 (2018), pp. 70–90.

[31] Asheesh Kumar Singh et al., « Deep learning for plant stress phenotyping: trends and future perspectives », *in*: *Trends in plant science* 23.*10* (2018), pp. 883–898.

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, « Imagenet classification with deep convolutional neural networks », *in*: *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[33] Forrest N. Iandola et al., « DenseNet: Implementing Efficient ConvNet Descriptor Pyramids », *in*: *ArXiv* abs/1404.1869 (2014).

[34] Joseph Redmon and Ali Farhadi, « YOLOv3: An Incremental Improvement », *in*: *ArXiv* abs/1804.02767 (2018).

[35] Shaoqing Ren et al., « Faster r-cnn: Towards real-time object detection with region proposal networks », *in*: *Advances in neural information processing systems*, 2015, pp. 91–99.

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, « U-net: Convolutional networks for biomedical image segmentation », *in*: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[37] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, « Segnet: A deep convolutional encoder-decoder architecture for image segmentation », *in*: *IEEE transactions on pattern analysis and machine intelligence* 39.*12* (2017), pp. 2481–2495.

[38] Ashish Vaswani et al., « Attention is all you need », *in*: *Advances in neural information processing systems* 30 (2017).

[39] Clément Douarre et al., « Novel data augmentation strategies to boost supervised segmentation of plant disease », *in*: *Computers and electronics in agriculture* 165 (2019), p. 104967.

[40] Kamlesh Golhani et al., « A review of neural networks in plant disease detection using hyperspectral data », *in*: *Information Processing in Agriculture* 5.*3* (2018), pp. 354–371.

[41] Stefan Thomas et al., « Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective », *in*: *Journal of Plant Diseases and Protection* 125.*1* (2018), pp. 5–20.

[42] Intel®, *Intel RealSense Documentation: Intel RealSense Depth Tracking Cameras*, Accessed July 07, 2022.

[43] Hadhami Garbouge, Pejman Rasti, and David Rousseau, « Enhancing the Tracking of Seedling Growth Using RGB-Depth Fusion and Deep Learning », *in*: *Sensors* 21.*24* (2021), p. 8425.

[44] A C McCormac, P D Keefe, S R Draper, et al., « Automated vigour testing of field vegetables using image analysis. », *in*: *Seed Science and Technology* 18.*1* (1990), pp. 103–112.

[45] Y Sako et al., « A system for automated seed vigour assessment », *in*: *Seed science and technology* 29.*3* (2001), pp. 625–636.

[46] A L Hoffmaster et al., « An automated system for vigor testing three-day-old soybean seedlings », *in*: *Seed Science and Technology* 31.*3* (2003), pp. 701–713.

[47] Julio Marcos-Filho et al., « Assessment of melon seed vigour by an automated computer imaging system compared to traditional procedures », *in*: *Seed Science and Technology* 34 (July 2006), pp. 485–497.

[48] Julio Marcos Filho, Ana Lúcia Pereira Kikuti, and Liana Baptista de Lima, « Procedures for evaluation of soybean seed vigor, including an automated computer imaging system », *in*: *Revista Brasileira de Sementes* 31.*1* (2009), pp. 102–112.

[49] Ronny V. L. Joosen et al., « germinator: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination », *in*: *The Plant Journal* 62.*1* (2010), pp. 148–159.

[50] É Belin et al., « Thermography as non invasive functional imaging for monitoring seedling growth », *in*: *Computers and electronics in agriculture* 79.*2* (2011), pp. 236–240.

[51] Landry Benoit et al., « Computer vision under inactinic light for hypocotyl–radicle separation with a generic gravitropism-based criterion », *in*: *Computers and Electronics in Agriculture* 111 (2015), pp. 12–17.

[52] Julio Marcos Filho, « Seed vigor testing: an overview of the past, present and future perspective », *in*: *Scientia Agricola* 72.*4* (2015), pp. 363–374.

[53] Friederike Gnädinger and Urs Schmidhalter, « Digital counts of maize plants by unmanned aerial vehicles (UAVs) », *in*: *Remote sensing* 9.*6* (2017), p. 544.

[54] Pouria Sadeghi-Tehran et al., « Automated method to determine two critical growth stages of wheat: heading and flowering », *in*: *Frontiers in plant science* 8 (2017), p. 252.

[55] Pejman Rasti et al., « Low-cost vision machine for high-throughput automated monitoring of heterotrophic seedling growth on wet paper support. », *in*: *BMVC*, 2018, p. 323.

[56] Ruizhi Chen et al., « Monitoring cotton (Gossypium hirsutum L.) germination using ultrahigh-resolution UAS images », *in*: *Precision agriculture* 19.*1* (2018), pp. 161–177.

[57] Biquan Zhao et al., « Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery », *in*: *Frontiers in plant science* 9 (2018), p. 1362.

[58] Yu Jiang et al., « DeepSeedling: deep convolutional network and Kalman filter for plant seedling detection and counting in the field », *in*: *Plant methods* 15.*1* (2019), p. 141.

[59] Salma Samiei et al., « Deep learning-based detection of seedling development », *in*: *Plant Methods* 16.*1* (2020), pp. 1–11.

[60] Charles Nock et al., « Assessing the potential of low-cost 3D cameras for the rapid measurement of plant woody structure », *in*: *Sensors* 13.*12* (2013), pp. 16216–16233.

[61] David Rousseau et al., « Multiscale imaging of plants: current approaches and challenges », *in*: *Plant methods* 11.*1* (2015), pp. 1–9.

[62] Joan R Rosell-Polo et al., « Kinect v2 sensor-based mobile terrestrial laser scanner for agricultural outdoor applications », *in*: *IEEE/ASME Transactions on Mechatronics* 22.*6* (2017), pp. 2420–2427.

[63] Adar Vit and Guy Shani, « Comparing rgb-d sensors for close range outdoor agricultural phenotyping », *in*: *Sensors* 18.*12* (2018), p. 4413.

[64] Rodrigo Méndez Perez, Fernando Auat Cheein, and Joan R Rosell-Polo, « Flexible system of multiple RGB-D sensors for measuring and classifying fruits in agri-food Industry », *in*: *Computers and Electronics in Agriculture* 139 (2017), pp. 231–242.

[65] Jorge Martinez-Guanter et al., « Low-cost three-dimensional modeling of crop plants », *in*: *Sensors* 19.*13* (2019), p. 2883.

[66] Michaela Servi et al., « Metrological Characterization and Comparison of D415, D455, L515 RealSense Devices in the Close Range », *in*: *Sensors* 21.*22* (2021), p. 7770.

[67] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency, « Multimodal machine learning: A survey and taxonomy », *in*: *IEEE transactions on pattern analysis and machine intelligence* 41.*2* (2018), pp. 423–443.

[68] Pradeep K Atrey et al., « Multimodal fusion for multimedia analysis: a survey », *in*: *Multimedia systems* 16.*6* (2010), pp. 345–379.

[69] Dhanesh Ramachandram and Graham W Taylor, « Deep multimodal learning: A survey on recent advances and trends », *in*: *IEEE signal processing magazine* 34.*6* (2017), pp. 96–108.

[70] Abhinav Valada et al., « Deep Multispectral Semantic Scene Understanding of Forested Environments Using Multimodal Fusion », *in*: *2016 International Symposium on Experimental Robotics*, ed. by Yoshihiko Kulić Danaand Nakamura, Oussama Khatib, and Gentiane Venture, Cham: Springer International Publishing, 2017, pp. 465–477.

[71] Andreas Eitel et al., « Multimodal deep learning for robust RGB-D object recognition », *in*: *International Conference on Intelligent Robots and Systems (IROS)* (2015), pp. 681–687.

[72] Jordi Sanchez-Riera et al., « A comparative study of data fusion for RGB-D based visual recognition », *in*: *Pattern Recognition Letters* 73 (2016), pp. 1–6.

[73] Anran Wang et al., « Large-Margin Multi-Modal Deep Learning for RGB-D Object Recognition », *in*: *IEEE Transactions on Multimedia* 17.*11* (2015), pp. 1887–1898.

[74] Ran Bezen, Yael Edan, and Ilan Halachmi, « Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms », *in*: *Computers and Electronics in Agriculture* 172 (2020), p. 105345.

[75] Nitish Srivastava and Ruslan Salakhutdinov, « Learning representations for multimodal data with deep belief nets », *in*: *29th International Conference Machine Learning (Workshop)* (2012).

[76] Yu Cao et al., « Medical image retrieval: a multimodal approach », *in*: *Cancer informatics* 13 (2014), CIN–S14053.

[77] Ian Lenz, Honglak Lee, and Ashutosh Saxena, « Deep learning for detecting robotic grasps », *in*: *The International Journal of Robotics Research* 34.*4-5* (2015), pp. 705–724.

[78] Ashesh Jain et al., « Recurrent neural networks for driver activity anticipation via sensory-fusion architecture », *in*: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2016, pp. 3118–3125.

[79] Xinhang Song et al., « Learning Effective RGB-D Representations for Scene Recognition », *in*: *Trans. Img. Proc.* 28.*2* (2019), pp. 980–993.

[80] Yanhua Cheng et al., « Semi-Supervised Multimodal Deep Learning for RGB-D Object Recognition », *in*: *Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16)*, 2016, pp. 3345–3351.

[81] Li Sun et al., « A Novel Weakly-Supervised Approach for RGB-D-Based Nuclear Waste Object Detection », *in*: *IEEE Sensors Journal* 19.*9* (2019), pp. 3487–3500.

[82] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, « Speech recognition with deep recurrent neural networks », *in*: *2013 IEEE international conference on acoustics, speech and signal processing*, IEEE, 2013, pp. 6645–6649.

[83] Hadhami Garbouge, Pejman Rasti, and David Rousseau, « Deep learning-based detection of seedling development from indoor to outdoor », *in*: *International conference on systems,signals and image processing(IWSSIP)*, vol. 1, IEEE, 2021, pp. 1–11.

[84] Massimo Minervini et al., « Phenotiki: An open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants », *in*: *The Plant Journal* 90.*1* (2017), pp. 204–216.

[85] Miguel Granados et al., « Background Inpainting for Videos with Dynamic Objects and a Free-Moving Camera », *in*: 2012, pp. 682–695.

[86] Camille Couprie et al., « Indoor Semantic Segmentation using depth information », *in*: *First International Conference on Learning Representations (ICLR 2013)*, Scottsdale, AZ, United States, 2013, pp. 1–8.

[87] Wenpeng Yin et al., « Comparative study of CNN and RNN for natural language processing », *in*: *arXiv preprint arXiv:1702.01923* (2017).

[88] Kun Zhou et al., « Time Series Forecasting and Classification Models Based on Recurrent with Attention Mechanism and Generative Adversarial Networks », *in*: *Sensors* 20.*24* (2020), p. 7211.

[89] Yuan Yuan and Lei Lin, « Self-Supervised Pre-Training of Transformers for Satellite Image Time Series Classification », *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2020).

[90] Vivien Sainte Fare Garnot et al., « Satellite image time series classification with pixel-set encoders and temporal self-attention », *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12325–12334.

[91] Alexey Dosovitskiy et al., « An image is worth 16x16 words: Transformers for image recognition at scale », *in*: *arXiv preprint arXiv:2010.11929* (2020).

[92] Gustavo Scalabrini Sampaio, Leandro Augusto da Silva, and Maurıcio Marengoni, « 3D Reconstruction of Non-Rigid Plants and Sensor Data Fusion for Agriculture Phenotyping », *in*: *Sensors* 21.*12* (2021), p. 4115.

[93] Jonghoon Jin et al., « Tracking with deep neural networks », *in*: *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, IEEE, 2013, pp. 1–5.

[94] Deepti Srivastava et al., « Role of circadian rhythm in plant system: An update from development to stress response », *in*: *Environmental and Experimental Botany* 162 (2019), pp. 256–271.

[95] Gembloux Agro-Bio Tech, *Principaux stades repères de la végétation des céréales*, Accessed Janvier 16 2022, 2008.

[96] B. Boser, I. Guyon, and V. Vapnik, « Pattern recognition system using support vectors », *in*: *United States Patent and Trademark Office* (1997).

[97] Vladimir N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1999.

[98] Laleh Armi and Shervan Fekri-Ershad, « Texture image analysis and texture classification methods - A review », *in*: (2019).

[99] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein, « Textural Features for Image Classification », *in*: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.*6* (1973), pp. 610–621.

[100] David Harwood et al., « Texture classification by center-symmetric auto-correlation, using Kullback discrimination of distributions », *in*: *Pattern Recognition Letters* 16.*1* (1995), pp. 1–10.

[101] T. Ojala, M. Pietikainen, and T. Maenpaa, « Multiresolution gray-scale and rotation invariant texture classification with local binary patterns », *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.*7* (2002), pp. 971–987.

[102] Pejman Rasti et al., « Supervised Image Classification by Scattering Transform with Application to Weed Detection in Culture Crops of High Density », *in*: *Remote Sensing* 11.*3* (2019).

[103] Hadhami Garbouge, Pejman Rasti, and David Rousseau, « Deep Learning-Based Detection of Seedling Development from Indoor to Outdoor », *in*: *Systems, Signals and Image Processing*, ed. by Gregor Rozinaj and Radoslav Vargic, Cham: Springer International Publishing, 2022, pp. 121–131.

[104] Hadhami Garbouge et al., « Deep learning-based detection of seedling development from controlled environment to field », *in*: *International Horticultural Congress* (2022).

[105] Salma Samiei et al., « Toward Joint Acquisition-Annotation of Images with Egocentric Devices for a Lower-Cost Machine Learning Application to Apple Detection », *in*: *Sensors* 20.*15* (2020), p. 4173.

[106] Natalia Sapoukhina et al., « Data Augmentation From RGB to Chlorophyll Fluorescence Imaging Application to Leaf Segmentation of Arabidopsis Thaliana From Top View Images », *in*: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2563–2570.

[107] Clément Douarre et al., « Transfer learning from synthetic data applied to soil–root segmentation in x-ray tomography images », *in*: *Journal of Imaging* 4.5 (2018), p. 65.

[108] Liangqiong Qu et al., « Deshadownet: A multi-context embedding deep network for shadow removal », *in*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4067–4075.

[109] Bin Ding et al., « Argan: Attentive recurrent generative adversarial network for shadow detection and removal », *in*: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10213–10222.

[110] Hieu Le and Dimitris Samaras, « Shadow removal via shadow image decomposition », *in*: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8578–8587.

[111] Joseph W Goodman, *Speckle phenomena in optics: theory and applications*, Roberts and Company Publishers, 2007.

[112] Kyunghyun Cho et al., « On the properties of neural machine translation: Encoder-decoder approaches », *in*: *arXiv preprint arXiv:1409.1259* (2014).

[113] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, « Deep learning », *in*: *nature* 521.*7553* (2015), pp. 436–444.

[114] Celso M. de Melo et al., « Next-generation deep learning based on simulators and synthetic data », *in*: *Trends in Cognitive Sciences* 26.*2* (2022), pp. 174–187.

[115] Artzai Picon et al., « Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets », *in*: *Computers and Electronics in Agriculture* 194 (2022), p. 106719.

[116] Amreen Abbas et al., « Tomato plant disease detection using transfer learning with C-GAN synthetic images », *in*: *Computers and Electronics in Agriculture* 187 (2021), p. 106279.

[117] Gábor Horváth et al., « Sunflower inflorescences absorb maximum light energy if they face east and afternoons are cloudier than mornings », *in*: *Scientific Reports* 10.*1* (2020), pp. 1–15.

[118] Joseph Redmon et al., « You only look once: Unified, real-time object detection », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[119] Christophe Pradal et al., « OpenAlea: a visual programming and component-based software platform for plant modelling », *in*: *Functional plant biology* 35.*10* (2008), pp. 751–760.

[120] Sebastian Stenglein, « Fusarium poae: A pathogen that needs more attention », *in*: *Journal of Plant Pathology* 91 (Mar. 2009), pp. 25–36.

[121] Christine Schwake-Anduschus et al., « Distribution of deoxynivalenol, zearalenone, and their respective modified analogues in milling fractions of naturally contaminated wheat grains », *in*: *World Mycotoxin Journal* 1 (Mar. 2015).

[122] Frédéric Serre et al., « Phénotypage au champ des céréales pour la fusariose de l'épi (FHB) par analyse automatique d'images », *in*: (2015).

[123] Ruicheng Qiu et al., « Detection of Fusarium Head Blight in Wheat Using a Deep Neural Network and Color Imaging », *in*: *Remote Sensing* 11.*22* (2019).

[124] Wen-Hao Su et al., « Automatic Evaluation of Wheat Resistance to Fusarium Head Blight Using Dual Mask-RCNN Deep Learning Frameworks in Computer Vision », *in*: *Remote Sensing* 13.*1* (2021).

[125] Tagel Aboneh et al., « Computer Vision Framework for Wheat Disease Identification and Classification Using Jetson GPU Infrastructure », *in*: *Technologies* 9.*3* (2021).

[126] Ning Zhang et al., « Development of Fusarium head blight classification index using hyperspectral microscopy images of winter wheat spikelets », *in*: *Biosystems Engineering* 186 (2019), pp. 83–99.

[127] E. Bauriegel et al., « Early detection of Fusarium infection in wheat using hyperspectral imaging », *in*: *Computers and Electronics in Agriculture* 75.*2* (2011), pp. 304–312.

[128] Dongyan Zhang et al., « Development and Evaluation of a New Spectral Disease Index to Detect Wheat Fusarium Head Blight Using Hyperspectral Imaging », *in*: *Sensors* 20.*8* (2020).

[129] Jonas Franke and Gunter Menz, « Multi-temporal wheat disease detection by multi-spectral remote sensing », *in*: *Precision Agriculture* 8.*3* (2007), pp. 161–172.

[130] Jayme G.A. Barbedo, Casiane S. Tibola, and José M.C. Fernandes, « Detecting Fusarium head blight in wheat kernels using hyperspectral imaging », *in*: *Biosystems Engineering* 131 (2015), pp. 65–76.

[131] S.R. Delwiche et al., « Estimating percentages of fusarium-damaged kernels in hard wheat by near-infrared hyperspectral imaging », *in*: *Journal of Cereal Science* 87 (2019), pp. 18–24.

[132] « The development of a hyperspectral imaging method for the detection of Fusarium-damaged, yellow berry and vitreous Italian durum wheat kernels », *in*: *Biosystems Engineering* 115.*1* (2013), pp. 20–30.

[133] Dong-Yan Zhang et al., « Integrating spectral and image data to detect Fusarium head blight of wheat », *in*: *Computers and Electronics in Agriculture* 175 (2020), p. 105588.

[134] Xiu Jin et al., « Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field », *in*: *Remote Sensing* 10.*3* (2018).

[135] Elke Bauriegel, Antje Giebel, and Werner B Herppich, « Hyperspectral and chlorophyll fluorescence imaging to analyse the impact of Fusarium culmorum on the photosynthetic integrity of infected wheat ears », *in*: *Sensors* 11.*4* (2011), pp. 3765–3779.

[136] Qiong Zheng et al., « Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages », *in*: *Sensors* 19.*1* (2019).

[137] *Performance and value of new plant varieties*, `https://www.geves.fr/recherche-et-developpement/activites-de-recherche/evaluation-varietes-environnement/`, Accessed July 19,2022.

[138] J.M. Roger et al., « CovSel: Variable selection for highly multivariate and multi-response calibration. Application to IR spectroscopy », *in*: *Chemometrics and Intelligent Laboratory Systems* 106.*2* (2010), 27 p.

[139] G. Mclachlan, « Mahalanobis Distance », *in*: *Resonance* 4 (June 1999), pp. 20–26.

[140] S Dhanabal and SJIJCA Chandramathi, « A review of various k-nearest neighbor query processing techniques », *in*: *International Journal of Computer Applications* 31.*7* (2011), pp. 14–22.

[141] MA. Hearst et al., « Support vector machines », *in*: *IEEE Intelligent Systems and their Applications* 13.*4* (1998), pp. 18–28.

[142] David Powers and Ailab, « Evaluation: From precision, recall and F-measure to ROC, informedness, markedness correlation », *in*: *J. Mach. Learn. Technol* 2 (Jan. 2011), pp. 2229–3981.

[143] Karl-Heinz Dammer et al., « Detection of head blight (Fusarium ssp.) in winter wheat by color and multispectral image analyses », *in*: *Crop Protection* 30.*4* (2011), pp. 420–428.

[144] Aaron Carass et al., « Evaluating white matter lesion segmentations with refined Sørensen-Dice analysis », *in*: *Scientific reports* 10.*1* (2020), pp. 1–19.

[145] Jeffrey W White et al., « Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards », *in*: *Computers and Electronics in Agriculture* 96 (2013), pp. 1–12.

[146] L. Li, Q. Zhang, and D Huang, « A review of imaging techniques for plant phenotyping », *in*: *Frontiers in Plant Science* 14(11) (2014), pp. 20078–20111.

[147] Z. Chunjiang et al., « Crop phenomics: current status and perspectives », *in*: *Frontiers in Plant Science* 10 (2019), p. 714.

[148] John Atanbori, Andrew P French, and Tony P Pridmore, « Towards infield, live plant phenotyping using a reduced-parameter CNN », *in*: *Machine Vision and Applications* 31.*1* (2020), p. 2.

[149] *Field Book*, `https://play.google.com/store/apps/details?id=com.fieldbook.tracker&hl=en_US`, Accessed July 07,2022.

[150] Muhammad Ayaz et al., « Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk », *in*: *IEEE Access* 7 (2019), pp. 129551–129583.

[151] Hadhami Garbouge et al., « Toward Numerical Practices in Variety Testing: a Rationale to Select the Most Promising Traits », *in*: *Report of the annual meeting of UPOV* (2020).

[152] Etienne David et al., « Global Wheat Challenge 2020: Analysis of the competition design and winning models », *in*: *arXiv preprint arXiv:2105.06182* (2021).

[153] Fayao Liu et al., « Learning depth from single monocular images using deep convolutional neural fields », *in*: *IEEE transactions on pattern analysis and machine intelligence* 38.*10* (2015), pp. 2024–2039.

[154] Salma Samiei et al., « Toward a Computer Vision Perspective on the Visual Impact of Vegetation in Symmetries of Urban Environments », *in*: *Symmetry* 10.*12* (2018), p. 666.

[155] Gopal K Kanji, *100 statistical tests*, Sage, 2006.

[156] Laurent Girin et al., « Dynamical variational autoencoders: A comprehensive review », *in*: *arXiv preprint arXiv:2008.12595* (2020).

[157] Peter Room, Jim Hanan, and Przemyslaw Prusinkiewicz, « Virtual plants: new perspectives for ecologists, pathologists and agricultural scientists », *in*: *Trends in Plant Science* 1.*1* (1996), pp. 33–38.

[158] Jibitesh Mishra and Sarojananda Mishra, *L-system Fractals*, Elsevier, 2007.

[159] Frédéric Boudon et al., « L-Py: an L-system simulation framework for modeling plant architecture development based on a dynamic language », *in*: *Frontiers in plant science* 3 (2012), p. 76.

Titre : Deep learning appliqué à l'imagerie multicomposante pour des problématiques de test de variétés

Mot clés : Tests des variétés végétales, vision par ordinateur, apprentissage profond, Depth-RGB, fusion, transfer, imagerie à faible coût.

Résumé : La thèse propose des contributions méthodologiques originales basées sur la vision par ordinateur et des méthodes apprentissage automatique pour le domaine de tests des variétés. Les systèmes d'imagerie pour les plantes sont développée ces dernières années en direction du phénotypage pour des expérimentations en milieu contrôlé ainsi que pour le domaine de l'agriculture. Le domaine de tests des variétés consiste à réaliser des mesures pour valider la qualité et l'originalité de toute nouvelle variété avant d'autoriser sa commercialisation. Jusqu'ici, il a été peu étudié au moyen d'outils numériques et les tests actuels sont essentiellement le résultat des inspections visuelles. Les travaux de la thèse se sont concentrés sur le test des variétés pour des grandes cultures. Sur un plan méthodologique, nous investiguons l'usage des systèmes d'imageries multicomposantes avec des capteurs à bas-coût et des méthodes d'apprentissage par réseaux de neurones profonds.

Dans une première partie, nous explorons le potentiel de capteurs multicomposantes RGB-Depth en test des variétés qui fournissent une information de trichromacie et de distance des plantes à la caméra. Les fusions précoce, intermédiaire et tardive de ces composantes dans un réseau de neurones par convolution ou à mémoire locale ou de type « transformer » sont examinées. Nous montrons la valeur ajoutée de la carte de distance notamment pour estimer les cinétiques des stades de développement individuels de plantules le jour comme la nuit. Ensuite, nous explorons les mêmes approches d'imagerie RGB-Depth pour la détection de stades de développement collectifs dans des petites parcelles sous la forme de textures.

Dans une seconde partie, nous abordons la question du possible transfert de connaissance de traits mesurés en milieux contrôlés (chambre de culture, phytotron) vers des milieux moins contrôlés (serres ou champs). Nous revisitons pour ce faire la détection de stades de développement de plantules. Une méthode d'augmentation de données simulant des ombres est proposée et montre son intérêt pour des approches d'apprentissage par transfert en serres comme au champ. Une ouverture vers l'usage de données de synthèse pour de l'apprentissage par transfert est proposée.

Dans une troisième partie, nous développons une imagerie multispectrale optimisée pour la détection et quantification de pathologies dans des tests de résistance aux maladies. Chaque étape est détaillée et validée sur des expérimentations qui s'étalent sur plus de trois saisons. Un pipeline complet est présenté incluant à nouveau des éléments d'apprentissage profond et d'apprentissage machine classique. Une ouverture vers la détermination automatique de protocole d'acquisition est proposée en annexe.

En plus de nos contributions méthodologiques, nous avons fourni des outils informatiques. Nous avons développé un logiciel pour traiter les séquences d'images RGB-Depth pour détection des stades de développement, mesurés la hauteur des plantes en temps réel, séparés les génotypes automatiquement, etc. Et aussi nous avons mis en disposition des bases de données annotées et des modèles entrainés.

Title: Deep learning applied to multi-component imagery for variety testing problems

Keywords: Variety testing, computer vision, deep learning, Depth-RGB, fusion, transfer, low cost imaging.

Abstract: The thesis proposes original methodological contributions based on computer vision and machine learning techniques to the variety testing research. Variety testing consists in performing measurements to validate the quality and originality of any new variety before allowing its commercialization. The current tests are essentially the result of visual inspections, and digital phenotyping is not common. The work of this PhD has focused on developing automated variety testing methods for crops. We build new multicomponent imaging systems based on low-cost sensors and deep learning networks.
In the first part, we explore the potential of multi-component RGB-Depth sensors in variety testing, which provide trichromacy and distance informations from the plants to the camera. Early, intermediate, and late fusions on these components are examined using a convolutional, local memory and transformer neural network. We demonstrate the benefits of the distance map, especially for estimating the kinetics of the individual developmental stages of seedlings during the daytime and nighttime. Then, we explore the same approaches of RGB-Depth imaging for detecting developmental stages in small plots as textures.
In the second part, we address the issue of the possible transfer learning of traits measured in controlled environments (growth chamber, phytotron) to less controlled environments (greenhouses or fields). To do so, we revisit the detection of seedling development stages. Furthermore, a data augmentation method simulating shadows is proposed and shows interest in transfer learning approaches in greenhouses and the field. Lastly, using synthetic data for transfer learning is proposed.

In the third part, we develop an optimized multispectral camera for detecting and quantifying disease in plant resistance tests. Each step is detailed and validated on an image dataset acquired during three seasons. A global pipeline is presented, including deep learning and classical machine learning methods. An opening toward the automatic determination of acquisition protocol is proposed in Annex A.

Additionally to our methodological contributions, we provided various computer tools. We developed software to process RGB-Depth image sequences for stage detection, measuring plant heights in real-time, etc. Furthermore, we provided annotated databases and generated trained models.