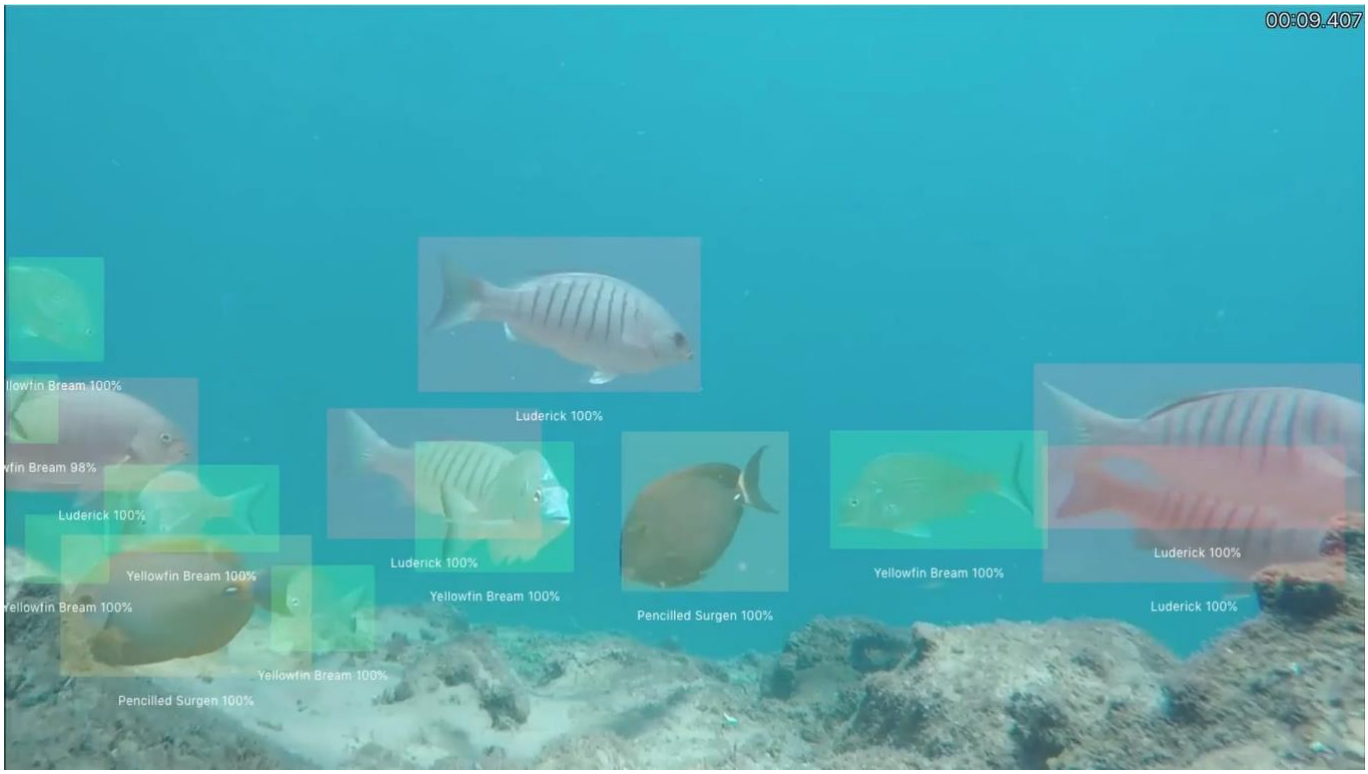
TER 2021-071 – Development

Title:

Deep Learning System for animal recognition in underwater images and videos from Mediterranean Sea



Students:

Giuseppe Alessio Dimonte	Master 2 EIT Digital (DSC)
Arturo Adan Pinar	Master 2 EIT Digital (DSC)
Yueqiao Huang	MAM5

Supervisors:

Frederic Precioso

Diane Lingrand

1. Executive summary

The main objective of the project is the development of a Deep Learning model able to recognize fishes and corals in images and videos of the Mediterranean Sea recorded from an underwater drone.

To train the model the creation of a custom dataset is needed. The process to generate a useful dataset consists in three main steps:

1) Adaptation of Megadetector[1] to the underwater scenario using models trained on the SeaCLEF[2] dataset.
2) Implementation of temporal consistency and tracking to reduce the number of images with the same subject.
3) Development of a GUI (Graphical User Interface) through which experts can annotate the images and assign a label to each bounding box.

Once the dataset is ready, it is possible to implement the various CNNs (Convolutional Neural Networks) for:

1) Object detection: in the literature there are many algorithms that can be employed for this purpose (R-CNN, SPP-net, YOLO, SSD). For this reason, they will be tested and compared in order to choose the best configuration based on MAP[3] (Mean Average Precision) and IoU[4] (Intersection over Union).
2) Instance segmentation: using Mask R-CNN it is possible to exploit the first stage of R-CNN and then add a layer[5] to predict the mask.
3) Counting: it is possible to use the counting by detection mode, exploiting the segmented instances inferred from the instance segmentation part.

Finally, the outputs are combined showing the information for each single frame.

2. Description of the project

Technological context

The technological context is the set of systems, devices, and services that the users have available.

For a Deep Learning project, GPUs (Graphical Processing Units) are fundamental: they are powerful chips, that allow the usage of neural networks in a faster way than CPUs, often making the training phase more than 100x faster.

Université Côte d'Azur provided the access to a VM (Virtual Machine), available through VPN and SSH and equipped with the GPU model "NVIDIA Quadro RTX 6000".

Fanny Simoes gave access to the repository containing the "Mercantour project" on GitHub and, starting from there, a way to understand how to implement and use the Megadetector tool.

Finally, all the source code of the project will be made available on GitHub, keeping track of the code development, and allowing to parallelize tasks for team collaboration.

[1] Megadetector is a software that can locate species on every image, returning the coordinates of the bounding boxes

[2] SeaCLEF is a dataset containing videos and images of fish species in their natural coral reef environment

[3] MAP is a popular metric for measuring the performance of a model

[4] IoU is an evaluation metric used to measure the accuracy of an object detector on a particular dataset

[5] In Deep Learning a layer is a structure in the model architecture, which takes information from the previous layers and passes information to the next layer

Motivations

- **Problem**: identification and counting of different fishes and corals in underwater images and videos.
- **Involved parts**: ecological surveillance, biodiversity monitoring, marine ecosystem analysis are all sciences benefiting from this kind of project.
- **Use**: the project will provide experts a software able to help them to keep track of the Mediterranean Sea fauna and flora.
- **Goal 1**: provide a tool able to identify and count different fishes and corals in underwater images and videos.
- **Goal 2**: provide a UI to experts for giving them the possibility of correctly annotating images containing fishes and/or corals.

Identified risks (and countermeasures)

- Underwater images are blurrier than air images: the CNNs can struggle during the training phase to understand the content of images and for this reason image preprocessing must be used in order to mitigate the noise and return clearer images.
- Fishes move in unusual and unexpected way compared to other animals: the trajectory followed by fishes can change rapidly and, because of being in a fluid-environment, they can have a wider range of motion. To overcome this aspect an accurate data augmentation part needs to be implemented to train the network to recognize these patterns.
- The underwater drone records video while moving. In this scenario the movement of the objects and the camera are mixed, resulting in a complex problem. To face the situation, it is possible to perform motion clustering and classification, comparing two successive image frames[6].

Scenarios

The project faces two different scenarios, as described in Figure 1 (see Annex):

1. The main scenario of the project is the one aligned with the main objective of the project, namely provide to the users a program able to identify, count and segment different fishes and corals for unseen underwater images and videos provided by them.
2. The secondary scenario concerns the activity of annotation: fish and coral experts provide their knowledge through the designed UI for contributing in putting the correct labels on the boundary boxes, creating a custom dataset used for training the following CNNs.

3. Implementation

The implementation plan is detailed with a GANTT chart (see Annex, figure 2), displaying the time allocation for each activity and keeping in mind dependencies on other tasks to achieve the final result. The plot defines 24 major activities to be performed.

1) The light blue bars indicate the steps performed by the team.
2) The dark blue bars indicate the actions that need to be performed by the expert team.
3) The green bar represents Christmas holidays.
4) The red bars indicate the projects' deliveries.

Yueqiao, which has experience with image processing, is able to improve the quality of the input images before using them for training the CNNs. Arturo instead has previous experience with UI development and will be in charge of the design of the GUI. Alessio, which has already experience with object detection thanks to a previous project, will work on shifting the Megadetector to the recognition of bounding boxes for fishes and corals. After those first steps, each team member will focus on one specific CNN among the three, maintaining an imperative communication through the GitHub repository and the validation with the supervisors.

[6] https://www.researchgate.net/publication/221127238_Moving_object_detection_under_free-moving_camera

4. ANNEX

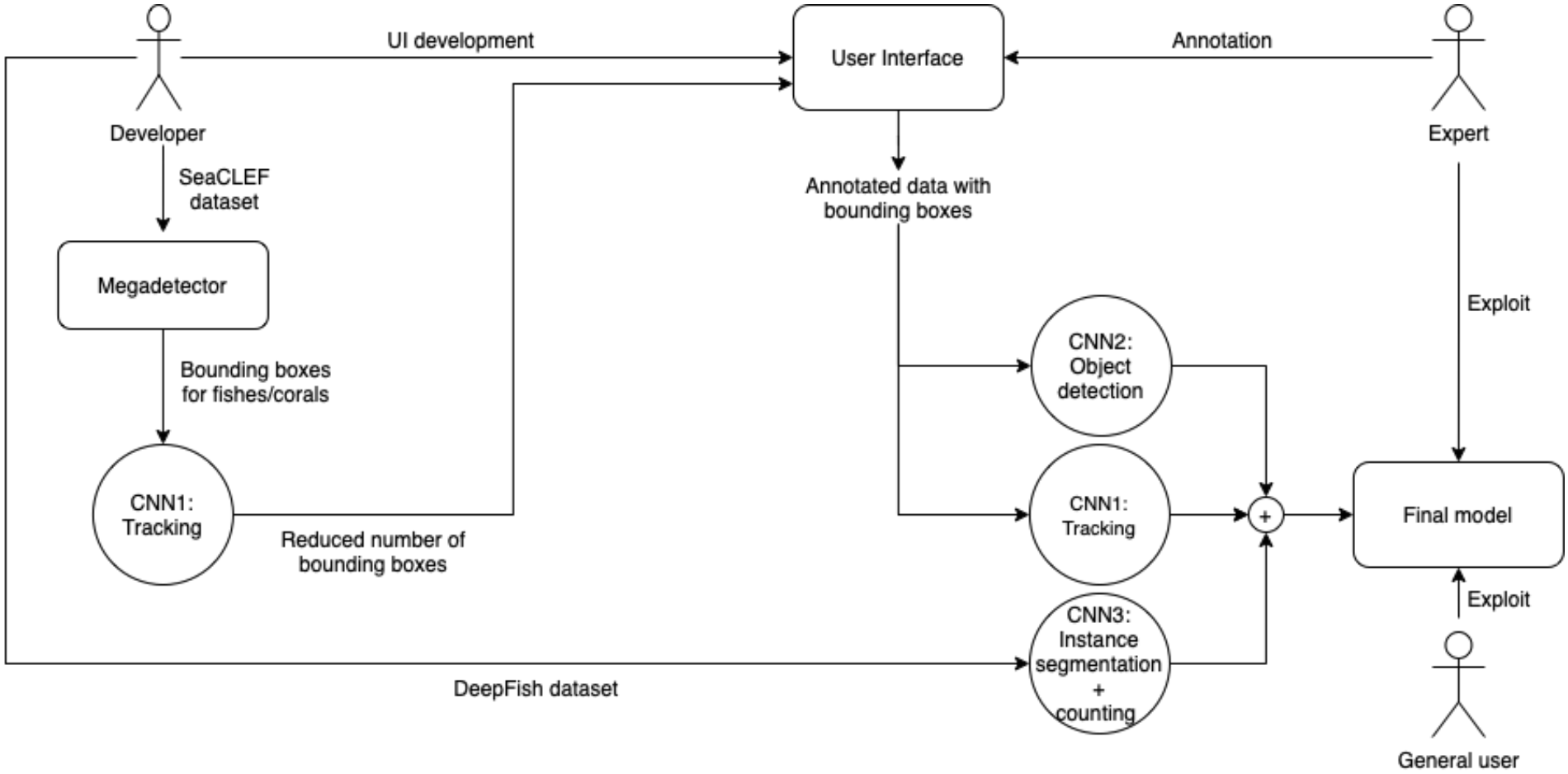
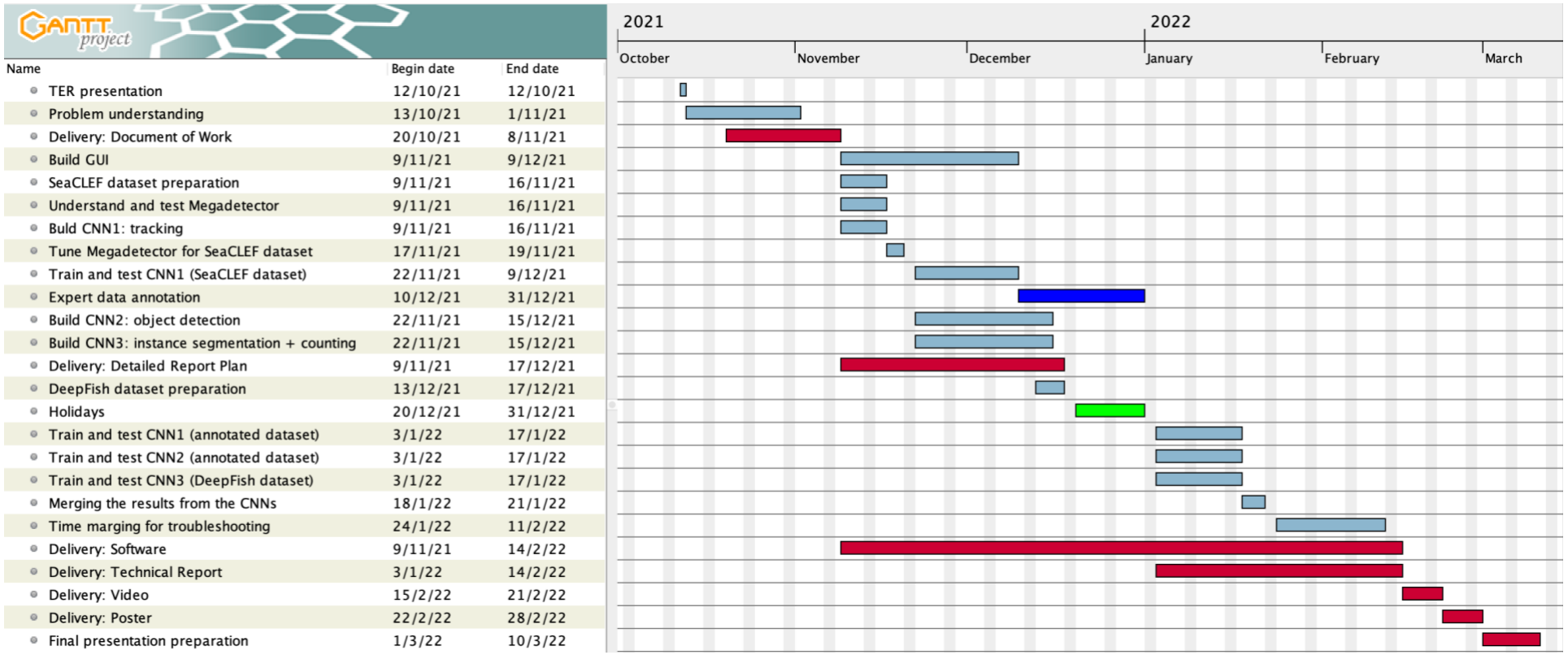


Figure 1: use-case diagram



Name	Begin date	End date
TER presentation	12/10/21	12/10/21
Problem understanding	13/10/21	1/11/21
Delivery: Document of Work	20/10/21	8/11/21
Build GUI	9/11/21	9/12/21
SeaCLEF dataset preparation	9/11/21	16/11/21
Understand and test Megadetector	9/11/21	16/11/21
Buld CNN1: tracking	9/11/21	16/11/21
Tune Megadetector for SeaCLEF dataset	17/11/21	19/11/21
Train and test CNN1 (SeaCLEF dataset)	22/11/21	9/12/21
Expert data annotation	10/12/21	31/12/21
Build CNN2: object detection	22/11/21	15/12/21
Build CNN3: instance segmentation + counting	22/11/21	15/12/21
Delivery: Detailed Report Plan	9/11/21	17/12/21
DeepFish dataset preparation	13/12/21	17/12/21
Holidays	20/12/21	31/12/21
Train and test CNN1 (annotated dataset)	3/1/22	17/1/22
Train and test CNN2 (annotated dataset)	3/1/22	17/1/22
Train and test CNN3 (DeepFish dataset)	3/1/22	17/1/22
Merging the results from the CNNs	18/1/22	21/1/22
Time marging for troubleshooting	24/1/22	11/2/22
Delivery: Software	9/11/21	14/2/22
Delivery: Technical Report	3/1/22	14/2/22
Delivery: Video	15/2/22	21/2/22
Delivery: Poster	22/2/22	28/2/22
Final presentation preparation	1/3/22	10/3/22

Figure 2: GANTT chart