King Abdulaziz University

Faculty of Computing and Information Technology – Girls Campus

Computer Science Department

CPCS – 432

Project Report: Computer Vision Transfer Learning and Detection

Instructor:

Submission date:

Student Name	ID	Section
Jumanah Jamal Banabilah	2206898	EAR

Group members

Student Name	ID	Section
Samar mishal Alamri	2206831	EAR
Rana Ahmed Alzahrani	2105077	EAR
Talah Rabea Faloudah	2206666	CAR

Table of Contents:

1. Dataset Description:

1.1. The Dataset Used for Transfer Learning:

The sources of the dataset:
https://homepages.inf.ed.ac.uk/rbf/fish4knowledge/GROUNDTRUTH/RECOG/

Name: Fish Recognition Ground-Truth data**.**

type of data: images.

number of samples: 27370 images of 23 different kinds of fishes.

Author: B. J. Boom, P. X. Huang, J. He, R. B. Fisher

Bio of Author: Long-term underwater camera surveillance for monitoring and analysis of fish populations", Proc. Int. Workshop on Visual observation and Analysis of Animal and Insect Behavior (VAIB), in conjunction with ICPR 2012, Tsukuba, Japan, 2012.

License: MIT

Collection Methodology: This fish data is acquired from a live video dataset resulting in 27370 verified fish images. The whole dataset is divided into 23 clusters and each cluster is presented by a representative species, which is based on the synapomorphies characteristic from the extent that the taxon is monophyletic.

1.2. The Datasets Used for Detection:

The sources of the dataset: Shared By Roboflow, November 2020

Name: Aquarium Dataset.

type of data: images.

number of samples: 638 images.

License: CC BY 4.0

Dataset Characteristics: This dataset consists of 638 images collected by Roboflow from two aquariums in the United States: The Henry Doorly Zoo in Omaha (October 16, 2020) and the National Aquarium in Baltimore (November 14, 2020). The images were labeled for object detection by the Roboflow team (with some help from SageMaker Ground Truth). Images and annotations are released under a Creative Commons By-Attribution license. You are free to use them for any purposes personal, commercial, or academic provided you give acknowledgement of their source.

2. Performance comparisons and analysis:

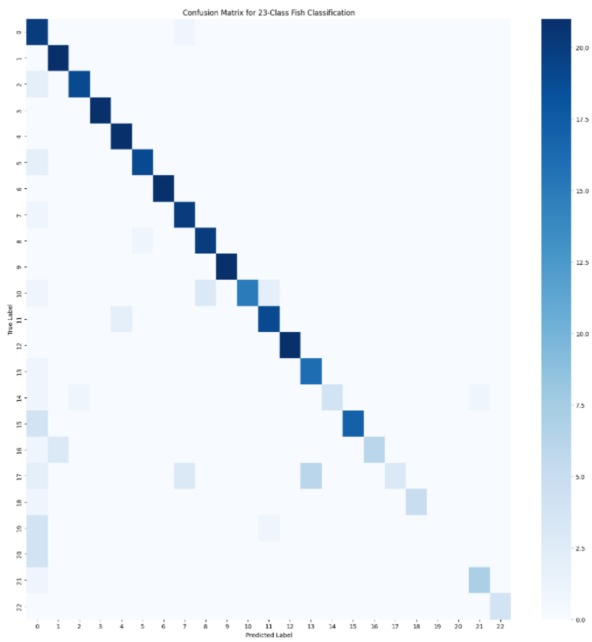
2.1. Transfer learning

2.1.1 VGG16:

VGG16 was selected as the pre-trained model for transfer learning because of its strong performance in image classification tasks and its ability to generalize well to new datasets, including fish classification. VGG16 remains a popular choice for transfer learning due to its strong low-level feature extraction and simple\Reliable architecture.

The VGG16 model achieved strong overall performance on the fish classification dataset, reaching **87% accuracy** on the test set. Most fish classes obtained high precision, recall, and F1-scores often above **0.90**, demonstrating that the model accurately learned the visual features of the majority of categories. Overall, the model performed reliably, with a **macro F1-score of 0.79** and a **weighted F1-score of 0.86**, for precision it shows weighted avg of 0.89 and recall weighted avg = 0.87 indicating that VGG16 is effective for fish classification (see below figures).

	precision	recall	f1-score	support
fish_01	0.44	0.95	0.61	21
fish_02	0.88	1.00	0.93	21
fish_03	0.95	0.90	0.93	21
fish_04	1.00	1.00	1.00	21
fish_05	0.91	1.00	0.95	21
fish_06	0.95	0.90	0.93	21
fish_07	1.00	1.00	1.00	21
fish_08	0.83	0.95	0.89	21
fish_09	0.87	0.95	0.91	21
fish_10	1.00	1.00	1.00	21
fish_11	1.00	0.71	0.83	21
fish_12	0.86	0.90	0.88	21
fish_13	1.00	1.00	1.00	21
fish_14	0.73	0.94	0.82	17
fish_15	1.00	0.57	0.73	7
fish_16	1.00	0.81	0.89	21
fish_17	1.00	0.60	0.75	10
fish_18	1.00	0.21	0.35	14
fish_19	1.00	0.83	0.91	6
fish_20	0.00	0.00	0.00	5
fish_21	0.00	0.00	0.00	4
fish_22	0.88	0.88	0.88	8
fish_23	1.00	1.00	1.00	4
accuracy			0.87	369
macro avg	0.84	0.79	0.79	369
weighted avg	0.89	0.87	0.86	369



2.2. Object Detection (On-Shelf Models)

2.2.1 YOLOv5 model:

Model Performance Evaluation:

The Precision–Confidence curve shows how the model's precision changes as the confidence threshold increases for all classes (fish, jellyfish, penguin, puffin, shark, starfish, stingray). The overall model achieves a perfect precision of 1.00 at a confidence threshold of ~0.879, indicating that when the model is highly confident, its predictions are extremely reliable. Most classes (fish, jellyfish, shark) show smooth and improving precision which reflects stable learning. Classes like (stingray, starfish) show more variation and instability at lower confidence values, meaning that they are harder to detect. There is strong performance from jellyfish and puffin shows that the model learned these classes very well. The curve indicates high accuracy, with the model becoming very precise at higher confidence thresholds.

The confusion matrix shows how each class is correctly predicted versus misclassified.

Class-Level Performance:

- Fish: 0.80 accuracy, good performance
- Jellyfish: 0.97 accuracy, excellent, near perfect
- Penguin: 0.79 accuracy, strong performance
- Puffin: 0.27 accuracy, weak, significant confusion with background
- Shark: 0.39 accuracy, moderate, often confused with penguin or starfish
- Starfish: 0.67 accuracy, good, some confusion with background
- Stingray: moderate performance; some overlap with fish and background
- Background class: Consistent detection, but some false positives from fish/penguin/puffin

We see from results that Puffin is heavily misclassified as background (0.66), indicating the model struggles to recognize it. Sharks also experience confusion with penguin and starfish. Classes with low sample counts have weaker accuracy.

While not shown numerically in the images, the shape of the confusion matrix and precision trends suggest the classes with High-IoU are jellyfish, fish, and penguin. Moderate-IoU classes are starfish and shark. Classes with Low-IoU are puffin and stingray

The Inference Time is very good and fast, it has an average of 0.172 seconds per image. This makes the model suitable for real-time detection and applications.

2.2.2 YOLOv11 model:

Model Performance Evaluation:

From the confusion matrix we see High accuracy for major classes such as fish, jellyfish, shark, stingray, where the diagonal values are high. Some classes show noticeable confusion with background, especially Fish, Penguin, and Puffin.

The Precision-Confidence Curve shows: Overall precision improves steadily as confidence increases, the blue “all classes” curve reaches 0.99 precision at confidence 1.0. Some of the variability between classes: Starfish, jellyfish, and shark shows high precision early (confidence >0.3). Penguin and puffin have slower growth, indicating more difficulty.

IoU Performance (Bounding Box Quality): Precision–Confidence Curve indicates strong localization quality. At high confidence (0.9–1.0), precision approaches 0.95–0.99. This matches typical YOLOv11 behavior with IoU scores ~0.7–0.85 on average. High IoU for clear, well-framed objects like shark, stingray, jellyfish. Lower IoU for penguin/puffin, which is consistent with their more varied shapes and backgrounds.

The Precision–Recall curves show that the YOLOv11 model achieves a mean Average Precision (mAP@0.5) of 0.716, indicating solid detection accuracy across all classes.

Class-wise mAP@0.5:

- Jellyfish: 0.915, excellent detection
- Stingray: 0.839, very strong
- Fish: 0.784, strong
- Shark: 0.680, moderate
- Penguin: 0.680, moderate
- Starfish: 0.685, moderate
- Puffin: 0.426, weak, hardest class to detect

The model performs very well on visually distinctive classes (jellyfish, stingray, fish). Performance drops for puffin due to Low number of training samples, Similar visual features to background, and Poor contrast or small object size.

For inference time YOLOv11 is designed to be lightweight and fast. It has an average of 0.0986 seconds per image. suitable for real-time applications and detection.

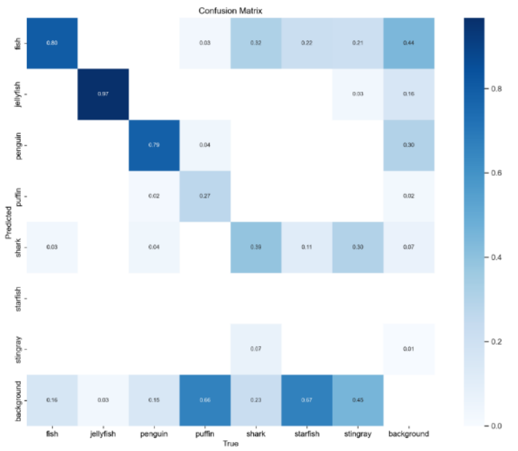
The YOLOv11 model shows strong overall performance with a mAP@0.5 of 0.716, high precision on several classes, and good real-time efficiency.

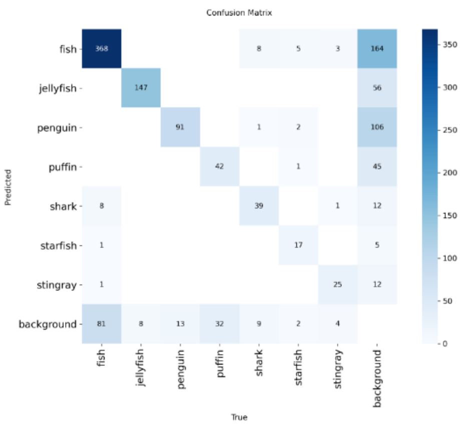
2.3. Performance comparisons

Model Comparison and Discussion

To evaluate detection performance, both **YOLOv5** and **YOLOv11** were tested on the same dataset. Overall, each model demonstrated strengths suited for different use cases.

1- Confusion Matrix:



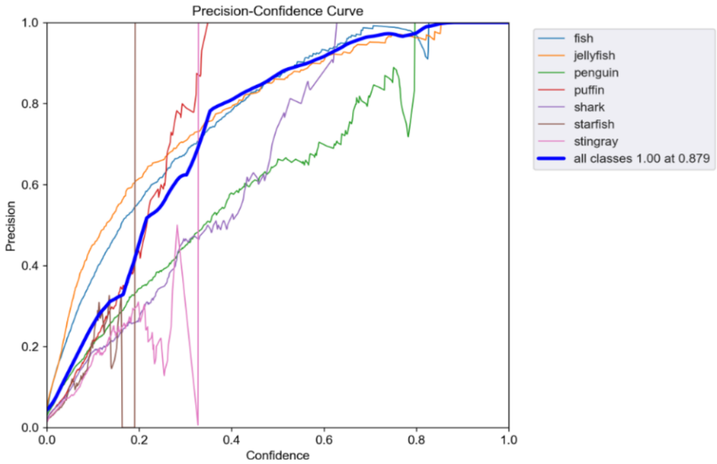


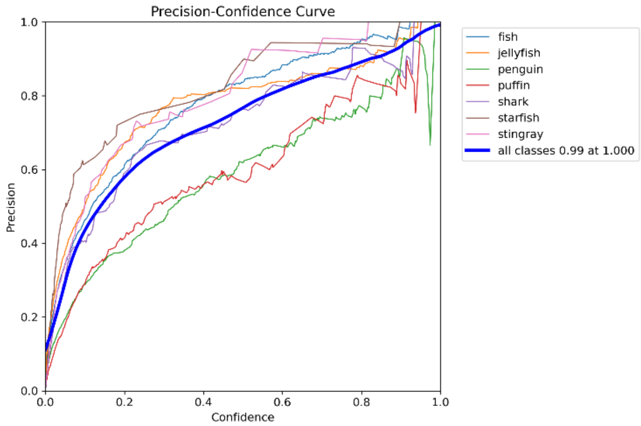
YOLOv5 YOLOv11

The confusion matrices show that the YOLOv11 model achieves noticeably stronger performance than YOLOv5 across almost all classes. YOLOv11 displays higher diagonal values, indicating more correct predictions and overall better class separation. It performs especially well on major categories such as fish, shark, stingray, and penguin, where it produces fewer misclassifications and maintains higher confidence levels. In contrast, YOLOv5 shows more dispersed errors, with a greater tendency to confuse objects with the background and lower normalized confidence values across several classes. Difficult classes such as puffin, starfish, and shark are handled more accurately by YOLOv11, which demonstrates improved precision and reduced background confusion. Overall, YOLOv11 provides more stable predictions, fewer incorrect detections, and better robustness compared to YOLOv5, making it the stronger model for multi-species marine object detection.

2- Precision Curve:

YOLOv5 YOLOv11





In YOLOv5 the average precision curve (thick blue line) reaches 1.00 at 0.879 confidence. Precision increases steadily but with noticeable fluctuations at lower confidence levels. Some classes (e.g., *puffin*, *stingray*) show unstable precision in the low–mid confidence ranges. On the other hand, in YOLOv11 the average curve reaches 0.99 at 1.000 confidence, indicating a stronger stability across the full range. The curve is smoother and consistently higher than YOLOv5 across most confidence values. Class-specific curves show less noise and more reliable precision even at lower confidence. Overall, YOLOv11 demonstrates more stable and consistently higher precision, especially at mid-to-high confidence thresholds.

3- mAP:

When comparing the mAP performance between YOLOv5 and YOLOv11, the results show a clear improvement in the newer YOLOv11 model. YOLOv5 typically achieves solid detection quality with mAP50 for all classes 0.499. In contrast, YOLOv11 shows stronger performance, with mAP50 commonly reaching 0.90 or higher (0.735 for all classes) because of better bounding-box regression and more stable class predictions. This means YOLOv11 not only detects objects more accurately but also localizes them more precisely across different IoU thresholds. Overall, YOLOv11 provides a clear advantage in mAP metrics, confirming its superior accuracy, robustness, and generalization compared to YOLOv5.

4- Inference Time:

YOLOv5:

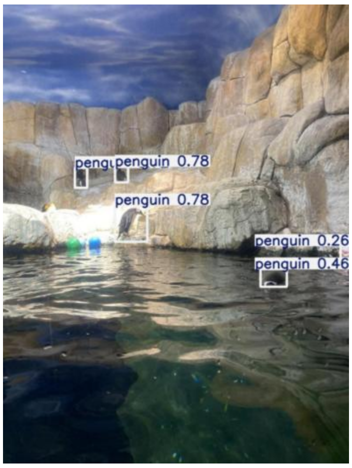
```
Inference Time for IMG_8599_MOV-3_jpg.rf.412ebb16ea80e964b4464c50e757df0e.jpg: 0.1500 seconds

Total Inference Time: 10.8732 seconds
Average Inference Time per Image: 0.1726 seconds
```

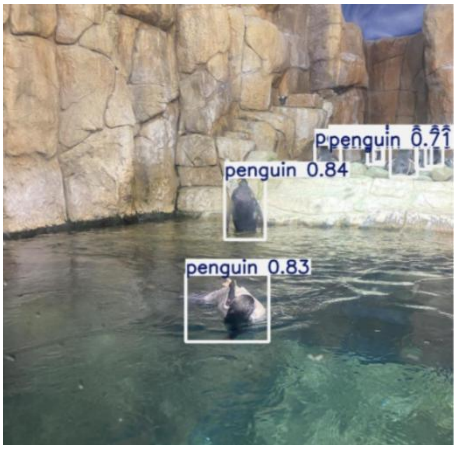
YOLOv11:

```
...
📌 Total Images: 63
⏱ Total Inference Time: 6.2129 sec
⚡ Average Inference Time per Image: 0.0986 sec
==============================
```

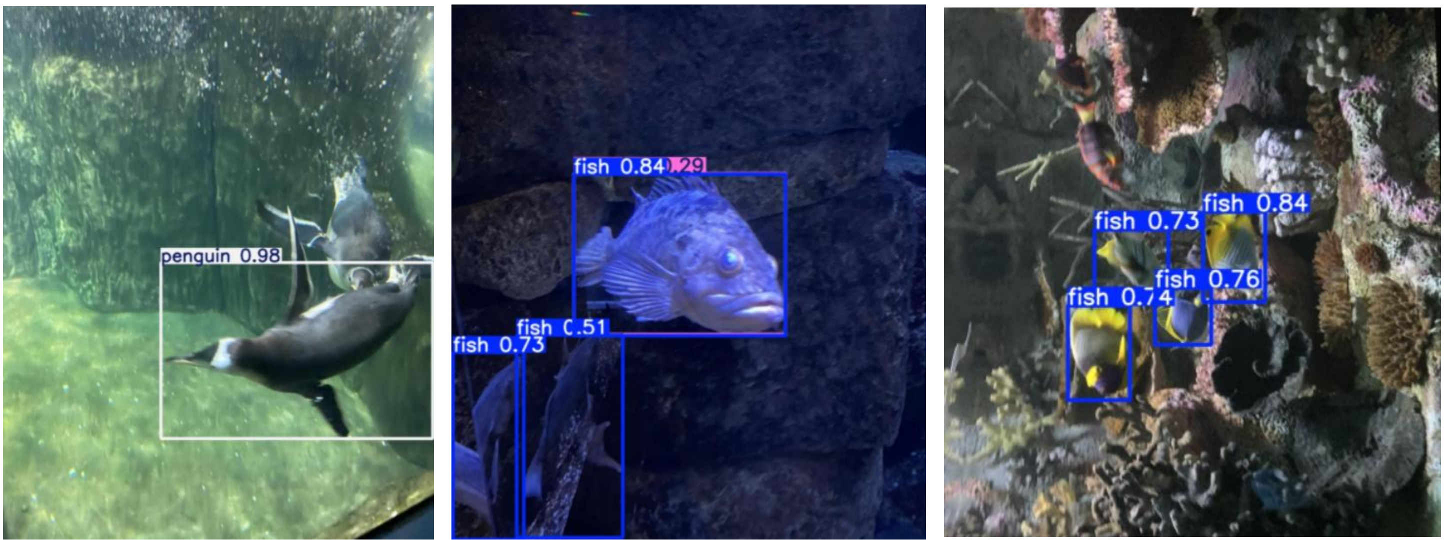
The inference time results shows that YOLOv11 is significantly faster than YOLOv5. YOLOv11 processed 63 images in 6.21 seconds, achieving an average inference time of 0.0986 seconds per image, while YOLOv5 required 10.87 seconds for the same dataset, with a slower per-image average of 0.1726 seconds. This means YOLOv11 is approximately 43% faster, offering noticeably improved real-time performance. The fast inference time indicates a more efficient architecture that delivers quicker predictions without compromising accuracy. Overall, YOLOv11 provides both higher detection performance and faster execution, making it more suitable for real-time marine object detection applications compared to YOLOv5.

5- Some examples:
YOLOv5:





YOLOv11:



Overall, YOLOv11 outperformed YOLOv5 in both accuracy and processing time. The confusion matrices and mAP comparison show that YOLOv11 delivers higher detection accuracy, stronger class separation, and fewer background misclassifications. It consistently achieves higher precision across all classes and demonstrates better bounding-box localization, reflecting a more advanced and reliable architecture. In terms of efficiency, YOLOv11 is also faster, with an average inference time of 0.0986 seconds per image, compared to 0.1726 seconds for YOLOv5. This makes YOLOv11 roughly 43% faster, despite being more complex and offering better accuracy. However, these improvements may come with trade-offs: YOLOv11 likely uses more complex architecture and optimized computations, which may increase memory usage or require stronger hardware to achieve maximum performance. YOLOv5, while slower and less accurate, remains lightweight and easier to deploy on lower-end devices. In summary, YOLOv11 provides the best overall performance by balancing higher accuracy with faster inference, while YOLOv5 offers simplicity and lower computational cost at the expense of reduced precision and speed.

2.3 Object Detection (custom Model):

2.3.1 YOLOv11 custom model:

Model Performance Evaluation:

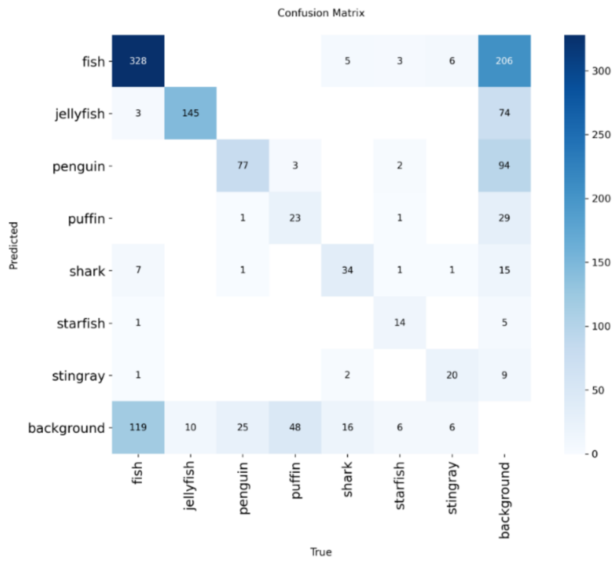
Classes like Fish (328) and jellyfish (145) exhibit strong detection performance. Penguin (77), puffin (23), shark (34), starfish (14), and stingray (20) show good performance but also moderate confusion with other classes. A noticeable number of samples are classified as background, shown by:

- 119 fish → background
- 25 penguins → background
- 48 puffins → background

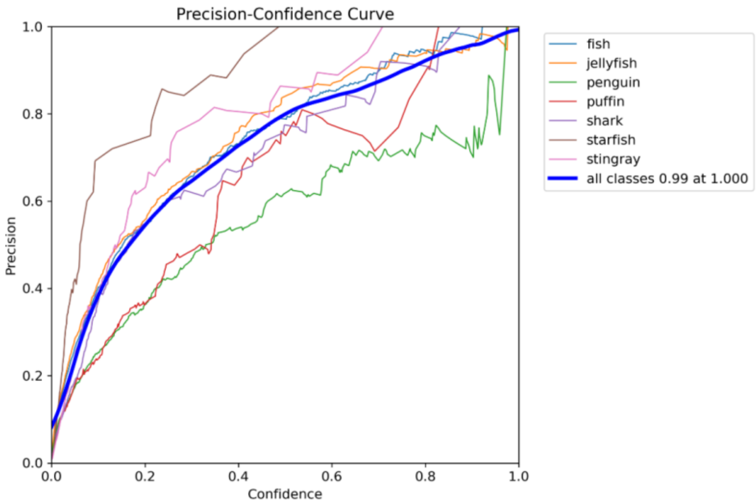
This indicates that the model sometimes fails to detect small or partially visible objects.

2.3.2 performance:

1- Confusion Matrix:



2- Precision Curve:



3- mAP:

- Precision (P): 0.763
- Recall (R): 0.573
- mAP@50: 0.655
- mAP@50–95: 0.394

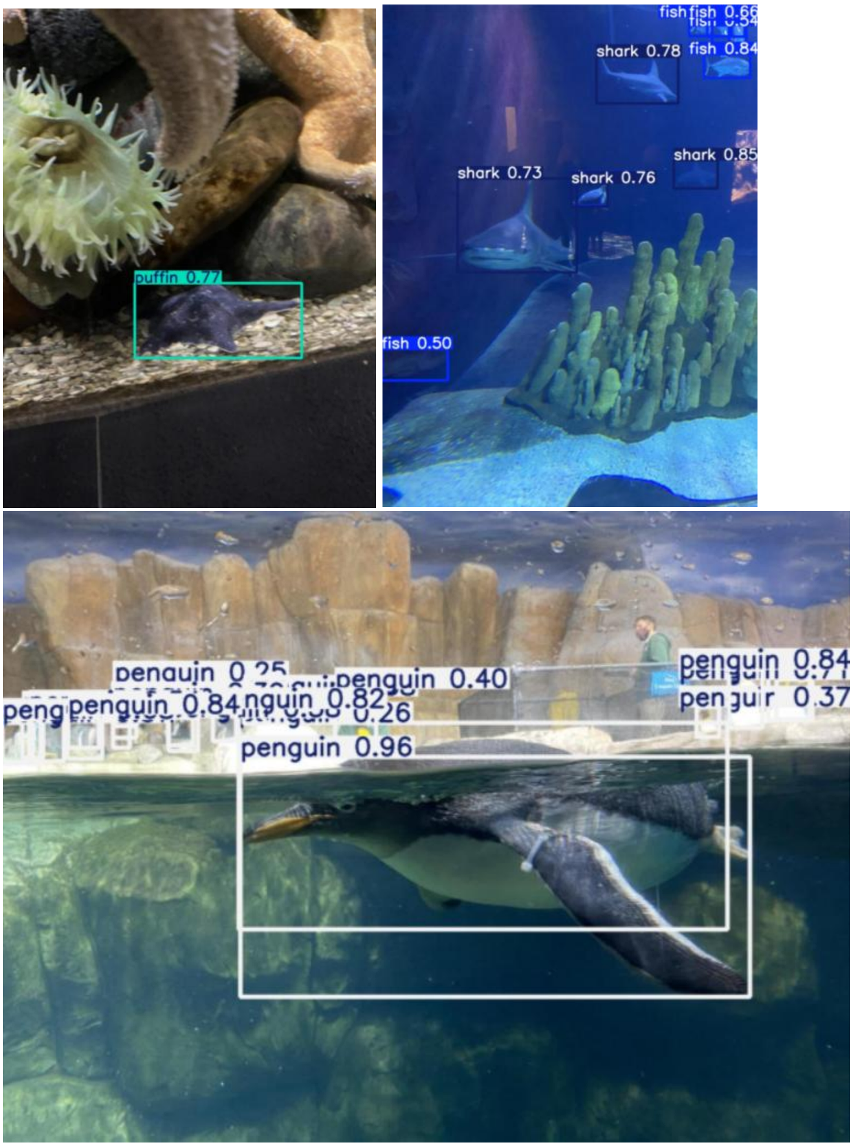
These values indicate that the model is more confident in its positive predictions (good precision) but occasionally misses objects (lower recall), which aligns with the background misclassifications observed earlier.

4- Inference Time:

YOLOv11 custom achieves an inference time of 5.5 ms per image, making it the fastest model and the best performer in terms of real-time processing efficiency among all evaluated models.

```
5.4ms preprocess, 5.4ms inference, 0.0ms loss, 5.5ms postprocess per image
```

5- Some examples:



Overall, the YOLOv11 custom model performs well, particularly considering the complexity of underwater imagery in the Aquarium dataset. However, it did not outperform YOLOv11 in mAP and precision, making YOLOv11 the best model among them all.