UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Monitoring Tunas and Sharks Using YOLO Models in the Galápagos Islands

Proyecto de Titulación

Diego Santiago Morales Arcos

Noel Pérez Pérez, Ph.D.

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magíster en Inteligencia Artificial

Quito, 18 de diciembre de 2024

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN Monitoring Tunas and Sharks Using YOLO Models in the Galápagos Islands

Diego Santiago Morales Arcos

Nombre del Director del Programa:	Felipe Grijalva
Título académico:	Ph.D. en Ingeniería Eléctrica
Director del programa de:	Inteligencia Artificial
Nombre del Decano del colegio Académico:	Eduardo Alba
Título académico:	Doctor en Ciencias Matemáticas
Decano del Colegio:	Ciencias e Ingenierías
Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Doctor en Física

Quito, diciembre 2024

© DERECHOS DE AUTOR

Nombre del estudiante:	Diego Santiago Morales Arcos
Código de estudiante:	00339647
C.I.:	2100614383
Lugar y fecha:	Quito, 18 de diciembre de 2024.

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

Los tiburones y las atunes desempeñan un papel fundamental en los ecosistemas marinos, pero sus poblaciones están disminuyendo debido a la sobrepesca y la pérdida de hábitat. Se necesitan urgentemente métodos de monitoreo precisos y no invasivos para guiar estrategias de conservación efectivas. En este estudio, proponemos un sistema de detección automatizada basado en YOLO diseñado para identificar con precisión tiburones (específicamente tiburones sedosos y tigres) y atunes en videos submarinos grabados en las Islas Galápagos. Nuestro conjunto de datos de entrenamiento se construyó a partir de dos clips de video de un minuto, uno centrado en tiburones sedosos y el otro en tiburones tigre y atunes, obteniendo 229 imágenes anotadas. Usamos el 90% de estas imágenes para el entrenamiento y el 10% para la prueba, aplicando un procedimiento de validación cruzada de 5 pliegues. Cada modelo fue entrenado durante 30 épocas, y se evaluaron varias arquitecturas de YOLO (por ejemplo, YOLOv8 Medium y YOLOv9 Medium) basándose en la precisión promedio (mAP@50) y la velocidad de inferencia. Entre las configuraciones probadas, YOLOv9 Medium logró la mayor mAP@50 (95.83%), mientras que YOLOv8 Medium proporcionó un buen equilibrio entre precisión y eficiencia computacional, alcanzando un mAP@50 de 94.20%. Al ajustar la tasa de procesamiento de fotogramas (por ejemplo, de 20 fotogramas por segundo a 1 fotograma por segundo), el sistema se puede optimizar para monitoreo en tiempo real o casi en tiempo real. Para evitar la contaminación de datos, el entrenamiento y la evaluación se realizaron en clips de video distintos. Nuestros resultados indican que los marcos de detección basados en YOLO pueden facilitar un monitoreo eficiente y confiable de tiburones y atunes, proporcionando una herramienta poderosa para esfuerzos de conservación informados y la gestión sostenible de áreas marinas protegidas.

Palabras clave: YOLO, Visión por Computadora, Tiburones, Atunes, Detección de Especies, Ecosistemas Marinos, Detección de Objetos.

ABSTRACT

Sharks and tunas play a pivotal role in marine ecosystems, yet their populations are declining due to overfishing and habitat loss. Accurate, non-invasive monitoring methods are urgently needed to guide effective conservation strategies. In this study, we propose a YOLO-based automated detection system designed to accurately identify sharks (specifically silky and tiger sharks) and tunas in underwater videos recorded in the Galápagos Islands. Our training dataset was constructed from two one-minute video clips—one focusing on silky sharks and the other on tiger sharks and tunas—yielding 229 annotated images. We used 90% of these images for training and 10% for testing, applying a 5-fold cross-validation procedure. Each model was trained for 30 epochs, and multiple YOLO architectures (e.g., YOLOv8 Medium and YOLOv9 Medium) were evaluated based on mean Average Precision (mAP@50) and inference speed. Among the tested configurations, YOLOv9 Medium achieved the highest mAP@50 (95.83%), while YOLOv8 Medium provided a strong balance between accuracy and computational efficiency, attaining a mAP@50 of 94.20%. By adjusting the frame processing rate (e.g., from 20 frames per second to 1 frame per second), the system can be optimized for real-time or near real-time monitoring. To avoid data contamination, training and evaluation were conducted on distinct video clips. Our results indicate that YOLO-based detection frameworks can facilitate efficient, reliable monitoring of sharks and tunas, providing a powerful tool for informed conservation efforts and sustainable management of marine protected areas.

Key words: YOLO, Computer Vision, Sharks, Tunas, Species Detection, Marine Ecosystems, Object Detection.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

Monitoring Tunas and Sharks Using YOLO Models in the Galápagos Islands

Diego Morales, Noel Pérez Pérez Colegio de Ciencias e Ingenierías "El Politécnico",
Universidad San Francisco de Quito USFQ,
Campus Cumbayá, Casilla Postal 17-1200-841, Quito, Ecuador
Email: dsmorales@estud.usfq.edu.ec, nperez@usfq.edu.ec

***Abstract*—Sharks and tunas play a pivotal role in marine ecosystems, yet their populations are declining due to overfishing and habitat loss. Accurate, non-invasive monitoring methods are urgently needed to guide effective conservation strategies. In this study, we propose a YOLO-based automated detection system designed to accurately identify sharks (specifically silky and tiger sharks) and tunas in underwater videos recorded in the Galápagos Islands. Our training dataset was constructed from two one-minute video clips—one focusing on silky sharks and the other on tiger sharks and tunas—yielding 229 annotated images. We used 90% of these images for training and 10% for testing, applying a 5-fold cross-validation procedure. Each model was trained for 30 epochs, and multiple YOLO architectures (e.g., YOLOv8 Medium and YOLOv9 Medium) were evaluated based on mean Average Precision (mAP@50) and inference speed. Among the tested configurations, YOLOv9 Medium achieved the highest mAP@50 (95.83%), while YOLOv8 Medium provided a strong balance between accuracy and computational efficiency, attaining a mAP@50 of 94.20%. By adjusting the frame processing rate (e.g., from 20 frames per second to 1 frame per second), the system can be optimized for real-time or near real-time monitoring. To avoid data contamination, training and evaluation were conducted on distinct video clips. Our results indicate that YOLO-based detection frameworks can facilitate efficient, reliable monitoring of sharks and tunas, providing a powerful tool for informed conservation efforts and sustainable management of marine protected areas.**

***Index Terms*—YOLO, Computer Vision, Sharks, Tunas, Species Detection, Marine Ecosystems, Object Detection.**

I. Introduction

Object detection and tracking have become essential tools in various real-world scenarios such as surveillance [4], assistive technologies, microscopy, and notably, marine species monitoring [1]. In the marine environment, apex predators like sharks and economically valuable species such as tunas play a vital role in maintaining the balance and health of marine ecosystems. Their population dynamics influence prey communities and overall biodiversity. However, sharks and tunas are under increasing pressure from overfishing and habitat loss, leading to alarming declines in their populations [2].

Silky sharks (*Carcharhinus falciformis*) and tiger sharks (*Galeocerdo cuvier*) are particularly susceptible to these threats. Both species inhabit the waters around the Galápagos Islands—a UNESCO World Heritage site—and contribute significantly to the ecological balance of this marine ecosystem [5]. Despite the ecological importance and conservation status of these species, their monitoring has traditionally relied on labor-intensive and time-consuming methods such as manual counting and tagging. Such approaches are not only costly and prone to human error but may also disturb the animals and potentially alter their natural behaviors [4].

To address these limitations, recent advances in computer vision and deep learning have enabled the development of automated detection and tracking systems capable of analyzing large volumes of underwater imagery. Among these, YOLO (You Only Look Once) models have emerged as a popular choice due to their real-time inference speeds and high detection accuracy. YOLO-based frameworks have been successfully applied to detect marine fauna, offering a promising avenue for non-invasive, scalable, and efficient data collection [6].

In this study, we focus on the application of advanced YOLO models (YOLOv8 and YOLOv9) for detecting silky sharks, tiger sharks, and tunas in underwater video footage from the Galápagos Islands. By leveraging a dataset derived from carefully curated video clips, we implement a 5-fold cross-validation protocol and train for multiple epochs to ensure robust model performance. Our goal is to identify a YOLO-based detection system

that not only achieves high accuracy (measured by mean Average Precision, mAP) but can also operate at inference speeds that support real-time or near real-time monitoring. This approach facilitates the large-scale, continuous monitoring of shark and tuna populations, providing vital information that can guide conservation strategies, support sustainable fisheries management, and ultimately help maintain the ecological integrity of marine protected areas.

II. Materials and Methods

A. YOLO-based Detection Methods

YOLO (You Only Look Once) is a single-stage object detection framework that jointly learns object localization and classification. Instead of employing separate region proposal and classification steps, YOLO divides the input image into a grid and directly predicts bounding boxes, objectness scores, and class probabilities. This approach enables real-time inference speeds, which is advantageous for continuous underwater video analysis, where large volumes of data must be processed efficiently.

B. YOLOv8 and YOLOv9 Architectures and Configurations

Recent YOLO variants, such as YOLOv8 and YOLOv9, incorporate architectural enhancements aimed at improving detection accuracy and robustness in challenging conditions. Key innovations include Cross Stage Partial Networks (CSPNet) for efficient gradient propagation, Feature Pyramid Networks (FPN), and Path Aggregation Networks (PANet) for effective multi-scale feature fusion, as well as advanced activation functions like Mish. These enhancements provide improved sensitivity to small, partially occluded objects, which is particularly advantageous for detecting marine species in underwater environments.

In this study, we evaluate several configurations of YOLOv8 and YOLOv9 models. Each configuration differs in terms of model capacity and complexity, reflecting a trade-off between accuracy and inference speed. Smaller variants (e.g., Nano, Tiny) are optimized for faster inference and reduced computational overhead, making them suitable for real-time applications on constrained hardware. Larger variants (e.g., Medium, Large, X-Large) have increased parameter counts and deeper architectures, often yielding higher accuracy but requiring more computational resources.

Table I provides an overview of the specific YOLOv8 and YOLOv9 model configurations used in this paper, along with their approximate number of parameters. These parameter counts serve as a guideline for understanding the resource requirements and potential performance differences between models.

Table I
YOLOv8 and YOLOv9 Model Configurations and Approximate Number of Parameters

Model	Configuration	Approx. #Parameters
YOLOv8-n	Nano	$\sim$ 3.2M
YOLOv8-s	Small	$\sim$ 11.2M
YOLOv8-m	Medium	$\sim$ 25.9M
YOLOv8-l	Large	$\sim$ 46.4M
YOLOv8-x	X-Large	$\sim$ 68.2M
YOLOv9-t	Tiny	$\sim$ 3.3M
YOLOv9-s	Small	$\sim$ 12.0M
YOLOv9-m	Medium	$\sim$ 27.5M
YOLOv9-c	Compact	$\sim$ 24.4M
YOLOv9-e	Efficient	$\sim$ 20.0M

C. Non-Maximum Suppression (NMS)

Multiple bounding boxes often overlap around the same object. To refine these raw predictions, Non-Maximum Suppression (NMS) is applied. NMS selects the bounding box with the highest confidence score for each detected object and discards overlapping, redundant boxes based on Intersection-over-Union (IoU) thresholds. This process yields cleaner and more accurate final detections.

D. Proposed Method

The proposed method leverages YOLOv8 and YOLOv9 architectures to detect silky sharks (*Carcharhinus falciformis*), tiger sharks (*Galeocerdo cuvier*), and tunas in underwater video footage from the Galápagos Islands. The pipeline comprises:

label=0)

1) **Input Preprocessing:** Underwater video clips are sampled at a chosen frame rate to generate individual frames for analysis.
2) **Feature Extraction and Detection:** The selected YOLO models (YOLOv8 or YOLOv9) receive frames as input, extracting

features across multiple scales and predicting bounding boxes and class probabilities.

3) **Post-processing:** NMS is applied to remove duplicate detections and yield final bounding boxes for each identified shark or tuna.

E. Dataset and Annotation

The training dataset was derived from two one-minute underwater video clips recorded in the Galápagos Islands. One video primarily featured silky sharks, while the other contained tiger sharks and tunas. From these videos, a total of 229 frames were extracted. Each frame was annotated using RoboFlow, assigning bounding boxes and class labels (silky shark, tiger shark, tuna) to all visible targets.

To ensure a robust evaluation and prevent data contamination, no frames from the test videos were included in the training process. Specifically, 90% of the 229 annotated images were used for training, while the remaining 10% constituted the test set. This split ensured that the training and testing processes were isolated, allowing reliable assessment of model generalization.

F. Training Protocol and Experimental Setup

A 5-fold cross-validation strategy was implemented to enhance model reliability and reduce overfitting. The training dataset was partitioned into five folds, with four folds used for training and one for validation in each iteration. This process was repeated such that each fold served as a validation set once.

Each model (YOLOv8 and YOLOv9 configurations) was trained for 30 epochs. The Adam optimizer with a cosine annealing scheduler was employed, and weight decay regularization was applied to promote stable convergence and prevent overfitting. After training, the best model weights were selected based on validation metrics, such as mean Average Precision (mAP) at 0.50 IoU threshold (mAP@50).

Two distinct sets of videos (10-second and 15-second clips), not used in training, were reserved for testing. These clips enabled a realistic evaluation of model performance under field conditions, assessing the potential for false positives, overcounting, or missed detections.

G. Inference Speed and Real-time Considerations

Initial model evaluations were conducted at 20 frames per second (fps) to analyze short test videos (10–15 seconds). However, the frame sampling rate can be reduced to 1 fps for real-time or near real-time deployments, allowing the system to run efficiently on moderate hardware. Adjusting the frame rate provides flexibility: high fps for short clips and peak activity detection, and lower fps for continuous long-term monitoring with reduced computational cost.

H. Hardware and Software Configuration

All experiments were carried out on a workstation equipped with an NVIDIA GeForce GTX 1660 Ti GPU (16 GB RAM) and an Intel Core i7 processor. The models were implemented using Python 3.8, PyTorch 1.8, OpenCV, and related libraries (NumPy, Matplotlib). This hardware-software configuration ensured a balance between computational efficiency and cost, enabling both training and inference tasks.

I. Evaluation Metrics

Model performance was assessed using mAP@50, which measures the accuracy of predicted bounding boxes at an IoU threshold of 0.50. Additionally, the more stringent mAP@50-95 metric was considered, providing a comprehensive view of model performance across multiple IoU thresholds.

Per-video analyses examined detection counts, inference times, and occurrence of false positives or overcounting. Confusion matrices, precision-recall curves, and example detection frames were also generated to provide qualitative insights into model behavior, guiding the selection of the best-performing YOLO configuration for practical deployment.

III. Results and Discussion

Following the established experimental setup, the detection performance of the proposed YOLO-based method was validated under conditions intended to simulate real-time monitoring scenarios. A 5-fold cross-validation procedure was employed on the training dataset, and multiple configurations of YOLOv8 and YOLOv9 were examined. The evaluation metrics included the mean Average

Precision at a 0.50 IoU threshold (mAP@50) and at multiple IoU thresholds (mAP@50-95), providing both a broad and detailed perspective on detection accuracy and localization quality.

A. Performance of the Proposed Method

Table II presents the mAP@50 and mAP@50-95 results for various YOLOv8 and YOLOv9 configurations. YOLOv9 Medium achieved the highest mAP@50 (95.83%), indicating strong localization capabilities. When considering mAP@50-95, YOLOv8 Medium achieved the best score (67.03%), suggesting it provides a balanced trade-off between detection accuracy and robustness across different IoU thresholds.

Table II
mAP Scores for YOLOv8 and YOLOv9 Models.

Model	mAP@50 (%)	mAP@50-95 (%)
YOLOv9-t	93.26	67.14
YOLOv9-s	95.68	66.69
YOLOv9-m	**95.83**	66.61
YOLOv9-c	95.37	65.99
YOLOv9-e	94.97	62.93
YOLOv8-n	94.16	64.95
YOLOv8-s	95.07	65.53
YOLOv8-m	94.20	**67.03**
YOLOv8-l	94.55	66.60
YOLOv8-x	95.04	66.00

Confusion matrices (Figs. 1 and 2) for YOLOv8 Medium and YOLOv9 Medium provide insights into class-specific performance. The matrices are structured for four classes: Silky Shark, Tiger Shark, Tuna, and Background. Darker shades appear along the diagonal for the shark and tuna classes, indicating correct classifications. For the Background class, the darkest cell occurs at the Background-Tuna intersection, reflecting the underlying frequency and class distribution in the dataset rather than a systematic model bias.

Precision-Recall (PR) curves (Figs. 3 and 4) show the relationship between precision and recall, allowing the selection of optimal confidence thresholds depending on conservation goals. Higher precision thresholds reduce false positives, critical for accurate population estimates, while higher recall thresholds ensure that most individuals are detected, even at the expense of occasional misclassifications.

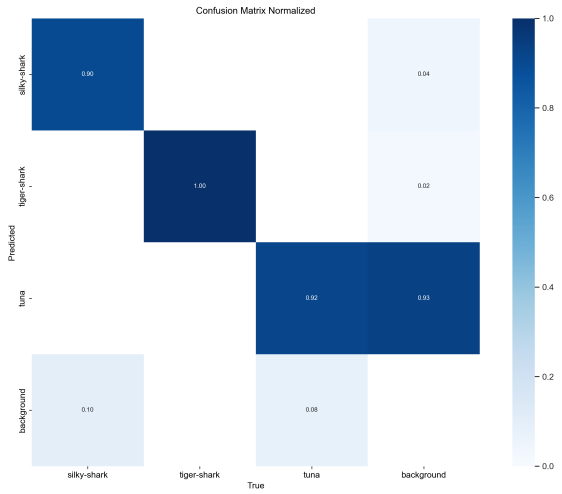

Figure 1. Confusion matrix for YOLOv8 Medium.

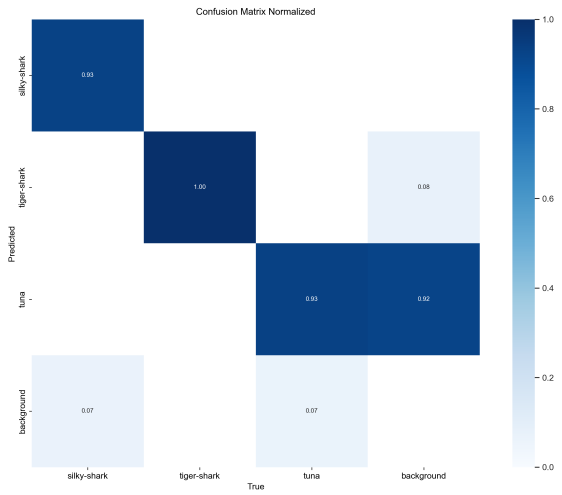

Figure 2. Confusion matrix for YOLOv9 Medium.

B. Performance of the Proposed Method in the Experimental Setup

To evaluate real-world applicability, the models were tested on two short, unseen videos (10s and 15s), described in Table III. While YOLOv8 Medium and YOLOv9 Medium closely matched expected species counts without introducing spurious results, certain other configurations produced false positives or overcounts. These differences underscore the importance of evaluating models on realistic scenarios rather than relying solely on aggregate metrics.

Qualitative analyses further supported these findings. Figure 5 illustrates scenarios where some models introduced false positives or overcounting, reinforcing the need for careful model selection.

Table III
PER-VIDEO DETECTION COUNTS AND DETECTION TIMES FOR YOLOV8 AND YOLOV9 MODELS.

max width=

2*Model	test_tiger_tunas.mp4				test_silkies.mp4				2*False Positive
	Tuna	Tiger Shark	Silky Shark	Time (s)	Tuna	Tiger Shark	Silky Shark	Time (s)	
Expected	>20	1	0	–	0	0	2	–	-
YOLOv9 Tiny	23	1	0	96.39	0	0	2	66.29	N
YOLOv9 Small	24	1	0	125.34	0	0	2	85.54	N
YOLOv9 Medium	27	1	0	182.78	0	0	2	108.97	N
YOLOv9 Compact	27	2	0	244.27	0	0	2	122.82	Yes (Ov
YOLOv9 Efficient	27	1	0	425.20	0	0	2	222.97	N
YOLOv8 Nano	20	1	0	85.96	0	0	2	41.66	N
YOLOv8 Small	25	1	1	122.65	1	0	2	54.46	Yes (False
YOLOv8 Medium	26	1	0	174.80	0	0	2	83.61	N
YOLOv8 Large	26	1	0	189.25	0	0	2	118.82	N
YOLOv8 X-Large	28	1	1	250.41	0	0	3	159.36	Yes (Overcount,

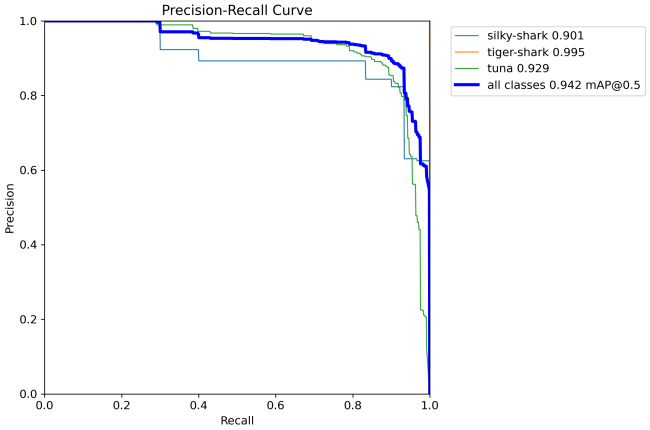

Figure 3. Precision-Recall curve for YOLOv8 Medium.

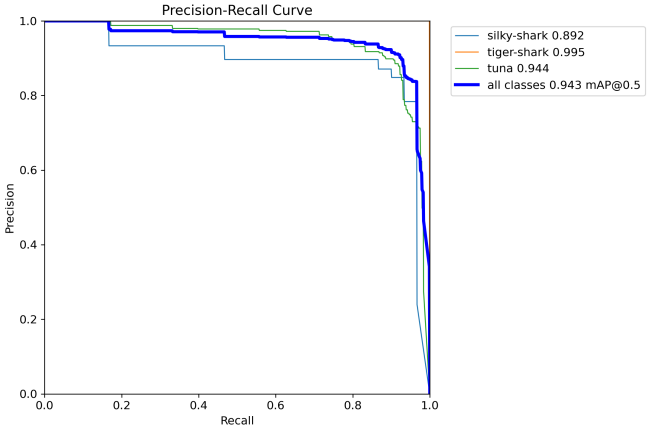

Figure 4. Precision-Recall curve for YOLOv9 Medium.

Conversely, successful detections are shown in Figs. 6 and 7, where YOLO-based models accurately identified silky sharks, tiger sharks, and tunas. These examples highlight the performance stability of top-performing configurations like YOLOv8 Medium and YOLOv9 Medium.

Temporal analyses of detections provide additional ecological insights. Figures 8, 9, and 10 show timeframes of species detections for YOLOv8 Medium and YOLOv9 Medium. In these figures, each model's detection results are coupled with a corresponding timeframe plot beneath the detection image. For YOLOv9 Medium, the timeframe plots are smoother, suggesting more stable detections over time. This stability can be crucial in identifying peak activity periods or seasonal patterns, informing more effective conservation measures.

Adjusting the frame sampling rate allows flexible deployment strategies. While initial evaluations were conducted at 20 frames per second (fps) for detailed inspection of short clips, reducing the sampling rate to 1 fps enables continuous, long-term monitoring with reduced computational overhead. At this lower frame rate, species remain within the field of view long enough to ensure reliable detections, making the method suitable for sustained monitoring efforts even on moderate hardware.

Overall, these results demonstrate that the latest YOLO architectures can effectively detect and track shark and tuna species in underwater footage. By providing high accuracy, stable detections over time, and adaptability in frame processing rates, the proposed YOLO-based approach emerges as a valuable non-invasive tool for marine ecosystem monitoring and conservation planning.

IV. CONCLUSION AND FUTURE WORK

This study presented a YOLO-based detection system capable of accurately identifying silky sharks, tiger sharks, and tunas in underwater video footage from the Galápagos Islands. By

evaluating multiple configurations of YOLOv8 and YOLOv9 architectures and employing a 5-fold cross-validation protocol, we identified models that combine high detection accuracy with robustness and computational efficiency. In particular, YOLOv8 Medium achieved a favorable balance, attaining a mAP@50 of 94.20%, while YOLOv9 Medium attained the highest mAP@50 (95.83%).

These models demonstrated reliable performance on previously unseen videos, maintaining stable detections over time, and avoiding systematic false positives or overcounting. Adjusting the frame processing rate allowed the system to operate in real-time or near real-time conditions on moderate hardware, providing flexibility for various deployment scenarios. The results underscore the potential of automated, computer vision-based approaches to support marine conservation efforts, offering a non-invasive, scalable, and cost-effective tool for monitoring species within marine protected areas.

Future work will focus on integrating advanced temporal modeling methods to improve multi-frame species tracking, enhancing the system's ability to follow individuals over time. Additionally, employing dedicated models for each species will allow customization based on distinct visual and behavioral characteristics, ensuring a flexible framework that can scale to include new species as needed. Finally, test deployments in various marine protected areas will provide invaluable insights into performance under a wide range of environmental conditions, further refining and validating the system's utility for conservation and management purposes.

References

[1] J. D. Stevens, R. Bonfil, N. K. Dulvy, and P. A. Walker, "The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems," *ICES Journal of Marine Science*, vol. 57, no. 3, pp. 476–494, 2000.

[2] N. K. Dulvy, S. L. Fowler, J. A. Musick, R. D. Cavanagh, P. M. Kyne, L. R. Harrison, J. K. Carlson, L. N. Davidson, S. V. Fordham, M. P. Francis *et al.*, "Extinction risk and conservation of the world's sharks and rays," *eLife*, vol. 3, p. e00590, 2014.

[3] J. K. Baum, R. A. Myers, D. G. Kehler, B. Worm, S. J. Harley, and P. A. Doherty, "Collapse and conservation of shark populations in the northwest atlantic," *Science*, vol. 299, no. 5605, pp. 389–392, 2003.

[4] K. A. Spyker and D. Pollard, "Approaches and methods for marine species monitoring in new south wales," *Marine species monitoring, New South Wales*, 2003.

[5] A. Hearn, J. K. Baum, A. C. Kubiszeski, S. M. Vincent, and P. G. Salinas-de León, "The decline of coastal apex shark populations in the galapagos marine reserve," *Marine Ecology Progress Series*, vol. 415, pp. 1–9, 2010.

[6] E. Ulloa, C. Jara, R. Vasquez, A. Peña, N. Perez-Perez, and M. Campos, "Hammerhead shark detection using faster r-cnn," in *2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*. IEEE, 2020, pp. 96–100.

[7] A. Peña, E. Ulloa, R. Vasquez, N. Perez-Perez, and M. Campos-Cabrera, "Species detection in marine protected areas using convolutional neural networks," in *2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*. IEEE, 2020, pp. 84–89.

[8] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[9] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[11] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 390–391.

[12] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.

[13] D. Misra, "Mish: A self regularized non-monotonic neural activation function," *arXiv preprint arXiv:1908.08681*, 2019.

[14] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *European conference on computer vision*. Springer, 2014, pp. 346–361.

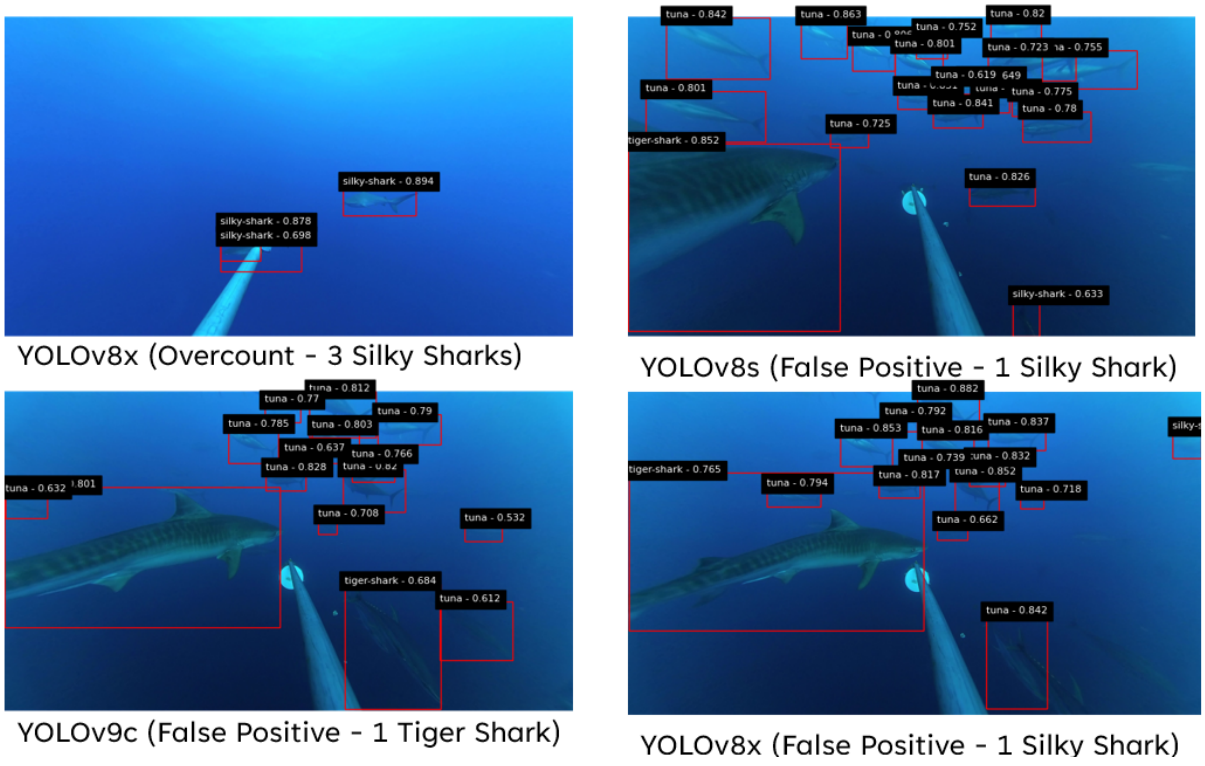

Figure 5. Examples where certain models introduce false positives or overcounting.

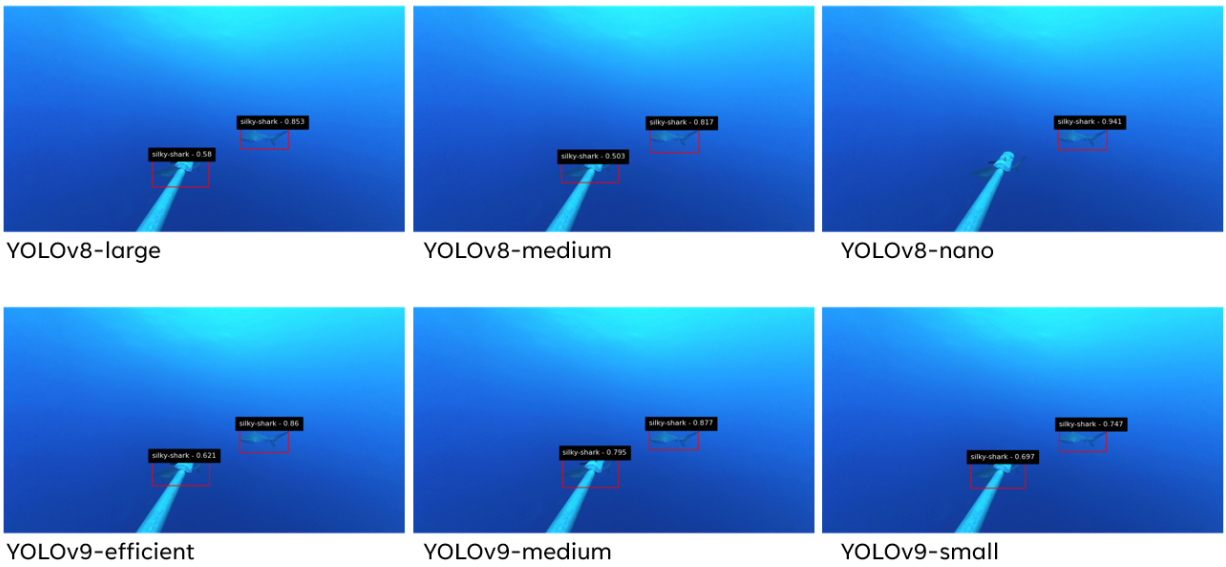

Figure 6. Successful detections of silky sharks across multiple YOLO-based configurations (YOLOv8-l, YOLOv8-m, YOLOv8-n, YOLOv9-e, YOLOv9-m, YOLOv9-s).

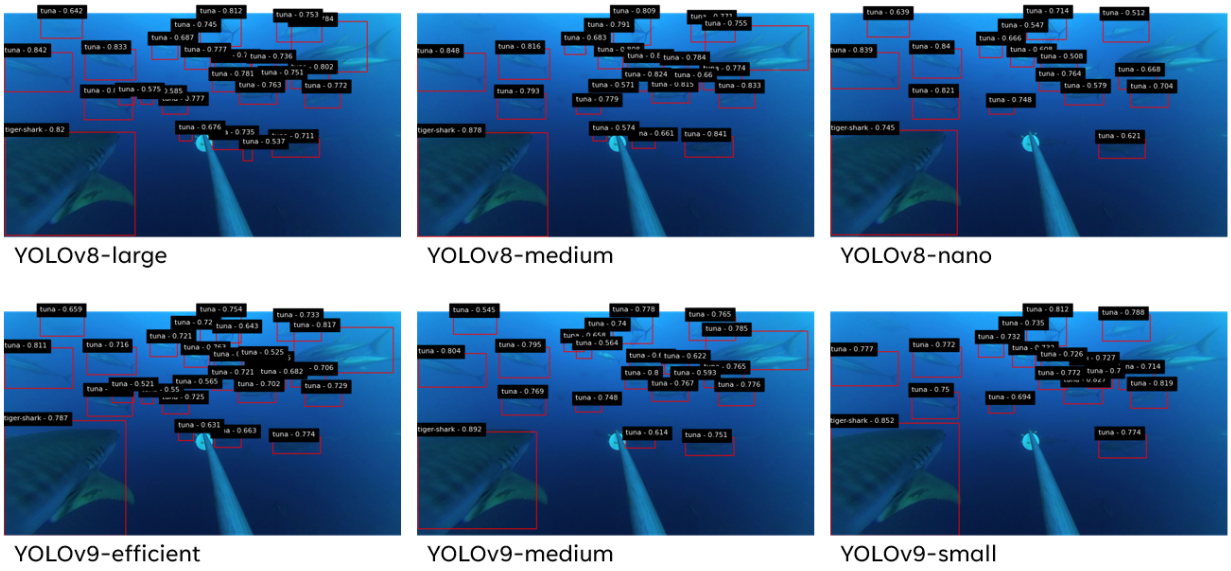

Figure 7. Successful detections of tiger sharks and tunas across various YOLO-based models (YOLOv8-l, YOLOv8-m, YOLOv8-n, YOLOv9-e, YOLOv9-m, YOLOv9-s).

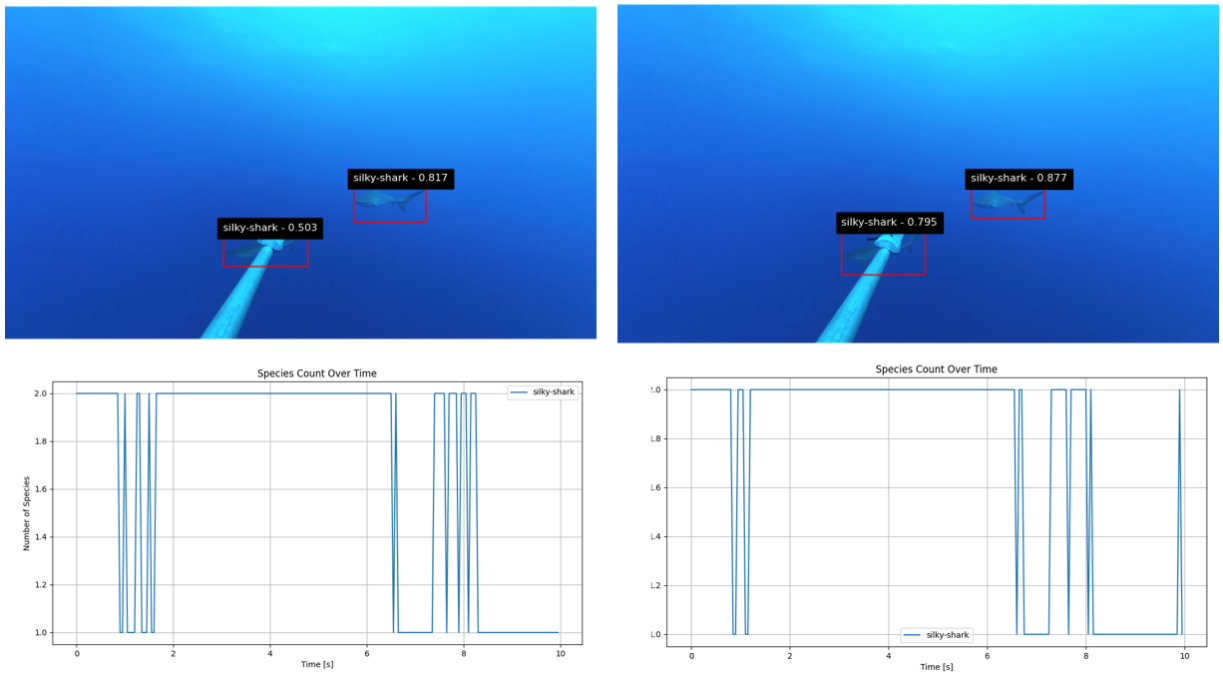

Figure 8. Comparative detection and timeframe analysis for silky sharks. Left: YOLOv8 Medium detection and corresponding timeframe. Right: YOLOv9 Medium detection and timeframe.

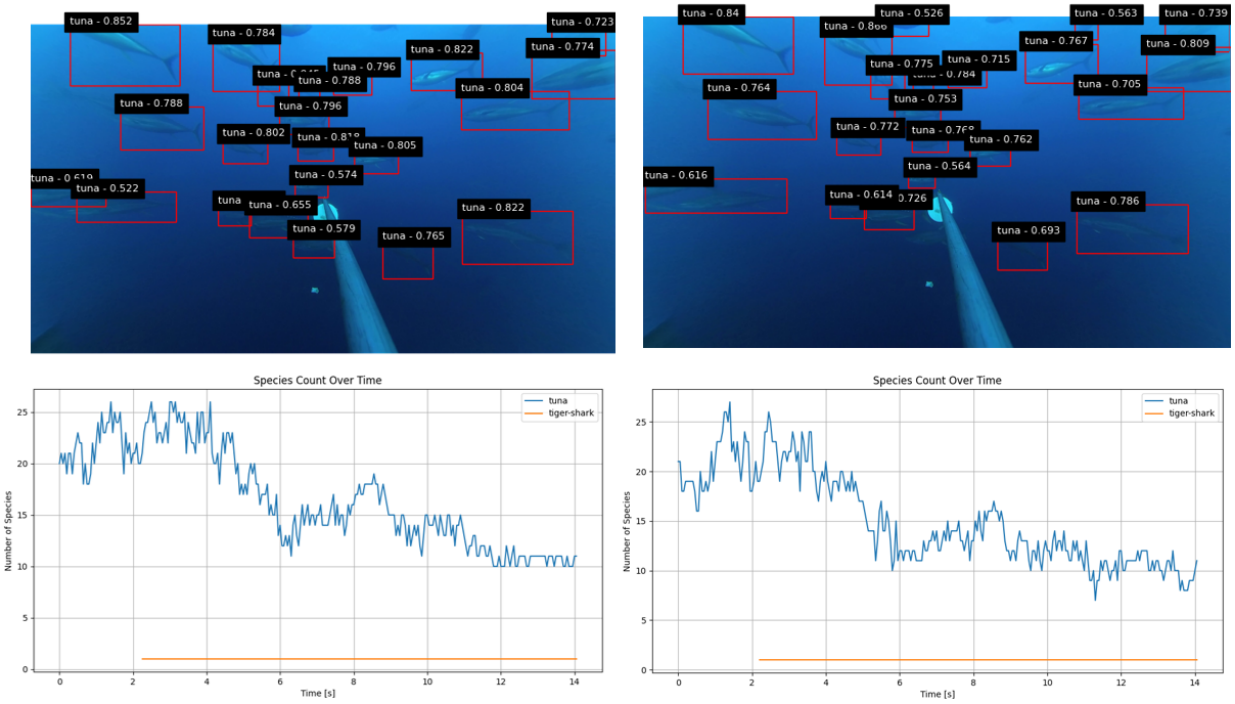

Figure 9. Comparative detection and timeframe analysis for tunas. Left: YOLOv8 Medium detection and corresponding timeframe. Right: YOLOv9 Medium detection and timeframe.

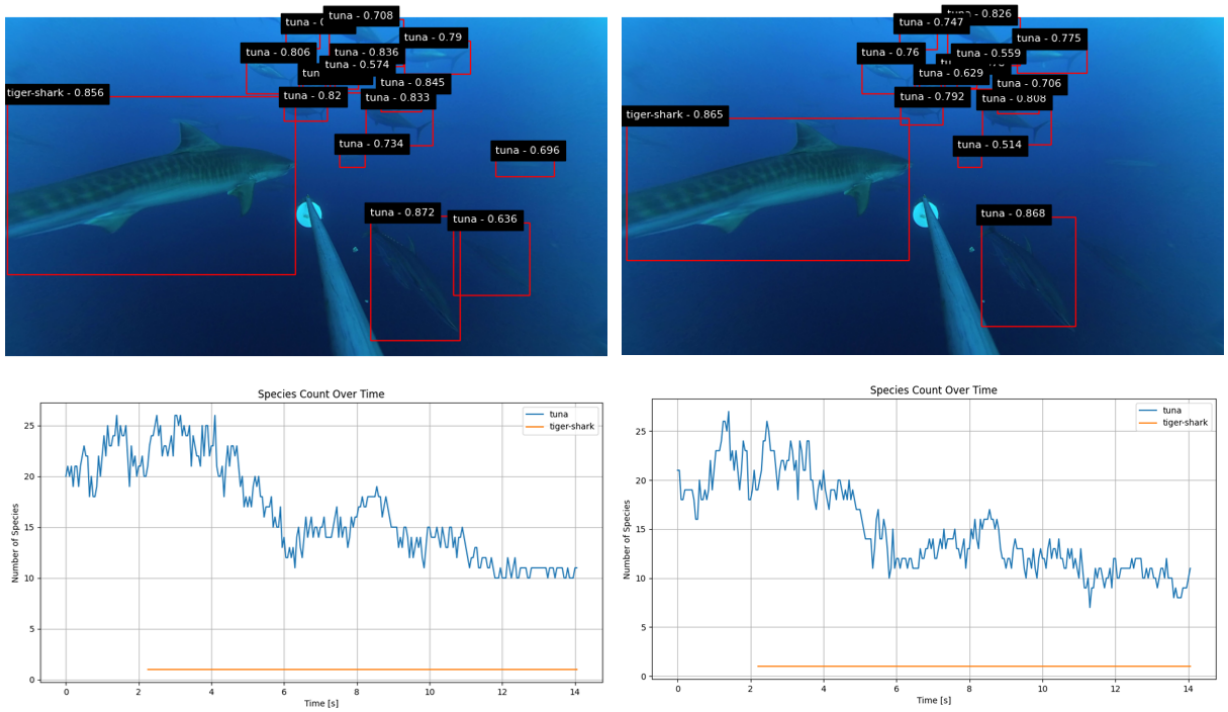

Figure 10. Comparative detection and timeframe analysis for tiger sharks. Left: YOLOv8 Medium detection and corresponding timeframe. Right: YOLOv9 Medium detection and timeframe.