**PEC 2 -Status Report of the thesis Underwater species detection in images and videos**

Ricard Fos Serdà

**Description of the progress on the project as of 09/05/2022**

The following tasks were planned to be finished by 09/05/2024:

* **Image data recollecting and preprocessing for the training of the models.**

Planned to finish by 07/03/2022 (delay)

This task was delayed but successfully finished by April 21st. I found a public dataset and preprocessed it, correcting some labeling errors from the original and converting it into a better format for my experiments, where each row of the dataset contains one image and all its boxes and segmentations, instead of having one row for each box.

* **Object Detection Implementation and evaluation**

Planned to finish by 25/04/2022. (Slight delay)

I finished training and doing inference on the first Faster R-CNN model by April 26 and finished the mAP evaluation algorithms by April 29th, achieving good results. But the following YOLOv5 training and evaluation was finished on May 8th. The implementation of each model was preceded by the careful study of each model by reading its research papers and other sources, and documenting each of them in my written TFM.

* **(Optional) Semantic Segmentation Implementation**

Planned to finish by 09/05/2022 (on time)

I finished training and evaluating a Instance Segmentation Mask R-CNN model on video successfully by May 1st achieving good results.

* **(Optional) Get extra info from detected species**

Planned to finish by 09/05/2022 (delay, possibly drop)

I had the idea of getting extra info like the position of the species in relation to the camera or their size. This optional task will probably be dropped, but I may try to implement Object Tracking algorithms to the YOLOv5 output if I have time or do it after finishing the TFM.

* **(Optional) Create new Dataset**

Planned to finish by 16/05/2022 (dropped)

I have dropped the task of creating and labelling my own dataset from public videos as it is very time consuming and has little value.

**Changes in the planning**

After the last status report, I had a delay in respect to the original planning as I started learning and practicing the Pytorch library for object detection on videos before training and using my own models and dataset.

I dropped the idea of getting extra info from detected species and implement object tracking instead if I have enough time. I have dropped the idea of creating my own dataset. I also dropped the idea of training a Transformer model (SWIN), which is just a new backbone model for Faster and Mask R-CNN that are not appropriate for real time detection, but I might include some documentation about it in the final text.

**List of partial Results**

* Documentation of 2-stage and1-stage object detection models.
* Data Preprocessing notebook for the luderick dataset
* Faster R-CNN training notebook
* Faster R-CNN inference and evaluation notebook
* Mask R-CNN training notebook
* Mask R-CNN inference and evaluation notebook
* YOLOv5 training notebook
* YOLOv5 inference and evaluation notebook
* Output video from Mask R-CNN with inference time fps printed (the video is 30fps)
* Output video from YOLOv5 with inference time fps printed (the video is 30fps)