M2 Internship Offer
Multi-Plant Agnostic Counting: Few Shot Learning Specialized to Multiple Predefined New Classes

1 Context

1.1 Overall Context

To reduce the use of synthetic inputs, such as pesticides and fertilizers, or even eliminate them altogether, the transition to **agroecology** involves a spatio-temporal diversification of agricultural stands. Among the possible strategies, **sowing different species in the same field**, a practice known as **intercropping**, presents multiple advantages, such as controlling disease epidemics [Bou13] and improving the usage of soil nutrients. The association between a cereal and a legume (*légumineuse* in French) is of particular interest because legumes can develop a symbiosis with soil microorganisms fixing atmospheric nitrogen, thus letting more soil nitrogen available for the cereal . However, virtually all breeding programs currently select varieties based on sole crop data only, under high input management, despite significant variety × cropping type interactions [DDF+22]. Trait-blind models can estimate the magnitude of these interactions [HME+], but only trait-based models can decipher the underlying mechanisms [VMG+23]. As these models require large amount of data to be calibrated, leveraging high-throughput phenotyping technologies, such as drone imaging, is a prerequisite.

1.2 The MoBiDiv and CoBreeding Projects

As part of the MoBiDiv project (ANR PPR CPA), a field trial was conducted on INRAE's experimental plots on the "plateau de Saclay". Each of **200 varieties of bread wheat** was cultivated as a sole crop as well as an **intercrop with each of 2 pea varieties**, each modality being replicated twice and spatially randomized. As part of the CoBreeding project (ANR PEPR Agroécologie & Numérique), the modalities involving a subset of forty wheat varieties and one pea variety were replicated. These additional microplots were used in the PhD of V. Freitas for destructive plant sampling and calibration of predictive models of various traits of interest based on drone images.

1.3 Scientific Environment

Research labs:
Inria TAU team (joint team between Inria, CNRS and Université Paris-Saclay)
GQE at IDEEV (INRAE, Université Paris-Saclay, CNRS and AgroParisTech)
ECOSYS in Palaiseau (INRAE and AgroParisTech)
Location: mostly TAU (LISN) (building 660 "Digiteo", at Université Paris-Saclay)
Supervision: François Landes, TAU team (francois.landes@universite-paris-saclay.fr)
Sylvain Chevallier, TAU team (sylvain.a.chevallier@inria.fr)
Timothée Flutre, GQE (IDEEV) (timothee.flutre@inrae.fr)
Jérôme Enjalbert, GQE (IDEEV) (jerome.enjalbert@inrae.fr)
Jean-Marc Gilliot, ECOSYS (jean_marc.gilliot@agroparistech.fr)

2 Problem statement: Data

The density of plants *per species* is a main driver of crop yield, but is also crucial to study plant-plant interactions and their genetic basis. Indeed, the balance between competition and facilitation depends on it, and plants evolved ways of perceiving the density of their neighbors.
The MoBiDiv+CoBreeding field trial contained around 1600 micro-plots of 1,5 m wide and 7 m long. All wheat micro-plots were sown at 300 (respectively, 150) seeds per m^2 in sole crop (resp., intercrop), while all pea micro-plots were sown at 40 seeds per m^2 for both cropping types. Such plant densities over so many micro-plots made it impossible to count all plants manually. Instead, two linear meters were located in each of the 160 CoBreeding micro-plots (+ 30 control micro-plots), and **plants were manually counted per species only along these two meters per micro-plot**, early in the cycle to still be able to distinguish them from one another. This sub-sampled manual data already indicated that the expected sowing densities were reached, but only on average, with **large fluctuations between plots** due to the sowing, germination and emergence processes, with strong spatial effects and potential influence of varieties.
In parallel, several drone flights were made, at two altitudes, to take RGB images of all micro-plots. Images taken at **high altitude** (25m) are 80%-overlapping so as to be merged into a single image called an orthomosaic, by a dedicated photogrammetric software. The **orthomosaic is georeferenced very accurately, but has a low precision**. At a **lower altitude** (6 m), a single image was taken at the vertical of each micro-plot, providing a **much better resolution but preventing the reconstruction of an accurately georeferenced orthomosaic** (see figure 1).
Concretely, **we are interested in giving a precise estimate of the count and location of the plants** on the pictures taken, that cover all the micro-plots. We may use the manual counts over one meter as labels, or annotate some pictures manually (we should need only very few for "training").
This year, a new wheat-pea trial, STICSMIX, was designed, with far fewer micro-plots (30), but with more sowing densities. As before, plant counting will be performed manually and drone images will be taken as before. Moreover, **additional high-resolution images** will also be taken using a smartphone handed by a long selfie stick, to compare this low-cost technique with the drone-based one. **This internship will help to confirm or infirm the usefulness of that approach, and more generally may help optimize the data acquisition protocol**.

Figure 1: Snapshot of a **low-altitude** drone image. The herb-looking like pairs of leaves correspond to young wheat plants, and the small button-like (less numerous) plants correspond to young pea plants.

3 Problem statement: Models

3.1 State of the art

In terms of Machine Learning (ML) and Computer Vision (CV), some models are readily available to perform what is called **Class Agnostic Counting** (CAC), in a zero-shot or few-shot fashion [HDZ+24, PLZK]. Concretely, one pre-trains a large model on carefully annotated data (where each instance to be counted has been pointed), thus allowing to generalize to new instances: new pictures of the same classes of course, but also new classes (types of objects unseen at training time). The zero-shot counting occurs either through text prompting (using CLIP) or by some automatic detection of the object of interest in the picture. The few-shot paradigm generally consists in pointing one or a few instances of the object (class) of interest in the picture, then have the model count all occurrences [HDZ+24].
A crude summary of the inner workings of these models is the following. The pre-trained backbone is used as a feature extractor, patch by patch of the input image. In the few-shot context, **exemplars** (objects pointed at by annotations) also go through the backbone. Comparing the matching score between each patch and the exemplar(s), one obtains a **density map**, pixel by pixel, or at the patch scale, of where are the instances of the target class. From this density map, heuristics allow to obtain a count. The heuristics can also be included into the ML model, so that its parameters are trained in an end-to-end fashion, to improve counting.
Plants have peculiarities compared to most objects one can think of: they are highly deformable, their size and color vary a lot over time, and at a given time, different instances of the same species may vary in size. Some models have tailored the counting heuristics, and to some extent, the backbone design, to address these issues [LYWY23, WXCL24, HLX+25], defining what is called **Plant Agnostic Counting**. Such models will be our starting point.

3.2 Directions for research

Here, we are interested in counting (and locating) wheat and pea plants in numerous pictures. It would be absurd to have to point to 2-3 wheat and pea plants in each picture, since we always count these 2 classes. Instead, we want to have a **larger number of exemplars** (e.g. 30 to 50) of each class **extracted once and for all**, then have the model count these 2 classes in many pictures without further supervision. Ideally, one would not sequentially count wheat, then count pea in the images, but rather **design a Loss in a contrastive approach**, using the fact that a given blob of green on a picture has to be either wheat, xor pea.
For this, one will first have to experiment a bit with the existing models. Without pre-training it should be possible to perform a rather minimal hacking of existing codes, to investigate how one can **efficiently use the information contained in numerous exemplars** (30-50). Still without pre-training, it should also be possible to **adapt the counting heuristics to produce 2 density maps** (wheat, pea) and normalize them to account for the (pea xor wheat) concept.
After these proofs of concept and depending on the quality of the results obtained, **one may also code a new training pipeline, with the appropriate loss design, to perform a pre-training optimized for this particular downstream task**. Thus, depending on the results, there will be more or less need for editing the full pipeline (backbone, heuristics and Loss design).

In case of high accuracies obtained sufficiently early, the intern should help the Inrae team to get familiar with the pipeline, optimize the code for production, leading to a software paper (and possibly a publication in an ML conference and/or ecology-methods journal).
In the case where more ML-related novel ideas are needed to obtain satisfactory results, publication in a major ML conference will become more likely, while delaying the perspective of a quick software paper.
In all cases, under the assumption of serious work, **there should be room for a publication**.

4 Expected skills

We aim to recruit motivated and talented students. The skills required are:

- Solid background in ML (representation learning, few-shot learning, etc)
- Scientific rigor
- Proficiency in python and pytorch
- Some appeal for applications / Data Science

References

[Bou13] Mark A Boudreau. Diseases in intercropping systems. *Annual review of phytopathology*, 51(1):499–519, 2013.

[DDF+22] Dereje T Demie, Thomas F Döring, Maria R Finckh, Wopke van der Werf, Jérôme Enjalbert, and Sabine J Seidel. Mixture× genotype effects in cereal/legume intercropping. *Frontiers in Plant Science*, 13:846720, 2022.

[HDZ+24] Zhizhong Huang, Mingliang Dai, Yi Zhang, Junping Zhang, and Hongming Shan. Point, Segment and Count: A Generalized Framework for Object Counting. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17067–17076, Seattle, WA, USA, June 2024. IEEE.

[HLX+25] Xiaonan Hu, Xuebing Li, Jinyu Xu, Abdulkadir Duran Adan, Letian Zhou, Xuhui Zhu, Yanan Li, Wei Guo, Shouyang Liu, Wenzhong Liu, and Hao Lu. TasselNetV4: A vision foundation model for cross-scene, cross-scale, and cross-species plant counting, September 2025. arXiv:2509.20857 [cs].

[HME+] B. Haug, M. M. Messmer, J. Enjalbert, I. Goldringer, E. Forst, T. Flutre, T. Mary-Huard, and P. Hohmann. Advances in Breeding for Mixed Cropping – Incomplete Factorials and the Producer/Associate Concept. 11.

[LYWY23] Dunlu Lu, Jianxiong Ye, Yangxu Wang, and Zhenghong Yu. Plant Detection and Counting: Enhancing Precision Agriculture in UAV and General Scenes. *IEEE Access*, 11:116196–116205, 2023.

[PLZK] Jer Pelhan, Alan Lukežicˇ, Vitjan Zavrtanik, and Matej Kristan. A Novel Unified Architecture for Low-Shot Counting by Detection and Segmentation. *NeurIPS 2024*.

[VMG+23] Rémi Vezy, Sebastian Munz, Noémie Gaudio, Marie Launay, Patrice Lecharpentier, Dominique Ripoche, and Eric Justes. Modeling soil-plant functioning of intercrops using comprehensive and generic formalisms implemented in the STICS model. *Agronomy for Sustainable Development*, 43(5):61, October 2023.

[WXCL24] Zhicheng Wang, Liwen Xiao, Zhiguo Cao, and Hao Lu. Vision Transformer Off-the-Shelf: A Surprising Baseline for Few-Shot Class-Agnostic Counting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6):5832–5840, March 2024.