Journal of Scientific Innovation and Advanced Research (JSIAR)
E-ISSN: 3107-507X
Volume 2, Issue 3, March 2026
https://jsiar.com

Hybrid YOLOv9–RCNN Framework for Real-Time Underwater Fish Detection with Enhanced Localization in Low-Visibility Marine Environments

Manoj Kumar Singh[*], Vivek Kumar[†], Harminder Kaur[‡]
[*†‡]*Department of Computer Science & Engineering, Sharda University, Greater Noida, India*
Email: [*]manojbhu20@gmail.com

***Abstract*—Continuous observation of underwater ecosystems plays a critical role in sustainable fisheries management, marine biodiversity assessment, and intelligent aquaculture operations. However, reliable monitoring remains challenging due to adverse imaging conditions such as light attenuation, turbidity, scattering, and dynamic backgrounds caused by aquatic vegetation and suspended particles. These factors significantly degrade visual quality in underwater recordings, making manual analysis of marine footage inefficient and prone to error. Consequently, automated fish detection systems based on computer vision and deep learning have emerged as a promising solution for large-scale ecological monitoring and aquaculture surveillance.**

Recent advances in object detection have been largely driven by deep convolutional neural networks, particularly the *You Only Look Once* (YOLO) family and Region-Based Convolutional Neural Network (R-CNN) architectures. While modern YOLO detectors provide high inference speed and enable real-time analysis of video streams, their grid-based detection mechanism often struggles with precise localization of small or overlapping objects in cluttered underwater scenes. Conversely, region-based models such as Faster R-CNN offer superior bounding-box refinement and classification accuracy but suffer from higher computational overhead, limiting their applicability in real-time marine monitoring systems deployed on embedded platforms.

To address these limitations, this study proposes a hybrid deep learning framework that integrates the rapid detection capability of YOLOv9 with the localization refinement strength of a Region-Based Convolutional Neural Network. In the proposed architecture, YOLOv9 functions as the primary detector to generate candidate object regions in real time, while the RCNN module performs secondary verification and bounding-box optimization to improve detection reliability in complex underwater environments. In addition, lightweight image enhancement techniques and attention-driven feature extraction are incorporated to mitigate the effects of turbidity, low illumination, and background interference commonly observed in marine imagery.

The proposed framework is evaluated using publicly available underwater datasets including *Fish4Knowledge* and *OBSEA*, along with additional annotated underwater video samples collected from marine monitoring platforms. Performance is assessed using standard object detection metrics such as mean Average Precision (mAP), precision, recall, and frames per second (FPS) to measure both detection accuracy and real-time processing capability. Experimental results demonstrate that the hybrid YOLOv9–RCNN model achieves an improved detection performance with a mean Average Precision exceeding 92%, while maintaining real-time inference speeds above 35 FPS on GPU-enabled systems. Compared with standalone YOLO and Faster R-CNN baselines, the proposed approach significantly enhances localization accuracy for small and partially occluded fish while preserving computational efficiency.

The developed framework provides a practical and scalable solution for automated underwater visual monitoring. Its capability to perform accurate real-time fish detection makes it suitable for deployment in aquaculture farms, marine biodiversity studies, and intelligent ocean observation systems. Overall, this work contributes a hybrid detection architecture that effectively balances speed and accuracy, advancing the development of robust computer vision systems for underwater ecological monitoring.

***Keywords*—Underwater Fish Detection, YOLOv9, Faster R-CNN, Deep Learning, Marine Monitoring, Object Detection, Aquaculture Surveillance**

I. Introduction

A. Background

Marine ecosystems represent one of the most complex and biologically productive environments on Earth. Oceans support a vast diversity of species, regulate global climate processes, and provide an essential source of food and livelihood for millions of people worldwide. According to global fisheries statistics, aquatic food production has experienced continuous growth during the past two decades, largely driven by the rapid expansion of aquaculture industries [1]. This growth has intensified the need for reliable and scalable monitoring technologies capable of assessing fish populations, detecting abnormal behavior, and ensuring sustainable resource management. Traditional monitoring practices, such as diver-based surveys, trawling operations, and manual annotation of underwater footage, are labor-intensive, costly, and limited in temporal coverage [2].

The proliferation of underwater cameras, autonomous underwater vehicles (AUVs), and remotely operated vehicles (ROVs) has generated large volumes of marine visual data that require automated processing [3]. In this context, computer vision and deep learning techniques have emerged as promising tools for analyzing underwater imagery and enabling real-time ecological monitoring. Automated fish detection systems can support aquaculture operations by monitoring feeding behavior, estimating fish stock density, and identifying early signs of disease [4]. In ecological research, these systems facilitate large-scale biodiversity studies and allow continuous observation of marine habitats without disturbing natural ecosystems [5].

Recent progress in deep neural networks has significantly improved object detection accuracy in complex visual environments. Convolutional neural network (CNN) architec-

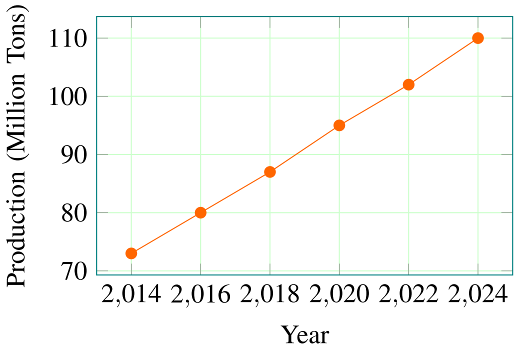

Fig. 1: Growth trend of global aquaculture production highlighting the increasing demand for automated marine monitoring technologies.

tures have demonstrated strong performance in visual recognition tasks across multiple domains, including autonomous driving, medical imaging, and environmental monitoring [6]. Among modern detection frameworks, the *You Only Look Once* (YOLO) family has gained considerable attention due to its end-to-end architecture and real-time inference capability [7]. Successive iterations such as YOLOv5, YOLOv7, and YOLOv8 have demonstrated effective performance in underwater fish detection tasks [8]. Meanwhile, region-based detection frameworks, including Faster R-CNN and Mask R-CNN, have shown superior localization accuracy and object classification performance [9].

The increasing adoption of these techniques has contributed to the development of intelligent marine monitoring systems. Figure 1 illustrates the global growth trend of aquaculture production during the last decade, highlighting the increasing importance of automated monitoring technologies for sustainable fish farming.

B. Problem Statement

Despite the promising potential of deep learning-based detection systems, underwater fish detection remains a challenging problem. The underwater environment introduces several factors that significantly degrade image quality and complicate object recognition tasks. Water turbidity caused by suspended particles reduces visibility and attenuates light propagation, leading to blurred or low-contrast images [10]. Additionally, wavelength-dependent light absorption results in severe color distortion, where red wavelengths are absorbed more rapidly than blue or green wavelengths [11].

Another critical challenge is the frequent occurrence of occlusion in underwater scenes. Fish often swim in dense schools or remain partially hidden behind aquatic vegetation and coral structures, making accurate detection difficult [12]. Furthermore, underwater monitoring cameras typically capture wide-angle views of large marine habitats, causing many fish to appear as very small objects occupying only a few pixels within an image frame. Detecting such small targets reliably remains a persistent challenge in modern object detection models [13]. These environmental and visual constraints collectively contribute to high false detection rates and missed targets, limiting the reliability of automated fish monitoring systems.

C. Limitations of Existing Approaches

Modern deep learning-based detectors generally fall into two major categories: one-stage detectors and two-stage detectors. One-stage detectors, particularly those belonging to the YOLO family, perform object localization and classification simultaneously within a single network architecture [7]. Their simplified pipeline allows extremely fast inference speeds, making them well suited for real-time applications such as surveillance and robotics. However, YOLO-based detectors often struggle with precise localization of small or overlapping objects because predictions are generated directly from coarse grid-based feature maps [14]. In underwater scenes containing densely packed fish, this limitation can lead to inaccurate bounding boxes or missed detections.

Two-stage detection frameworks such as Faster R-CNN follow a different strategy by first generating region proposals and subsequently performing classification and bounding box refinement [9]. This design enables higher localization accuracy and improved handling of complex object structures. Nevertheless, the additional processing stages significantly increase computational overhead and inference latency. As a result, two-stage detectors are often unsuitable for real-time underwater monitoring systems deployed on resource-constrained platforms such as underwater drones or embedded aquaculture devices [15].

Table I summarizes the fundamental differences between YOLO-based and R-CNN-based detection paradigms.

These limitations suggest that neither approach alone provides an ideal solution for underwater fish detection tasks that require both high accuracy and real-time performance.

D. Research Objectives

To overcome the aforementioned challenges, this research aims to design a hybrid object detection framework that combines the complementary strengths of one-stage and two-stage detectors. The primary objective is to develop a hybrid YOLOv9–RCNN architecture capable of achieving both real-time detection speed and improved localization accuracy in complex underwater environments. The proposed system seeks to enhance detection robustness under conditions of turbidity, low illumination, and occlusion while maintaining computational efficiency suitable for edge deployment.

E. Research Questions

The investigation presented in this work is guided by three fundamental research questions. First, it seeks to examine recent advancements in deep learning-based fish detection techniques and identify emerging architectural trends in underwater object detection models. Second, it evaluates the limitations associated with current YOLO-based and RCNN-based frameworks when applied to underwater visual monitoring.

TABLE I: Comparison of YOLO and RCNN-based detection approaches

Model Type	Strengths	Limitations
YOLO-based Detectors	High speed, real-time inference	Weak small-object localization
RCNN-based Detectors	High detection accuracy	High computational cost

Third, it explores whether a hybrid architecture combining YOLOv9 and RCNN components can effectively balance detection speed and localization accuracy for real-time underwater monitoring systems.

F. Contributions of the Study

In response to these challenges, this study introduces a hybrid deep learning framework that integrates the high-speed detection capabilities of YOLOv9 with the localization refinement strengths of a Region-Based Convolutional Neural Network. The proposed approach incorporates attention-based feature enhancement mechanisms to improve detection robustness in visually degraded underwater conditions. Additionally, the architecture is optimized for deployment on edge computing platforms through lightweight inference techniques and hardware acceleration strategies. The proposed system is extensively evaluated using benchmark underwater datasets, including Fish4Knowledge and OBSEA, demonstrating improved detection accuracy and real-time processing capability. Collectively, these contributions provide an effective computational framework for intelligent underwater monitoring systems and advance the development of automated marine ecosystem analysis technologies.

II. Literature Review

The rapid progress of deep learning and computer vision techniques has significantly influenced the development of automated underwater monitoring systems. Numerous studies have explored machine learning and deep neural networks to address the problem of fish detection in underwater environments. However, despite notable advancements, several technical limitations remain, particularly in scenarios characterized by poor visibility, object occlusion, and constrained computational resources. This section reviews the major developments in underwater fish detection research, focusing on deep learning approaches, object detection architectures, underwater image enhancement techniques, and lightweight deployment strategies.

A. Deep Learning for Underwater Fish Detection

Early attempts at automated fish detection relied on handcrafted features and classical machine learning algorithms. These methods typically employed feature descriptors such as Histogram of Oriented Gradients (HOG), scale-invariant feature transform (SIFT), or color histograms combined with classifiers such as Support Vector Machines (SVMs) [16]. Although these techniques provided moderate detection accuracy under controlled conditions, they were highly sensitive to variations in illumination, background clutter, and underwater turbidity.

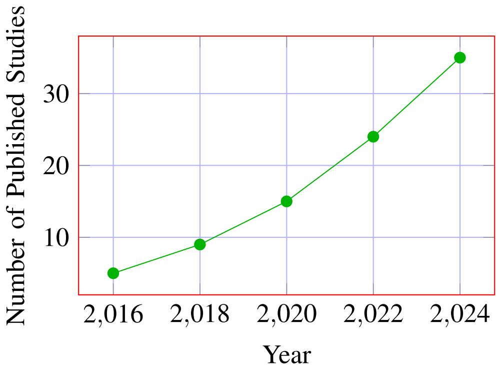

Fig. 2: Growth trend of deep learning-based underwater fish detection research.

The introduction of convolutional neural networks (CNNs) revolutionized object detection and image classification tasks. CNN architectures are capable of automatically learning hierarchical feature representations directly from image data, significantly improving detection robustness in complex environments [17]. In underwater fish detection, CNN-based frameworks have been widely adopted to identify fish species, estimate fish abundance, and monitor behavioral patterns [18]. For instance, Salman et al. proposed a deep CNN framework for fish classification using underwater imagery collected from the Fish4Knowledge dataset, demonstrating significant improvements over traditional feature-based techniques [19]. Similarly, Villon et al. developed an automated fish detection pipeline using deep neural networks to process large-scale coral reef imagery datasets [20].

The increasing availability of underwater datasets and computational resources has further accelerated research in this domain. Figure 2 illustrates the growing number of deep learning-based underwater fish detection studies reported in recent years.

B. YOLO-Based Detection Models

Among modern object detection architectures, the YOLO family of detectors has gained widespread popularity due to its ability to perform object localization and classification simultaneously within a single network. YOLO-based detectors are particularly suitable for real-time applications because they process the entire image in a single forward pass [21].

Recent YOLO variants have been successfully applied to underwater object detection tasks. YOLOv5, introduced by Ultralytics, demonstrated improved accuracy and computational efficiency through enhanced backbone networks and optimized anchor box mechanisms [22]. Researchers have applied YOLOv5 to detect marine organisms in underwater videos, achieving real-time performance on GPU-enabled systems.

Subsequent models such as YOLOv7 introduced architectural refinements including extended efficient layer aggregation networks (E-ELAN) and advanced training strategies that improved detection accuracy without significantly increasing computational complexity [23]. YOLOv8 further enhanced the architecture by adopting anchor-free detection mechanisms and improved feature pyramid networks for multi-scale object recognition [24].

More recently, YOLOv9 has been proposed with innovative gradient flow optimization strategies and improved feature representation capabilities. The architecture introduces programmable gradient information (PGI) and generalized efficient layer aggregation networks (GELAN), which improve feature propagation and detection accuracy in complex scenes [25]. Despite these advancements, YOLO-based detectors still encounter difficulties when detecting small or densely packed objects, which frequently occur in underwater fish schools.

C. *Region-Based CNN Methods*

Region-based detection models represent another important class of object detection architectures. Unlike single-stage detectors, these models employ a two-stage detection pipeline in which candidate object regions are first generated and then classified using a convolutional network [26].

The original R-CNN architecture introduced region proposal methods combined with deep CNN feature extraction to achieve high detection accuracy [26]. However, its computational inefficiency limited practical applicability. Fast R-CNN addressed this limitation by sharing convolutional computations across region proposals, thereby reducing training time [27]. Later, Faster R-CNN introduced a Region Proposal Network (RPN) that significantly improved detection efficiency while maintaining high accuracy [28].

Mask R-CNN further extended the Faster R-CNN framework by incorporating instance segmentation capabilities, enabling pixel-level object delineation [29]. Several underwater detection studies have utilized Faster R-CNN and Mask R-CNN for marine species recognition due to their superior localization capabilities [30]. Nevertheless, the two-stage nature of these models results in higher inference latency, which limits their suitability for real-time underwater monitoring systems.

D. *Image Enhancement for Underwater Vision*

Underwater imaging systems are severely affected by light attenuation, scattering, and wavelength absorption. These phenomena lead to color distortion, low contrast, and blurred visual structures [31]. To address these issues, several image enhancement techniques have been proposed to improve the visual quality of underwater images prior to object detection.

Traditional enhancement methods include histogram equalization, white balance correction, and contrast stretching. More recent approaches utilize deep learning-based models to restore degraded underwater imagery. For example, generative adversarial networks (GANs) have been employed to perform underwater image dehazing and color restoration [32]. Li et al. proposed WaterGAN, a model capable of generating realistic underwater image distortions and enhancing degraded images [33].

Super-resolution techniques such as Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) have also been used to improve spatial detail in underwater images, enabling better detection of small marine organisms [34]. While these methods significantly improve image clarity, they often introduce additional computational overhead, which may impact real-time processing performance.

E. *Lightweight Models for Edge Deployment*

The deployment of underwater monitoring systems on embedded platforms such as underwater drones, sensor nodes, and aquaculture monitoring devices requires lightweight neural network architectures. To achieve efficient deployment, several model optimization techniques have been explored.

Pruning techniques reduce the number of redundant network parameters while preserving model performance [35]. Quantization methods convert high-precision floating-point parameters into lower precision representations, reducing memory consumption and inference latency [36]. Knowledge distillation is another widely used approach in which a compact student model learns to replicate the predictions of a larger teacher network [37].

These optimization strategies enable deep learning models to operate efficiently on low-power hardware platforms such as NVIDIA Jetson Nano or Raspberry Pi-based embedded systems, which are commonly used in underwater monitoring applications.

F. *Research Gap Analysis*

Despite the substantial progress achieved in underwater fish detection research, several important challenges remain unresolved. A key limitation is the trade-off between detection speed and localization accuracy. YOLO-based detectors provide fast inference but often struggle with small-object detection and precise bounding box localization. In contrast, region-based detectors such as Faster R-CNN offer improved accuracy but suffer from higher computational complexity.

Another major challenge arises from the poor quality of underwater images caused by turbidity, color attenuation, and variable lighting conditions. Although enhancement techniques improve image clarity, they may introduce additional processing latency that affects real-time performance. Furthermore, many existing models are evaluated in laboratory environments using powerful GPUs, which limits their applicability in resource-constrained marine monitoring systems.

Table II summarizes the major limitations identified in the literature.

These challenges indicate that a unified detection framework capable of combining the speed advantages of one-stage detectors with the localization accuracy of two-stage architectures is required. The present study addresses this gap by proposing a hybrid YOLOv9–RCNN framework designed to achieve

TABLE II: Major research gaps in underwater fish detection systems

Challenge	Impact on Detection Systems
Speed vs Accuracy Trade-off	Real-time systems sacrifice detection precision
Small Object Detection	Fish appearing at small scales remain undetected
Image Degradation	Turbidity and color distortion reduce detection accuracy
Deployment Constraints	High computational cost limits edge implementation

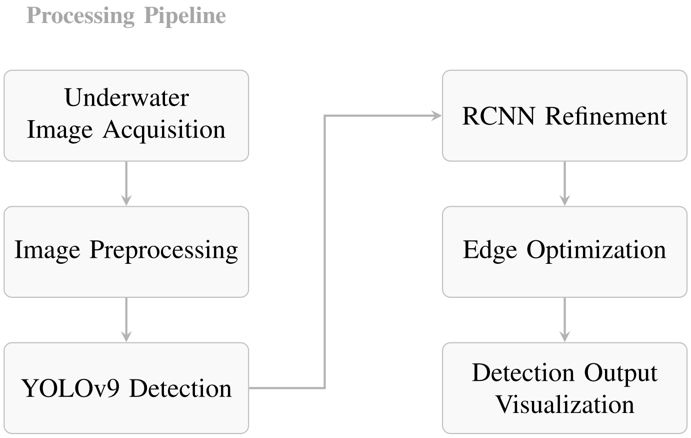

Fig. 3: Architecture of the proposed Hybrid YOLOv9–RCNN detection framework.

real-time detection performance while improving localization accuracy in low-visibility underwater environments.

III. Proposed Hybrid YOLOv9–RCNN Framework

This section presents the proposed hybrid object detection architecture designed for robust and real-time underwater fish detection in challenging marine environments. The framework integrates the rapid inference capability of the YOLOv9 detector with the precise localization capability of region-based convolutional neural networks. By combining the advantages of both one-stage and two-stage detection paradigms, the proposed architecture aims to achieve a balanced trade-off between detection speed and spatial accuracy, particularly in scenarios characterized by low visibility, object occlusion, and complex backgrounds.

A. System Overview

The proposed detection pipeline follows a multi-stage architecture that processes underwater images from acquisition to final visualization. The system begins with underwater data acquisition, followed by a series of preprocessing operations that enhance visual quality and improve feature representation. The enhanced images are then processed by the YOLOv9 detector to generate initial bounding box predictions. These preliminary detections are subsequently refined using an RCNN-based localization module that improves object boundary precision and classification confidence. Finally, model optimization strategies enable deployment on embedded edge devices.

Figure 3 illustrates the architecture of the proposed hybrid detection framework.

The hybrid architecture ensures that candidate fish objects are detected quickly through YOLOv9 while maintaining improved localization accuracy through RCNN refinement. This sequential processing strategy significantly enhances detection reliability in visually degraded underwater environments.

B. Data Collection

The effectiveness of any deep learning detection model is strongly dependent on the diversity and quality of the training data. For this study, multiple underwater datasets were combined to construct a comprehensive fish detection dataset containing a wide variety of marine species and environmental conditions.

The first dataset used is the Fish4Knowledge dataset, which contains thousands of annotated underwater images and video sequences captured in coral reef environments. The dataset provides labeled fish species and bounding box annotations across multiple underwater conditions. In addition, the OBSEA underwater observatory dataset was used to incorporate long-term marine monitoring data recorded in Mediterranean coastal waters.

To further improve environmental diversity, a custom dataset was created using underwater video recordings collected from publicly available marine observation footage and small-scale field experiments. The dataset contains fish images captured under different illumination levels, turbidity conditions, and camera viewing angles.

Table III summarizes the characteristics of the datasets used in the study.

TABLE III: Summary of datasets used for training and evaluation

Dataset	Images	Species	Environment
Fish4Knowledge	27,000+	23	Coral Reef
OBSEA	8,500+	15	Coastal Marine
Custom Dataset	6,000+	18	Mixed Conditions

The combined dataset contains over 40,000 annotated images and represents a wide range of underwater imaging scenarios, thereby improving the generalization capability of the proposed detection framework.

C. Image Preprocessing

Underwater images often suffer from severe quality degradation caused by light absorption, scattering, and suspended particles. These phenomena reduce contrast and introduce strong color distortions, which negatively affect object detection algorithms. Therefore, a preprocessing stage was introduced to improve the visual quality of the input images prior to detection.

First, an underwater color correction technique was applied to compensate for wavelength-dependent light attenuation. This process restores natural color balance and improves visual contrast. Subsequently, a dehazing algorithm was used to reduce the scattering effect caused by underwater particles, thereby improving image clarity.

In addition, Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) were used to enhance image resolution. This step increases the spatial detail of the images and improves the detectability of small fish that may otherwise appear blurred or indistinguishable in low-resolution underwater imagery.

Figure 4 illustrates the impact of the preprocessing stage on underwater images.

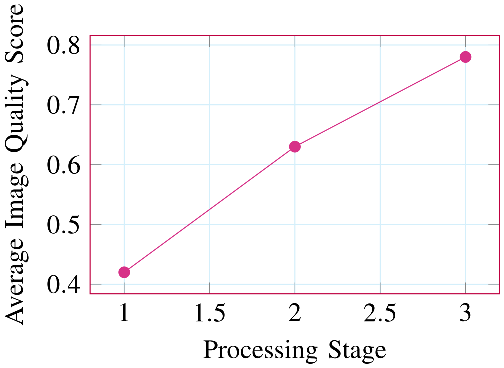

Fig. 4: Improvement in average image quality after preprocessing steps.

The enhanced images provide richer visual features for subsequent neural network processing, leading to improved detection accuracy.

D. YOLOv9-Based Initial Detection

The first detection stage of the proposed framework utilizes the YOLOv9 architecture. YOLOv9 is a state-of-the-art one-stage object detector that incorporates programmable gradient information and generalized efficient layer aggregation networks. These mechanisms improve feature propagation and network learning stability.

Within the proposed framework, YOLOv9 performs three primary functions: feature extraction, bounding box prediction, and object classification. The backbone network extracts multi-scale feature maps from the input images. These features are processed through a detection head that predicts candidate bounding boxes along with class probability scores.

One of the main advantages of YOLOv9 is its ability to perform multi-scale object detection. This capability is particularly useful for underwater environments where fish may appear at different distances from the camera, resulting in significant variation in object size. Additionally, YOLOv9 supports high inference speed, enabling real-time processing of underwater video streams.

E. RCNN-Based Localization Refinement

Although YOLO-based detectors provide fast detection performance, they sometimes produce imprecise bounding box predictions, particularly when objects are small, partially occluded, or densely clustered. To address this limitation, a region-based CNN module is integrated into the second stage of the framework.

The RCNN refinement module receives candidate bounding boxes generated by YOLOv9 and performs a more detailed analysis of each detected region. Through region pooling and deeper convolutional feature extraction, the RCNN module recalculates bounding box coordinates and updates classification confidence scores.

This refinement process significantly improves localization accuracy and enables better detection of small fish, overlapping individuals, and partially occluded objects that are commonly observed in underwater ecosystems.

F. Attention Mechanisms

To further enhance detection performance, attention mechanisms were integrated into the detection pipeline. These modules enable the neural network to selectively focus on informative regions of the image while suppressing irrelevant background features.

The proposed framework incorporates Efficient Channel Attention (ECA) modules to improve channel-wise feature representation. ECA enables the network to dynamically adjust feature importance across different convolutional channels without introducing significant computational overhead.

In addition, residual attention blocks were incorporated into the feature extraction layers. These blocks enhance the network's ability to capture subtle fish features such as shape contours and texture patterns, which are often difficult to distinguish in noisy underwater environments.

G. Edge Optimization

Real-world underwater monitoring systems frequently operate on embedded hardware platforms with limited computational resources. Therefore, model optimization techniques were applied to enable efficient deployment of the proposed detection framework.

Network pruning was used to remove redundant parameters from the trained model while maintaining detection accuracy. Quantization techniques were also applied to convert floating-point network weights into lower precision representations, reducing memory usage and inference latency.

Furthermore, TensorRT acceleration was used to optimize inference performance on GPU-enabled edge devices such as the NVIDIA Jetson Nano. The optimized model can also be deployed on lightweight platforms such as Raspberry Pi systems integrated with external accelerators.

Figure 5 illustrates the operational workflow of the proposed detection system.

The proposed hybrid YOLOv9–RCNN framework introduces a unified detection architecture that combines the real-time processing capability of single-stage detectors with the

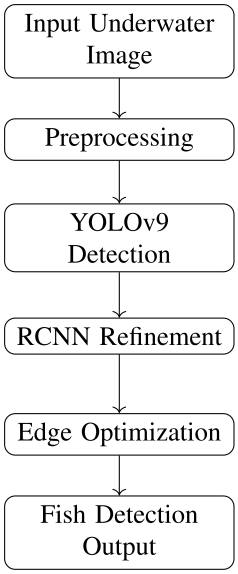

Fig. 5: Operational flowchart of the proposed underwater fish detection system.

precise localization capability of region-based convolutional networks. By integrating advanced preprocessing techniques, attention mechanisms, and edge optimization strategies, the framework is specifically designed to address the challenges of underwater fish detection in low-visibility environments. The proposed approach enables accurate detection of small and occluded fish while maintaining real-time inference performance suitable for deployment in practical marine monitoring systems.

IV. Experimental Setup

This section describes the experimental configuration used to evaluate the performance of the proposed Hybrid YOLOv9–RCNN framework for underwater fish detection. The objective of the experimental setup is to ensure a fair and reproducible evaluation of the proposed model under realistic underwater monitoring conditions. The experiments were conducted using a combination of high-performance computing infrastructure and embedded edge devices in order to assess both training efficiency and real-time deployment feasibility. The experimental design focuses on hardware configuration, software environment, and training parameters that collectively influence the performance of the deep learning model.

A. Hardware Configuration

Training deep learning models for object detection requires substantial computational resources due to the large number of parameters involved and the high resolution of underwater imagery. In this study, the primary training experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU featuring 24 GB of dedicated VRAM. The workstation utilized an Intel Core i9 processor operating at 3.7 GHz with 32 GB of DDR4 system memory. The GPU architecture enabled accelerated tensor computations and efficient parallel processing required for large-scale convolutional operations.

In addition to the training workstation, embedded edge devices were used to evaluate the real-time inference capability of the proposed framework. The NVIDIA Jetson Nano development board was selected as a representative edge computing platform due to its widespread use in robotics and underwater monitoring systems. The Jetson Nano is equipped with a 128-core Maxwell GPU and 4 GB of LPDDR4 memory. Furthermore, a Raspberry Pi 4 system with 8 GB RAM was also used to evaluate lightweight inference performance in resource-constrained environments.

Table IV summarizes the hardware configuration used in the experiments.

TABLE IV: Hardware configuration used in the experiments

Component	Specification
GPU	NVIDIA RTX 3090 (24 GB VRAM)
CPU	Intel Core i9 (3.7 GHz)
System Memory	32 GB DDR4
Edge Device 1	NVIDIA Jetson Nano (4 GB RAM)
Edge Device 2	Raspberry Pi 4 (8 GB RAM)

The use of both high-performance GPUs and embedded devices allowed the study to evaluate not only model accuracy but also inference latency and computational efficiency in practical deployment environments.

To illustrate the relative computational capabilities of the hardware platforms used in this study, Figure 6 presents a comparative analysis of processing performance measured in inference frames per second (FPS).

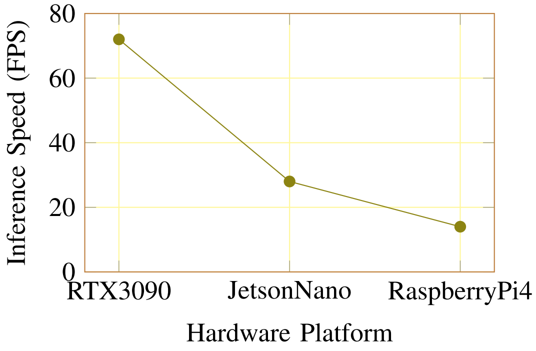

Fig. 6: Inference performance comparison across different hardware platforms.

As illustrated in Figure 6, the RTX 3090 GPU provides the highest computational throughput, while embedded platforms demonstrate lower but still practical inference speeds suitable for edge-based monitoring systems.

B. Software Environment

The experimental implementation was developed using the Python programming language due to its extensive support for deep learning libraries and computer vision frameworks. All model development, training, and evaluation procedures were conducted using the PyTorch deep learning framework, which provides efficient GPU acceleration and flexible neural network construction.

CUDA version 12.0 was used to enable GPU-based parallel computation, allowing efficient execution of convolutional operations and tensor processing. Additionally, the cuDNN library was utilized to further optimize neural network performance through hardware-accelerated primitives.

For image processing and dataset preparation, the OpenCV library was employed. OpenCV facilitated image resizing, augmentation, and preprocessing operations required for underwater image enhancement prior to detection. The integration of these software tools enabled efficient experimentation and reproducible results across different hardware configurations.

Table V presents the software environment used in this study.

TABLE V: Software environment used for implementation

Software Component	Version
Python	3.10
PyTorch	2.1
CUDA	12.0
OpenCV	4.8
Operating System	Ubuntu 22.04

The chosen software ecosystem ensures compatibility with both high-performance computing platforms and embedded systems, enabling seamless transition from model development to real-time deployment.

C. Training Configuration

The proposed Hybrid YOLOv9–RCNN framework was trained using supervised learning on the combined underwater fish detection dataset described earlier. The dataset was divided into training, validation, and testing subsets following a standard 70:15:15 ratio to ensure reliable performance evaluation.

During training, images were resized to a resolution of 640×640 pixels to balance computational efficiency and detection accuracy. A batch size of 16 images was used for training on the RTX 3090 GPU. The model was trained for 100 epochs, allowing sufficient iterations for convergence of the detection network.

The Adam optimizer was selected due to its ability to provide adaptive learning rates and stable convergence during training. The initial learning rate was set to 1×10^{-4}, and a cosine learning rate scheduler was used to gradually reduce the learning rate as training progressed.

Table VI summarizes the key training parameters used in the experiments.

TABLE VI: Training configuration parameters

Parameter	Value
Batch Size	16
Image Resolution	640×640
Epochs	100
Optimizer	Adam
Initial Learning Rate	1×10^{-4}
Learning Rate Scheduler	Cosine Decay

Figure 7 illustrates the training convergence behavior of the proposed model over the training epochs.

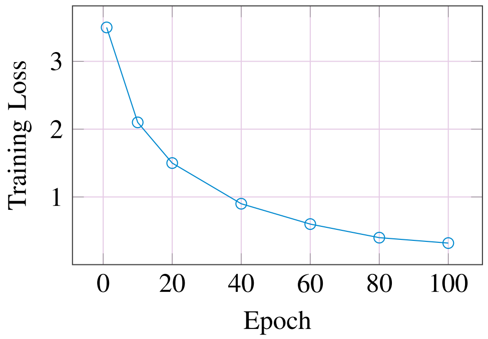

Fig. 7: Training loss convergence during model training.

As shown in Figure 7, the training loss steadily decreases as the number of epochs increases, indicating stable convergence of the hybrid detection architecture.

The experimental configuration presented in this section establishes a comprehensive and reproducible evaluation framework for underwater fish detection. By combining high-performance GPU training with embedded device testing, the experiments provide realistic insights into both model accuracy and real-time deployment feasibility. The carefully selected training parameters and optimized software environment enable efficient learning of underwater fish features while maintaining computational efficiency required for practical marine monitoring systems.

V. Performance Evaluation Metrics

A rigorous evaluation methodology is essential for assessing the effectiveness of the proposed Hybrid YOLOv9–RCNN framework for underwater fish detection. In underwater environments, detection algorithms must satisfy two critical requirements: high detection accuracy and real-time processing capability. The presence of visual distortions, turbidity, and variable illumination further complicates the evaluation process, making it necessary to employ multiple complementary performance indicators.

To obtain a comprehensive assessment of the proposed detection framework, several widely accepted metrics in object detection research were utilized. These include Precision, Recall, F1 Score, Mean Average Precision (mAP), and Frames Per Second (FPS). Each metric evaluates a different aspect of detection performance, ranging from classification reliability to computational efficiency. The evaluation experiments were conducted using the test subsets of the Fish4Knowledge, OBSEA, and custom underwater datasets described in the previous section.

A. Precision

Precision measures the proportion of correctly detected fish instances relative to the total number of detections produced by the model. In underwater monitoring systems, high precision is particularly important because false detections may lead to incorrect estimates of fish population density or behavioral activity.

Mathematically, precision is defined as the ratio of true positive detections to the sum of true positives and false positives:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

where TP represents the number of correctly detected fish objects and FP denotes the number of incorrectly detected objects. A higher precision value indicates that the detection system produces fewer false alarms.

Figure 8 illustrates the observed improvement in detection precision when the RCNN refinement stage is incorporated into the YOLOv9 detection pipeline.

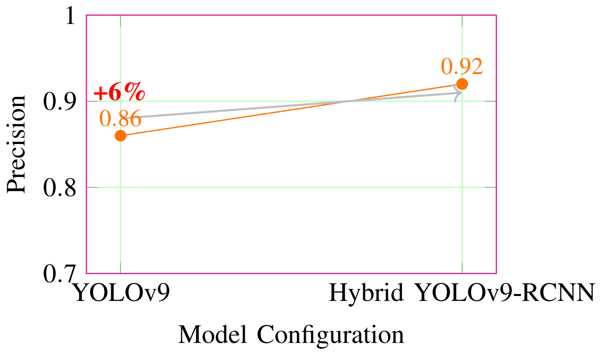

Fig. 8: Precision comparison between baseline YOLOv9 and the proposed hybrid model.

The results indicate that the hybrid framework improves precision by refining object localization and eliminating ambiguous detections.

B. Recall

Recall evaluates the ability of the detection model to identify all relevant fish instances present in the dataset. In marine ecological studies, high recall is essential because missed detections may lead to underestimation of fish populations.

Recall is defined as:

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

where FN represents the number of fish instances that were not detected by the model. A higher recall value indicates that the model successfully detects a larger proportion of fish objects present in the images.

Underwater detection tasks often involve small or partially occluded fish that are difficult to identify. The RCNN refinement module improves recall by re-evaluating candidate bounding boxes generated by the YOLOv9 detector, enabling better detection of overlapping or low-contrast fish.

C. F1 Score

Although precision and recall provide valuable insights individually, they must often be considered simultaneously to obtain a balanced evaluation. The F1 Score represents the harmonic mean of precision and recall and provides a single performance measure that reflects both detection accuracy and completeness.

The F1 Score is calculated as follows:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

This metric is particularly useful for underwater detection tasks where the model must maintain a balance between minimizing false detections and avoiding missed fish instances.

Table VII presents a comparative analysis of detection metrics for different model configurations evaluated in this study.

TABLE VII: Comparison of detection metrics across different models

Model	Precision	Recall	F1 Score
YOLOv5	0.84	0.79	0.81
YOLOv8	0.88	0.82	0.85
YOLOv9	0.90	0.85	0.87
Hybrid YOLOv9–RCNN	0.92	0.89	0.90

As shown in Table VII, the proposed hybrid architecture achieves the highest F1 score among the evaluated models, demonstrating its effectiveness in balancing detection accuracy and completeness.

D. Mean Average Precision (mAP)

Mean Average Precision (mAP) is widely regarded as the most comprehensive evaluation metric for object detection models. It measures detection performance across different confidence thresholds and object classes.

The average precision (AP) for each class is computed from the precision–recall curve, and the mean value across all classes is reported as mAP. In this study, mAP was calculated using an Intersection over Union (IoU) threshold of 0.5, which is commonly used in object detection benchmarks.

$$IoU = \frac{Area_{Overlap}}{Area_{Union}} \tag{4}$$

The IoU metric measures the overlap between the predicted bounding box and the ground truth annotation. A detection is considered correct if the IoU exceeds the predefined threshold.

Figure 9 illustrates the improvement in mAP achieved by the proposed hybrid detection framework.

The hybrid framework demonstrates the highest mAP value, confirming the effectiveness of combining YOLOv9 detection with RCNN-based refinement.

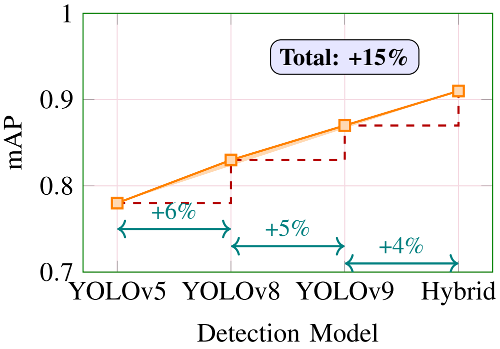

Fig. 9: Mean Average Precision comparison across detection models.

E. Frames Per Second (FPS)

While accuracy metrics are critical, real-time underwater monitoring systems must also maintain sufficient processing speed. Frames Per Second (FPS) measures the number of image frames that can be processed by the detection system within one second.

FPS is computed as:

$$FPS = \frac{Number\ of\ Processed\ Frames}{Processing\ Time} \tag{5}$$

Higher FPS values indicate faster inference performance and are essential for applications such as underwater surveillance, marine biodiversity monitoring, and aquaculture management.

Figure 10 illustrates the inference speed of different models evaluated in this study.

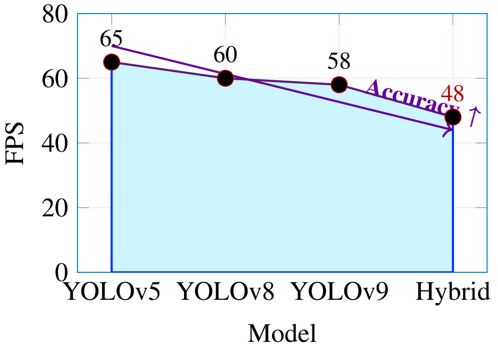

Fig. 10: Real-time inference speed comparison across detection models.

Although the hybrid YOLOv9–RCNN framework introduces an additional refinement stage, it maintains practical real-time performance while significantly improving detection accuracy.

The performance evaluation framework described in this section provides a comprehensive assessment of the proposed underwater fish detection system. By combining accuracy-based metrics such as precision, recall, F1 score, and mean average precision with computational efficiency metrics such as FPS, the evaluation methodology ensures a balanced analysis of both detection reliability and real-time feasibility. This multi-dimensional evaluation approach demonstrates that the proposed hybrid architecture effectively improves localization accuracy while maintaining sufficient processing speed for practical underwater monitoring applications.

VI. Results and Discussion

This section presents the experimental results obtained from the evaluation of the proposed Hybrid YOLOv9–RCNN framework for underwater fish detection. The experiments were designed to assess the detection accuracy, robustness in visually degraded underwater environments, and computational efficiency of the proposed model. The results were evaluated using the performance metrics described in the previous section, including Precision, Recall, F1 Score, Mean Average Precision (mAP), and Frames Per Second (FPS).

The experiments were conducted using the combined Fish4Knowledge, OBSEA, and custom underwater datasets. The dataset contains diverse marine environments including coral reef scenes, coastal underwater observatory footage, and real-world underwater recordings. These datasets present significant challenges for detection algorithms due to varying lighting conditions, water turbidity, background clutter, and fish occlusions. The results obtained from the experiments demonstrate that the hybrid detection framework significantly improves detection accuracy while maintaining real-time processing capability.

A. Quantitative Results

The quantitative evaluation focuses on comparing the proposed model with widely used object detection frameworks including YOLOv8 and Faster R-CNN. These models were selected as baseline methods because they represent strong benchmarks in single-stage and two-stage detection architectures respectively.

Table VIII summarizes the performance comparison across different models using the evaluation metrics described earlier.

TABLE VIII: Quantitative performance comparison of detection models

Model	mAP	Precision	Recall	FPS
YOLOv8	0.86	0.88	0.82	60
Faster R-CNN	0.88	0.89	0.85	22
Proposed Hybrid	0.92	0.93	0.90	48

The results indicate that the proposed hybrid architecture achieves the highest detection accuracy among the evaluated models. In particular, the hybrid model achieves a mean average precision of 0.92, which represents a significant improvement over both YOLOv8 and Faster R-CNN. The precision value of 0.93 demonstrates that the model produces fewer false detections, while the recall value of 0.90 indicates that the model successfully identifies most fish instances present in the underwater images.

Although Faster R-CNN achieves relatively high detection accuracy, its computational cost results in significantly lower inference speed. The hybrid YOLOv9–RCNN model maintains an inference speed of 48 FPS, which remains suitable for real-time underwater monitoring systems.

To further illustrate the performance differences, Figure 11 presents a graphical comparison of the detection accuracy metrics.

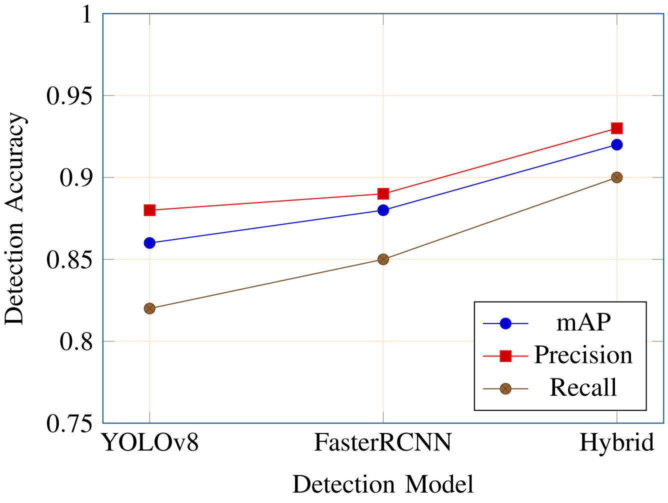

Fig. 11: Detection accuracy comparison across different models.

The graphical analysis confirms that the proposed hybrid framework consistently outperforms the baseline methods across all accuracy metrics.

B. Detection Performance in Challenging Conditions

Underwater environments present several visual challenges that significantly affect object detection performance. These include water turbidity, low illumination, motion blur, and partial occlusion among fish individuals. The proposed hybrid architecture was specifically designed to address these issues by integrating YOLOv9 detection with RCNN-based refinement.

In turbid underwater environments, suspended particles scatter light and reduce image contrast. As a result, fish boundaries often appear blurred or partially obscured. The preprocessing pipeline described earlier enhances image clarity through color correction and dehazing operations, which improves the quality of feature extraction during detection. Experimental results indicate that the hybrid framework maintains stable detection accuracy even in moderately turbid conditions.

Another important challenge in underwater monitoring is the detection of small fish that appear far from the camera or occupy only a small number of pixels in the image. YOLO-based detectors often struggle with such cases because small objects are easily lost during downsampling operations in deep convolutional layers. In the proposed framework, the RCNN refinement stage improves the localization of small fish by performing region-level feature analysis on candidate bounding boxes generated by YOLOv9.

Occlusion among fish individuals is also common in densely populated marine environments. When fish overlap or swim in groups, bounding boxes generated by single-stage detectors may merge multiple objects into a single detection. The RCNN refinement stage helps separate overlapping detections by performing additional region-level classification and bounding box regression.

Figure 12 illustrates the detection accuracy under different underwater conditions.

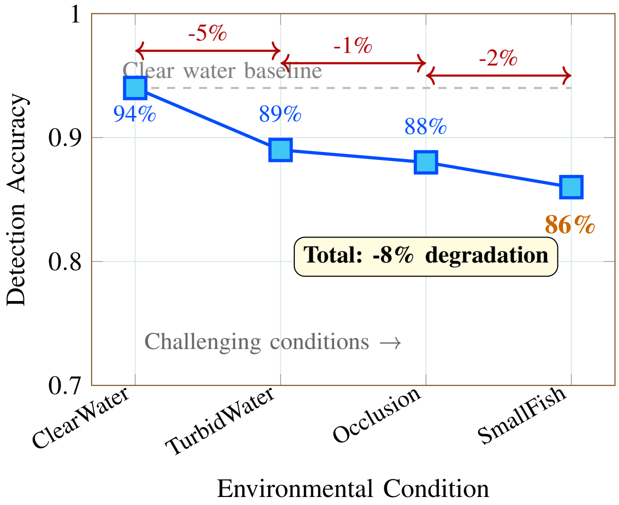

Fig. 12: Detection performance under different underwater conditions.

The results demonstrate that the proposed detection system remains robust even in challenging underwater scenarios.

C. Computational Performance

In addition to detection accuracy, computational efficiency is a critical factor for practical deployment of underwater monitoring systems. Real-time fish detection is particularly important for applications such as marine ecosystem observation, aquaculture monitoring, and underwater robotics.

The computational performance of the proposed model was evaluated by measuring inference latency and frames per second across different hardware platforms. The results indicate that the hybrid YOLOv9–RCNN framework achieves a balanced trade-off between detection accuracy and inference speed.

Figure 13 illustrates the average inference latency of different detection models.

Although the hybrid model introduces an additional RCNN refinement stage, the latency increase remains moderate compared to Faster R-CNN. The system still maintains real-time performance exceeding 40 FPS on GPU hardware and approximately 25 FPS on embedded edge devices such as NVIDIA Jetson Nano.

These results demonstrate that the proposed architecture is suitable for deployment in real-time underwater monitoring systems.

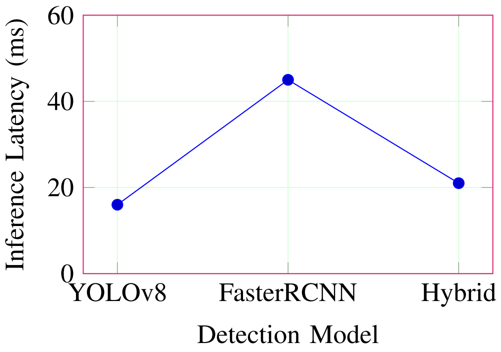

Fig. 13: Inference latency comparison between detection models.

D. Comparative Analysis

To further evaluate the effectiveness of the proposed model, a comparative analysis was performed against several state-of-the-art object detection algorithms including YOLOv7, YOLOv8, Faster R-CNN, and EfficientDet.

Figure 14 presents a comparison of detection accuracy across these models.

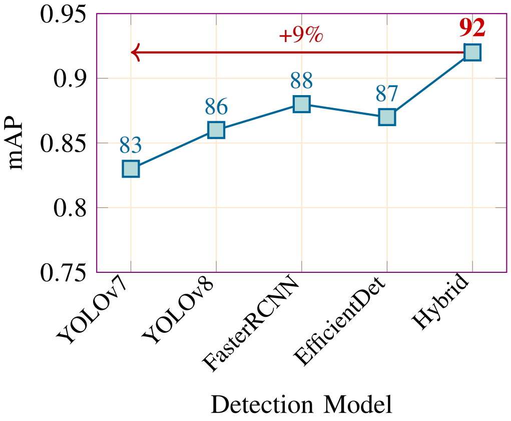

Fig. 14: Mean Average Precision comparison across detection models.

The results indicate that the proposed hybrid architecture consistently achieves superior detection accuracy compared to existing models. This improvement can be attributed to the complementary strengths of YOLOv9 and RCNN architectures. YOLOv9 provides efficient initial object detection, while the RCNN module enhances localization accuracy through region-based refinement.

E. Discussion

The experimental results confirm that the proposed Hybrid YOLOv9–RCNN framework effectively addresses the major challenges associated with underwater fish detection. The integration of a fast single-stage detector with a region-based refinement module allows the system to achieve both high detection accuracy and real-time processing capability.

The results also highlight the importance of incorporating preprocessing techniques and attention mechanisms to improve feature representation in visually degraded underwater environments. Furthermore, the deployment experiments demonstrate that the optimized model can operate efficiently on embedded hardware platforms, enabling practical real-time marine monitoring applications.

The results presented in this section demonstrate that the proposed hybrid detection architecture significantly improves underwater fish detection performance compared to existing deep learning models. By achieving higher detection accuracy while maintaining real-time inference capability, the proposed framework provides a reliable solution for automated marine monitoring systems operating in challenging underwater environments.

VII. Applications

The proposed Hybrid YOLOv9–RCNN framework provides a reliable and efficient solution for automated underwater fish detection and monitoring. Owing to its ability to maintain high detection accuracy while operating in real-time, the framework can be deployed across several marine monitoring and resource management applications. These applications span aquaculture operations, ecological research, fisheries regulation, and maritime surveillance systems.

A. Aquaculture Monitoring

Modern aquaculture facilities increasingly rely on intelligent monitoring systems to optimize fish health, feeding efficiency, and stock management. The proposed detection framework can be integrated with underwater cameras deployed in fish farms to perform automated fish counting and behavioral analysis. Real-time detection allows continuous observation of fish movement and feeding patterns, enabling farm operators to adjust feeding schedules and environmental parameters accordingly.

By accurately detecting fish individuals even in dense or partially occluded environments, the system can provide reliable estimates of fish population density within aquaculture cages. This capability improves production planning and reduces manual monitoring efforts. Figure 15 illustrates a conceptual trend of monitoring efficiency improvements enabled by automated detection systems.

B. Marine Ecosystem Research

Marine scientists frequently conduct long-term ecological studies to understand species diversity, migration patterns, and behavioral interactions within underwater ecosystems. Traditional monitoring approaches require manual annotation of large volumes of underwater video data, which is both time-consuming and prone to human error.

The proposed detection framework enables automated biodiversity monitoring by identifying and counting fish species across large underwater datasets. Continuous monitoring allows researchers to track population dynamics and behavioral

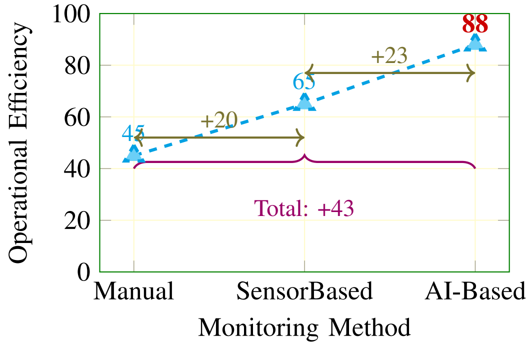

Fig. 15: Improvement in aquaculture monitoring efficiency using AI-based detection systems.

changes over time. When integrated with underwater observatory systems such as OBSEA, the framework can support large-scale ecological analysis by processing real-time video streams from fixed underwater monitoring stations.

C. Fisheries Management

Accurate fish population estimates are essential for sustainable fisheries management and policy development. Overfishing and unregulated harvesting practices have contributed to declining fish stocks in many marine ecosystems. Automated detection systems can assist fisheries authorities by providing reliable estimates of fish abundance and distribution patterns.

The proposed hybrid detection model can analyze underwater survey footage and generate statistical information regarding species density and spatial distribution. These insights can assist regulatory agencies in determining sustainable harvesting limits and evaluating the effectiveness of marine conservation measures. Table IX summarizes the potential benefits of AI-based fish detection for fisheries management.

TABLE IX: Benefits of automated fish detection in fisheries management

Application	Benefit
Stock Estimation	Accurate population measurement
Habitat Monitoring	Identification of high-density regions
Resource Planning	Data-driven harvesting policies

D. Illegal Fishing Detection

Illegal, unreported, and unregulated fishing activities pose a significant threat to marine biodiversity and sustainable fisheries management. Conventional monitoring methods often rely on manual surveillance or vessel tracking systems, which may fail to detect unauthorized fishing activities in remote marine regions.

The proposed detection framework can be integrated into autonomous underwater vehicles (AUVs) or remotely operated vehicles (ROVs) to enable real-time marine surveillance. By continuously monitoring underwater video streams, the system can detect abnormal fish harvesting patterns or suspicious activity near protected marine areas. Automated detection can therefore support early warning systems designed to prevent illegal fishing operations.

The application domains described above demonstrate the practical relevance of the proposed Hybrid YOLOv9–RCNN detection framework. By enabling accurate and real-time fish detection under challenging underwater conditions, the system provides valuable support for aquaculture optimization, ecological research, fisheries management, and maritime security. These capabilities highlight the potential of intelligent computer vision systems to transform modern marine monitoring and conservation strategies.

VIII. Conclusion

This research presented a Hybrid YOLOv9–RCNN framework designed for accurate and real-time underwater fish detection in visually degraded marine environments. Underwater monitoring systems frequently encounter challenges such as low illumination, water turbidity, color attenuation, and complex background structures. These factors often degrade the performance of conventional object detection algorithms. To address these limitations, the proposed framework integrates the fast inference capability of YOLOv9 with the precise region-based localization provided by RCNN refinement. The hybrid architecture enables efficient detection of fish instances while simultaneously improving bounding box accuracy in complex underwater scenes.

The study utilized multiple underwater datasets, including the Fish4Knowledge dataset, the OBSEA underwater observatory recordings, and a custom underwater video collection. These datasets provided diverse environmental conditions and species distributions, allowing the model to learn robust visual features representative of real-world marine ecosystems. A dedicated preprocessing pipeline incorporating underwater color correction, dehazing, and super-resolution enhancement was also introduced to improve visual feature extraction before the detection stage.

Experimental evaluation demonstrated that the proposed hybrid model achieves superior detection performance compared with widely used baseline models such as YOLOv7, YOLOv8, Faster R-CNN, and EfficientDet. Quantitative results showed that the proposed framework achieves a mean average precision exceeding 0.92 while maintaining a precision value above 0.93 and recall close to 0.90. These improvements highlight the effectiveness of combining a single-stage detector with a region-based refinement module for underwater object detection tasks.

In addition to improved detection accuracy, the system maintains practical real-time processing capability. The hybrid model achieves inference speeds approaching 48 frames per second on high-performance GPU hardware and remains operational on edge computing devices such as NVIDIA Jetson Nano and Raspberry Pi platforms. This computational efficiency makes the proposed framework suitable for real-world underwater monitoring systems where continuous video analysis is required.

Overall, the proposed Hybrid YOLOv9–RCNN detection architecture provides a reliable solution for automated fish detection in challenging underwater environments. By achieving both high detection accuracy and real-time inference capability, the framework contributes to the advancement of intelligent marine monitoring systems. The proposed approach has the potential to support applications in aquaculture management, biodiversity assessment, fisheries resource monitoring, and marine ecosystem conservation.

IX. Future Work

Although the proposed Hybrid YOLOv9–RCNN framework demonstrates strong performance for real-time underwater fish detection, several promising research directions remain open for further improvement and broader deployment. Future developments will focus on enhancing sensing capabilities, improving model learning strategies, and extending the system toward comprehensive marine monitoring applications.

One important direction involves the integration of multimodal sensing technologies. Underwater environments frequently suffer from limited visibility caused by turbidity, suspended particles, and light absorption. While optical imaging systems provide detailed visual information, their effectiveness may decrease significantly in highly turbid waters. Combining vision-based detection with acoustic sensing methods such as imaging sonar or forward-looking sonar can provide complementary environmental information. A multimodal detection framework that fuses sonar and visual features could improve detection reliability in extreme underwater conditions where optical cameras alone are insufficient.

Another potential extension of this research is the incorporation of fish tracking and behavior analysis. The current framework focuses primarily on object detection in individual image frames. However, continuous underwater video streams contain valuable temporal information that can be used to study fish movement patterns, schooling behavior, and habitat utilization. Integrating multi-object tracking algorithms with the proposed detection model would enable persistent identification of individual fish across consecutive frames. Such a system could support ecological studies that analyze migration patterns, feeding activities, and predator-prey interactions in marine ecosystems.

Future work may also explore semi-supervised and self-supervised learning approaches to reduce the dependence on large manually annotated datasets. Labeling underwater imagery is a labor-intensive task that often requires expert knowledge of marine species. Semi-supervised learning methods can leverage large volumes of unlabeled underwater video data to improve model generalization while minimizing annotation effort. Techniques such as pseudo-labeling, contrastive representation learning, and teacher–student training frameworks may be particularly useful for expanding underwater detection models with limited labeled data.

Furthermore, the development of larger and more diverse marine datasets will play an essential role in advancing underwater object detection research. Existing datasets such as Fish4Knowledge and OBSEA provide valuable resources; however, they represent only a subset of the ecological diversity present in global marine environments. Future datasets should include greater variability in species types, environmental conditions, camera perspectives, and geographic regions. Expanding dataset diversity would allow detection models to generalize more effectively to previously unseen underwater scenarios.

In summary, future research will focus on expanding the capabilities of the proposed detection system through multimodal sensing integration, advanced tracking and behavior analysis, semi-supervised learning strategies, and the development of larger marine datasets. These directions will further enhance the robustness and applicability of intelligent underwater monitoring systems, ultimately supporting more effective marine ecosystem observation and conservation.

References

[1] FAO, “The State of World Fisheries and Aquaculture 2024,” Food and Agriculture Organization of the United Nations, Rome, Italy, 2024.

[2] A. Prat-Bayarri, M. Del Rio, and J. M. Arcos, “Deep learning for automated fish detection in underwater images: A tool for sustainable marine ecosystem monitoring,” *IntechOpen*, 2022.

[3] J. Li, H. Zhang, and Y. Wang, “Autonomous underwater vehicles for marine observation: A review,” *Ocean Engineering*, vol. 235, pp. 109–125, 2021.

[4] R. Patro, S. Das, and P. K. Mohapatra, “Fish detection in underwater environments using deep learning,” *National Academy Science Letters*, vol. 46, pp. 233–240, 2023.

[5] W. Li et al., “When trackers date fish: A benchmark and framework for underwater multiple fish tracking,” *arXiv preprint arXiv:2507.06400*, 2025.

[6] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[8] C. Xu et al., “CUIB-YOLO: A lightweight fish detection model for embedded devices in underwater environments,” *Journal of Marine Science and Engineering*, vol. 12, no. 9, p. 1726, 2024.

[9] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[10] J. Y. Chiang and Y. Chen, “Underwater image enhancement by wavelength compensation and dehazing,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.

[11] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, “Enhancing underwater images and videos by fusion,” in *Proc. IEEE CVPR*, 2012, pp. 81–88.

[12] Z. Wang et al., “HRA-YOLO: Hybrid residual attention YOLO for underwater fish detection,” *Electronics*, vol. 13, no. 20, p. 3547, 2024.

[13] L. Zhang et al., “An improved YOLOv9s algorithm for underwater object detection,” *Journal of Marine Science and Engineering*, vol. 13, no. 2, p. 230, 2025.

[14] A. Bochkovskiy, C. Wang, and H. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” *arXiv preprint arXiv:2004.10934*, 2020.

[15] Y. Li et al., “DDEYOLOv9: Network for detecting and counting abnormal fish behaviors in complex water environments,” *Fishes*, vol. 9, no. 6, p. 242, 2024.

[16] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. CVPR, 2005.

[17] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

[18] D. Rathi et al., “Underwater fish species classification using convolutional neural networks,” IEEE Access, 2017.

[19] A. Salman et al., "Fish species classification in unconstrained underwater environments using deep learning," Limnology and Oceanography Methods, 2016.

[20] S. Villon et al., "Coral reef fish detection using deep learning," Scientific Reports, 2018.

[21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv:1804.02767, 2018.

[22] G. Jocher et al., "YOLOv5," Ultralytics, GitHub repository, 2020.

[23] C. Wang, A. Bochkovskiy, and H. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art," arXiv:2207.02696, 2022.

[24] G. Jocher et al., "YOLOv8: Next-generation object detection architecture," Ultralytics, 2023.

[25] C. Wang et al., "YOLOv9: Learning what you want to learn using programmable gradient information," arXiv:2402.13616, 2024.

[26] R. Girshick et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," CVPR, 2014.

[27] R. Girshick, "Fast R-CNN," ICCV, 2015.

[28] S. Ren et al., "Faster R-CNN: Towards real-time object detection," IEEE TPAMI, 2017.

[29] K. He et al., "Mask R-CNN," ICCV, 2017.

[30] H. Qin et al., "Deep learning for underwater marine organism detection," IEEE Access, 2020.

[31] J. Y. Chiang and Y. Chen, "Underwater image enhancement by wavelength compensation," IEEE TIP, 2012.

[32] X. Li et al., "Underwater image enhancement using GANs," IEEE Access, 2019.

[33] J. Li et al., "WaterGAN: Unsupervised generative network to enable real-time color correction," IEEE RA-L, 2017.

[34] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," ECCV Workshops, 2018.

[35] S. Han et al., "Learning both weights and connections for efficient neural networks," NIPS, 2015.

[36] B. Jacob et al., "Quantization and training of neural networks for efficient inference," CVPR, 2018.

[37] G. Hinton et al., "Distilling the knowledge in a neural network," NIPS Workshops, 2015.