RESEARCH PLAN

Topic:	**Multimodal plankton recognition**
Laboratory:	Computer Vision and Pattern Recognition Laboratory (CVPRL)
Resources:	Master's thesis worker Veikka Immonen
	Supervisor Assoc. Prof Tuomas Eerola
	Supervisor M.Sc. Joona Kareinen
	Supervisor D.Sc. Fedor Zolotarev
	Supervisor Prof. Lasse Lensu
	Supervisor Prof. Heikki Kälviäinen

Date	Author	Comments	Status
25.11.2024	Tuomas	First version	Draft

1 Background

Research on planktonic systems is hampered by the bottleneck of acquiring species-level information of the communities, consisting of hundreds of microorganism species with generation times in the order of hours. Concealed plankton community dynamics reflect changes in environmental forcing, growth traits of competing species, and multiple food web interactions. Recent technological advances have led to emergence of automated imaging instruments, with improving resolution and output rates, up to tens of thousands images per hour. It is becoming possible to produce real-time Big Data of plankton communities. This opens new horizons for testing core hypotheses in planktonic systems, derived from macroscopic realms: in community ecology, biodiversity research, and ecosystem functioning. The next bottleneck is obvious, however. No human will ever screen the millions of images. Big data calls for computer vision and machine learning approaches producing interoperable data across platforms and systems.

In this Master's thesis, multimodal plankton recognition is considered, that is utilizing additional measurement information for plankton image recognition. Certain plankton imaging instruments provide supplementary data besides images. For example, CytoSense instruments produce scatter and fluorescence profiles for all particles passing through the sensor (see Fig. 1). This information can be valuable for plankton recognition. Moreover, capturing images is generally slower than measuring optical properties. CytoSense captures profile data for all plankton particles but images only a small portion of them. Therefore, utilizing scatter and fluorescence profiles has the potential to provide more comprehensive information about plankton communities. Plankton image recognition is a well-established research topic, with state-of-the-art models achieving human-level accuracy. However, the classification of scatter and fluorescence profiles has not been widely studied. Ideally, a plankton recognition model should process both optical profiles and images, and when both modalities are available, use them simultaneously to improve recognition accuracy.

During this thesis project, a universal plankton recognition model for CytoSense data will be developed. The model should accept either images, optical profiles, or both as input. One possible method is to utilize cross-modality alignment. Similar to the widely-used CLIP (Contrastive Language-Image Pretraining), the idea is to use contrastive loss to align the corresponding feature vectors from each modality so that they appear close to each other in the feature space (see Fig. 2). However, unlike CLIP, where the second modality is text, profile data will be used in addition to images. A recognition model trained on such features can then be applied to either modality. When both modalities are available, more reliable feature vectors can potentially be obtained. The methodology should be general enough to apply to other multimodal plankton data besides CytoSense.

The objectives of this thesis work are as follows:

1. To implement and train a contrastive optical profile-image pretraining model to obtain modality-independent feature representations.
2. To build a plankton recognition model on top of the learned feature representations allowing to recognize plankton class (e.g. species) from either or both of the modalities.
3. To evaluate the model on CytoSense data.

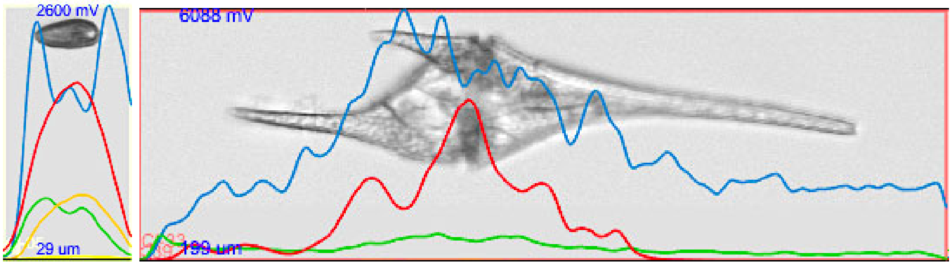

Figure 1: CytoSense data.

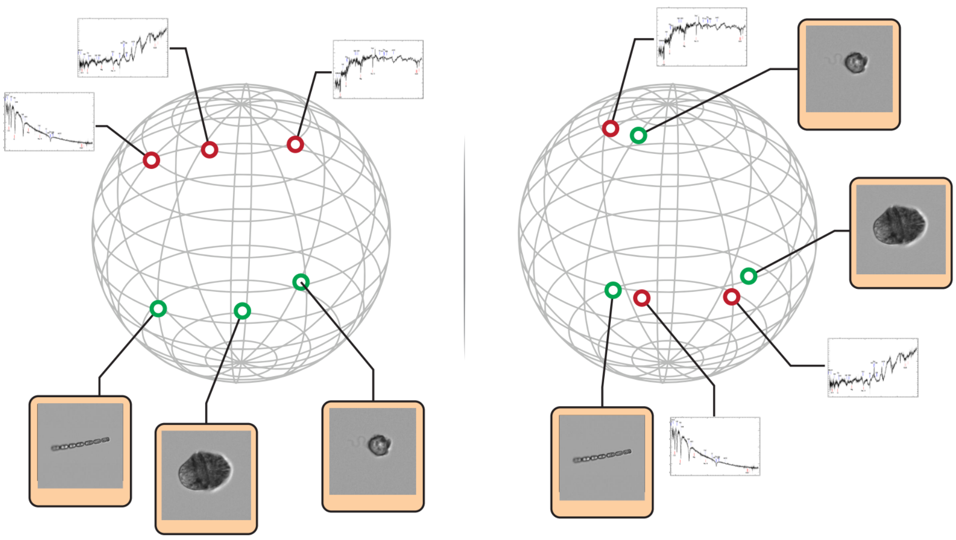

Figure 2: Contrastive optical profile-image pretraining.

2 Task description and schedule

The main steps of the work are as follows:

1.	Get acquainted with the FastVision and FastVision-plus projects, related publications and datasets.	25.11.-31.1.
2.	Get acquainted with the CytoSense instrument and the collected dataset.	
3.	Make a literature review on multimodal image recognition and modality alignment.	
4.	Together with the supervisors, select the model for feature alignments and train it on CytoSense data	1.1.-30.4.
5.	Visualize the feature space.	
6.	Build a plankton recognition model on top of the learned feature representations.	
7.	Study approaches to improve the recognition accuracy when both modalities (feature representations) are available.	
8.	Evaluate the method on CytoSense data.	
9.	Write the background, descriptions of the materials and methods, performed experiments and their results, discussion, and conclusion for the thesis.	1.5.-27.5.
10.	Get comments from the supervisors.	
11.	Revise the thesis according to the comments.	
12.	Write a manuscript to a scientific conference/journal about the achieved results.	28.5.-15.6.
13.	Get comments from the supervisor(s).	
14.	Revise the manuscript according to the comments.	
15.	Send the manuscript to a review process.	