Goal Document of Master's Degree Project

Working Title:	Detection and Tracking of Soil Microorganisms using Deep Learning
Thesis Worker:	Jingmo Bai & Zuoyi Yu
Worker Email:	ji8255ba-s@student.lu.se & zu2043yu-s@student.lu.se
Start Date:	2024-02-05
Preliminary End Date:	2024-06-01
Supervisor:	Bo Bernhardsson & Edith Hammer & Hanbang Zou
Supervisor Email:	bo.bernhardsson@control.lth.se & edith.hammer@biol.lu.se & hanbang.zou@biol.lu.se
Examiner:	Johan Eker
Examiner Email:	johan.eker@control.lth.se

1 Background

Plastic pollution has become a global environmental issue, with significant threats to ecosystems. As plastic residues reside in the soil, their interactions with soil microorganisms including bacteria, fungi and protists have emerged as a critical study area. With the help of microfluidics technology, we build artificial microhabitats where microorganisms live and we can observe them in real time with the help of a microscope, providing a method to investigate their interactions with nanoplastics.

2 Motivation

Understanding how plastic residues influence microbial communities and their functions in soil ecosystems is crucial for assessing the long-term environmental consequences. This project aims to utilize Deep Learning techniques to help investigate the relationships between plastic residues and soil microorganisms by detecting and tracking different species of microorganisms and analyzing their behaviour and properties. Mathematical analysis provides insights into the influence of plastic residues on microorganisms in soil ecosystems.

3 Overall Objectives

The main goal of this thesis is to build a model based on Deep Learning techniques to help with easier mathematical microbiology research. This work is divided into two sections, which are model development and data analysis respectively. For the first section, data labelling and enhancements are needed before training. Then the major task in this section is to train a model to track, classify and segment different types of protists in the videos. The analysis section involves investigating the effects of nanoplastics on protists from a mathematical perspective, for example, moving speed, orientation and size.

4 Research Questions

1. How can data augmentation be performed if there is a lack of the dataset? Is pre-processing yielding a better accuracy?
2. How to deal with the protists with dynamic shapes? For example, the amoeba do not have cell walls, which allows for free movements by using pseudopods.
3. How to analyze and compare the effects of nanoplastics on protists? How can these be quantified? How does this help with further microbiology research?

5 Preliminary Methodology

We plan to start by reading articles related to microbes and protists to obtain some fundamental knowledge about the background of our project. Learning how to distinguish different types of protists is the main goal in this phase. With enough knowledge, we will begin to label the dataset. As it may occur a lack of data problem, data argumentation may be applied. After preparing enough data for training, the second phase is to train a model which can detect, track, classify and segment all the protists in the videos. A preliminary plan to achieve these functions is to use YOLOv8 or Mask R-CNN to build the model. Some pre-processing methods, such as image enhancement and normalization, are also planned for better performance. After achieving the most favourable model, some mathematical analysis is required to research how nanoplastics affect the habits of the protists. We plan to analyze at least three aspects: speed and orientation measured by tracking, and size measured by segmentation.

6 Previous Work

The background knowledge and related thesis can be found through Google Scholar, and Micaela's PhD thesis[3] is also really helpful. The existing models we plan to try, YOLOv8[2] and Mask R-CNN[1], are open-source and published on GitHub and their websites.

7 Resources

7.1 Datasets

The datasets provided by the department of biology comprised videos and images of different species of protists obtained by observing the microfluidics chips via the microscope[3].

7.2 Supercomputer resources

The Alvis cluster is a national NAISS resource dedicated to Artificial Intelligence and Machine Learning research. The system is built around Graphical Processing Units (GPUs) accelerator cards and consists of several types of compute nodes with multiple NVIDIA GPUs.

8 Timeline

1. **Preliminary Work (2.05 - 2.18)**
   1.1 Read related paper, setup virtual environment
   1.2 Label dataset
2. **Development Phase (2.19 - 3.31)**
   2.1 Protists detection
   2.2 Classification
   2.3 Tracking in the videos
3. **Analysis Phase (4.01 - 4.14)**
4. **Paper (4.15 - 5.31)**

References

[1] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[2] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, January 2023.

[3] P Micaela Mafla-Endara. *Encounters at the microscale: Unraveling soil microbial interactions with nanoplastics.* PhD thesis, Lund University, 2023.