Proceeding Paper

YOLO-Based Fish Detection in Underwater Environments †

Mohammed Yasser Ouis *,‡ **and Moulay Akhloufi** ‡ [ID]

Perception, Robotics and Intelligent Machines Research Lab (PRIME), Department of Computer Science, Universite de Moncton, 18 Antonine-Maillet Ave, Moncton, NB E1A 3E9, Canada; moulay.akhloufi@umoncton.ca

* Correspondence: emo7847@umoncton.ca

† Presented at the 5th International Electronic Conference on Remote Sensing, 7–21 November 2023; Available online: https://ecrs2023.sciforum.net/.

‡ These authors contributed equally to this work.

Abstract: In this work, we present a comprehensive study on fish detection in underwater environments using sonar images from the Caltech Fish Counting Dataset (CFC). We use the CFC dataset, initially designed for tracking purposes, to optimize and evaluate the performance of YOLO v7 and YOLO v8 models in fish detection. Our findings demonstrate the high performance of these deep learning models in accurately detecting fish species in sonar images. In our evaluation, YOLO v7 achieved an average precision of 68.3% (AP50) and 62.15% (AP75), while YOLO v8 demonstrated an even better performance with an average precision of 72.47% (AP50) and 66.21% (AP75) across the test dataset of 334,017 images. These high-precision results underscore the effectiveness of these models in fish detection tasks under various underwater conditions. With a dataset of 162,680 training images and 334,017 test images, our evaluation provides valuable insights into the models performance and generalization across diverse underwater conditions. This study contributes to the advancement of underwater fish detection by showcasing the suitability of the CFC dataset and the efficacy of YOLO v7 and YOLO v8 models. These insights can pave the way for further advancements in fish detection, supporting conservation efforts and sustainable fisheries management.

Keywords: fish detection; Caltech Fish Counting Dataset (CFC); underwater environments; YOLO v7; YOLO v8

1. Introduction

The exploration and understanding of underwater ecosystems have garnered significant attention due to their crucial role in maintaining the balance of our planet's biodiversity and climate regulation. Within these aquatic environments, the study of fish populations plays an essential role in unraveling the mysteries of marine life and its complex interactions. Detecting and monitoring fish within underwater environments is important for numerous applications, ranging from ecological research and conservation efforts to fisheries management and aquaculture. In recent years, advances in computer vision and deep learning have brought about innovative solutions to automate fish detection, revolutionizing the way we comprehend and manage underwater ecosystems.

Numerous studies [1–4] have been conducted to address the challenge of fish detection in underwater environments, recognizing its importance in facilitating non-invasive monitoring and analysis. However, many of these earlier endeavors were constrained by the limited availability of large and diverse datasets [3]. This limitation led to methods that, while effective on their respective training data, were often difficult to generalize across different underwater environments. The outcome was methodologies that exhibited bias towards the specific dataset used for training, thus restricting the applicability and adaptability of the proposed models to real-world scenarios.

The turning point in the field of fish detection came with the release of the Caltech Fish Dataset [1]. Comprising an extensive collection of 162,680 training images and 334,017 test

images, this dataset presents a comprehensive representation of fish in varied underwater contexts. This novel dataset offers researchers and practitioners a wealth of diverse samples, enabling the development and evaluation of fish detection models that can surpass the limitations of prior approaches. The volume and diversity of data in the Caltech Fish Dataset [1] hold the promise of enabling more efficient training of models and fostering enhanced generalization capabilities across a broader spectrum of underwater scenarios.

In addition to providing a challenging benchmark in a novel application domain, the dataset allows for detailed study in three areas that have received limited attention from the computer vision community and lack supporting benchmarks: multiple-object tracking, video-based counting, and generalization of tracking and counting methods. The introduction of these aspects adds a new dimension to the field of fish detection, fostering pathways for more comprehensive research and innovation.

This paper delves into the domain of fish detection in underwater environments, building upon deep learning techniques. Specifically, we explore the utilization of YOLO (You Only Look Once) model variants for fish detection tasks. The objective is to investigate the effectiveness of these models in harnessing the power of the Caltech Fish Dataset [1] to achieve robust and generalized fish detection capabilities. By leveraging the ample data provided by this dataset, we aim to bypass the constraints of earlier approaches and pave the way for more adaptable and transferable fish detection solutions.

In the following sections, we will delve into the details of our methodology, the experimental setup, and the results obtained from our investigation. We will highlight the advancements made possible by the Caltech Fish Dataset [1] and discuss how our findings contribute to the broader field of underwater ecosystem analysis. Ultimately, this study contributes to the growing body of knowledge that seeks to harness cutting-edge technologies to better comprehend and manage our complex underwater world.

2. Materials and Methods

In our investigation of effective fish detection within underwater environments, we employed two state-of-the-art deep learning models, namely YOLO v7 [5] and YOLO v8 [6], which are known for their remarkable performance in object detection tasks. These models were fine-tuned on the Caltech Fish Counting Dataset [1], a comprehensive resource designed for fish detection, tracking, and counting. This dataset comprises manually marked sonar videos, extracted to yield 1233 clips from the Kenai River (4300 fish), 262 clips from the Elwha River (884 fish), and 72 clips from the Nushagak River (3070 fish). It includes bounding box annotations for fish tracks, totaling 515,933 annotations across 8254 tracks from seven cameras. Moreover, the dataset supports algorithmic evaluation, featuring a training set of 162,680 images (1762 tracks, 132,220 bounding boxes) and a validation set of 30,518 images (207 tracks, 18,565 bounding boxes), allowing for comprehensive performance assessment. Figure 1 illustrates the dataset images and the challenges they present in the context of underwater fish detection.
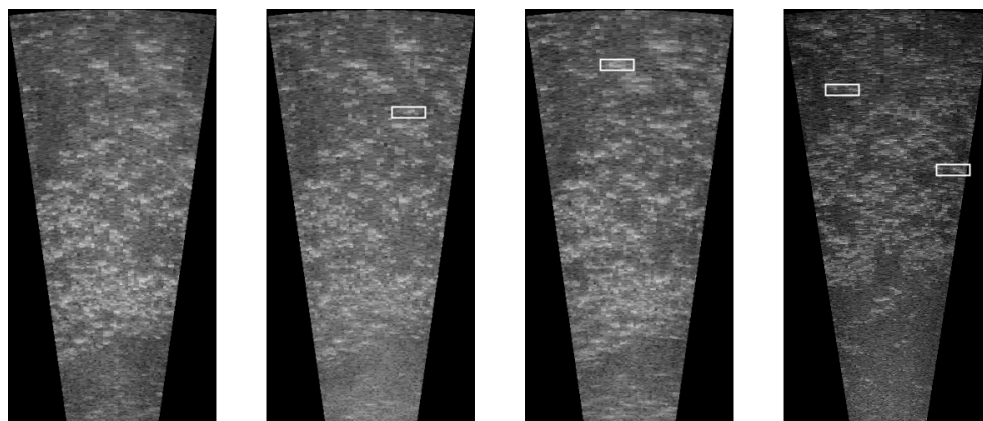

Figure 1. Sample images from the Caltech Fish Counting Training Dataset.

Compared to other detectors, YOLO v7 [5] has undergone key changes, including the Extended Efficient Layer Aggregation Network (E-ELAN) and Model Scaling for Concatenation-based Models, contributing to its superior performance. It reduces parameters and computations by about 40% and 50%, respectively. Meanwhile, on the other hand, YOLO v8 [6] is assumed to be the new state of the art due to its higher mAPs and lower inference speed on the COCO dataset [7]. Other notable enhancements include a novel anchor-free detection system, modifications to the convolutional blocks within the model, and the incorporation of Mosaic augmentation during training, which is deactivated in the final 10 epochs. These improvements collectively contribute to YOLO v8's [6] advanced object detection capabilities and training efficiency.

3. Results

For our work, we employed pre-trained YOLO v7 [5] and YOLO v8 [6] models, which were initially trained on the COCO dataset [7]. We further fine-tuned these models on the Caltech dataset [1], a key component of our research. The training process extended over 100 epochs, with a batch size of 32, facilitated by the high-performance NVidia A100 GPU provided by Compute Canada.

Table 1 summarizes the fish detection results achieved using different models, with performance metrics represented as Average Precision (AP), namely AP50 and AP75. Notably, YOLO v8 [6] demonstrated the highest performance, achieving an AP50 of 72.47% and an AP75 of 66.21%. YOLO v7 also achieved remarkable performance, with an AP50 of 68.3% and an AP75 of 62.15%. In comparison to other models, including Faster R-CNN + Resnet101, ScaledYOLOv4 CSP, and YOLOv5m, our YOLO-based models outperformed them in terms of AP50. However, it is essential to highlight that these results, while promising, indicate the need for further research and development to achieve optimal performance in fish detection within underwater environments.

Table 1. Fish detection results using different models. AP50 and AP75 represent average precision at 50% and 75% IoU thresholds, respectively.

Model Name	Average Precision (AP50)	Average Precision (AP75)
YOLO v8	72.47%	66.21%
YOLO v7	68.3%	62.15%
Faster R-CNN + Resnet101 [1]	65.9%	-
ScaledYOLOv4 CSP [1]	65.3%	-
YOLOv5m [1]	66.4%	-

4. Discussion

Our study has demonstrated the effectiveness of utilizing YOLO v7 [5] and v8 [6] for fish detection in underwater videos. These advanced versions of YOLO, which continue to refine and optimize object detection, have showcased promising results in detecting fish species within challenging underwater environments. We have thoroughly analyzed the performance of these models using the average precision metric, considering their ability to handle diverse lighting conditions and underwater complexities.

However, our exploration also revealed that there is still considerable potential for improvement in this models. The complexity and diversity of underwater environments, combined with the unique challenges they present, require continued research and development. Future work should focus on the fine-tuning of YOLO v7 [5] and YOLO v8 [6], as well as the exploration and integration of more sophisticated architectures and techniques specific to underwater conditions.

Furthermore, there is an opportunity to leverage the increasing availability of underwater datasets, such as the Caltech Fish Dataset [1], to facilitate the training and evaluation of these more advanced models using average precision as the primary performance measure. The continuous development of underwater vision technologies holds the potential to contribute significantly to our understanding and conservation efforts in aquatic ecosystems.

In summary, as we work towards achieving greater accuracy and efficiency in fish detection within underwater environments, more work is still needed. This future work would include enhancing deep learning architectures, including YOLO v7 [5] and YOLO v8 [6], and pushing the boundaries of what is achievable in the field of underwater computer vision. Through these ongoing efforts, we aim to make substantial contributions to the field, ultimately enabling better management and preservation of our invaluable underwater ecosystems.

5. Conclusions

In conclusion, our work underscores the potential of YOLO v7 [5] and YOLO v8 [6] for fish detection in underwater environments. However, to fully realize their capabilities, ongoing research is imperative, involving fine-tuning of these models and exploration of advanced underwater-specific architectures. This ongoing effort is crucial for advancing our understanding and conservation of aquatic ecosystems through improved computer vision techniques.

References

1. Kay, J.; Kulits, P.; Stathatos, S.; Deng, S.; Young, E.; Beery, S.; Van Horn, G.; Perona, P. The Caltech Fish Counting Dataset: A Benchmark for Multiple-Object Tracking and Counting. In Proceedings of the 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; pp. 290–311.
2. Knausgård, K.M.; Wiklund, A.; Sørdalen, T.K.; Halvorsen, K.T.; Kleiven, A.R.; Jiao, L.; Goodwin, M. Temperate fish detection and classification: A deep learning based approach. *Appl. Intell.* **2022**, *52*, 6988–7001. [CrossRef]
3. Xu, W.; Matzner, S. Underwater fish detection using deep learning for water power applications. *CSCI* **2018**, *10*, 313–318
4. Mahoro, E.; Akhloufi, M.A. Automated fish detection and classification on sonar images using detection transformer and YOLOv7. *SPIE* **2023**, *12749*, 9–16.
5. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 17–24 June 2023; pp. 7464–7475.
6. Ultralytics. YOLO v8. Available online: https://ultralytics.com/yolov8 (accessed on 10 August 2023).
7. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. *ECCV* **2014**, 740–755.