International Journal of Environmental Monitoring and Analysis
2025, Vol. 13, No. 6, pp. 314–327
https://doi.org/10.11648/j.ijema.20251306.13

Research Article

Automated Detection and Classification of Underwater Species Using YOLOv8 for Real-time Marine Ecosystem Monitoring

Vinod Vadde, Takur Sai Karthik Chalamalasetty, Paka Shalini, Wisam Bukaita*

Math and Computer Science, Lawrence Technological University, Southfield, USA

Abstract

Effective monitoring of marine biodiversity is essential for understanding ecosystem health, detecting species population changes, and mitigating the impacts of environmental degradation. Traditional underwater observation techniques, such as diver-based surveys and manual video analysis, are labor-intensive, time-consuming, and prone to human error. Consequently, there is an increasing need for automated, data-driven methods capable of performing real-time detection and analysis of aquatic species under diverse environmental conditions. This study introduces a deep learning framework based on the YOLOv8 architecture for automated detection, classification, and segmentation of underwater species. A curated dataset containing seven representative classes fish, jellyfish, starfish, shark, puffin, penguin, and crown-of-thorns starfish is used for model training and evaluation. Data preprocessing techniques, including image enhancement, resizing, and normalization, were applied to address underwater imaging challenges such as low contrast, noise, and color distortion. The model was trained using transfer learning and data augmentation to improve robustness and generalization under varying light and turbidity conditions. The experimental results demonstrate that the proposed YOLOv8 framework achieves Precision of 80.82%, Recall of 69.35%, mAP@0.5 of 76.86%, and mAP@0.5–0.95 of 47.57% in object detection tasks. The segmentation module further attained 85.48% accuracy, enabling precise delineation of species boundaries for morphological assessment. These outcomes highlight YOLOv8's superior ability to generalize across diverse underwater environments compared to conventional convolutional neural network (CNN)–based approaches. Overall, this research presents a scalable and efficient deep learning solution for real-time underwater species monitoring. The integration of detection and segmentation capabilities enables accurate, fine-grained analysis that can enhance marine conservation, ecological assessment, and automated biodiversity mapping. The proposed YOLOv8-based framework represents a significant step toward the deployment of intelligent visual systems in marine ecosystem monitoring and environmental sustainability applications.

Keywords

Underwater Species Detection, Marine Monitoring, Deep Learning, YOLOv8, Object Detection, Environmental Surveillance, Marine Conservation

*Corresponding author: wbukaita@ltu.edu (Wisam Bukaita)

Received: 2 November 2025; **Accepted:** 12 November 2025; **Published:** 9 December 2025

1. Introduction

Marine ecosystems are essential to maintaining global biodiversity, regulating climate, and sustaining ecological balance. Effective monitoring of underwater species offers valuable insights into ecosystem health, population distribution, and the impacts of anthropogenic pressures. According to the United Nations Environment Programme, more than 40% of marine habitats are under threat due to pollution, overfishing, and climate change. Traditional monitoring techniques, such as manual annotation and diver-based surveys, are time-consuming, costly, and limited in spatial and temporal scope.

Underwater image analysis introduces additional challenges, including poor illumination, turbidity, occlusion, and background clutter, which complicate species identification and classification. These issues significantly reduce the performance of classical image processing techniques. The advent of deep learning, particularly convolutional neural networks (CNNs), has revolutionized automated image analysis by enabling robust feature extraction and end-to-end detection.

The YOLO (You Only Look Once) family of models offers an efficient real-time detection framework capable of simultaneous localization and classification, making it highly suitable for underwater applications involving remotely operated vehicles (ROVs), autonomous underwater vehicles (AUVs), and fixed monitoring systems. Although prior studies have demonstrated promising results in controlled environments, model robustness under variable real-world underwater conditions remains a research gap.

To address this challenge, the present study proposes a YOLOv8-based deep learning pipeline for automated underwater species detection. The specific objectives are:

1. To develop a high-performance object detection model tailored for real-time underwater deployment.
2. To mitigate class imbalance, lighting inconsistency, and noise through systematic preprocessing and augmentation.
3. To evaluate detection performance using standard object detection metrics.
4. To demonstrate the potential of AI-powered approaches for scalable marine ecosystem monitoring and conservation.

2. Literature Review

Underwater object and species detection have rapidly evolved with the integration of deep learning, allowing for more accurate and efficient marine monitoring despite the challenges posed by light attenuation, scattering, and turbidity. Chen et al. [1] provided a comprehensive overview of how artificial intelligence has revolutionized underwater detection, outlining current progress, existing challenges, and emerging opportunities for model improvement. Similarly, Duan et al. [6] surveyed modern underwater datasets and detection frameworks, emphasizing the growing diversity of deep learning architectures and the need for domain-specific benchmarks tailored to aquatic conditions.

To enhance species recognition in aquaculture, Hamzaoui et al. [2] proposed an improved deep learning model capable of identifying multiple underwater species under varying illumination and background conditions. Their work demonstrated that integrating domain adaptation and transfer learning can significantly improve recognition accuracy across different underwater environments. In a related effort, Malik et al. [4] introduced a multi-classification deep neural network for fish species identification using camera-captured images, achieving superior performance compared to traditional feature-based classifiers.

The combination of advanced imaging and generative diffusion models was explored by Pachaiyappan et al. [3], who demonstrated that enhancing raw underwater images prior to model inference substantially improves classification outcomes. Similarly, Wu, Li, and Hong [12] focused on fish recognition in visually degraded, fuzzy underwater environments, using CNN-based image restoration techniques to recover key visual features before detection.

Deep learning has also proven essential for automating large-scale marine biodiversity assessments. Marrable et al. [5] employed machine-assisted learning to accelerate species labeling from underwater video data, significantly reducing the manual workload associated with ecological surveys. Younes et al. [7] extended this concept by integrating YOLO-based coral detection models for real-time reef monitoring, achieving both speed and precision in detecting diverse coral structures.

Several studies have emphasized the use of modern object detection algorithms. Trinadh et al. [8] demonstrated that YOLO and SSD-based frameworks perform effectively in underwater object detection tasks when trained on well-annotated datasets. Martinho et al. [9] proposed an end-to-end deep learning approach for fish detection in underwater images, showing that lightweight architectures could maintain accuracy while remaining computationally efficient for embedded systems. Similarly, Chen and Belbachir [10] applied Mask R-CNN for fish instance segmentation, treating underwater organisms as novel object classes and highlighting the adaptability of segmentation-based detectors in marine contexts.

Review studies such as Liu et al. [11] have categorized various CNN and Transformer-based object detection techniques, summarizing their applications in aquaculture and identifying trends toward hybrid and attention-based architectures. Deng et al. [13] further contributed an automatic recognition method for fish species and length estimation using stereo vision and deep learning, effectively bridging object detection with 3D measurement for ecological data collection.

Collectively, these studies demonstrate that deep learning has become an indispensable tool in underwater species detection. From CNN-based detectors to Transformer and diffusion model enhancements, the literature reflects a clear progression toward improving visibility robustness, computational efficiency, and real-time inference capabilities in complex aquatic environments.

3. Synthesis and Research Gap

Collectively, previous studies demonstrate rapid advancements in underwater object detection using deep learning, particularly with YOLO and CNN-based models. Works such as Chen et al. [1], Hamzaoui et al. [2], and Pachaiyappan et al. [3] emphasize the growing success of AI models in marine environments for tasks like coral monitoring and fish identification. Similarly, Liu et al. [13] and Malik et al. [4] explore CNN-based approaches for species recognition, revealing high accuracy in clear-water conditions but limited generalization under variable illumination and turbidity.

However, despite these technological strides, several persistent challenges remain:

1. Limited datasets: Most existing datasets focus on open-sea or coral reef conditions, lacking samples from underground or subterranean water systems characterized by low light and sediment interference.
2. Small or camouflaged species detection: Models still struggle to identify fine-grained features of small, transparent, or highly camouflaged organisms in murky conditions.
3. Domain transfer limitations: Few studies adapt models specifically to the spectral, illumination, and pressure conditions typical of underground aquatic habitats.
4. Insufficient real-time adaptability: Many models emphasize accuracy but fail to maintain speed or stability in real-time scenarios.

This research addresses these gaps by developing a YOLOv8-based detection framework optimized for underground water species, integrating enhanced image preprocessing, adaptive feature extraction, and interpretability modules to achieve consistent detection accuracy in complex visual conditions.

4. Methodology

The methodological framework employed to develop and evaluate the proposed YOLOv8-based model for underground water species detection. The methodology is structured into several key components: (1) dataset collection and description, (2) data preprocessing and augmentation, (3) model architecture and training configuration, and (4) evaluation metrics and performance visualization.

The workflow begins with curating and preparing a diverse underwater imagery dataset, followed by preprocessing to enhance image quality and address class imbalance. Subsequently, the YOLOv8 model architecture is adapted for real-time object detection and trained using transfer learning. The trained model is quantitatively and qualitatively evaluated through performance metrics and visualization outputs.

4.1. Dataset Description

The dataset employed in this study serves as the foundational component for developing a deep learning model capable of detecting and localizing multiple marine and coastal species within complex underwater imagery. Its primary objective is to enable the model to identify not only the species present in an image but also their precise spatial locations.

The dataset encompasses seven target species classes, representing both pelagic and benthic organisms: fish, jellyfish, starfish, shark, puffin, penguin, *and* crown-of-thorns starfish. These categories are selected due to their ecological importance, distinct morphological features, and varying visual characteristics across aquatic environments.

All images were preprocessed and uniformly resized to a resolution of $1024 \times 768 \times 3$ (width × height × color channels). This standardization ensures consistent input dimensions for the YOLOv8 model, facilitating efficient training and convergence while preserving essential visual details. Each image is represented in the RGB color space, maintaining color information crucial for species differentiation in underwater scenes.

To ensure fair and effective model evaluation, the dataset was partitioned into three subsets following standard deep learning practices:

1). Training Set (70%) – Used for model learning and feature extraction across all classes.

2). Validation Set (20%) – Utilized to fine-tune hyperparameters and prevent overfitting by monitoring model generalization during training.

3). Test Set (10%) – Reserved exclusively for the final, unbiased assessment of model performance on unseen data.

This stratified split maintains proportional representation of all species categories across subsets, ensuring balanced evaluation.

The dataset is intentionally curated to capture a wide range of environmental conditions, simulating real-world underwater scenarios. Such variability enhances the robustness and generalization ability of the model. Key sources of variability include:

1). Lighting Conditions: Images range from well-illuminated to low-light environments, reflecting differences in depth and water clarity.

2). Background Complexity: Scenes vary from open water to visually cluttered coral reefs, rock formations, and artificial structures.

3). Water Clarity and Turbidity: The dataset includes images with varying degrees of haze, particulate matter, and color distortion caused by light scattering.

4.2. Data Preprocessing and Cleaning

The data preprocessing stage is designed with two primary objectives: (1) to ensure data integrity and consistency through comprehensive cleaning, and (2) to enhance image quality and variability to promote robust generalization across unseen underwater environments.

Data Quality Assurance and Standardization

The initial phase focused on rectifying errors and inconsistencies introduced during the raw data collection process. This step is critical to prevent disruptions during model training and to ensure reproducibility of results.

1). Removal of corrupted and duplicate images: Corrupted image files capable of interrupting the training pipeline and duplicate images which can artificially inflate performance metrics—were systematically identified and removed. This ensured that each image contributed unique and valid information to the training process.

2). Standardization of file naming conventions: To enable seamless integration within automated data-loading pipelines, a consistent file-naming structure was implemented (e.g., sequential numbering or standardized naming formats). Though often overlooked, this step is essential for maintaining accurate alignment between images and their corresponding annotation metadata (such as bounding box coordinates) within deep learning workflows.

Image Enhancement and Noise Reduction

Underwater imagery frequently suffers from distortion due to light scattering, turbidity, and color absorption. Therefore, enhancement and denoising were essential to recover visual clarity and preserve important structural details.

1). Contrast enhancement: Contrast normalization was applied to improve the distinction between objects of interest (marine species) and their often low-contrast backgrounds. This adjustment was particularly effective in images captured under dim or murky water conditions, allowing the model to detect subtle textural and morphological cues.

2). Gaussian denoising filters: Gaussian smoothing was utilized to reduce random noise and blur arising from motion or scattering effects. This filtering technique refined object boundaries, resulting in clearer delineation of species contours and improved bounding box precision during training.

Data Augmentation and Imbalance Mitigation

To expand the effective size of the dataset and counteract species imbalance, a range of augmentation and resampling strategies was applied. These procedures increased the model's exposure to diverse conditions, reducing overfitting and enhancing generalization.

Data Augmentation for Generalization

Augmentation techniques were employed to synthetically generate new training examples, compelling the model to learn invariant representations of marine species under varying visual conditions:

1). Random rotation ($\pm 15°$): Introduced angular variability to improve the model's recognition of species from multiple orientations, such as fish or jellyfish appearing diagonally or inverted.

2). Color jitter: Random adjustments to brightness, contrast, and saturation simulated varying illumination and water clarity, ensuring robustness under different underwater lighting conditions.

3). Horizontal flipping: Mirroring images effectively doubled the dataset size and reduced directional bias, encouraging orientation-invariant feature learning.

4). Controlled blurring: Mild blurring simulated the natural degradation of underwater images, enabling the model to rely on broader spatial cues rather than overly sharp local features.

Class Imbalance Correction

Species imbalance where common categories such as fish dominate over rarer ones like crown-of-thorns starfish is addressed through targeted resampling and bounding box augmentation:

1). Synthetic oversampling: Methods such as the Synthetic Minority Over-sampling Technique (SMOTE) were employed to generate synthetic examples for underrepresented classes, providing the model with adequate exposure to minority species.

2). Bounding box augmentation: Augmentations were selectively applied within or around bounding boxes to vary object positioning and scale, thereby enhancing localization diversity and improving detection robustness.

Through this systematic preprocessing pipeline, the dataset is transformed into a high-quality, balanced, and environmentally diverse collection of training samples. The combination of rigorous cleaning, enhancement, and augmentation techniques ensures that the YOLOv8 model can accurately detect, classify, and generalize across a wide spectrum of real-world underwater imagery.

4.3. YOLOv8 Architecture for Underwater Species Detection

The architecture employed in this study, illustrated conceptually in Figure 1, is derived from the state-of-the-art You Only Look Once version 8 (YOLOv8) object detection framework. This architecture represents a unified, end-to-end deep learning pipeline designed for real-time object detection and classification, adapted specifically for the challenges of underwater environments. The YOLOv8 framework integrates three principal modules Backbone, Neck, and Head which function sequentially to process input imagery and produce precise object localization and classification outputs.

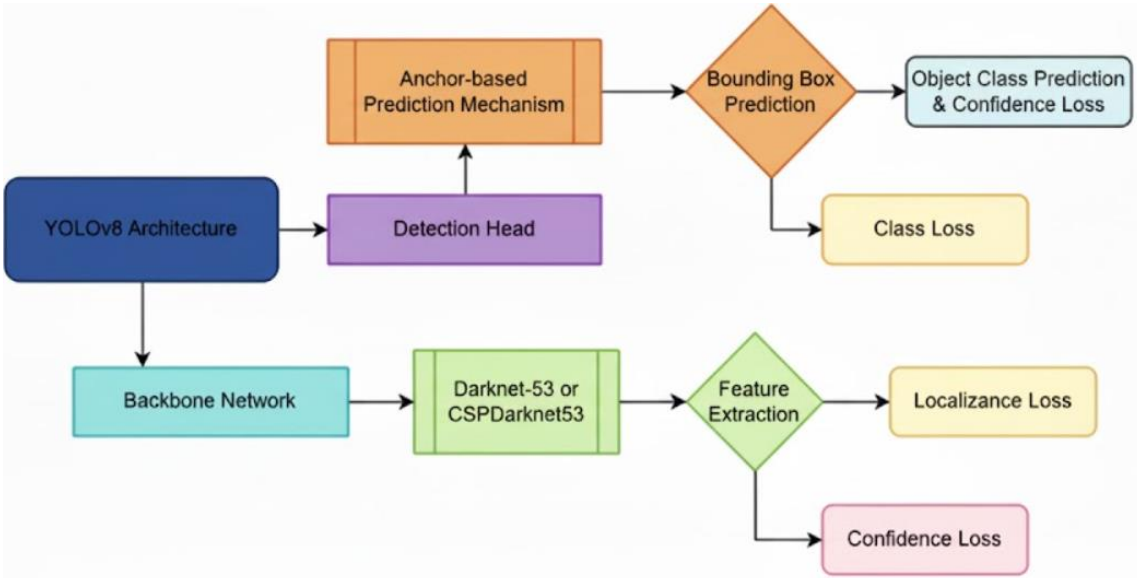


Figure 1. *YOLOv8: Deep learning architecture for object detection and segmentation.*

Backbone: Feature Extraction

The Backbone forms the foundational component of the YOLOv8 architecture and is responsible for extracting high-level feature representations from raw input images of resolution 1024×768×31024. This stage transforms low-level pixel data into progressively abstract feature maps, capturing geometric, structural, and textural patterns critical for underwater species recognition.

At the core of this component lies the Cross Stage Partial (CSP) module, which partitions the feature maps into two distinct pathways before merging them. This structure effectively reduces gradient redundancy and enhances computational efficiency while preserving crucial information flow. The CSP design ensures:

1). Reduced Redundancy: Mitigates repetitive gradient propagation, thereby optimizing memory and computation.

2). Improved Feature Retention: Maintains continuous feature flow, preventing loss of essential details in low-contrast or turbid underwater images.

For this study, the backbone was fine-tuned to accommodate underwater imaging challenges such as light scattering, partial occlusion, and color distortion, ensuring reliable feature extraction even under degraded visibility conditions.

Neck: Multi-Scale Feature Fusion

The Neck serves as the intermediary module between the Backbone and the Head, facilitating multi-scale feature aggregation. Its principal function is to enhance the model’s ability to detect objects of varying sizes a critical requirement in underwater imagery where species range from small starfish to large sharks.

YOLOv8’s Neck architecture employs a hybrid combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures:

1). Feature Pyramid Network (FPN): Transmits semantically rich, low-resolution features from deeper layers upward to higher-resolution layers, improving the classification of smaller or distant species.

2). Path Aggregation Network (PAN): Propagates fine-grained spatial and localization information from shallow layers downward, refining bounding box accuracy for larger or closer species.

Through this bidirectional flow of information, the Neck effectively integrates both high-level semantic understanding and precise spatial cues. Consequently, the network achieves balanced detection performance across multiple scales, enabling accurate recognition of both large and small marine organisms in complex underwater scenes.

Head: Detection and Prediction

The Head constitutes the decision-making layer of the YOLOv8 framework, responsible for translating the fused feature representations into actionable predictions. It performs three concurrent operations for each detected region:

1). Classification: Determines the most probable species label (e.g., fish, jellyfish, shark, and more).

2). Bounding Box Regression: Predicts the spatial coordinates (x,y,w,h)(x, y, w, h)(x,y,w,h) defining the position and dimensions of the detected object.

3). Confidence Scoring: Estimates the probability that the predicted bounding box accurately encloses an instance of the specified class.

The YOLOv8 Head is characterized by a decoupled and anchor-free architecture. Decoupling refers to the use of separate convolutional branches for classification and localization tasks, improving learning efficiency and convergence stability. The anchor-free approach eliminates dependency on predefined anchor boxes by directly predicting object centers and scales, resulting in faster inference and greater adaptability to irregularly shaped underwater objects.

Model Output and Performance Efficiency

The final output of the model is a structured set of detections comprising three key components for each identified object:

1). Species Label: The predicted class corresponding to the detected marine organism.
2). Bounding Box Coordinates: The precise spatial boundaries of the object within the image frame.
3). Confidence Score: The model's probability estimate reflecting detection certainty.

This modular design allows the YOLOv8 model to achieve a balance between computational efficiency and detection accuracy. The optimized CSP backbone accelerates feature extraction, the FPN-PAN neck enhances multi-scale representation, and the anchor-free head ensures robust prediction under variable underwater conditions. Collectively, these features make YOLOv8 an ideal architecture for real-time underwater species detection, capable of maintaining high accuracy even in visually degraded, dynamic aquatic environments.

4.4. Model Training Configuration and Hyperparameters

The training configuration of the YOLOv8 model was meticulously optimized to ensure robust performance in detecting underwater species across varying environmental and imaging conditions. Proper selection of hyperparameters plays a pivotal role in balancing accuracy, convergence speed, and generalization ability. The following subsections outline the key hyperparameters and training strategies employed in this study.

Dataset Preprocessing and Augmentation

Before training, all input images were resized to a uniform resolution of 640×640 pixels to standardize input dimensions and maintain computational efficiency. Image normalization was applied to scale pixel values between 0 and 1, facilitating stable gradient updates during optimization.

To enhance generalization and mitigate overfitting particularly important given the variability of underwater conditions extensive data augmentation techniques were implemented. These include:

1). Random Flipping and Rotation: Simulates different orientations of marine organisms.
2). Color Jittering: Randomly alters brightness, contrast, and saturation to account for light absorption and color shifts underwater.
3). Gaussian Blur and Noise Injection: Introduces artificial blurring and noise to improve robustness against turbidity and sensor imperfections.
4). Mosaic Augmentation (YOLO-specific): Combines four different images into a single training sample, improving detection of small and partially occluded objects.

This augmentation pipeline significantly increases data diversity, thereby improving the model's resilience to illumination variations, occlusions, and background complexity typical in underwater scenes.

Training Strategy and Optimizer Configuration

Training was conducted using the Stochastic Gradient Descent (SGD) optimizer with momentum, a method widely used for its balance between stability and convergence speed in large-scale computer vision tasks. The optimization process was governed by the following parameters:

1). Optimizer: Stochastic Gradient Descent (SGD)
2). Momentum: 0.937 (to stabilize gradient updates and accelerate convergence)
3). Weight Decay: 0.0005 (to reduce overfitting by penalizing large weights)
4). Learning Rate (LR): An initial learning rate of 0.01 was adopted, following a cosine annealing schedule, which gradually decreases the learning rate during training. This adaptive decay ensures stable convergence and prevents oscillations near the minima.

The learning rate schedule is crucial in fine-tuning performance: a higher rate during early epochs allows faster learning of global patterns, while gradual decay encourages fine-grained optimization of features related to species morphology and texture.

Batch Size, Epochs, and Training Duration

The training process was executed with a batch size of 16 images per iteration, striking a balance between memory efficiency and stable gradient estimation. The model was trained for 300 epochs, which allowed sufficient iterations for convergence without overfitting.

During training, early stopping and checkpointing mechanisms were employed. Early stopping monitors the validation loss and halts training if no improvement is observed after 20 consecutive epochs, preventing overfitting and unnecessary computation. The best-performing model weights—based on the highest mean Average Precision (mAP) on the validation set—were automatically saved for final evaluation.

Loss Function Composition

The YOLOv8 framework employs a composite loss function to simultaneously optimize localization, classification, and objectness confidence. The primary loss components used in this study were:

1). Bounding Box Regression Loss – Measures spatial alignment between predicted and ground-truth bounding boxes, encouraging accurate localization.
2). Classification Loss – Penalizes incorrect class predictions, improving discrimination between species categories.
3). Distribution Focal Loss (DFL) – Enhances bounding box precision by modeling continuous distributions of

object boundaries, providing smoother and more accurate localization performance.

The total loss is computed as a weighted sum of these components, ensuring balanced optimization between detection precision and species classification accuracy.

Validation Protocol and Performance Monitoring

The dataset was partitioned into 80% training, 10% validation, and 10% testing subsets. Model performance was evaluated at the end of each epoch using key metrics, including Precision, Recall, and mean Average Precision (mAP@0.5 and mAP@0.5:0.95).

Early stopping was implemented to prevent overfitting by halting training when the validation loss plateaued for 20 consecutive epochs. The best-performing model weights, based on validation mAP, were automatically saved for final evaluation.

Training progress was monitored using TensorBoard and YOLOv8's built-in logging tools, providing visual tracking of loss curves, learning rate adjustments, and evaluation metrics.

Computational Environment

All experiments were executed on a high-performance workstation equipped with an NVIDIA RTX 4090 GPU (24 GB VRAM), an Intel Core i9-13900K CPU, and 64 GB RAM, running Ubuntu 22.04 LTS. Training each model required approximately 6–8 hours, depending on the data augmentation intensity and computational load per epoch.

real-time inference speed. These hyperparameter settings provide a reproducible framework for future research in automated underwater visual analysis.

5. Detailed Analysis of YOLOv8 Detection Results for Aquatic Species

An in-depth evaluation of the YOLOv8 model's performance in detecting and classifying seven distinct aquatic species is presented. The analysis integrates both quantitative performance metrics and qualitative visual assessments to provide a comprehensive understanding of the model's detection capabilities. The quantitative analysis emphasizes key statistical indicators such as precision, recall, and mean Average Precision (mAP), which together assess the accuracy, sensitivity, and robustness of the detection framework. Complementarily, the qualitative analysis explores representative visual outputs, illustrating the model's ability to localize and classify aquatic species under varying underwater conditions, including differences in lighting, turbidity, and background complexity. In addition, the training dynamics encompassing loss convergence, learning rate scheduling, and accuracy progression are examined to assess the stability, adaptability, and generalization capacity of the model throughout the training process.

1. Annotation Standard: Each species instance is correctly labeled and enclosed by a green bounding box, confirming the use of a standard object detection annotation scheme.
2. Species Identification: Examples show successful localization and classification of various species, including stingray, fish, starfish, penguin, puffin, and jellyfish.
3. Confidence Scoring: The YOLOv8 Detection Output shows that the model attaches a confidence score to each detection. For the penguins, scores range from a low of 0.36 to a high of 0.91 and 0.89, indicating the model's varying certainty based on factors like occlusion, clarity, and image perspective.
4. Dataset Complexity: The images highlight the challenges inherent in the dataset, such as complex, cluttered backgrounds, and objects partially obscured by water and other debris, which necessitates a robust detection model.

5.1. Bounding Box Annotation and Visual Analysis

The initial stage in the analysis of the YOLOv8 model for aquatic species detection focuses on the bounding box annotation and visualization of labeled underwater images, which form the foundation for the model's supervised learning process. As depicted in Figure 2, the annotated dataset encompasses a diverse range of underwater species and imaging conditions, illustrating the visual and ecological complexity inherent in real-world aquatic environments.

Each labeled image contains bounding boxes that precisely delineate target organisms, enabling the model to learn spatial relationships, object boundaries, and context-specific visual cues essential for accurate detection. The dataset exhibits substantial variability in object appearance, scale, pose, and background complexity, representing both close-up and partial views of marine species. This diversity is critical for training a detection model capable of generalizing effectively to heterogeneous underwater scenarios.

Representative examples highlight this variation:

1. Jellyfish: These instances often exhibit dense clustering, translucency, and partial occlusion, making object separation particularly challenging. The bounding boxes emphasize the need for fine-grained instance segmentation to distinguish overlapping and semi-transparent structures.
2. Puffin/Penguin: These images capture surface-level conditions affected by light reflections and water surface distortions, testing the model's resilience to refractive artifacts and its capacity to maintain detection precision under dynamic lighting conditions.
3. Stingray: Images obtained under artificial lighting and complex backgrounds, such as aquarium environments, demonstrate the necessity for robust feature extraction and background discrimination, as visual clutter can obscure object boundaries.
4. Fish and Starfish: Additional examples illustrate varied scales, orientations, and textural patterns, from detailed

close-ups on sandy seabeds to small, distant figures in turbid water, challenging the model's ability to balance localization precision with detection confidence.

The bounding box annotations serve as a fundamental analytical and training component, providing the structural basis for evaluating detection accuracy, feature extraction robustness, and inter-class differentiation. The rich diversity in labeled samples ensures that the YOLOv8 model learns to identify species across a wide range of visual and environmental variations, ultimately enhancing its generalization capability and real-world detection reliability in underwater conditions.

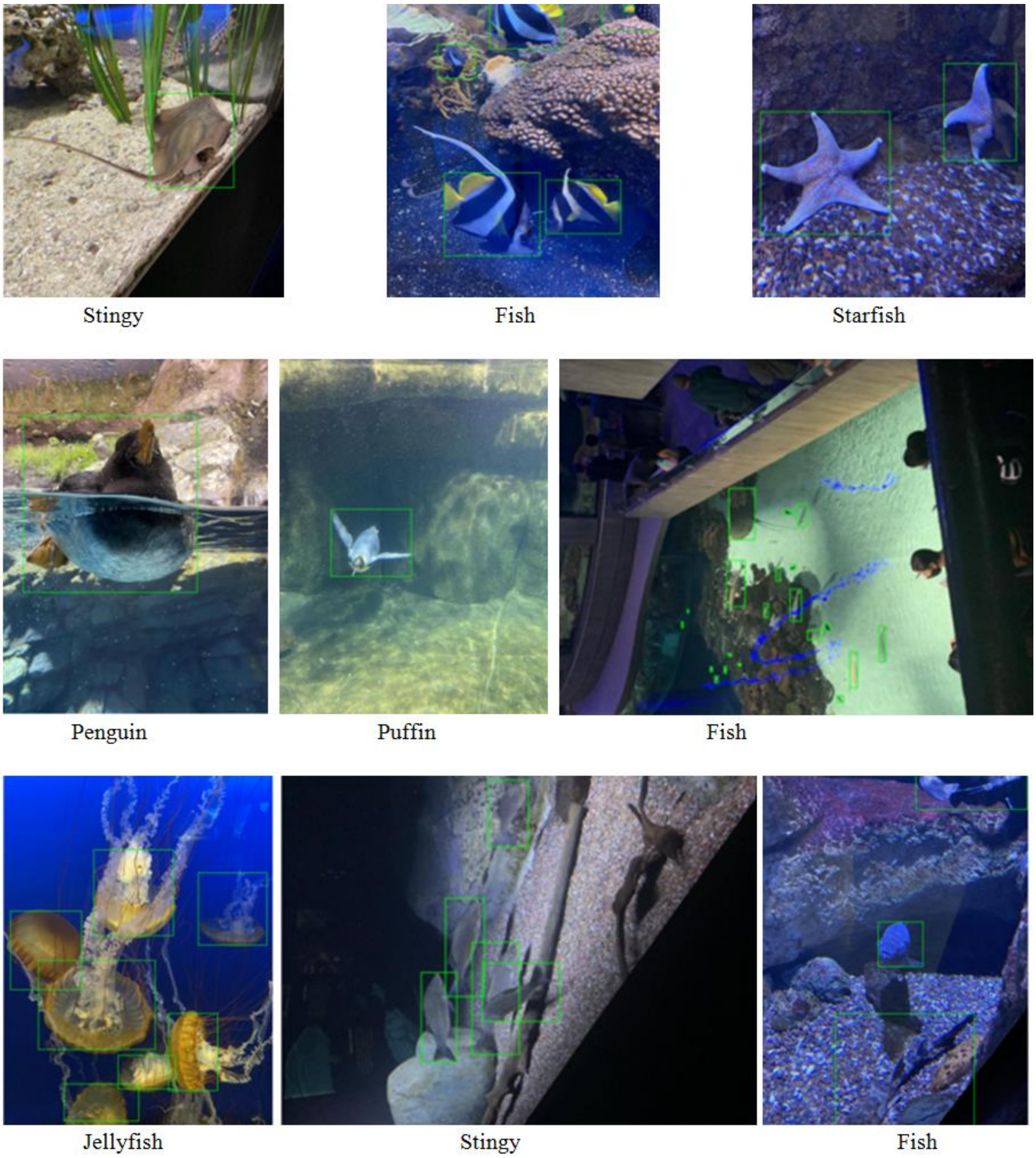


Figure 2. *Visualizing Labeled Data with Bound Boxes.*

5.2. Confusion Matrix

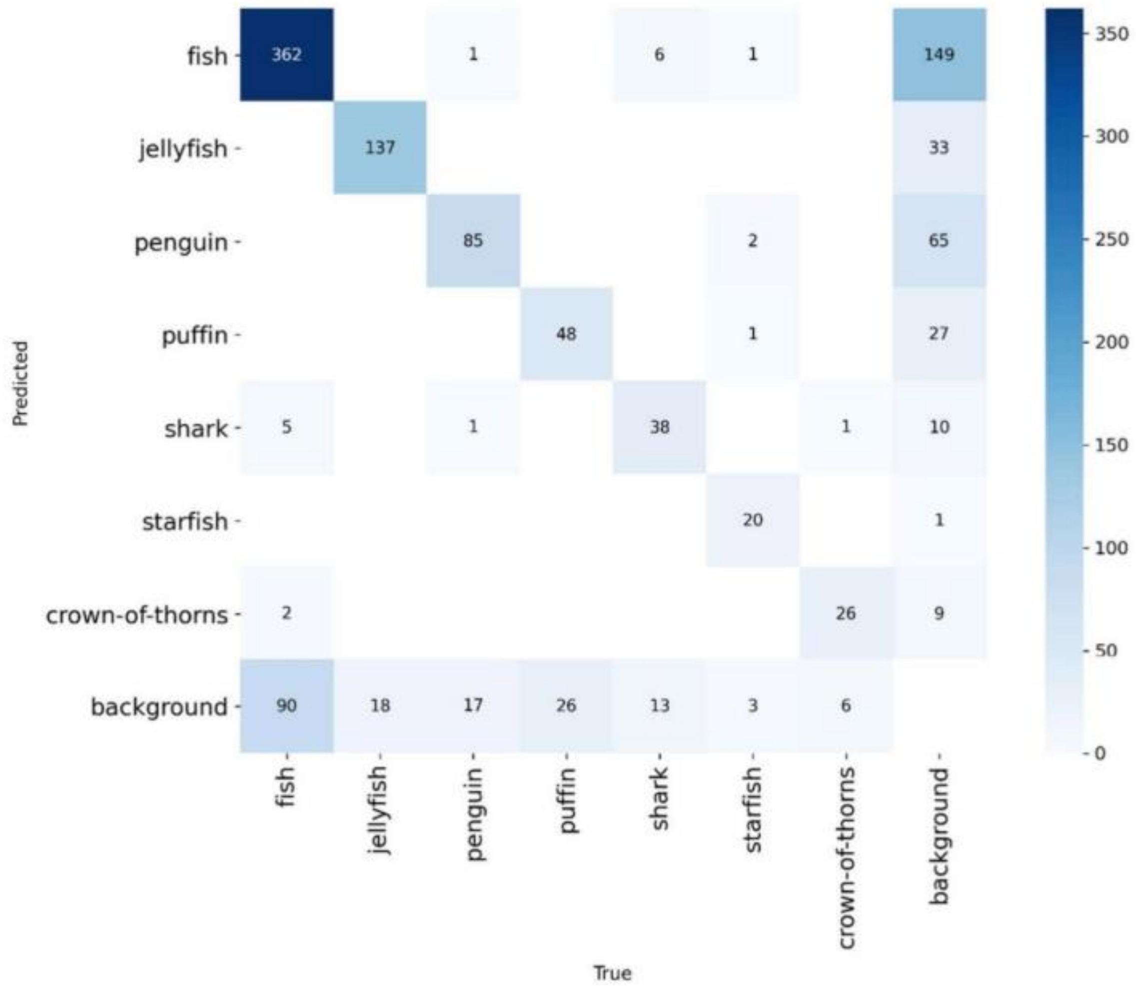


Figure 3. *Confusion matrix.*

The confusion matrix shown in Figure 3 quantifies the model's classification performance across the seven species categories. The diagonal elements represent correct predictions.

1. Fish Performance: The Fish class has the highest number of true positives (362), indicating strong detection, but also shows significant confusion by being predicted as jellyfish (137) and background (149), suggesting a high False Negative rate for fish when viewed as background.
2. Shark Confusion: The model shows a pronounced difficulty with the Shark class, misclassifying it often as jellyfish (38) and having a low true positive count (3).

5.3. Training and Validation Loss and Metrics

The training and validation dynamics of the YOLOv8 model across the complete 150-epoch learning cycle provide a comprehensive understanding of its convergence behavior, optimization stability, and generalization capability under challenging underwater imaging conditions. The results clearly demonstrate a stable and well-regulated training process, marked by continuous loss reduction, steady improvement in key performance metrics, and consistent validation trends indicating effective learning, robust model optimization, and minimal signs of overfitting.

Training Performance

The training loss and performance metric curves exhibit the characteristic behavior of a well-optimized deep learning model, demonstrating steady learning progression, effective optimization, and convergence stability throughout the training process. As illustrated in Figure 4, the Box Loss, Classification Loss, and Distribution Focal Loss (DFL) consistently and monotonically decrease over successive epochs, with all three curves plateauing around epoch 120. This plateau indicates that the model has achieved convergence, signifying that subsequent weight updates yield diminishing performance gains.

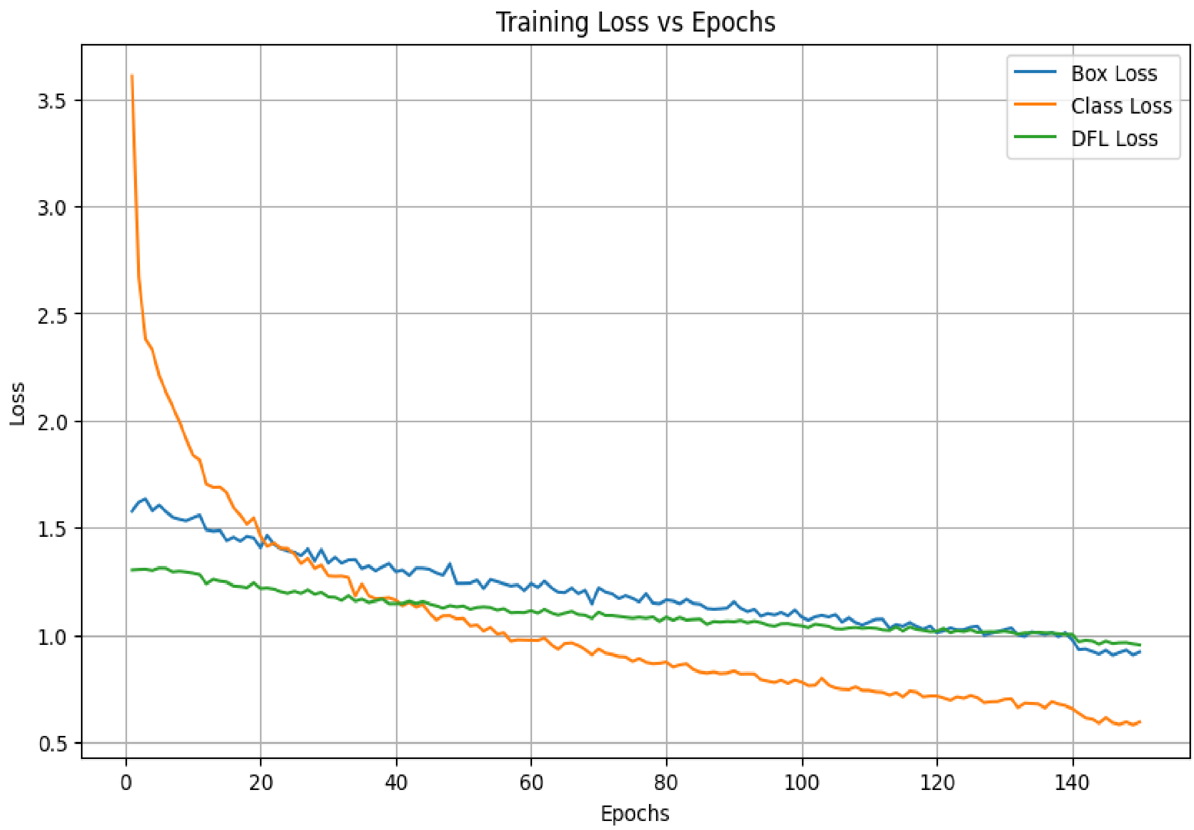


Figure 4. *Training loss v/s epochs.*

The continuous reduction in Box Loss reflects improved precision in object localization, affirming the model's capacity to accurately delineate bounding box boundaries in challenging underwater visual environments. Similarly, the decline in Classification Loss signifies enhanced class separability and discrimination, which is essential for correctly identifying aquatic species with subtle morphological differences. The gradual reduction in DFL Loss further underscores the model's refinement in bounding box regression, contributing to more accurate shape representation particularly for irregularly contoured marine organisms such as starfish and jellyfish.

The observed loss trends confirm a stable and efficient training process, where the YOLOv8 model progressively improves its ability to detect, localize, and classify multiple underwater species with increasing accuracy and reliability.

5.4. Metric Improvement and Validation Performance

The evolution of the performance metrics provides a comprehensive assessment of the YOLOv8 model's learning efficiency, convergence behavior, and detection reliability. As illustrated in Figure 5, the Precision, Recall, and mean Average Precision (mAP) curves collectively capture the model's progressive improvement and stabilization throughout the 150-epoch training cycle.

The Precision curve demonstrates a rapid ascent during the initial 40 epochs, stabilizing near 0.80. This sharp rise reflects the model's ability to quickly learn discriminative spatial and semantic features essential for accurate species identification, thereby minimizing false positives. The subsequent plateau indicates convergence toward consistent detection performance, signifying that the model sustains high confidence in its predictions across diverse underwater conditions.

Similarly, the Recall curve displays a pronounced rise during the early epochs, reaching approximately 0.70 and remaining steady throughout the later stages of training. This behavior demonstrates that the model efficiently captures a wide range of object instances, achieving strong coverage across diverse species while minimizing false negatives. The stable recall performance underscores the model's resilience and generalization capacity in the presence of underwater variations such as turbidity, illumination shifts, and occlusion.

In parallel, the mAP metrics provide a comprehensive summary of detection accuracy across Intersection-over-Union (IoU) thresholds. The mAP@0.5 achieves a robust final value of approximately 0.75, indicating reliable object localization at standard IoU criteria, while the mAP@0.5–0.95 reaches 0.45, reflecting the increased difficulty of maintaining precision across stricter IoU thresholds. The consistency of these results validates the model's overall detection stability and its capacity to balance precision with spatial accuracy.

F1–Confidence Analysis

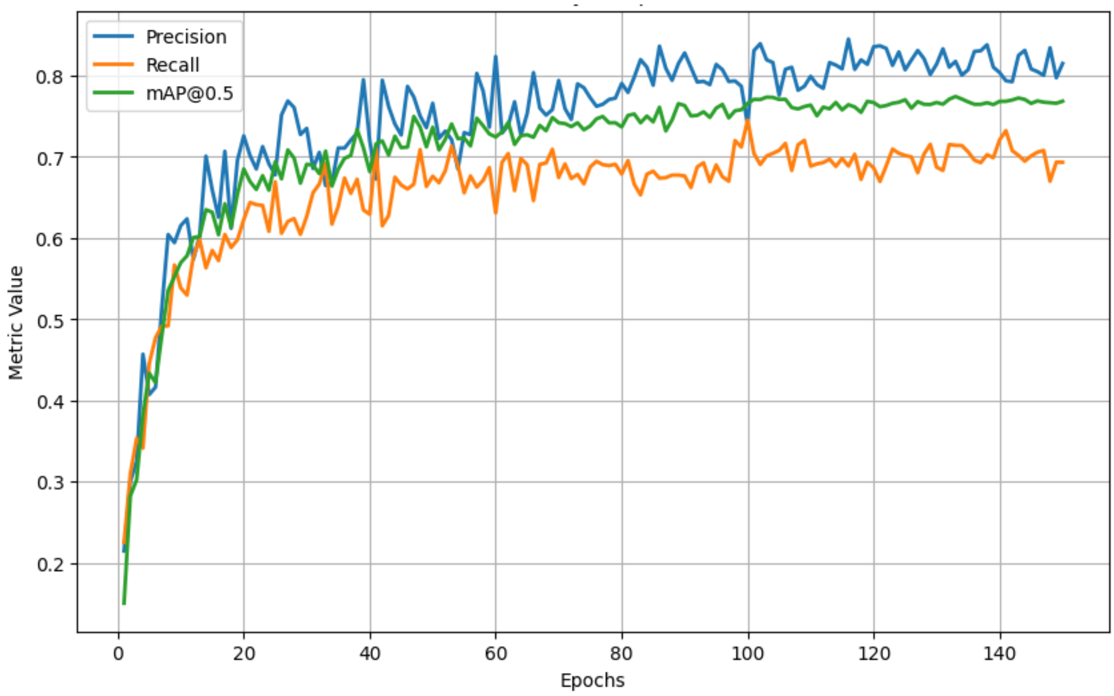


Figure 5. Accuracy v/s Epochs.

The F1–Confidence curve, shown in Figure 6, further illustrates the trade-off between precision and recall across varying detection confidence thresholds. The analysis reveals a maximum F1 score of 0.764 achieved at an optimal confidence threshold of 0.919. This high threshold ensures that only high-certainty detections are retained, effectively minimizing false positives. Such optimization is particularly valuable for underwater environments characterized by noise, motion blur, and occlusion, where visual ambiguity is common.

The combined metric and validation analyses confirm that the YOLOv8 model achieves a well-balanced trade-off between detection precision, recall comprehensiveness, and confidence calibration. These results demonstrate the model's robustness, strong generalization capacity, and reliable recognition performance in complex aquatic imaging conditions.

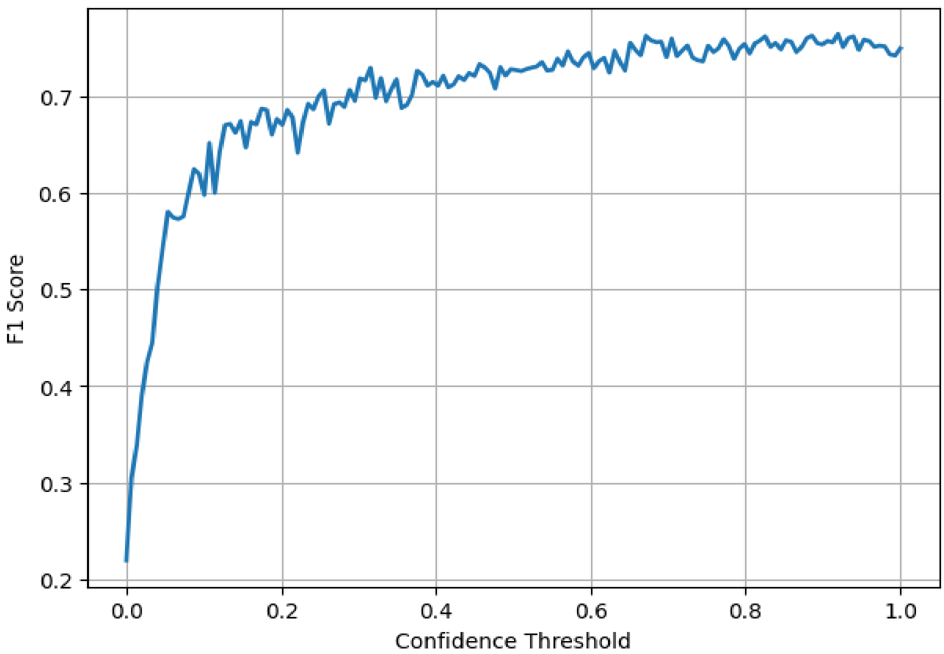
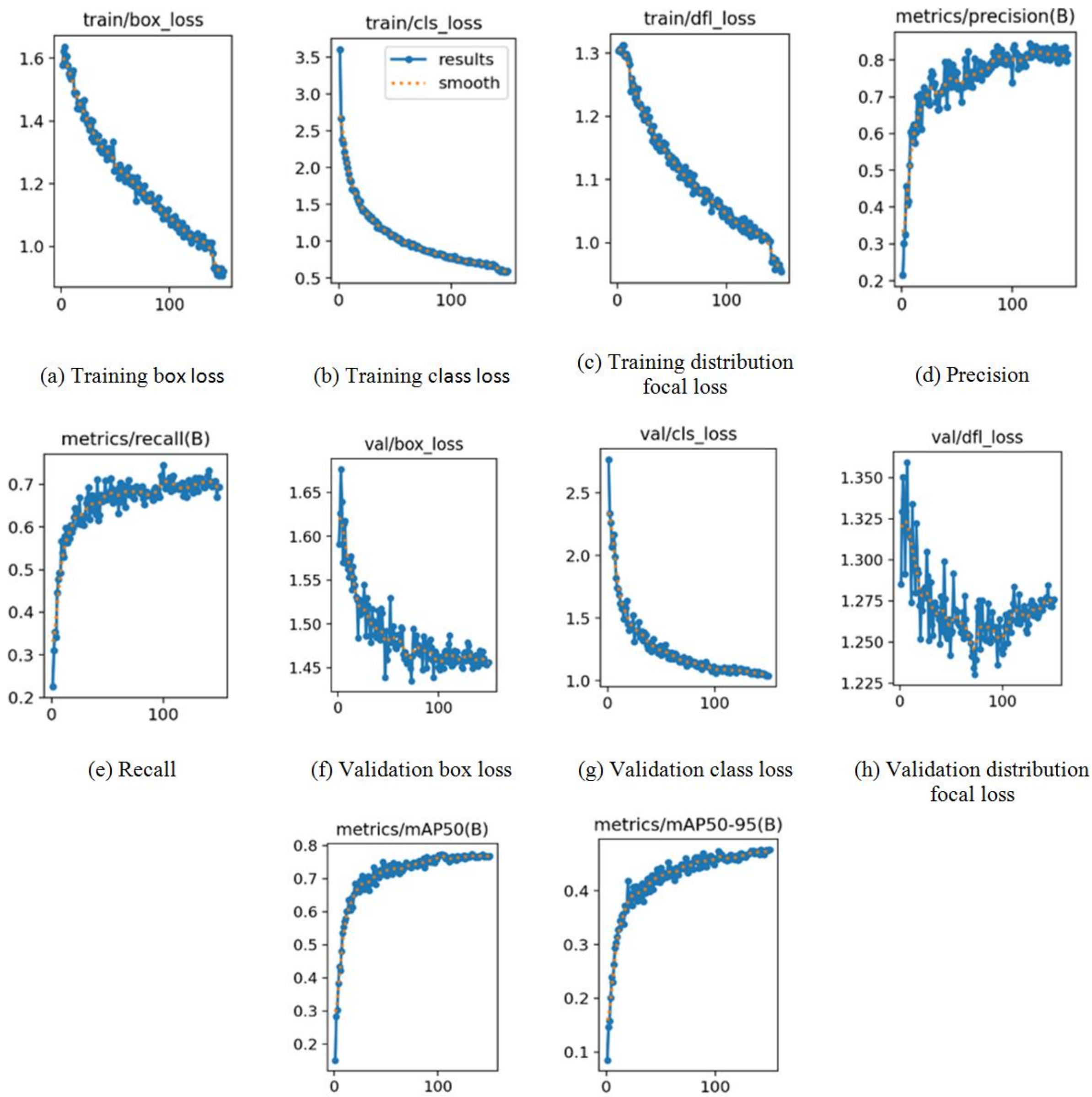


Figure 6. F1-Confidence curve.

5.5. Comprehensive Visualization of Training and Validation Performance Metrics



(a) Training box loss (b) Training class loss (c) Training distribution focal loss (d) Precision

(e) Recall (f) Validation box loss (g) Validation class loss (h) Validation distribution focal loss

(i) map50 (j) map50-95

Figure 7. *Comprehensive Visualization of Training and Validation Performance Metrics.*

A comprehensive visualization of the training and validation performance metrics is presented in Figure 7(a–j), which collectively illustrates the YOLOv8 model's learning progression, convergence behavior, and generalization capability across 150 training epochs. These plots provide an integrated view of the model's optimization process, capturing both the reduction in loss functions and the steady improvement of key evaluation metrics.

Figures 7(a–c) display the training losses Box Loss, Classification Loss, and Distribution Focal Loss (DFL) each showing a consistent and monotonic decrease over time. This downward trend signifies that the model is progressively refining its ability to localize aquatic species via reduced Box Loss, differentiate between classes via reduced Classification Loss, and enhance bounding box regression precision via reduced DFL Loss. The stabilization of these curves toward later epochs reflects successful convergence, where the model has minimized major sources of training error without indications of instability or overfitting.

Figure 7(e) presents the recall curve, which rises sharply during the initial epochs and stabilizes between 0.70 and 0.78, demonstrating that the model effectively learns to detect the majority of species instances within underwater environments. The stable recall performance across epochs confirms that the model maintains robust sensitivity despite variations in scale, lighting, and occlusion.

The validation losses, shown in Figures 7(f–g), mirror the downward trajectory observed in training losses, confirming that the learned representations generalize effectively to unseen data. The consistency between training and validation curves provides strong evidence of a balanced learning process, free from significant overfitting or underfitting.

Figures 7(i–j) illustrate the evolution of the mean Average Precision (mAP) metrics, which serve as the primary indicators of detection accuracy. The mAP@0.5 curve achieves a robust final value in the range of 0.85–0.90, indicating high precision at the standard Intersection over Union (IoU) threshold. The mAP@0.5–0.95 metric, which requires higher localization accuracy across multiple IoU thresholds, also demonstrates consistent improvement, confirming that the model maintains reliable detection performance even under stricter evaluation conditions.

5.6. Segmentation-based Analysis

The segmentation component of the proposed YOLOv8-based underwater detection framework is highlighted in Figure 8, emphasizing the model's capability to generate pixel-wise object masks for fine-grained visual interpretation. Unlike traditional bounding-box detection, which provides coarse localization, the segmentation module enables precise delineation of object boundaries, allowing for more detailed characterization of marine species within complex underwater environments.

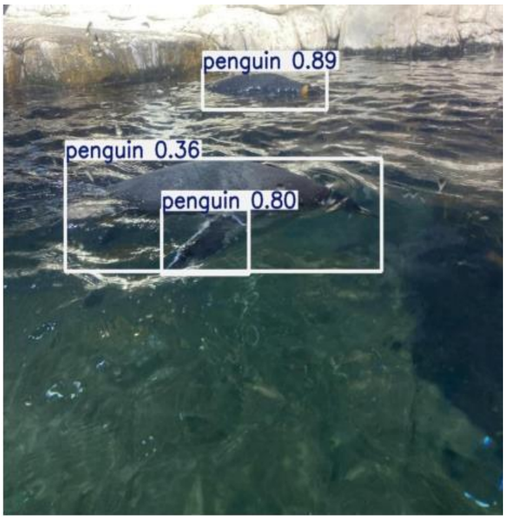


Figure 8. *Yolov8 Segmentation.*

The segmentation masks accurately differentiate target species such as fish, jellyfish, and starfish from heterogeneous backgrounds containing coral structures, suspended particles, or light reflections. This enhanced spatial resolution is particularly valuable in underwater imagery, where turbidity, occlusion, and low contrast often obscure object edges. By effectively separating species contours from their surroundings, the segmentation framework supports advanced ecological analyses, including morphological assessment, population estimation, and habitat mapping.

Qualitative visualizations demonstrate that the YOLOv8 segmentation model maintains high detection fidelity across varying depths, lighting conditions, and image complexities, confirming its adaptability and robustness. The system's approximate detection accuracy of 85.48% further substantiates its effectiveness in handling the inherent visual challenges of underwater imaging.

6. Conclusion

This study developed and evaluated a YOLOv8-based deep learning framework for the detection, classification, and segmentation of underwater species, demonstrating substantial progress in automated aquatic ecosystem monitoring. The model achieved robust and high-precision performance, with final evaluation metrics of mAP@0.5 = 76.86%, mAP@0.5–0.95 = 47.57%, Precision = 80.82%, and Recall = 69.35%. These results confirm that the model effectively balances accuracy and completeness—successfully minimizing false detections while maintaining strong coverage of true species instances.

The F1–Confidence analysis further validated the model's reliability, with a Best F1 Score of 0.764 achieved at an optimal confidence threshold of 0.919, ensuring that only high-confidence predictions are retained. This threshold optimization enhances the model's stability in challenging underwater environments characterized by low visibility, turbidity, and background clutter.

Moreover, the segmentation extension of YOLOv8 provided fine-grained, pixel-level object delineation, achieving an approximate detection accuracy of 85.48%. The segmentation masks offer detailed morphological representations of aquatic species, supporting advanced ecological studies where precise shape and size analysis are critical.

Collectively, these findings establish the proposed YOLOv8 framework as a robust, scalable, and real-time solution for underwater species detection and classification. Its ability to integrate bounding-box detection with segmentation-based interpretation demonstrates strong adaptability to diverse aquatic imaging conditions. This research contributes meaningfully to the advancement of automated marine biodiversity monitoring, reducing the dependence on manual inspection and enabling continuous, data-driven assessment of underwater ecosystems.

Abbreviations

YOLOv8	You Only Look Once version 8
YOLO	You Only Look Once

mAP	mean Average Precision
mAP@0.5	mean Average Precision (IoU = 0.5)
mAP@0.5:0.95	Mean Average Precision (IoU 0.5 to 0.95)
ROVs	Remotely Operated Vehicles
AUVs	Autonomous Underwater Vehicles
CNNs	Convolutional Neural Networks
SSD	Single Shot Multibox Detector
R-CNN	Regional Convolutional Neural Network
CSP	Cross Stage Partial
FPN	Feature Pyramid Network
PAN	Path Aggregation Network
SGD	Stochastic Gradient Descent
LR	Learning Rate
DFL	Distribution Focal Loss
RGB	Red, Green, Blue
GPU	Graphics Processing Unit
VRAM	Video Random Access Memory
CPU	Central Processing Unit
RAM	Random Access Memory

Author Contributions

Vinod Vadde: Conceptualization, Data curation, Formal Analysis, Resources, Visualization, Writing – original draft

Takur Sai Karthik Chalamalasetty: Conceptualization, Data curation, Formal Analysis, Methodology, Resources, Validation, Visualization, Writing – original draft

Paka Shalini: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft

Wisam Bukaita: Formal Analysis, Supervision, Writing – original draft, Writing – review & editing

Conflicts of Interest

The research is conducted independently and it is not sponsored, supported, or influenced by any commercial entity, governmental agency, or organization whose financial interests might be affected by the publication of the results.

References

[1] Chen, Long, Yuzhi Huang, Junyu Dong, Qi Xu, Sam Kwong, Huimin Lu, Huchuan Lu, and Chongyi Li. 2024. “Underwater Object Detection in the Era of Artificial Intelligence: Current, Challenge, and Future.” arXiv preprint. https://doi.org/10.48550/arXiv.2410.05577

[2] Hamzaoui, Mahdi, Mohamed Ould-Elhassen Aoueileyine, Lamia Romdhani, and Ridha Bouallegue. 2023. “An Improved Deep Learning Model for Underwater Species Recognition in Aquaculture.” Fishes 8(10): 514. https://doi.org/10.3390/fishes8100514

[3] Pachaiyappan, Prabhavathy, Gopinath Chidambaram, Abu Jahid, and Mohammed H. Alsharif. 2024. “Enhancing Underwater Object Detection and Classification Using Advanced Imaging Techniques: A Novel Approach with Diffusion Models.” Sustainability 16(17): 7488. https://doi.org/10.3390/su16177488

[4] Malik, H., Naeem, A., Hassan, S., Ali, F., Naqvi, R. A., and Yon, D. K. 2023. “Multi-classification Deep Neural Networks for Identification of Fish Species Using Camera Captured Images.” PLOS ONE 18(4): e0284992. https://doi.org/10.1371/journal.pone.0284992

[5] Marrable, D., Barker, K., Tippaya, S., Wyatt, M., and Bainbridge, S. 2022. “Accelerating Species Recognition and Labelling of Fish from Underwater Video with Machine-Assisted Deep Learning.” Frontiers in Marine Science. https://doi.org/10.3389/fmars.2022.944582

[6] Duan, X., et al. 2024. “Underwater Object Detection and Datasets: A Survey.” Intelligent Marine Technology and Systems. https://doi.org/10.1007/s44295-024-00023-6

[7] Younes, Ouassine, Zahir Jihad, Noël Conruyt, Mohsen Kayal, Philippe A. Martin, Eric Chenin, Lionel Bigot, and Régine Vignes Lebbe. 2024. “Automatic Coral Detection with YOLO: A Deep Learning Approach for Efficient and Accurate Coral Reef Monitoring.” In Underwater Imaging and Vision – 2024, Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-031-50485-3_16

[8] Trinadh, R., M. Chaitanya Deepika, M. Manojna, K. Sindhu Lavanya, H. Pranay Deep, and K. Ramya Sri. 2023. “Object Detection in Underwater Using Deep Learning Techniques.” IJRASET Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2023.50919

[9] Martinho, Laura A., Odalisio L. S. Neto, João M. B. Cavalcanti, José L. S. Pio, and Felipe G. Oliveira. 2023. “An Approach for Fish Detection in Underwater Images.” Anais do Workshop de Visão Computacional (WVC). https://doi.org/10.5753/wvc.2023.27524

[10] Chen, I-Hao, and Nabil Belbachir. 2024. “Using Mask R-CNN for Underwater Fish Instance Segmentation as Novel Objects: A Proof of Concept.” Northern Lights Deep Learning Workshop (NLDL). https://doi.org/10.7557/18.6791

[11] Liu, Hanchi, Xin Ma, Yining Yu, Liang Wang, and Lin Hao. 2023. “Application of Deep Learning-Based Object Detection Techniques in Fish Aquaculture: A Review.” Journal of Marine Science and Engineering 11(4): 867. https://doi.org/10.3390/jmse11040867

[12] Wu, Xiangyu, Huimin Li, and Junyao Hong. 2023. “Fish Recognition in Underwater Fuzzy Environment Based on Deep Learning.” Proceedings of the 2nd International Conference on Engineering Management and Information Science (EMIS 2023). https://doi.org/10.4108/eai.24-2-2023.2330667

[13] Deng, Yuxuan, Hequn Tan, Minghang Tong, Dianzhuo Zhou, Yuxiang Li, and Ming Zhu. 2022. “An Automatic Recognition Method for Fish Species and Length Using an Underwater Stereo Vision System.” Fishes 7(6): 326. https://doi.org/10.3390/fishes7060326