Aqua Vision: Few-Shot Learning Based Efficient Fish Identification in Challenging Aquatic Habitats

Vijaya J[1*], Bhomika Ratna Mandavi[2], Akshat Srivastava[3], Debashish Padhy[4]

[1]Assistant Professor/ Dept DSAI, IIITNR, Raipur, Chhattisgarh, India, 493661

[2]PG Student/ IIITNR, Raipur, Chhattisgarh, India, 493661

[3,4]UG Student/ IIITNR, Raipur, Chhattisgarh, India, 493661

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2025.12110121

Received: 29 November 2025; Accepted: 08 December 2025; Published: 18 December 2025

ABSTARCT

Aquatic ecosystems play a vital role in marine biodiversity and coastal protection, yet monitoring these habitats remains a significant challenge due to the scarcity of labeled data for training robust detection models. Traditional approaches often rely on extensive labeled datasets, which are costly and time-consuming to obtain, leading to a critical research gap in effective fish detection methodologies. This study introduces an innovative approach to fish detection by leveraging few-shot learning and pseudo-labeling techniques. We employ SimCLR, a contrastive learning framework, to pre-train a ResNet50-based encoder on unlabeled Deep Fish images, thereby extracting robust feature representations. These features are then utilized to train a Faster R-CNN object detection model using a limited set of labeled sea grass images. To further enhance the model's performance, we incorporate pseudo-labeling, a semi-supervised learning technique that generates additional training data from unlabeled images based on a confidence threshold. Our methodology demonstrates significant improvements in fish detection accuracy. The final model achieves an average precision of 0.8167 and recall of 0.7967, outperforming other state-of-the-art models such as YOLOv5 and RetinaNet. These results highlight the effectiveness of combining few-shot learning with pseudo-labeling in addressing the challenge of limited labeled data, paving the way for more efficient and accurate marine ecosystem monitoring.

Keywords: Fish detection, Few-shot learning, Pseudo- labeling, SimCLR, Faster R-CNN, Marine ecosystem monitoring

INTRODUCTION

The sustainable protection of marine ecosystems has become globally important due to increasing environmental changes and human pressures. Marine habitats such as coral reefs, wetlands, sea grass meadows, and mangroves support biodiversity, contribute to carbon sequestration, and protect coastal regions. They also sustain fisheries and livelihoods, playing a key role in food security. However, these ecosystems are increasingly threatened by habitat degradation, pollution, and climate change.

Fish, being highly sensitive to environmental variations, serve as important bioindicators of marine health. Traditional monitoring methods are labor-intensive, costly, and limited in scope, making efficient fish identification in challenging aquatic environments essential **[1-4].**

Recent advancements in artificial intelligence (AI), and in particular deep learning techniques, have dramatically transformed the landscape of ecological monitoring by automating the detection and classification of marine species in image and video data **[5-6]**. Convolutional neural networks (CNNs)

and emerging transformer-based models have demonstrated exceptional capabilities in visual recognition tasks across diverse domains including medical diagnostics, autonomous driving, and remote sensing **[7,8].** In the context of marine ecology, these models enable automated processing of large-scale underwater imagery to identify fish species and behaviors, facilitating continuous and high-resolution biodiversity assessments **[9]**. Nonetheless, the underwater domain presents unique challenges such as low lighting, high turbidity, occlusions caused by vegetation or other organisms, and color distortions, all of which significantly hamper model performance and generalizability **[10].**

A critical limitation in deploying deep learning for underwater applications is the scarcity of well-annotated datasets. Labeling marine species requires expert knowledge and is time-consuming due to ambiguous object boundaries, camouflage, and overlapping individuals in complex habitats [11]. To mitigate this bottleneck, recent research has explored techniques such as self-supervised learning (SSL) to exploit large volumes of unlabeled data by learning useful feature representations through contrastive or predictive tasks [12]. Complementing SSL, few-shot learning (FSL) frameworks have been developed to enable models to recognize novel fish species from only a handful of labeled examples by leveraging metric learning or meta-learning paradigms [13]. Additionally, semi-supervised learning (Semi-SL) methods like Fix Match utilize pseudo-labeling strategies to further harness unlabeled images and improve model robustness without incurring high annotation costs [14].

Building on these advances, we introduce **AquaVision**, a comprehensive detection and segmentation framework specifically tailored to the challenges of underwater fish identification in resource-constrained scenarios. AquaVision synergistically combines self-supervised representation learning via SimCLR [15], few-shot fine-tuning inspired by Prototypical Networks [16], and semi-supervised refinement using pseudo-labeling techniques [17] to maximize performance while minimizing the dependency on large labeled datasets. By initially learning robust and generalizable features from unlabeled underwater imagery, AquaVision adapts effectively to new fish species and habitats with very limited annotated data. The iterative semi-supervised learning process further enhances detection accuracy by leveraging confident predictions on unlabeled data.

Extensive evaluation against state-of-the-art object detectors such as YOLOv5 **[18]** and RetinaNet **[19]** demonstrates that AquaVision surpasses these baselines in precision, recall, and overall robustness, especially in challenging underwater conditions. This makes AquaVision a promising low-resource solution for scalable and automated marine biodiversity monitoring, contributing towards more effective conservation management and fostering the integration of AI technologies within ecological research.

LITERATURE REVIEW

The application of artificial intelligence, particularly deep learning, to marine visual analysis has gained considerable momentum in recent years. Researchers have leveraged these techniques for diverse tasks such as fish species classification, coral reef health assessment, detection of marine debris, and identification of underwater anomalies **[20,21].** Several benchmark datasets including Fish4Knowledge **[22],** the Underwater Robot Picking Contest (URPC) dataset **[23],** and DeepFish [24] have played a pivotal role in advancing algorithm development by providing annotated underwater images for training and evaluation. Despite this progress, a persistent challenge remains in translating these models to real-world marine environments due to factors such as domain shifts, variability in water conditions, and image degradation caused by factors like turbidity and lighting inconsistencies. A fundamental obstacle in underwater visual recognition is the limited availability of high-quality annotated data. Accurate labeling is often hindered by occlusions, cryptic coloration of marine organisms, and the requirement of expert taxonomic knowledge **[25].**

Zhang et al. [26] highlighted the intensive labor and potential for human error involved in manual annotation processes, particularly in ecologically complex habitats such as seagrass meadows and coral reefs. Although conventional approaches such as data augmentation, domain adaptation, and transfer learning have been employed to mitigate data scarcity, their effectiveness is often habitat-specific and

does not generalize well across diverse underwater settings.

To address the annotation bottleneck, self-supervised learning (SSL) has emerged as a promising paradigm. SSL techniques like SimCLR [15], MoCo [27], and BYOL [28] allow models to learn meaningful visual representations from large collections of unlabeled images by contrasting augmented views of the same sample. These approaches have shown considerable promise in underwater scenarios; where the cost and complexity of acquiring labeled datasets pose significant barriers. By extracting generalizable features, SSL methods enable downstream tasks such as classification and detection with fewer labeled samples.

Complementary to SSL, few-shot learning (FSL) strategies have been introduced to enable efficient recognition of new categories using only a handful of annotated instances. Techniques such as Prototypical Networks [16], Matching Networks [29], and Model-Agnostic Meta-Learning (MAML) [30] have demonstrated strong performance on low-data classification challenges. Specifically, Lu et al. [31] applied metric-based few-shot learning for marine biodiversity monitoring, achieving high accuracy in identifying fish species from limited labeled examples. FSL approaches thus provide a practical solution for adapting models to novel species or environments with minimal annotation effort. Semi-supervised learning (Semi-SL) methods further extend model capabilities by leveraging unlabeled data alongside limited labeled samples. Frameworks such as FixMatch [14] and Mean Teacher [17] utilize pseudo-labeling and consistency regularization to iteratively improve model predictions without requiring additional manual labels. These techniques are particularly advantageous in marine contexts where large volumes of unlabeled underwater video footage exist but annotations remain scarce. Incorporating such semi-supervised approaches enables models to achieve enhanced robustness and accuracy in diverse underwater conditions.

Traditional detectors like YOLOv5 [18] and RetinaNet [19] are fast and accurate on general datasets but struggle with underwater challenges such as low visibility, occlusions, and dynamic lighting. **AquaVision** overcomes these limitations by combining a robust feature extractor trained with contrastive self-supervised learning, refined through few-shot and semi-supervised fine-tuning. While SSL, FSL, and Semi-SL have been applied individually, unified frameworks are lacking. AquaVision fills this gap with an end-to-end pipeline that adapts to new fish species, diverse habitats, and changing conditions with minimal labeled data, enhancing AI-assisted marine ecosystem monitoring and conservation.

Proposed HAR Model

The proposed framework is specifically designed to address the unique challenges of underwater fish classification, particularly under low-resource conditions where annotated data is limited. Our proposed method leverages a multi-stage deep learning pipeline that integrates self-supervised representation learning, few-shot object detection, and semi-supervised training with pseudo-labels to ensure enhanced performance, adaptability to domain shifts, and robustness to visual noise common in underwater imagery.

The architecture comprises four core components:

1. Self-supervised feature extraction using SimCLR on unlabeled underwater datasets,
2. Task-specific fine-tuning via few-shot object detection using a Faster R-CNN backbone,
3. Semi-supervised refinement via pseudo-labeling of unlabeled instances, and
4. Joint retraining using both true and pseudo annotations to maximize learning signal.

An overview of the system architecture is illustrated in Figure 1, capturing the data flow and functional modules.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

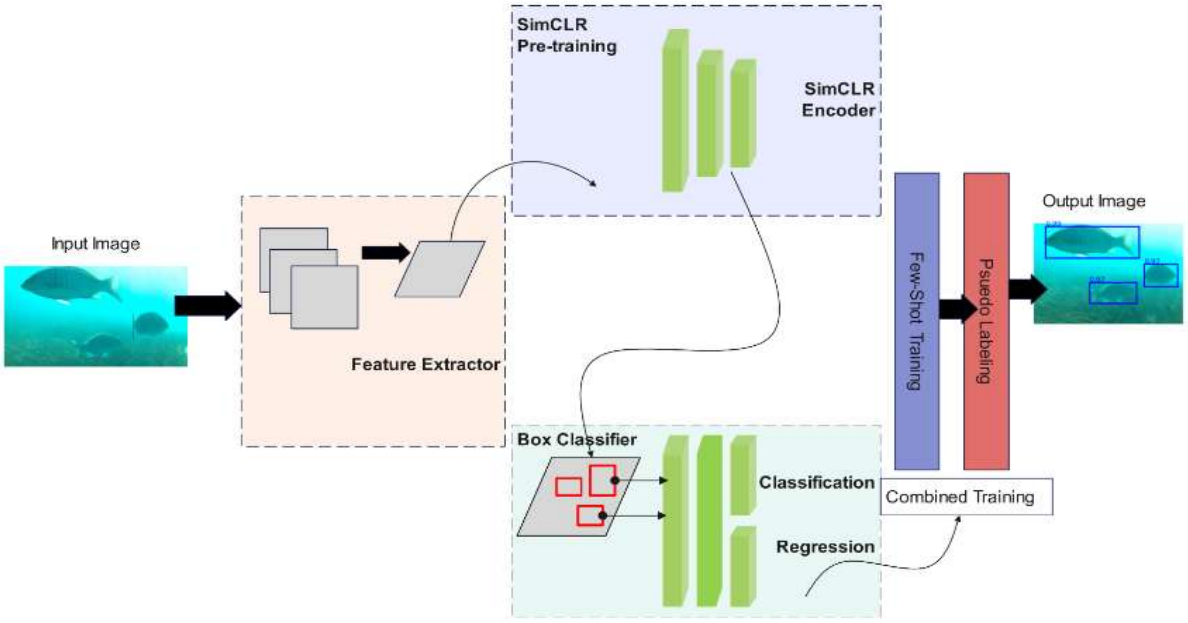


Figure 1: Architecture of the proposed AquaVision pipeline showing sequential integration of self-supervised learning, few-shot detection, and pseudo-label based semi-supervised refinement.

Self supervised Representation learning

To address the scarcity of annotated data, we begin by learning generalizable visual features using SimCLR, a self-supervised contrastive learning framework. We employ a ResNet-50 encoder as the backbone and train it on the unlabeled DeepFish dataset. The network is trained to maximize similarity between different augmented views of the same image, encouraging the model to learn semantically meaningful representations without the need for manual labels.

Given a batch of N images, two stochastic augmentations are applied to each sample, resulting in 2N transformed instances. Let $(\mathbf{z}_i, \mathbf{z}_j)$ represent a positive pair (i.e., two views of the same image). The objective is to bring them closer in the latent space while pushing apart other pairs. The loss function employed is the Normalized Temperature-scaled Cross Entropy (NT-Xent):

$$\mathcal{L}_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)} \qquad (1)$$

Here, sim(·, ·) denotes cosine similarity, τ is the temperature parameter, and the denominator sums over all possible negatives in the batch. This stage produces a feature encoder capable of capturing rich, domain-relevant semantic patterns without any labeled supervision.

Few short Fine Turning Using Faster R-CNNS

The pretrained encoder is then integrated into a two-stage object detection frame- work, Faster R-CNN, to facilitate fish localization and classification in underwater imagery. We employ a limited set of annotated samples from the Seagrass dataset for fine-tuning. This few-shot training step enables the network to adapt its general representations to the target task and data distribution which is shown in Algorithm 1.

The Faster R-CNN model outputs both bounding box coordinates and object class probabilities. The total loss comprises a classification term L_{cls} and a bounding box regression loss L_{reg}, balanced by a scalar hyper parameter λ:

$$L_{FRCNN} = L_{cls} + \lambda L_{reg} \qquad (2)$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Algorithm 1 FEW-SHOT FINE-TUNING OF FASTER R-CNN

Require: Pretrained encoder E, labeled few-shot dataset $\mathcal{D}_{\text{few}}$
Initialize Faster R-CNN model $\mathcal{M}$ with encoder E
for each epoch in training **do**
  for each mini-batch (x, y) from $\mathcal{D}_{\text{few}}$ **do**
    $\hat{y} \leftarrow \mathcal{M}(x)$ ▷ Run forward pass
    Compute classification loss: $\mathcal{L}_{\text{cls}}$
    Compute bounding box loss: $\mathcal{L}_{\text{reg}}$
    $\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{cls}} + \lambda \cdot \mathcal{L}_{\text{reg}}$
    Update model $\mathcal{M}$ via backpropagation
  end for
end for
return Fine-tuned detector $\mathcal{M}$

Semi Supervised Pseudo Labeling

To address the scarcity of annotated data in our fish identification task, we incorporate a semi-supervised learning strategy known as pseudo-labeling. This approach allows us to utilize a large pool of unlabeled images by treating certain model predictions as if they were ground truth labels. Following the initial training phase, where the object detection model (e.g., Faster R-CNN) is fine-tuned using a limited set of manually labeled images, the model is then used to infer bounding boxes and class labels on the remaining unlabeled dataset. For each prediction, the model also outputs a confidence score indicating the likelihood that the prediction is accurate. To ensure that only the most reliable predictions are used for further training, we define a confidence threshold =0.75. Any predicted object whose associated confidence score exceeds this threshold is considered a high-confidence prediction. These selected predictions are then designated as pseudo-labels. The core idea is to enrich the training dataset by incorporating these pseudo- labeled samples alongside the original labeled data. This extended dataset enables the model to learn from a broader variety of visual examples without requiring additional manual annotations. Although pseudo-labels may occasionally introduce noise, the use of a high confidence threshold helps mitigate this risk by filtering out uncertain predictions which is shown in Algorithm 2. This iterative process predicting on unlabeled data, selecting high-confidence pseudo-labels, and retraining the model with this augmented dataset enables the model to gradually improve its generalization capability. It effectively leverages both labeled and unlabeled data to enhance performance, particularly in scenarios where labeled data is scarce or costly to obtain.

Algorithm 2 PSEUDO-LABEL GENERATION FOR UNLABELED DATA

Require: Trained detector $\mathcal{M}$, unlabeled dataset $\mathcal{D}_{\text{unlabeled}}$, confidence threshold θ
Initialize pseudo-label set: $\mathcal{D}_{\text{pseudo}} \leftarrow \emptyset$
for each image $x \in \mathcal{D}_{\text{unlabeled}}$ **do**
  $\hat{y} \leftarrow \mathcal{M}(x)$ ▷ Predict bounding boxes
  for each detection $d \in \hat{y}$ **do**
    if confidence$(d) > \theta$ **then**
      Add pseudo-labeled pair (x, d) to $\mathcal{D}_{\text{pseudo}}$
    end if
  end for
end for
return Pseudo-labeled dataset $\mathcal{D}_{\text{pseudo}}$

Joint Retraining Using Few-Shot with Pseudo Label

To leverage both the manually labeled and pseudo-labeled datasets, we retrain the detection model in a unified supervised learning setup. The final training set is formed by merging the few-shot labeled dataset with the pseudo-labeled samples generated from the previous step:

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

$$D_{train} = D_{few} \cup D_{pseudo} \quad (3)$$

This combined dataset enables the model to benefit from the precision of high- quality human-labeled examples and the diversity of machine-generated annotations. The learning objective and loss function remain consistent with those used during initial fine-tuning. By training on this augmented dataset, the model is exposed to a broader distribution of visual features and object appearances, thereby enhancing its ability to generalize to unseen data. This joint training paradigm effectively exploits both strong (ground truth) and weak (pseudo-labeled) supervision to improve detection accuracy in low-resource settings is shown in Algorithm 3.

Algorithm 3 JOINT TRAINING WITH FEW-SHOT AND PSEUDO-LABELED DATA

Require: Initial model $\mathcal{M}$, few-shot dataset $\mathcal{D}_{\text{few}}$, pseudo-labeled dataset $\mathcal{D}_{\text{pseudo}}$

Merge datasets: $\mathcal{D}_{\text{train}} \leftarrow \mathcal{D}_{\text{few}} \cup \mathcal{D}_{\text{pseudo}}$
for each epoch in training **do**
  for each mini-batch $(x, y) \in \mathcal{D}_{\text{train}}$ **do**
    $\hat{y} \leftarrow \mathcal{M}(x)$ ▷ Inference step
    Compute $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{reg}}$
    $\mathcal{L}_{\text{combined}} \leftarrow \mathcal{L}_{\text{cls}} + \lambda \cdot \mathcal{L}_{\text{reg}}$
    Update model $\mathcal{M}$ via backpropagation
  end for
end for
return Final model $\mathcal{M}$

Hard Example Mining and Performance Analysis

To further enhance performance, hard example mining is implemented by identifying false positives and samples with low Intersection over Union (IoU) which is shown in Algorithm 4. These challenging samples can be recycled for adaptive re-training or curriculum learning strategies. For comprehensive evaluation, the model is assessed using widely recognized object detection metrics: precision, recall, F1-score, and mean Average Precision (mAP) . We benchmark AquaVision against several baselines including YOLOv5 and RetinaNet to highlight improvements in both low-shot and domain-specific performance.

Algorithm 4 HARD EXAMPLE MINING (HEM)

Require: Trained model $\mathcal{M}$, validation set $\mathcal{D}_{\text{val}}$, IoU threshold δ, classification confidence threshold α

Initialize hard set: $\mathcal{D}_{\text{hard}} \leftarrow \emptyset$
for each image-label pair $(x, y) \in \mathcal{D}_{\text{val}}$ **do**
  Predict: $\hat{y} \leftarrow \mathcal{M}(x)$
  for each ground truth object $y_i \in y$ **do**
    Match with predicted detection $\hat{y}_j$ using IoU
    if $\text{IoU}(y_i, \hat{y}_j) < \delta$ **or** $\text{confidence}(\hat{y}_j) < \alpha$ **then**
      Add (x, y_i) to $\mathcal{D}_{\text{hard}}$
    end if
  end for
end for
Increase sampling weight for $\mathcal{D}_{\text{hard}}$ in next training phase

Experimental setup

This section outlines the experimental environment, datasets used, preprocessing pipeline, and the evaluation metrics employed to assess the performance of Aqua- Vision. The experiments were designed to evaluate the robustness of the system under challenging underwater conditions using few-shot and semi-supervised learning techniques.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Data Set Description

DeepFish Dataset: The unlabeled dataset consists of a large collection of 40000 underwater images captured across diverse tropical marine habitats, similar to the labeled portion but without accompanying annotations such as fish count, location, or species information. These high-resolution images (1,920 × 1,080) are sourced from in-situ field recordings in 20 different coastal and near shore environments in tropical Australia [32].

The Sea grass subset: It is a focused portion of the Deep Fish dataset containing 3,255 images captured in sea grass-dominated marine habitats. It provides a limited set of manually annotated samples, including 16 segmentation-labeled images**,** making it well-suited for few-shot detection experiments. All annotations belong to a single class: fish**,** consistent with the broader Deep Fish labeling scheme. This subset reflects the visually complex nature of sea grass environments, where fish often appear partially occluded or camouflaged. Overall, it serves as a compact yet challenging benchmark for evaluating underwater fish detection models [33].

Data preprocessing

All images are first resized to a fixed resolution of 512×512 pixels to ensure uniformity across the pipeline. Preprocessing also includes:

- **Data Augmentation:** Random cropping, horizontal flipping, color jittering, and Gaussian blur are applied during SimCLR pretraining to create augmented views for contrastive learning.
- **Normalization:** Pixel intensities are normalized using the ImageNet mean and standard deviation to match pretrained model expectations.
- **Bounding Box Format Conversion:** For annotation compatibility, bounding box labels are converted to the format required by Faster R-CNN (i.e., [x min, y min, x max, y max]).
- **Confidence Thresholding:** During pseudo-labeling, detections below a confidence score of 0.75 are filtered out.

This preprocessing ensures consistent data quality and enables effective transfer learning from self-supervised features to detection tasks.

Training Configuration

In this section, we discuss the complete training setup used across all stages of the AquaVision pipeline. Each phase of the workflow started with self-supervised SimCLR pretraining, few-shot object detector fine-tuning, semi-supervised pseudo-label generation, and joint retraining was configured with carefully selected hyperparameters to ensure stable optimization, efficient learning, and strong generalization in underwater environments. The following tables [1-4] summarize the key parameters, optimization choices, data augmentations, and thresholds employed at each step of the pipeline.

Table 1: Self-Supervised Feature Extraction Using SimCLR

Parameter	**Value / Setting**	**Description**
Dataset	Unlabeled DeepFish	Source for contrastive pretraining
Encoder Backbone	ResNet-50	Initialized for SimCLR representation learning
Input Resolution	512 × 512	Fixed size for all SimCLR image pairs
Batch Size	32	Number of image pairs per batch

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Epochs	100	Total SimCLR pretraining iterations
Optimizer	Adam	Optimizer for contrastive learning
Learning Rate	10^{-4}	Base LR for SimCLR
LR Schedule	Cosine Annealing	Smooth LR decay over 100 epochs
Weight Decay	0.0001	Regularization to stabilize learning
Temperature (τ)	0.5	Scaling factor in NT-Xent loss
Data Augmentation	Random crop, flip, color jitter, Gaussian blur	Strong transforms required for SimCLR
Output	simclr_resnet50.pth	Pretrained encoder checkpoint

Table 2: Few-Shot Object Detection Fine-Tuning (Faster R-CNN)

Parameter	**Value / Setting**	**Description**
Dataset	Seagrass Few-Shot	Small labeled dataset for adaptation
Detector Architecture	Faster R-CNN	Two-stage detector for fish localization
Backbone	ResNet-50 (SimCLR pretrained)	Initialized from Stage 1
Input Resolution	512 × 512	Fixed resolution for detector training
Batch Size	8	Per-iteration processing for fine-tuning
Epochs	50	Number of supervised fine-tuning iterations
Optimizer	SGD (momentum = 0.9)	Standard for detection tasks
Learning Rate	10^{-4}	Base LR for fine-tuning
Weight Decay	0.0001	Regularization
LR Schedule	Cosine Annealing	Smooth learning rate decay
Confidence Threshold	0.75	Minimum score for predictions
IoU Threshold (HEM)	0.5	Hard-example mining threshold
Data Augmentation	Flip, crop, color jitter	Light augmentations to avoid label distortion
Output	fasterrcnn_fewshot.pth	Detector checkpoint

Table 3: Semi-Supervised Pseudo-Label Generation

Parameter	**Value / Setting**	**Description**
Dataset	Unlabeled DeepFish	Used to generate pseudo detections

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Detector Checkpoint	fasterrcnn_fewshot.pth	Model from Stage 2
Input Resolution	512 × 512	Same as training for consistency
Confidence Threshold	0.75	High-confidence detections only
NMS IoU Threshold	0.5	Removes duplicate bounding boxes
Small Box Removal	Enabled	Avoids noisy tiny detections
Minimum Area Ratio	0.005	Filters boxes <0.5% of image area
Output Pseudo Labels	pseudo_labels/	Directory storing generated annotations
Combined Dataset	combined_labeled_pseudo/	Merged true + pseudo dataset

Table 4: Joint Retraining Using True + Pseudo Labels

Parameter	**Value / Setting**	**Description**
Dataset	Labeled + Pseudo-Labeled	Unified training set
Input Resolution	512 × 512	Kept consistent across all stages
Batch Size	8	Detector training batch size
Epochs	50	Joint training iterations
Optimizer	SGD	For stable supervised learning
Momentum	0.9	Typical for detection
Learning Rate	10^{-4}	Base LR for retraining
Weight Decay	0.0001	Regularization
LR Schedule	Cosine Annealing	Ensures smooth convergence
Sampling Strategy	Labeled : Pseudo = 1 : 3	Balances real and pseudo examples
Pseudo-Loss Weight	0.5	Down-weights noisy pseudo labels
Pseudo Refresh	Disabled (optional)	Can be enabled every 10 epochs
Output	aquavision_final.pth	Final trained model

Evaluation Metrics

To quantitatively assess the performance of AquaVision, we use the following standard object detection metrics:

- **Precision and Recall**: Measure the accuracy and completeness of fish detections. High precision indicates fewer false positives, while high recall indicates fewer false negatives.
- **F1 Score**: Harmonic mean of precision and recall to capture the balance between both metrics.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

These metrics provide a comprehensive view of model performance across different conditions, including crowded, blurry, or occluded scenes common in underwater imagery.

RESULT ANALYSIS

This section presents an in-depth evaluation of the AquaVision framework, focusing on both qualitative and quantitative analyses. The performance of our proposed approach is assessed through visual comparisons and numerical metrics. Specifically, we evaluate the accuracy and robustness of fish detection in challenging underwater environments using the Seagrass dataset. Furthermore, we benchmark our results against three well established object detection models: Faster R-CNN, YOLOv5, and RetinaNet.

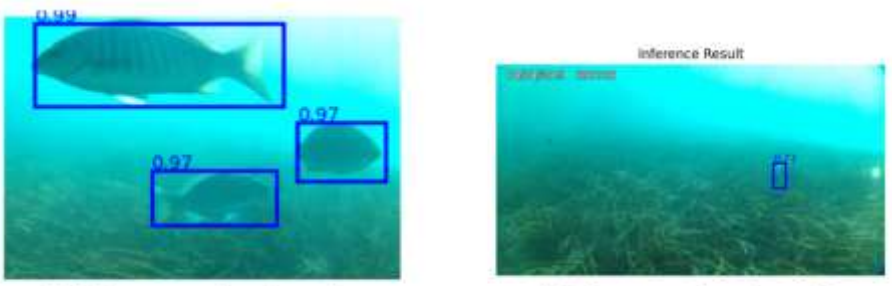
Qualitative Analysis

Visual inspection plays a crucial role in understanding the behavior of object detection models, particularly in complex natural environments. Figures 2a and Figure 2b showcase the outputs generated by the proposed AquaVision model on selected images from the Seagrass dataset. The model demonstrates a strong ability to identify multiple fish instances even under visually challenging conditions, including occlusions, shadows, and dense seagrass interference.

As depicted in Figure 2, the bounding boxes generated by the model are well- aligned with visible fish instances, even in scenarios where the targets are partially obscured or camouflaged. The model maintains high confidence in its detections, reflecting robustness against environmental noise and visual complexity. To further validate the model's performance, we provide a comparison with ground truth annotations. Figures 3 and 4 highlight the ground truth annotations and RetinaNet outputs respectively.

Following this, Figure5 and 6 present predictions from the used Faster R-CNN and YoloV5 for visual comparison: From these visual comparisons, we observe that the predicted bounding boxes exhibit strong spatial alignment with the ground truth labels.

Minor discrepancies are noted in heavily occluded or low-contrast regions, which can be attributed to the natural complexity of underwater imagery. Nevertheless, the model demonstrates a commendable ability to localize and differentiate fish against complex backgrounds.



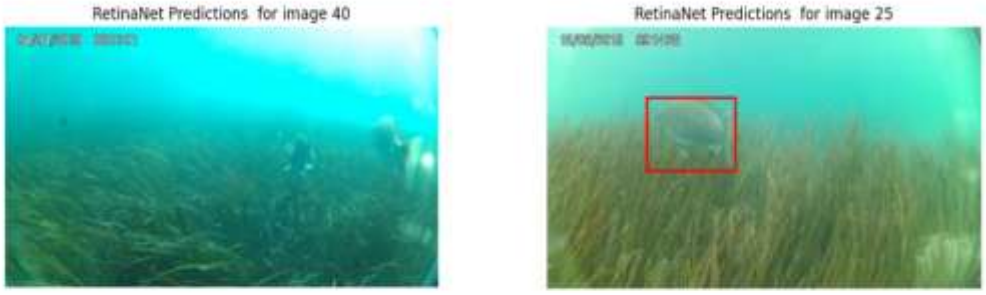
(a) Inference result on Image A (b) Inference result on Image B

Fig. 2: Predictions from the Aqua Vision model

(a) Ground truth for Image 40 (b) Ground truth for Image 25

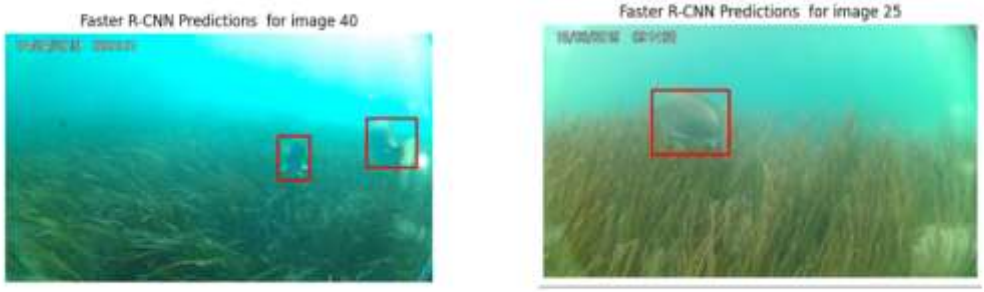
Fig. 3: Ground truth annotations for corresponding test images

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025



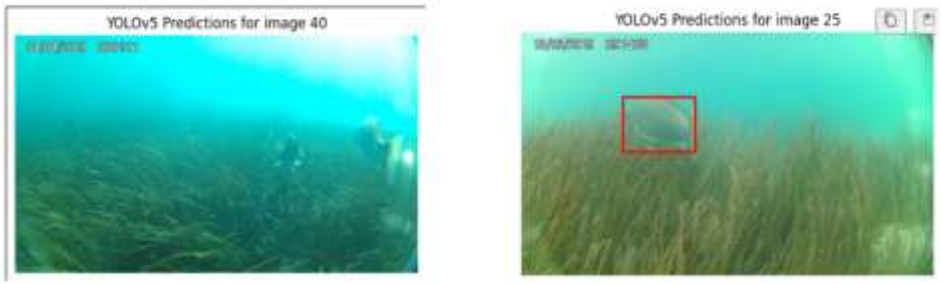
(a) RetinaNet prediction for Image 40 (b) RetinaNet prediction for Image 25

Fig. 4: Predicted outputs by RetinaNet



(a) Faster R-CNN prediction for Image 40 (b) Faster R-CNN prediction for Image 25

Fig. 5: Predicted outputs by Faster R-CNN



(a) YOLOv5 prediction for Image 40 (b) YOLOv5 prediction for Image 25

Fig. 6: Predicted outputs by YOLOv5

Quantitative Evaluation

For a more rigorous assessment, the performance of the proposed detection pipeline was evaluated on a test set comprising 50 images. We used standard object detection metrics such as Average Precision (AP), Average Recall (AR), and the F1-score to quantify model effectiveness. Table 2 summarizes the performance of our model in comparison with two widely used baselines, YOLOv5 and RetinaNet.

The results indicate that the proposed framework, which combines few-shot learning with pseudo-labeling strategies, consistently outperforms the baseline models across all evaluation metrics. Faster R-CNN, when enhanced with our training methodology, achieves the highest precision and recall, indicating superior detection accuracy and coverage.

The comparatively lower performance of YOLOv5 and RetinaNet highlights the advantage of our training strategy in limited-data regimes. These findings underscore the effectiveness of integrating semi-supervised learning techniques to improve model performance under real-world constraints, where acquiring labeled data

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

is often resource-intensive.

Model	Average Precision	Average Recall	F1-Score
Proposed Faster R-CNN	**0.8167**	**0.7967**	**0.8120**
YoloV5	0.6800	0.6733	0.6667
RetinaNet	0.4600	0.4467	0.4500

CONCLUSION

The AquaVision framework effectively showcases a novel and comprehensive approach by combining self-supervised learning, few-shot object detection, and semi-supervised refinement techniques to tackle the ongoing difficulties in accurately detecting fish within challenging underwater environments. The complexity of such environments characterized by factors like low visibility, occlusions, and intricate back- grounds poses significant obstacles for traditional detection methods. AquaVision addresses these challenges through an innovative training pipeline.

Central to the framework is the use of SimCLR-based self-supervised pretraining, which enables the model to extract robust and transferable feature representations from a large corpus of unlabeled DeepFish images. This pretraining step serves as a powerful foundation, allowing the system to learn general visual patterns without reliance on annotated data. Subsequently, this knowledge is fine-tuned on a limited, carefully labeled subset of Seagrass dataset images using the Faster R-CNN architecture. This few-shot learning approach ensures that the model adapts effectively to the target domain with minimal labeled examples.

To further enhance detection performance, the framework incorporates pseudo- labeling, a semi-supervised learning strategy that leverages unlabeled images by generating high-confidence predictions as additional training signals. This iterative refinement expands the effective training data and improves the model's ability to generalize across diverse underwater conditions. As a result, the enhanced AquaVision model significantly outperforms established object detection baselines such as YOLOv5 and RetinaNet in terms of average precision and recall metrics, achieving a precision and recall of approximately 0.82.

Beyond numerical metrics, qualitative evaluation reveals that AquaVision consistently localizes fish accurately even in visually complex scenes marked by heavy occlusion, dense seagrass, and variable lighting conditions. This highlights the model's robustness and practical utility in real-world scenarios

Declarations

Author contribution Information

All authors are equally contributed**.**

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper**.**

Funding

The authors does not receive any Fund

Data Availability statement

Throughout the research used the public available data set which is cited in 32 &33.

REFERENCES

1. Ahmed, R., & Tamim, M. T. R. (2025). Marine and Coastal Environments: Challenges, Impacts, and Strategies for a Sustainable Future. *International Journal of Science Education and Science*, *2*(1), 53-60.
2. Douglas, J., Niner, H., & Garrard, S. (2024). Impacts of marine plastic pollution on seagrass meadows and ecosystem services in Southeast Asia. *Journal of Marine Science and Engineering*, *12*(12), 2314.
3. Schmid, B., & Schöb, C. (2022). Biodiversity and ecosystem services in managed ecosystems. In *The ecological and societal consequences of biodiversity loss* (pp. 213-231). ISTE Ltd and John Wiley & Sons, Inc, London.
4. Hong, J. H., Semprucci, F., Jeong, R., Kim, K., Lee, S., Jeon, D., ... & Lee, W. (2020). Meiobenthic nematodes in the assessment of the relative impact of human activities on coastal marine ecosystem. *Environmental Monitoring and Assessment*, *192*(2), 81..
5. Lopez-Vazquez, V., Lopez-Guede, J. M., Marini, S., Fanelli, E., Johnsen, E., & Aguzzi, J. (2020). Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. *Sensors*, *20*(3), 726.
6. Jalal, A., Salman, A., Mian, A., Shortis, M., & Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, *57*, 101088.
7. Meena, T., Vijaya, J., & Harsha, B. (2025, February). Swin Transformers for Remote Sensing SAR Image Classification. In *2025 IEEE International Conference on Emerging Technologies and Applications (MPSec ICETA)* (pp. 1-6). IEEE.
8. Vijaya, J., Gopu, A., Suman, P., & Chaitanya, S. (2024, May). Revolutionising Image Enhancement Leveraging Power OF CNN'S. In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-6). IEEE.
9. Yassir, A., Andaloussi, S. J., Ouchetto, O., Mamza, K., & Serghini, M. (2023). Acoustic fish species identification using deep learning and machine learning algorithms: A systematic review. *Fisheries Research*, *266*, 106790.
10. Fu, C., Liu, R., Fan, X., Chen, P., Fu, H., Yuan, W., ... & Luo, Z. (2023). Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing*, *517*, 243-256.
11. Er, M. J., Chen, J., Zhang, Y., & Gao, W. (2023). Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. *Sensors*, *23*(4), 1990.
12. Li, J., Yang, W., Qiao, S., Gu, Z., Zheng, B., & Zheng, H. (2024). Self-supervised marine organism detection from underwater images. *IEEE Journal of Oceanic Engineering*.
13. Chungath, T. T., Nambiar, A. M., & Mittal, A. (2023). Transfer learning and few-shot learning based deep neural network models for underwater sonar image classification with a few samples. *IEEE Journal of Oceanic Engineering*, *49*(1), 294-310.
14. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., ... & Li, C. L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, *33*, 596-608.
15. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PmLR.
16. Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in neural information processing systems*, *30*.
17. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, *30*.
18. Li, L., Shi, G., & Jiang, T. (2023). Fish detection method based on improved YOLOv5. *Aquaculture International*, *31*(5), 2513-2530.
19. Shen, Z., & Nguyen, C. (2020, November). Temporal 3D RetinaNet for fish detection. In *2020 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-5). IEEE.

20. Vyshnav, K., Sooryanarayanan, R., & Madhav, T. V. (2024, April). Analysis of Underwater Coral Reef Health Using Neural Networks. In *OCEANS 2024-Singapore* (pp. 01-06). IEEE.
21. Chowdhury, A., Jahan, M., Kaisar, S., Khoda, M. E., Rajin, S. A. K., & Naha, R. (2024). Coral Reef Surveillance with Machine Learning: A Review of Datasets, Techniques, and Challenges. *Electronics*, *13*(24), 5027.
22. https://homepages.inf.ed.ac.uk/rbf/fish4knowledge/
23. https://www.kaggle.com/datasets/lywang777/urpc2020
24. https://datasetninja.com/deep-fish
25. Elmezain, M., Saoud, L. S., Sultan, A., Heshmat, M., Seneviratne, L., & Hussain, I. (2025). Advancing underwater vision: a survey of deep learning models for underwater object recognition and tracking. *IEEE Access*.
26. Zhang, F., Hu, J., & Sun, Y. (2025). Underwater fish image recognition based on knowledge graphs and semi-supervised learning feature enhancement. *Scientific Reports*.
27. Dalla Serra, F., Jacenków, G., Deligianni, F., Dalton, J., & O'Neil, A. Q. (2022, July). Improving Image Representations via MoCo Pre-training for Multimodal CXR Classification. In *Annual Conference on Medical Image Understanding and Analysis* (pp. 623-635). Cham: Springer International Publishing.
28. Niizumi, D., Takeuchi, D., Ohishi, Y., Harada, N., & Kashino, K. (2021, July). Byol for audio: Self-supervised learning for general-purpose audio representation. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
29. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., & Raffel, C. A. (2019). Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, *32*.
30. Fallah, A., Mokhtari, A., & Ozdaglar, A. (2020, June). On the convergence theory of gradient-based model-agnostic meta-learning algorithms. In *International Conference on Artificial Intelligence and Statistics* (pp. 1082-1092). PMLR.
31. Lu, J., Zhang, S., Zhao, S., Li, D., & Zhao, R. (2024). A metric-based few-shot learning method for fish species identification with limited samples. *Animals*, *14*(5), 755.
32. https://alzayats.github.io/DeepFish/
33. https://github.com/globalwetlands/luderick-seagrass