Can an AI agent configure a computer vision pipeline? Evaluating autonomous decision-making for agricultural image analysis

Starting date: as soon as possible

Duration: 6-12 months

Working Place: Agroscope Posieux

Language: French/English

Contact: Hassan-Roland Nasser, Helge Aasen

Email: roland.nasser@agroscope.admin.ch, helge.aasen@agroscope.admin.ch

Background

Building a detection model for a new agricultural dataset still requires an engineer to make a long series of decisions: which foundation model to use, how to prompt it, whether to tile the images, which thresholds to apply, how to filter the results, and when the labels are good enough to train on. Large language and vision-language models can now call tools, inspect intermediate results and plan multi-step workflows, and several recent papers use them to orchestrate annotation (Xia et al., 2025; Zhang et al., 2026). Almost all of this work is on text data or general-purpose images. Whether such an agent can make sound pipeline decisions on fine-grained agricultural imagery, where visually similar species and small objects are the norm, has not been studied.

Project

The thesis develops an agent that receives a raw drone image dataset and a task description, and must configure and run a detection pipeline built from existing components (SAM 3, Grounding DINO, embedding models, a lightweight detector). The agent is not the goal; it is the object of study. The research questions, to be refined with the student, are:

- Given only the data, does the agent reach the same pipeline configuration and accuracy as an experienced engineer, and on which datasets does it diverge?
- Which information does the agent need in order to decide well: raw images, model outputs, agreement between models, embedding clusters, sparse human feedback?
- When the agent declares a class finished or unlabelable, how often is it right against hidden expert labels? Where is it confidently wrong?
- Does a configuration learned on one dataset (pasture weeds) transfer to another (flowering field margins), and does the agent recognise when it should not?
- What does autonomy cost in compute and time, compared with the expert time it saves?

Development starts on RumexDrone, an existing Agroscope drone dataset of pasture weeds with full ground truth. The final experiments use drone images of flowering field margins collected in an Agroscope project on pollinator resources, for which expert labels are produced by a parallel Master thesis. The two students work in the same team and share tools, but the theses are independent.

Since this is an active research field, the exact scope and methods may be adjusted during the project as the literature evolves.

What we offer

- A well-defined experimental question in a field that is currently moving fast
- Existing datasets with ground truth, so experiments can start early
- Close supervision and a second student working on a complementary topic
- Access to GPU computing, model tracking and a budget for API calls
- Results intended for publication

Technical stack

Python, Git, PyTorch, an LLM orchestration library (LangGraph or similar), Weights & Biases, NVIDIA GPUs, Linux

Prior knowledge

Good proficiency in Python and basic knowledge of deep learning. Prior experience with LLM tool use or computer vision is a plus but can be learned during the project. The student should be comfortable designing experiments and reading recent papers, since much of the relevant work is less than a year old.

References

Xia et al., 2025. From Selection to Generation: A Survey of LLM-based Active Learning. ACL 2025.
Zhang et al., 2026. From Bounding Boxes to Visual Reasoning: An On-Policy Data Annotation Tool for Vision-Language Models. arXiv:2606.18846.
Oehme et al., 2026. Orchestrating segment anything models to accelerate segmentation annotation on agricultural image datasets. Frontiers in Artificial Intelligence 8:1748468.
Meta AI, 2025. SAM 3: Segment Anything with Concepts.