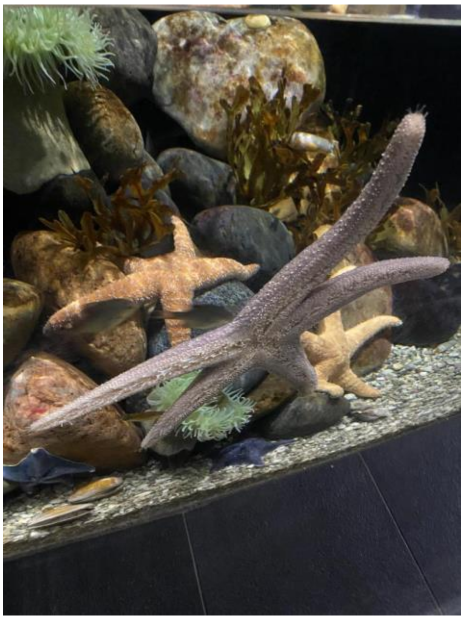
HW3 Object Detection + Data Augmentation Report

R11945005 郭庭沂

1) Image captioning

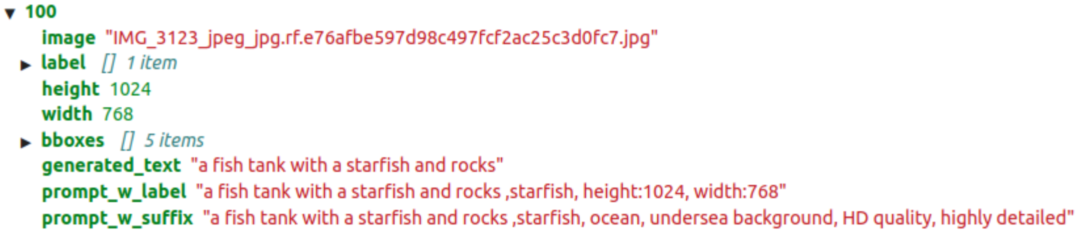
a. Compare three different pre-trained models



model	Generated text
Salesforce/blip2-opt-2.7b	this is a picture of a starfish in a tank
Salesforce/blip2-opt-6.7b-coco	this is a picture of a fish tank with a starfish and rocks
Salesforce/blip2-flan-t5-xl	this is a picture of a starfish in an aquarium

As the result, “Salesforce/blip2-opt-6.7b-coco” generated more details of the image than the others. Therefore, we will choose “Salesforce/blip2-opt-6.7b-coco” as the pretrained model for later problems.

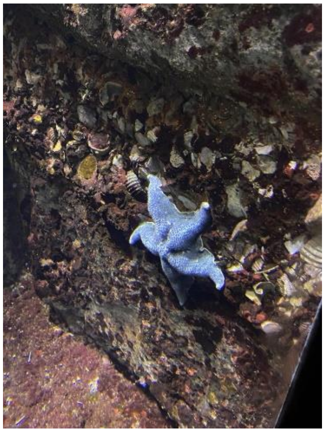
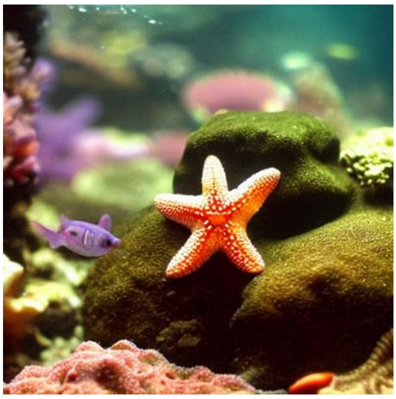
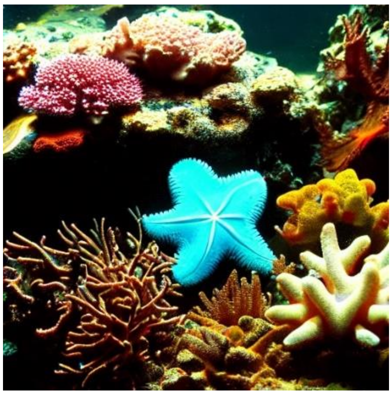
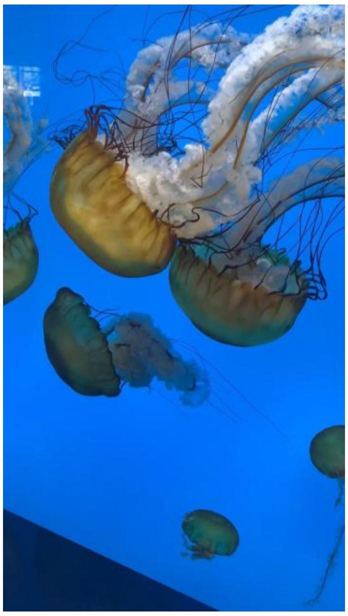
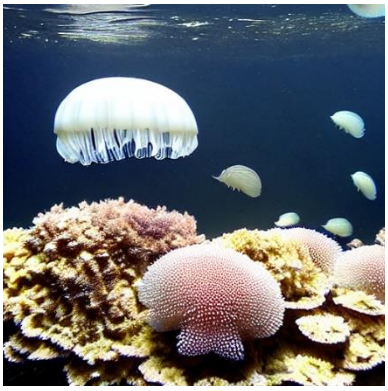
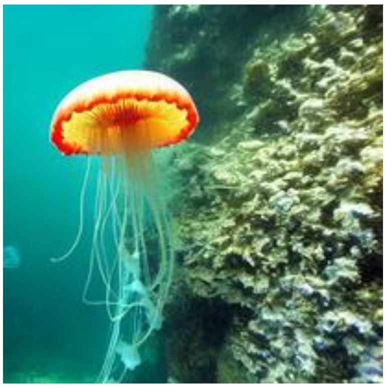
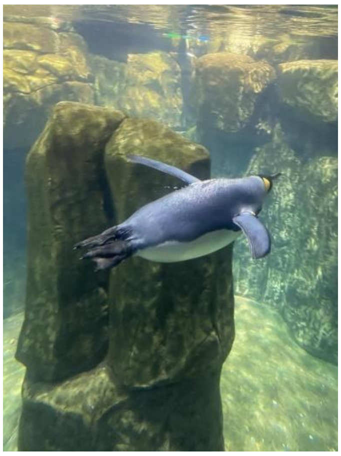
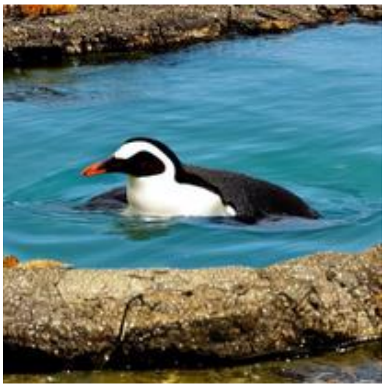
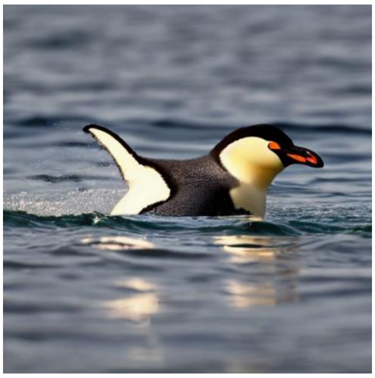
b. Design 2 templates of prompts for later generating comparison.



```
▼ 100
    image "IMG_3123_jpeg_jpg.rf.e76afbe597d98c497fcf2ac25c3d0fc7.jpg"
  ▶ label  [] 1 item
    height 1024
    width 768
  ▶ bboxes  [] 5 items
    generated_text "a fish tank with a starfish and rocks"
    prompt_w_label "a fish tank with a starfish and rocks ,starfish, height:1024, width:768"
    prompt_w_suffix "a fish tank with a starfish and rocks ,starfish, ocean, undersea background, HD quality, highly detailed"
```

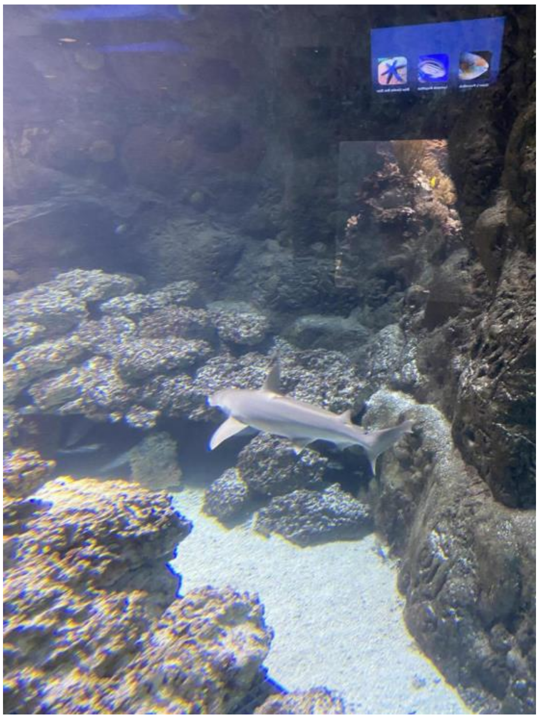
2) Text to image Generation

a. Use two kinds of generated prompts to generate images (text only)

Original	Prompt_w_label	Prompt_w_suffix
		
		
		

b. Select "Prompt_w_label" prompt, and perform image grounding generation. (text + image)

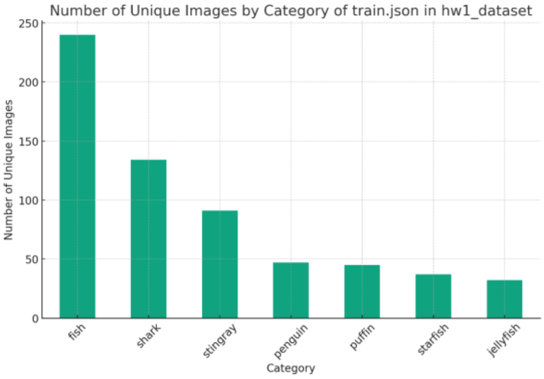
Original	"Prompt_w_label" + image

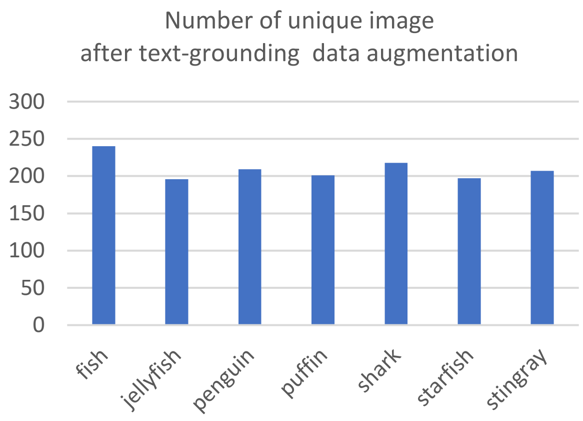
3) Table of performance based on FID

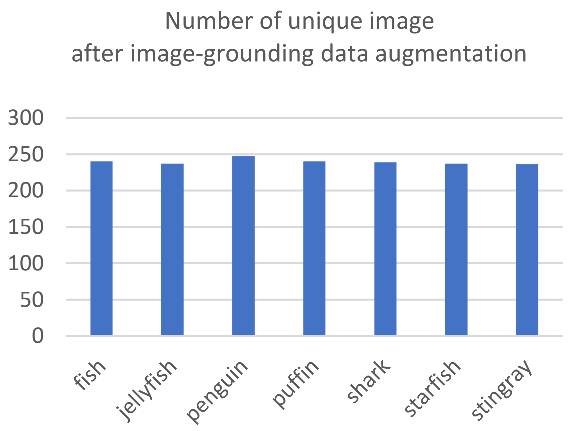
	Text Grounding		Image Grounding
Prompt	Prompt_w_label	Prompt_w_suffix	Prompt_w_label
FID	143	145	138

4) Table of the improvement of your detection model from HW1 after data augmentation



In order to solve the data imbalance problem, each category contains nearly 240 unique images by doing image grounding data augmentation, and each category contains nearly 200 unique images except for fish having original 240 images by doing text grounding data augmentation.





We use the checkpoint which is pretrained on COCO 2017 object detection dataset to train models on three datasets, which are “Before Data Augmentation”, “After text-grounding data augmentation”, and “After image-grounding data augmentation” individually. The performance of each model on validation dataset is listed in table below.

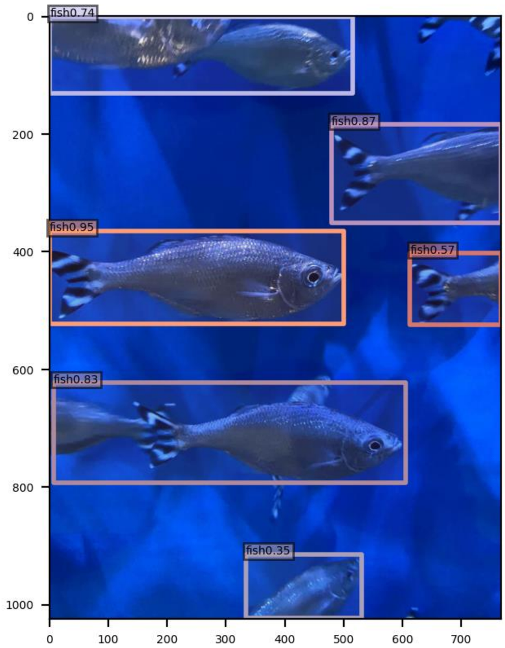
	Before Data Augmentation	After Data Augmentation (Text grounding)	After Data Augmentation (Image grounding)
AP	0.5213	0.4931	0.5018

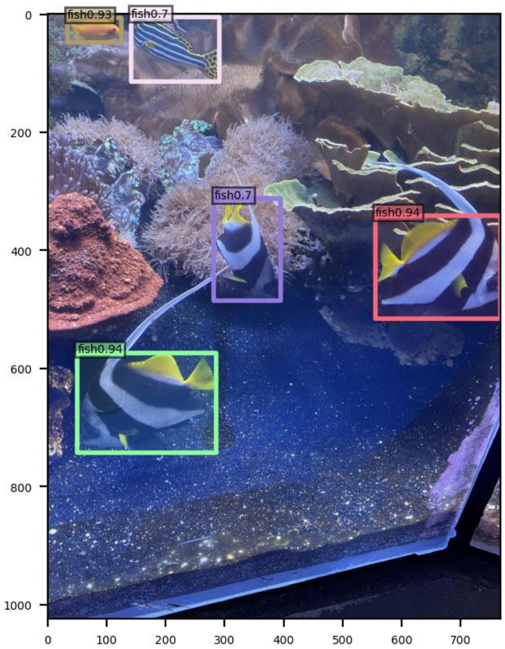
The results show that the performance of the models which are trained on augmented datasets are slightly degraded compared to non-augmented datasets.

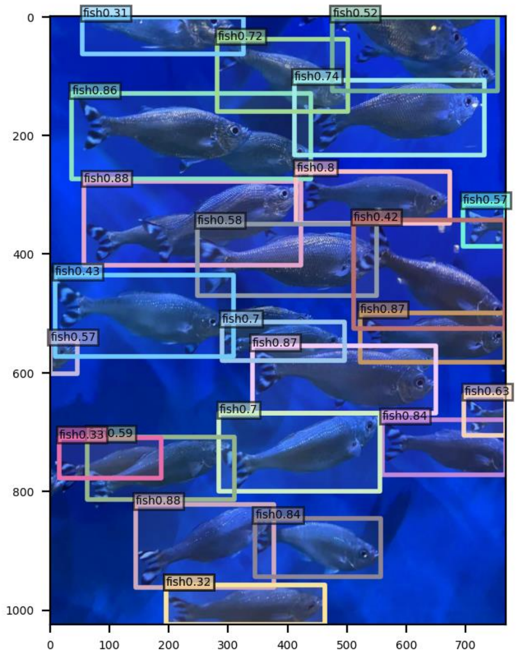
There are two reasons we consider why the performance doesn’t increase as expected. First, the original dataset contains 448 images, and the augmented dataset contains 1468, and 1676 images. It means that most images engaged in model training are generated images, therefore the performance of the model highly depends on the performance of the generated images. However, the data augmentation process includes image captioning and image generating, the quality of the generated images depend on the correctness of image captioning and the performance of the image generating model. We had found some miscaptioning text and some generated images seems unlikely to any categories. Also, the data imbalance of validation set might affect the evaluation of model performance too.

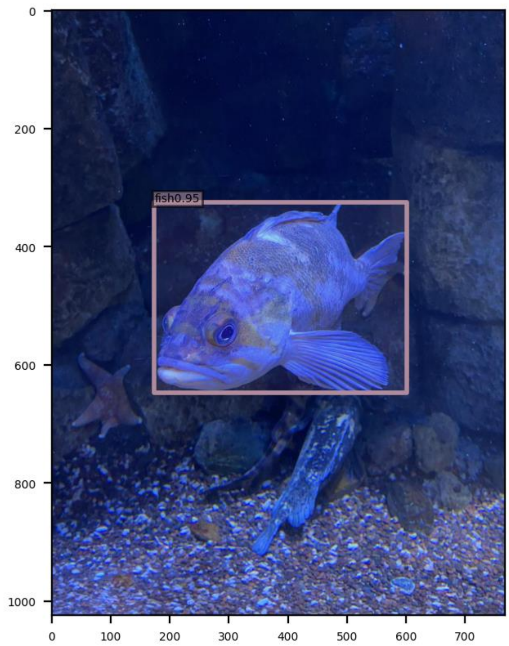
5) Visualization

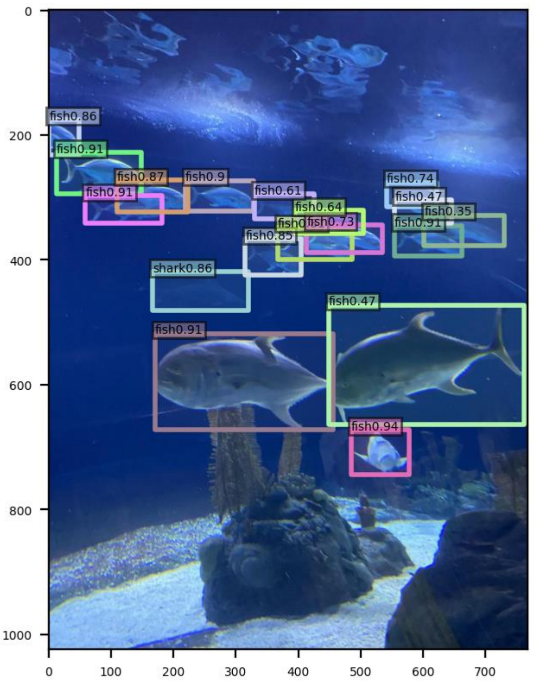
a. Fish



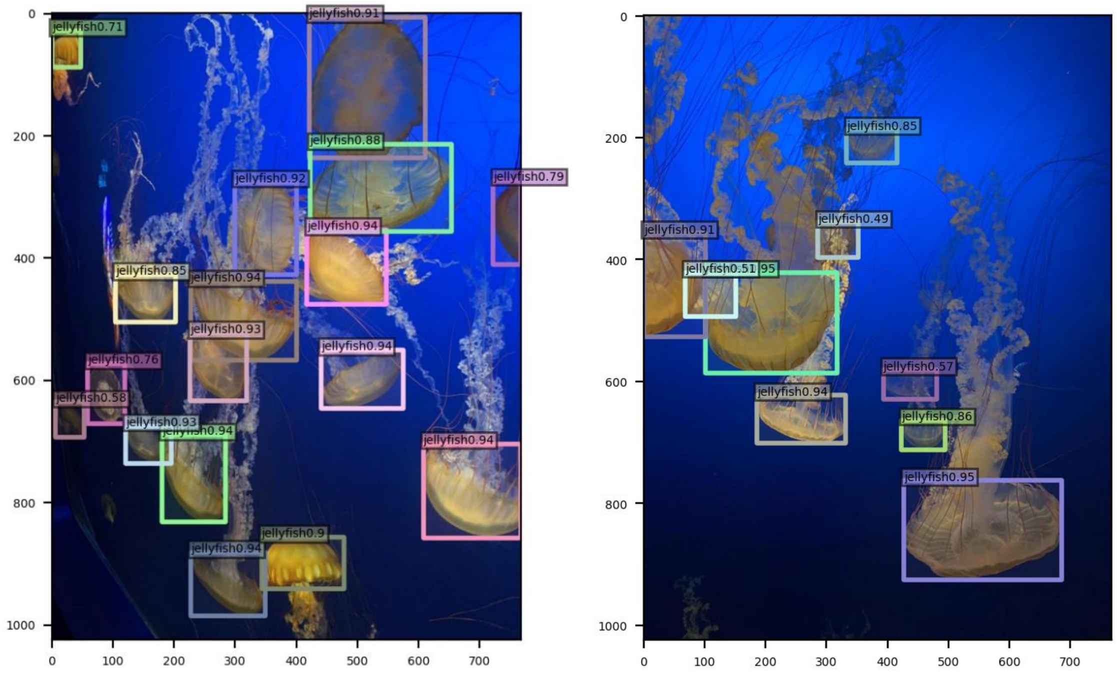


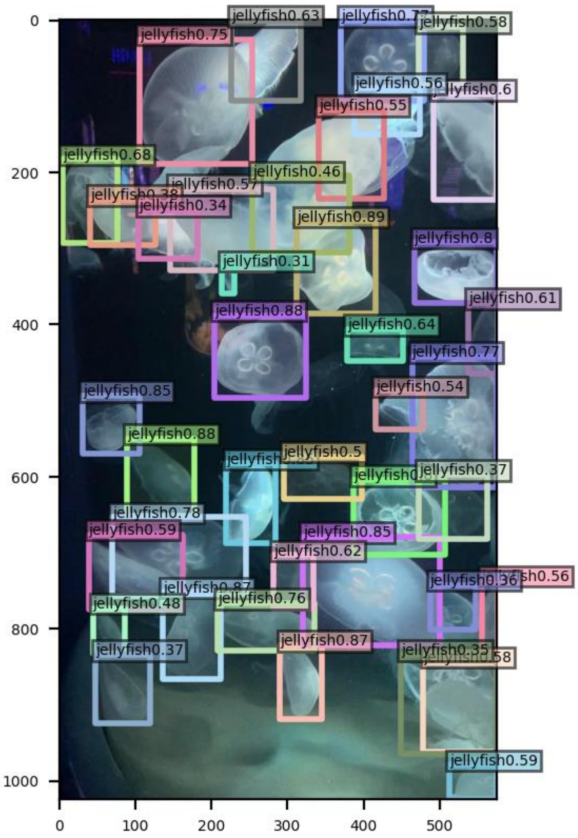


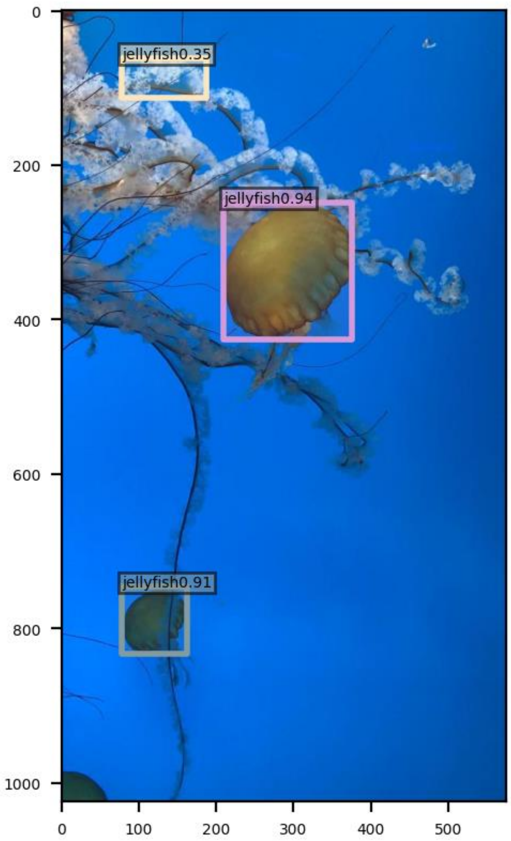


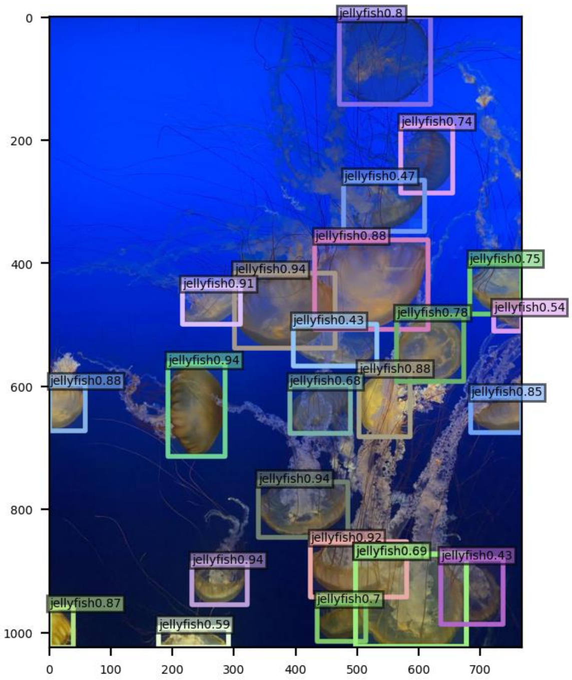

fish0.86
fish0.91
fish0.87
fish0.9
fish0.91
fish0.61
fish0.74
fish0.47
fish0.64
fish0.73
fish0.35
fish0.91
fish0.85
shark0.86
fish0.47
fish0.91
fish0.94
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

b. Jellyfish

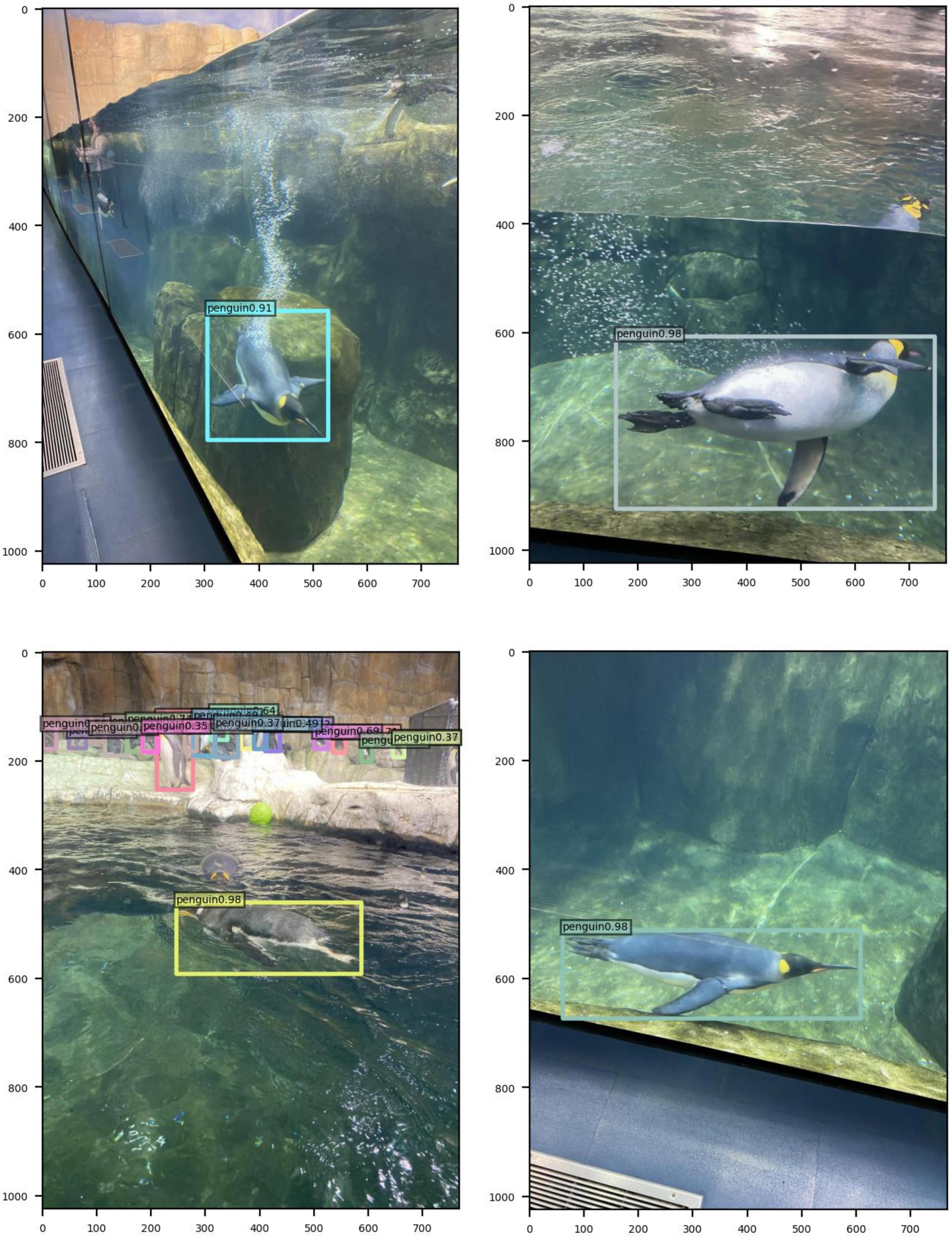

jellyfish0.71
jellyfish0.91
jellyfish0.88
jellyfish0.92
jellyfish0.79
jellyfish0.94
jellyfish0.85
jellyfish0.94
jellyfish0.93
jellyfish0.94
jellyfish0.76
jellyfish0.58
jellyfish0.93
jellyfish0.94
jellyfish0.94
jellyfish0.9
jellyfish0.94
jellyfish0.85
jellyfish0.91
jellyfish0.49
jellyfish0.51
95
jellyfish0.57
jellyfish0.94
jellyfish0.86
jellyfish0.95
0
200
400
600
800
1000
0
100
200
300
400
500
600
700


jellyfish0.63
jellyfish0.77
jellyfish0.58
jellyfish0.75
jellyfish0.56
llyfish0.6
jellyfish0.55
jellyfish0.68
jellyfish0.46
jellyfish0.57
jellyfish0.38
jellyfish0.34
jellyfish0.89
jellyfish0.8
jellyfish0.31
jellyfish0.61
jellyfish0.88
jellyfish0.64
jellyfish0.77
jellyfish0.85
jellyfish0.54
jellyfish0.88
jellyfish0.5
jellyfish0.37
jellyfish0.78
jellyfish0.59
jellyfish0.85
jellyfish0.62
jellyfish0.87
jellyfish0.36
sh0.56
jellyfish0.48
jellyfish0.76
jellyfish0.37
jellyfish0.87
jellyfish0.35
jellyfish0.58
jellyfish0.59
0
200
400
600
800
1000
0
100
200
300
400
500

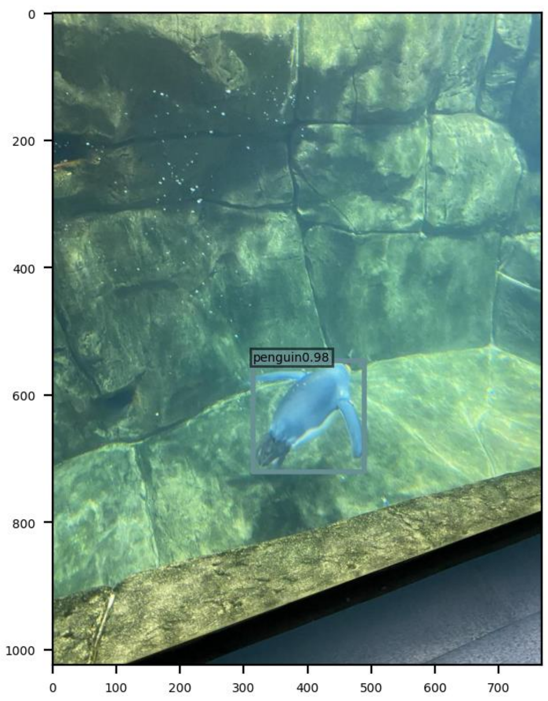

jellyfish0.35
jellyfish0.94
jellyfish0.91
0
200
400
600
800
1000
0
100
200
300
400
500


jellyfish0.8
jellyfish0.74
jellyfish0.47
jellyfish0.88
jellyfish0.75
jellyfish0.94
jellyfish0.91
jellyfish0.54
jellyfish0.43
jellyfish0.78
jellyfish0.94
jellyfish0.88
jellyfish0.88
jellyfish0.68
jellyfish0.85
jellyfish0.94
jellyfish0.92
jellyfish0.69
jellyfish0.43
jellyfish0.94
jellyfish0.87
jellyfish0.7
jellyfish0.59
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

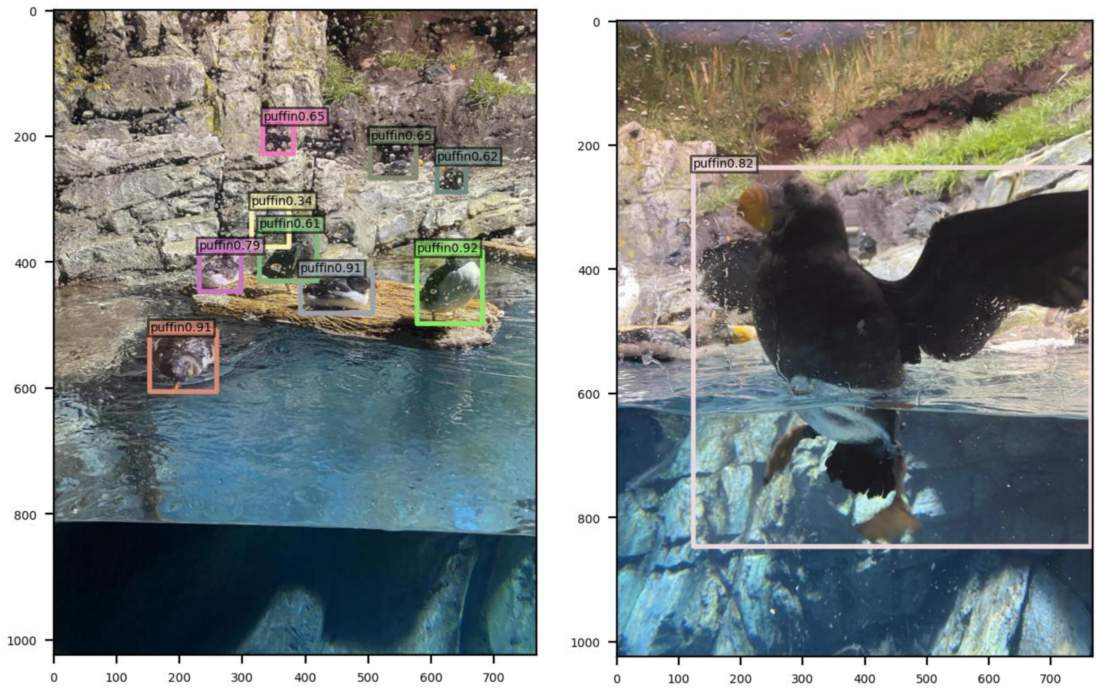
c. Penguin

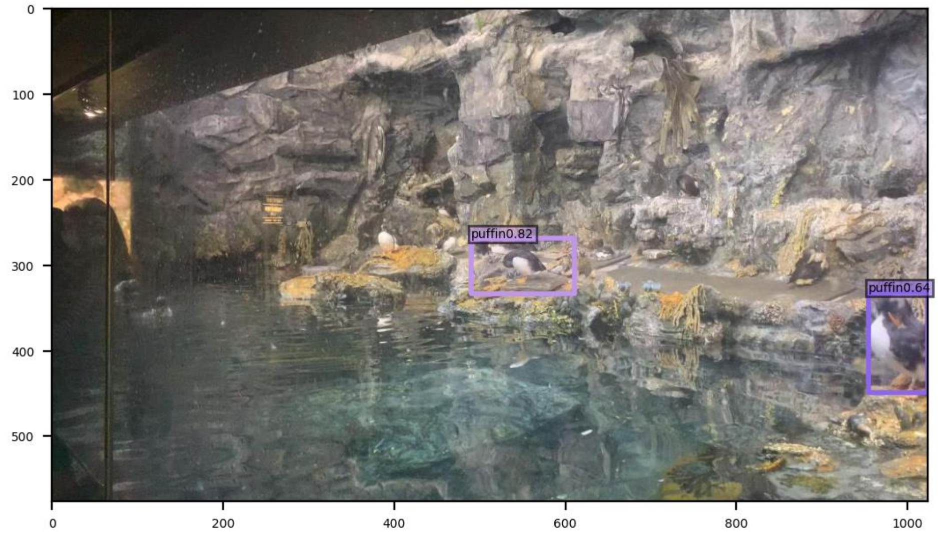


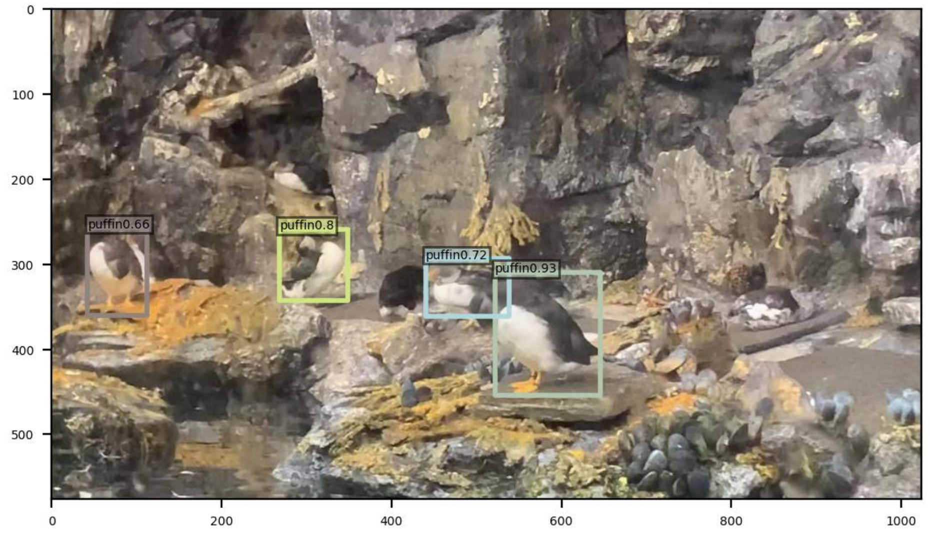
penguin0.91
penguin0.98
penguin0.98
penguin0.98
penguin0.35
penguin0.37
penguin0.69
penguin0.37
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

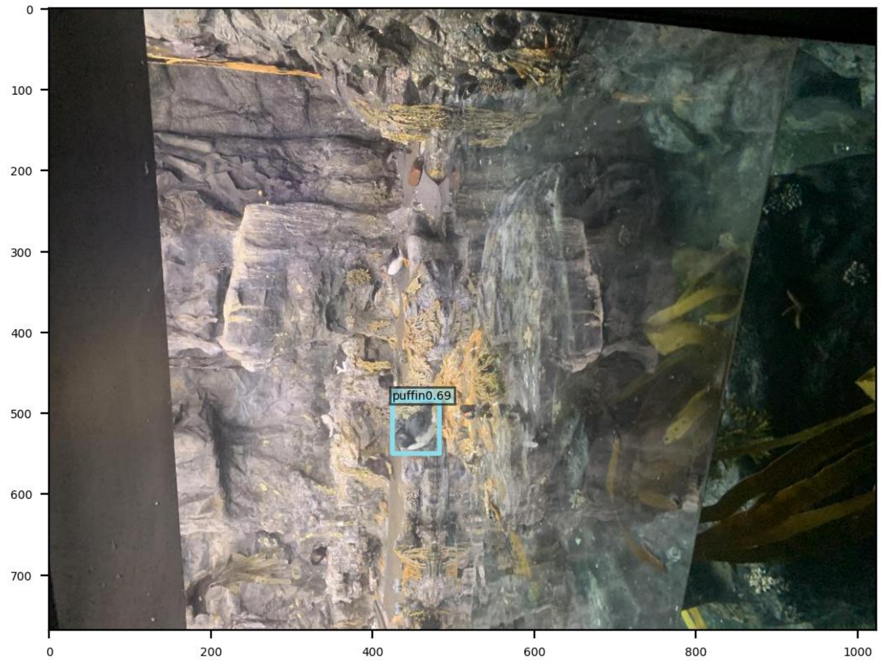

penguin0.98
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

d. Puffin

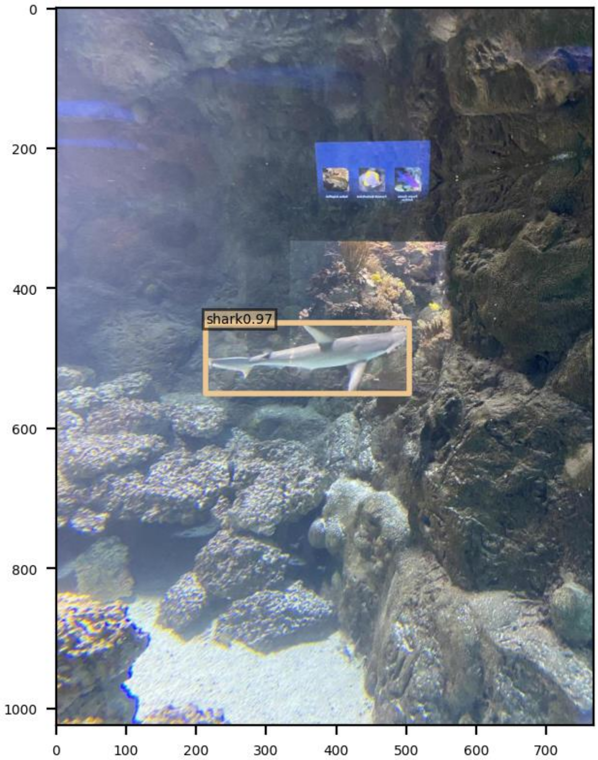

puffin0.65
puffin0.65
puffin0.62
puffin0.34
puffin0.61
puffin0.79
puffin0.91
puffin0.92
puffin0.91
puffin0.82
0
200
400
600
800
1000
0
100
200
300
400
500
600
700


puffin0.82
puffin0.64
0
100
200
300
400
500
0
200
400
600
800
1000

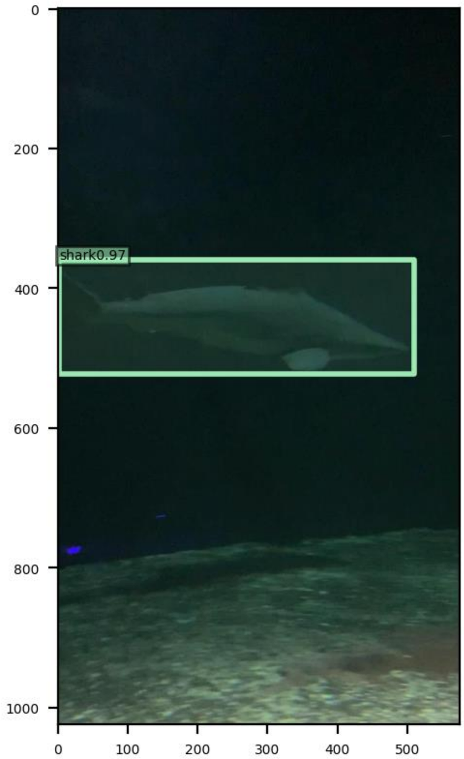

puffin0.66
puffin0.8
puffin0.72
puffin0.93
0
100
200
300
400
500
0
200
400
600
800
1000


puffin0.69
0
100
200
300
400
500
600
700
0
200
400
600
800
1000

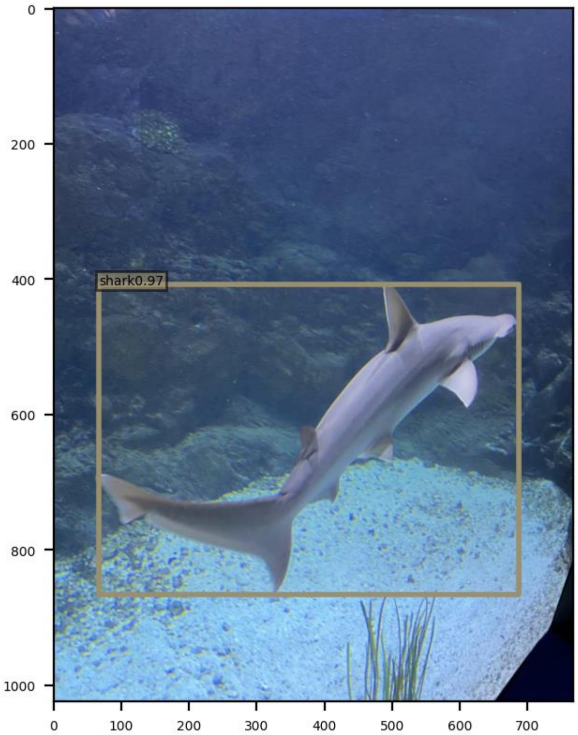
e. Shark



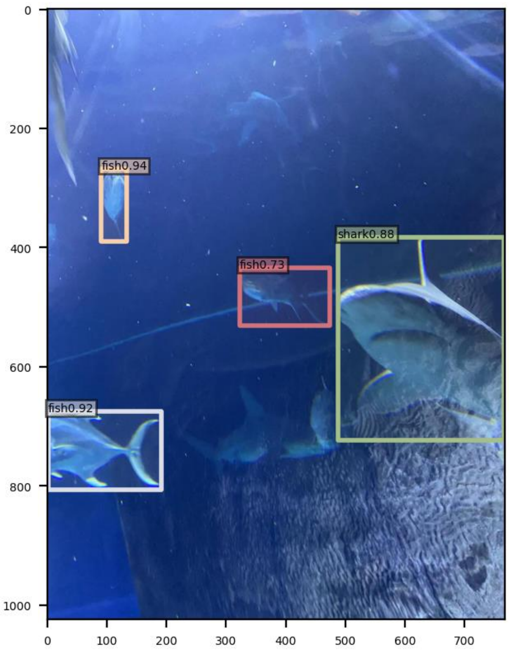
shark0.97
0
200
400
600
800
1000
0
100
200
300
400
500
600
700



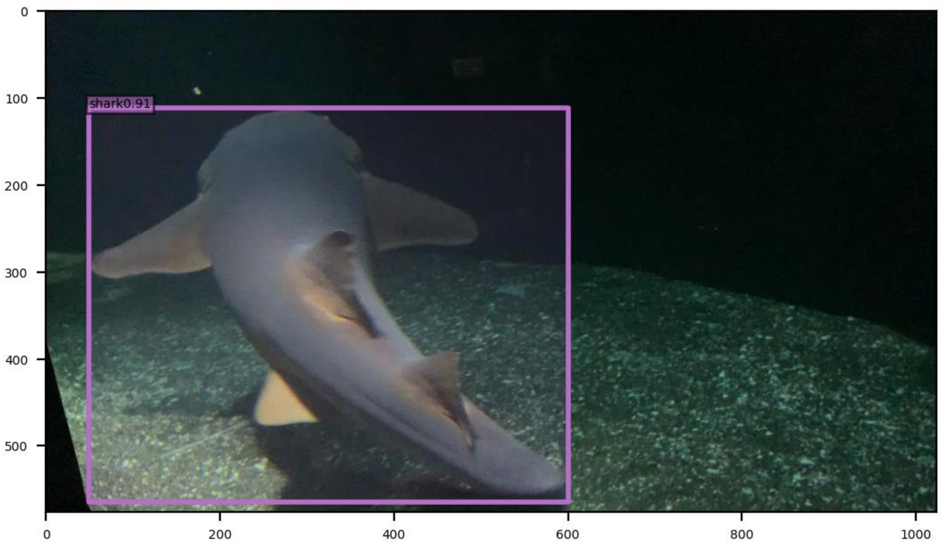
shark0.97
0
200
400
600
800
1000
0
100
200
300
400
500



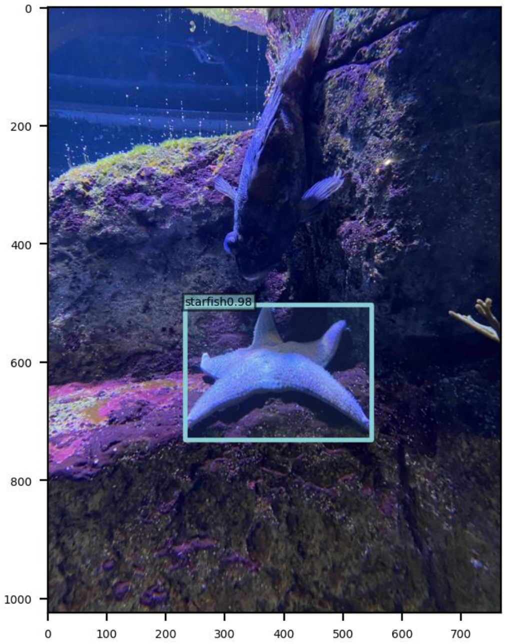
shark0.97
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

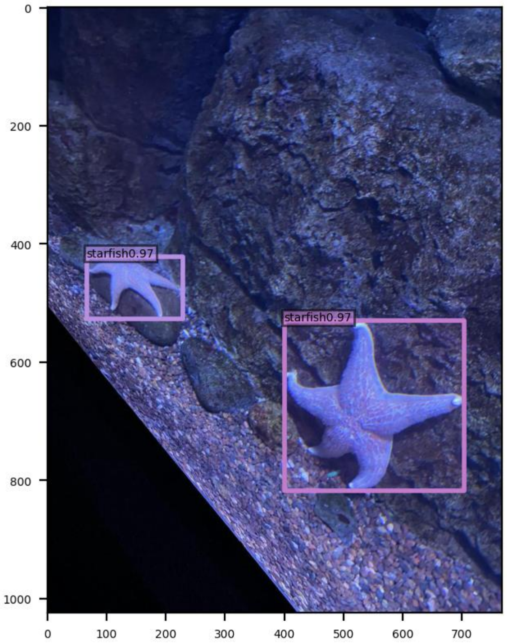


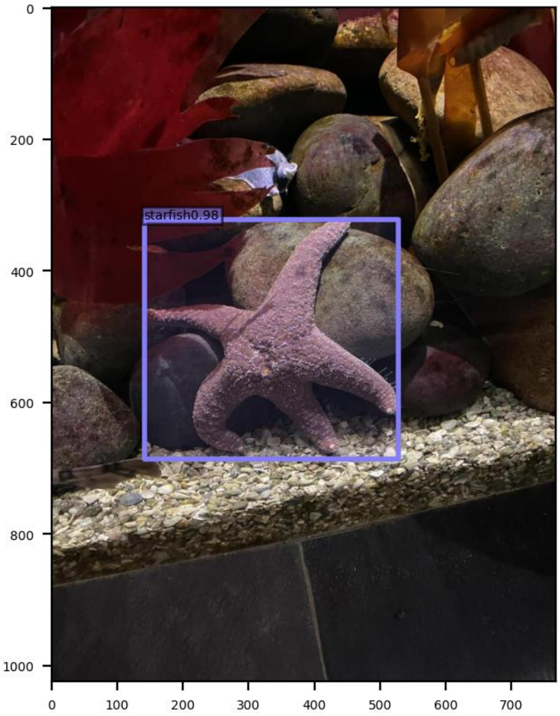
fish0.94
fish0.73
shark0.88
fish0.92
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

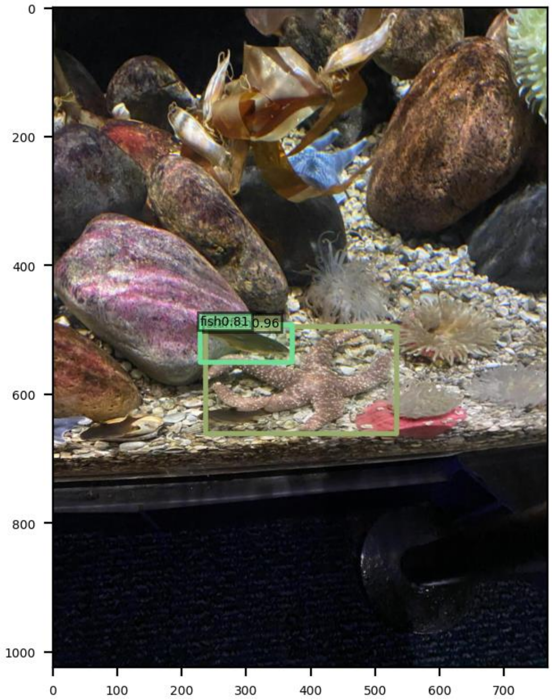

shark0.91
0
100
200
300
400
500
0
200
400
600
800
1000

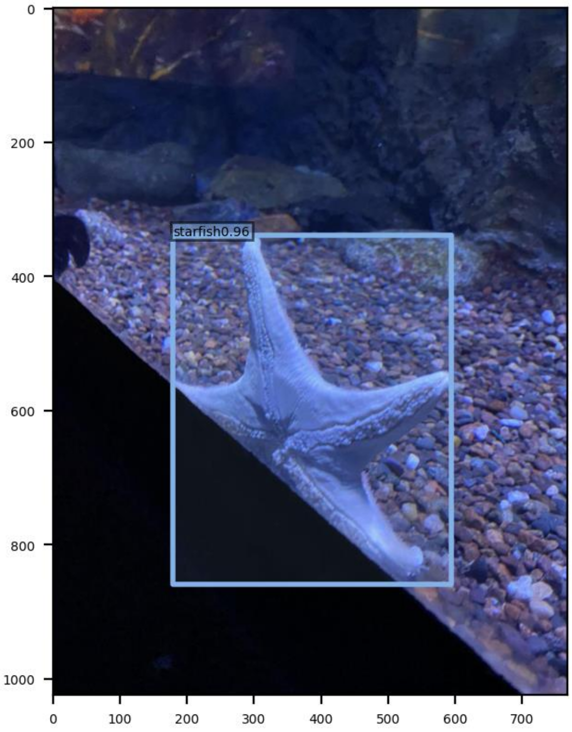
f. Starfish


starfish0.98
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

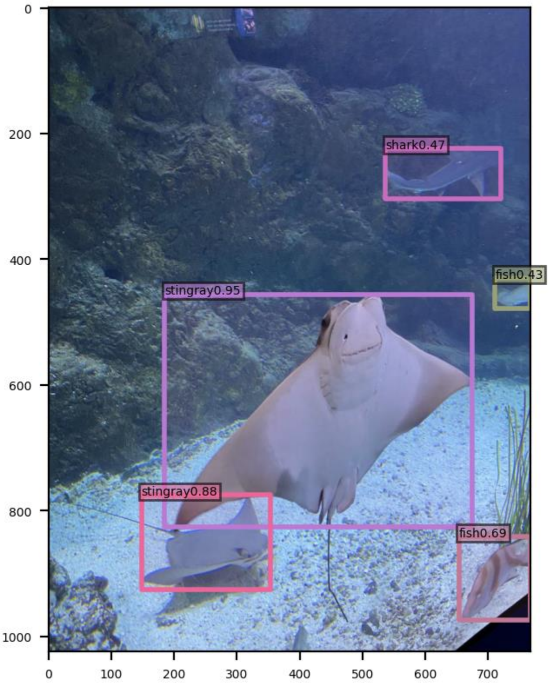

starfish0.97
starfish0.97
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

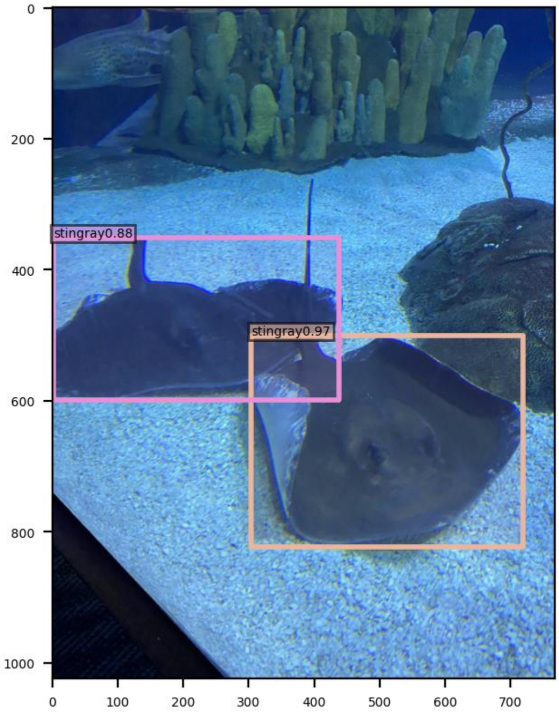

starfish0.98
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

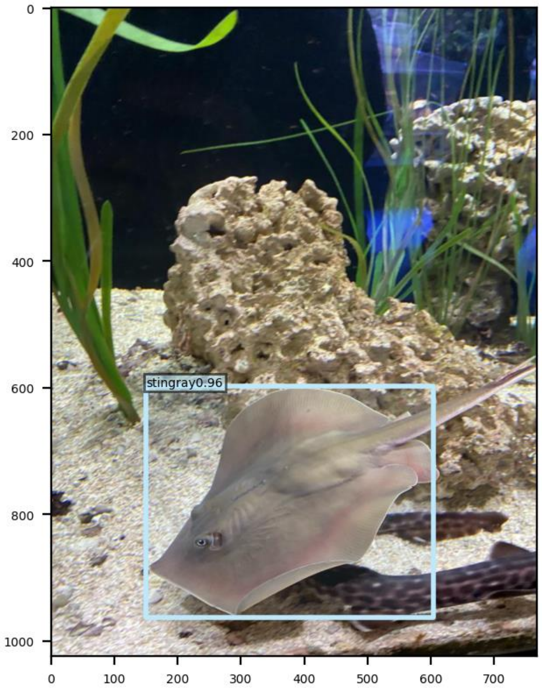

fish0.81
0.96
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

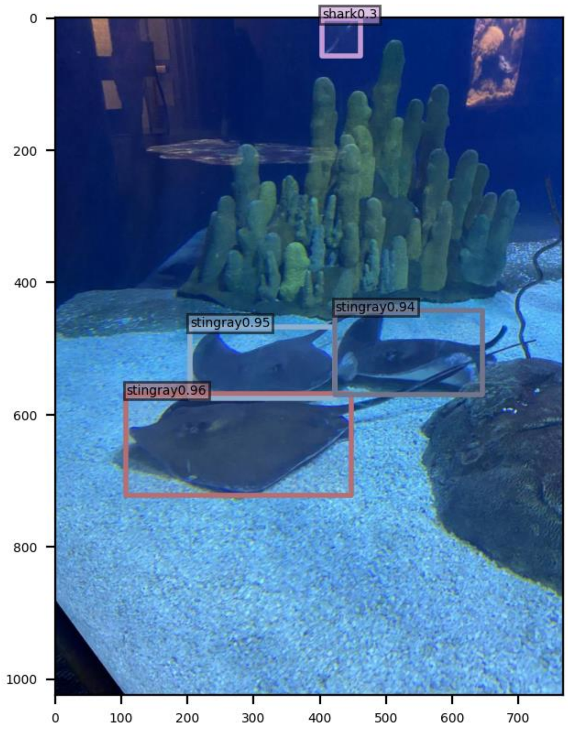

starfish0.96
0
200
400
600
800
1000
0
100
200
300
400
500
600
700

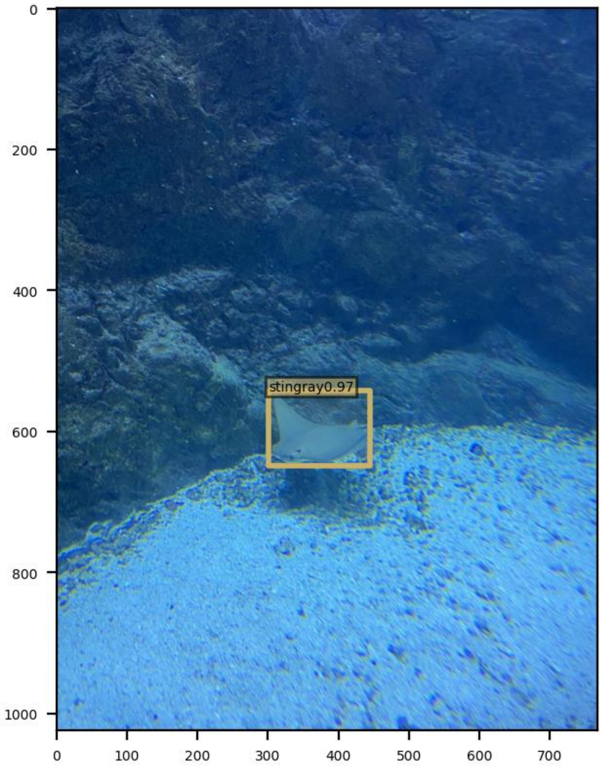
g. Stingray


shark0.47
fish0.43
stingray0.95
stingray0.88
fish0.69
0
200
400
600
800
1000
0
100
200
300
400
500
600
700


stingray0.88
stingray0.97
0
200
400
600
800
1000
0
100
200
300
400
500
600
700


stingray0.96
0
200
400
600
800
1000
0
100
200
300
400
500
600
700


shark0.3
stingray0.95
stingray0.94
stingray0.96
0
200
400
600
800
1000
0
100
200
300
400
500
600
700


stingray0.97
0
200
400
600
800
1000
0
100
200
300
400
500
600
700