

DEEP LEARNING

Real-Time AI-Based Detection of Fish and Humans in Underwater Environments

V. Anusha, M. Prarthana, R. Rakshitha, T B. Vathsala, S. Hanumanthappa

Department of Information Science and Engineering, Kalpataru Institute of Technology, Tiptur, Karnataka, India

Published October 2026

Abstract

Underwater visual monitoring is difficult because light attenuation, turbidity, color distortion, occlusion, and changing backgrounds can reduce the reliability of conventional object detection. This work presents an underwater human and fish detection system that combines a waterproof camera, image preprocessing, YOLO-based real-time object detection, and convolutional neural network (CNN) classification. The proposed system is intended to detect humans and fish in underwater video, classify fish by species and gender, and present the results through a monitoring dashboard. Human status is considered using observed movement and posture, while fish classification uses labeled underwater data such as the Fish4Knowledge dataset described in the project study. The implementation uses OpenCV and Python-based deep-learning components and provides real-time visualization with bounding boxes and class labels. The project report also defines tests for clear and low-visibility water, multiple objects, fast motion, partial occlusion, reflections, varying depth, debris, overlapping objects, and multi-frame operation. The demonstrated system provides a foundation for underwater surveillance, marine research, biodiversity monitoring, and search-and-rescue applications.

Keywords: computer vision; convolutional neural network; fish classification; underwater detection; YOLO

1 Introduction

Underwater environments contain varied marine life and support activities including fisheries, aquaculture, biodiversity observation, surveillance, diver safety, and search-and-rescue. Reviewing underwater video continuously by hand is labor-intensive and difficult to scale, creating a need for automated computer-vision methods. Fish are common targets in these scenes, while divers and swimmers may also require recognition for safety and monitoring.

Deep-learning methods have strengthened underwater visual analysis by learning image representations from training data. Convolutional neural networks (CNNs) support tasks such as recognizing fish, locating them, separating their regions, following their movement, and assigning classes. However, underwater images differ greatly from ordinary terrestrial images because depth, illumination, suspended particles, and the surrounding environment can alter visibility and object appearance.

As light passes through water, different wavelengths are absorbed and scattered, which can produce color shifts, weak contrast, haze, blur, and loss of fine details. Suspended particles may introduce additional noise, while fish and humans can appear at different scales or orientations. Vegetation, rocks, coral, bubbles, equipment, and other background elements can also partially hide the target.

These effects make reliable detection more difficult and create a need for suitable preprocessing and robust detection models.

The proposed project addresses these challenges through a combined underwater monitoring pipeline. A waterproof USB camera supplies the video, OpenCV is used for preprocessing, and YOLO is used for real-time object detection. A CNN-based stage is used for fish recognition, with the intended classification covering species and, when suitable labeled data are available, gender. Human detections are further considered using observed movement and posture information.

A unified detector is useful because an underwater camera may observe fish and people in the same scene rather than as separate tasks. Existing research has investigated fish analysis using labeled datasets such as DeepFish and underwater scene understanding using datasets such as SUIM, which includes fish or vertebrate categories and human divers [11], [14]. Human detection has also been studied with region-based detectors and lightweight networks [17], [18]. These studies show the value of considering both recognition capability and computational cost.

Real-time operation is especially important for underwater robots and embedded systems because available memory, processing power, energy, and communication bandwidth can be limited. YOLO is relevant to this requirement because detection and classification are handled within a single-stage framework. Even so, performance can change when deployment conditions differ from the training data, particularly because of differences in water transparency, lighting, camera characteristics, depth, background, object scale, and occlusion [12], [16].

The project also includes a monitoring dashboard that presents the camera stream, detected objects, bounding boxes, class labels, and alerts. The complete workflow therefore connects image acquisition, enhancement, object detection, classification, visualization, and notification. The intended application is practical underwater monitoring for marine observation, safety, and search-and-rescue support.

2 Related work

Underwater object-detection research has moved from manually designed image features toward machine-learning and deep-learning methods. Earlier approaches relied on color, edges, texture, shape, motion, thresholding, background subtraction, optical flow, and handcrafted descriptors. Such methods can work in controlled scenes, but changes in illumination, water clarity, background motion, and object appearance reduce their reliability. Deep networks became attractive because they can learn more complex visual representations from data [11], [12]. Hanumanthappa and Guruprakash [19], [20] proposed brain tumor detection using deep learning features.

Fish analysis is an important area of underwater computer vision. CNN models can learn layered visual features that help separate fish from complex backgrounds, while labeled datasets provide a basis for training and comparison. For example, DeepFish supplies labels covering fish location, instance-level segmentation, class recognition, and size estimation [14]. Such datasets are useful, although results obtained in one location or environmental condition may not transfer directly to another underwater setting.

Semantic understanding is also useful when more than one object category is present. The SUIM dataset contains underwater categories including fish or vertebrates and human divers, showing that different target types can be represented within a common underwater scene [15]. This supports the

motivation for a unified fish-and-human detection system instead of treating the two classes as completely separate problems.

YOLO-based detectors have received attention where rapid inference is required. Their single-stage design combines localization and classification in one detection process, making them suitable candidates for real-time underwater video and robotic platforms. Research has also examined ensemble YOLO methods for underwater fish detection [16]. Nevertheless, a general detector still has to cope with color degradation, low contrast, small targets, occlusion, and complex backgrounds before it can be expected to perform reliably underwater.

Human detection presents additional difficulties because divers and swimmers can change in appearance with posture, viewing direction, distance, lighting, bubbles, water movement, and partial occlusion. Faster R-CNN with data augmentation has been investigated for underwater people detection, while lightweight architectures have been proposed to reduce computational requirements [17], [18]. These approaches indicate that accuracy and processing cost need to be considered together when the target platform has limited resources.

Image enhancement is another important part of the underwater vision pipeline. Color correction, contrast improvement, dehazing, white balancing, Retinex-based processing, and learned restoration methods can improve the appearance of underwater images. However, a visually clearer image does not necessarily provide the features that a detector needs most; therefore, enhancement should be evaluated in relation to the final detection task [12]. Small and partially hidden objects create another challenge because downsampling may remove useful spatial information. Multi-scale features, feature fusion, attention, and suitable augmentation can help address these cases.

Domain variation remains a major research issue. Images collected with different cameras, at different depths, or in different geographical locations can vary in brightness, contrast, color distribution, and background. Transfer learning, richer training variation, and adaptation to deployment conditions can therefore improve model robustness [12]. At the same time, compact models are important for underwater robots and embedded devices where memory, energy, and computation are restricted [18].

The reviewed work indicates several practical gaps. Many studies focus primarily on either fish or human detection, while a camera may encounter both categories in one scene. Detection can also degrade because of turbidity, light scattering, color changes, suspended particles, small targets, and occlusion. Differences between training and deployment environments can further reduce generalization, and large models may be difficult to run in real time on embedded platforms. A unified framework that evaluates fish and human detection together while considering recognition quality and computational efficiency is therefore useful for underwater monitoring.

3 Methodology

3.1 Image acquisition and preprocessing

The waterproof camera captures real-time underwater frames. The report specifies a high-resolution camera with infrared capability for low-light or murky conditions. Before detection, frames can be resized and processed for noise, contrast, color, and haze-related effects.

3.2 YOLO-based detection

YOLO treats object detection as a single forward-pass prediction problem. An input frame is resized to the model input dimensions, object locations and class probabilities are predicted, and overlapping predictions are filtered using non-maximum suppression. This design supports the real-time requirement stated in the project.

3.3 CNN-based fish classification

The fish-classification stage uses CNN concepts including convolution, nonlinear activation, pooling, and dense classification layers. The report proposes training or fine-tuning with underwater datasets and using features such as shape, size, and visual patterns for species and gender classification.

3.4 Human status analysis

For human detections, the project describes using movement patterns and posture to distinguish active human subjects from motionless bodies. This output is intended to support rapid attention in safety and rescue scenarios; the report does not provide a validated medical or forensic classification protocol.

Existing approaches described in the project report may face difficulty when following moving targets, can lose detection accuracy in challenging underwater conditions, and can be affected by changes in the apparent size, shape, and color of objects. These limitations highlight the need for a detection pipeline that remains dependable when visibility and target appearance vary.

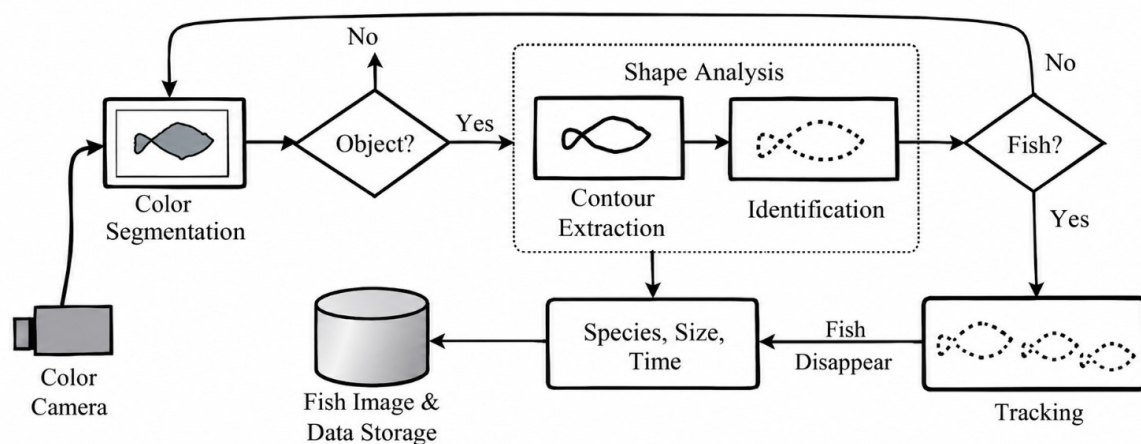


Figure 1. Architecture of the proposed underwater detection system.

Figure 1 shows that the proposed underwater detection architecture starts with the camera input and moves the captured frames through color segmentation, object validation, contour extraction, and shape identification. Valid targets are then processed for fish tracking and recording of species, size, and time details. The figure represents the overall workflow used for underwater monitoring.

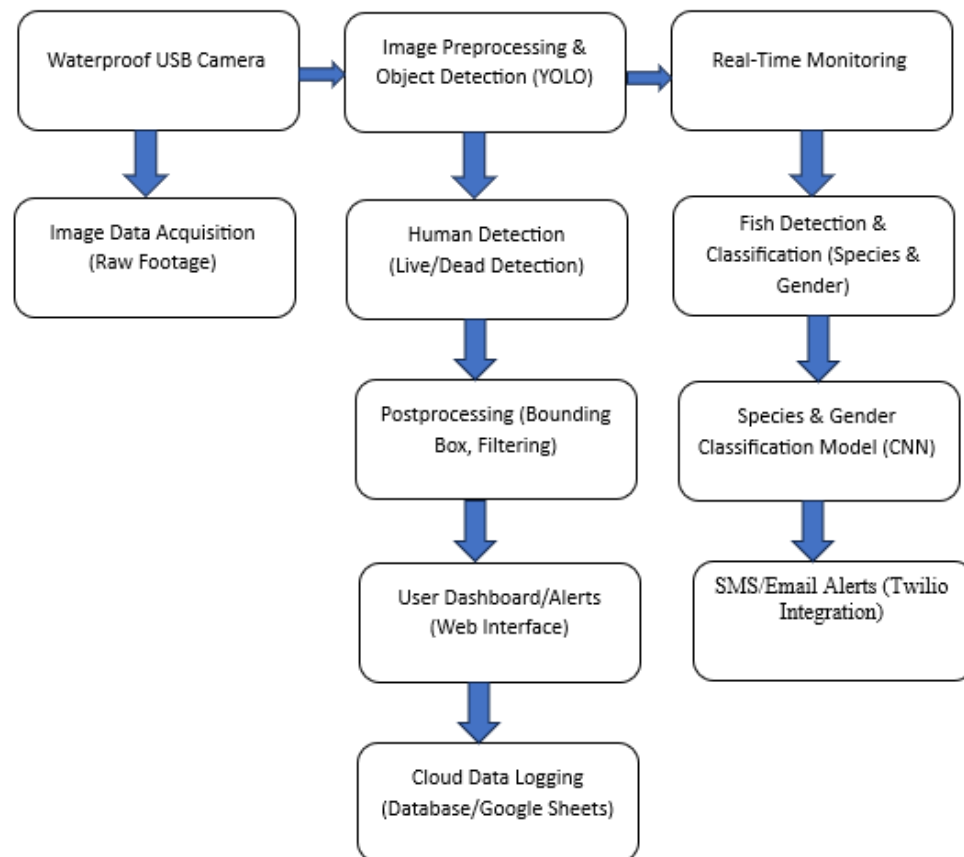


Figure 2. Data flow from image acquisition to dashboard and alerts.

Figure 2 shows that each acquired underwater frame passes through preprocessing and detection before the system separates human and fish objects into their respective analysis paths. Post-processing filters the outputs before they reach the monitoring dashboard. The resulting information can also be logged for observation, reporting, and alert generation.

The system receives underwater video from a waterproof USB camera and applies preprocessing before passing frames to the detection stage. Noise reduction, contrast adjustment, color correction, and image-enhancement operations are considered to compensate for visibility degradation.

YOLO performs real-time object detection by producing bounding boxes and confidence information. Detected humans are analyzed using movement and posture information, while detected fish are passed to a CNN-based classification stage. The classification output is intended to provide fish species and gender information when the available labeled data contain the required classes.

The resulting detections are filtered during post-processing and displayed through a web dashboard. The project design also describes notification through SMS or e-mail services and cloud or spreadsheet logging for maintaining historical detection information.

4 Implementation

The implementation uses Python together with OpenCV and deep-learning components. The project documentation also lists TensorFlow/Keras for CNN development, with optional PyTorch support.

The processing unit may be a laptop or PC with a CUDA-capable GPU, while lightweight deployment is proposed for devices such as NVIDIA Jetson or Raspberry Pi.

A dashboard presents the live stream and detection results. The demonstrated outputs in the supplied report include login, live webcam streaming, single- and multiple-fish detection, and single- and multiple-human detection. Bounding boxes and labels are used to make detections visible to the operator.

5 Result and discussion

The performance of the implemented underwater object detection system was examined using visual outputs generated from the system. The experiments focused on identifying fish and humans in the input images. The detection results are presented in Figures 3 and 4.



Figure 3. Single-fish detection output from the implemented system.

Figure 3 illustrates the system identifying a fish in the input image. The detection output illustrates how the implemented model locates an individual fish under the given testing conditions. This result demonstrates the system's ability to perform fish detection in a suitable underwater scene.

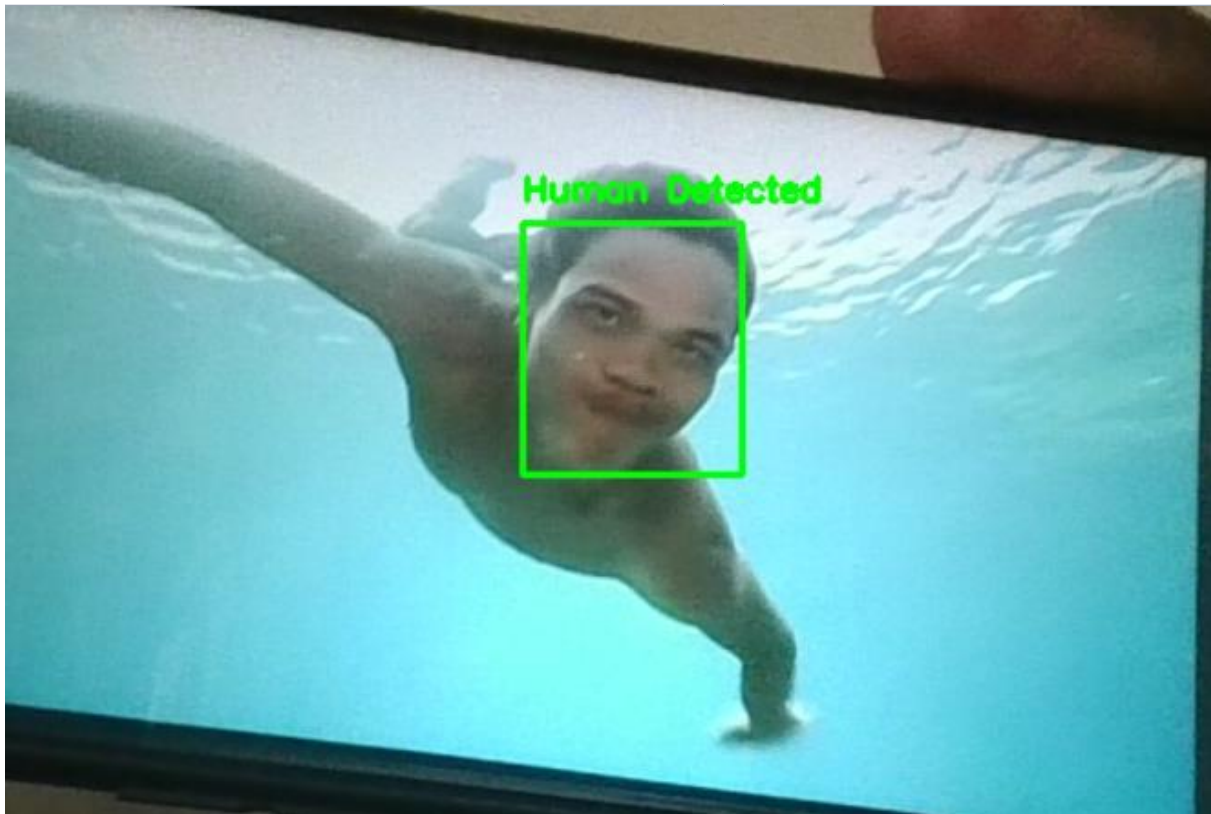


Figure 4. Single-human detection output from the implemented system.

Figure 4 shows the human detection using the implemented system. The output indicates that the system can identify an individual human in the tested image. This capability may be useful in underwater observation and monitoring applications where identifying the presence of people is important.

The detection outputs in Figures 3 and 4 illustrate the system's ability to recognize individual fish and human objects. These outputs provide a visual assessment of the implemented system's basic detection functionality. In practical underwater environments, detection reliability can be affected by several factors, including insufficient lighting, suspended particles, reflections, object movement, and partial obstruction. In particular, turbid water and poor illumination can reduce image quality, making it more difficult for the model to distinguish objects from their surroundings. In The figure 3 and 4 demonstrate selected detections but are not sufficient to establish the overall accuracy or efficiency of the system. A more detailed evaluation would require numerical measurements such as precision, recall, intersection over union (IoU), and frames per second (FPS).

6 Conclusion

This paper presents an end-to-end approach to underwater human and fish recognition using computer vision and deep learning. By integrating a waterproof camera, image preprocessing, YOLO-based object detection, CNN-based fish classification, dashboard visualization, and alert mechanisms, the system provides a practical framework for underwater monitoring. The experimental demonstrations illustrate the system's ability to detect both single and multiple humans and fish. However, underwater visibility and turbidity remain significant challenges that can affect detection accuracy and overall system performance. Future development should focus on expanding the dataset with more diverse underwater images, improving image enhancement techniques, optimizing models for

real-time inference, strengthening the experimental validation of gender classification, and integrating sonar or other complementary sensors to improve detection reliability.

References

- [1] Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. 2020.
- [2] Sadiq M, Sadat A, Mahmud F. *Underwater image enhancement using deep learning: A survey*. Journal of Imaging. 2021.
- [3] Robert LP, Jones GJ, [surname not supplied] MJG. *Fish4Knowledge: Collecting and analyzing massive coral reef fish video data*. Ecological Informatics. 2012.
- [4] Liu C, Xiao Y, Liu X. *Real-time human detection in low-light environments using infrared imaging and deep learning*. IEEE Access. 2020.
- [5] Adeel M, Iqbal A, Khalid R. *A review of convolutional neural networks for image classification*. Journal of Machine Learning Research. 2019.
- [6] Zhao Y, Zhang T, Xu Z. *Object detection in challenging environments: Applications and challenges*. Sensors. 2021.
- [7] Sabatini S, Pagliari P, Marini M. *Underwater acoustic and visual sensor fusion for human and object detection*. IEEE OCEANS. 2019.
- [8] Pathik G. *Underwater-fish-detection* [GitHub repository]. [cited year not supplied].
- [9] Tamim662. *YOLO-Fish* [GitHub repository]. [cited year not supplied].
- [10] Xahidbuffon. *Awesome underwater datasets* [GitHub repository]. [cited year not supplied].
- [11] Saleh A, Sheaves M, Azghadi MR. Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish and Fisheries*. 2022;23(4):977–999. <https://doi.org/10.1111/faf.12666>
- [12] [Authors not supplied]. A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing*. 2023;527:204–232. <https://doi.org/10.1016/j.neucom.2023.01.056>
- [13] [Authors not supplied]. Deep learning-based visual detection of marine organisms: A survey. *Neurocomputing*. 2023. <https://doi.org/10.1016/j.neucom.2023.02.018>
- [14] Garcia-d'Urso N, Galan-Cuenca A, Pérez-Sánchez P, et al. The DeepFish computer vision dataset for fish instance segmentation, classification, and size estimation. *Scientific Data*. 2022;9:287. <https://doi.org/10.1038/s41597-022-01416-0>
- [15] Islam MJ, Ho M, Sattar J. Semantic segmentation of underwater imagery: Dataset and benchmark. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; 2020. p. 1769–1776. <https://doi.org/10.1109/IROS45743.2020.9340821>
- [16] Mohankumar V, Anbalagan S. *A benchmark dataset and ensemble YOLO method for enhanced underwater fish detection*. ETRI Journal. [year not supplied]. <https://doi.org/10.4218/etrij.2024-0383>
- [17] Dulhare UN, Ali MH. Underwater human detection using Faster R-CNN with data augmentation. *Materials Today: Proceedings*. 2023;80:1940–1945. <https://doi.org/10.1016/j.matpr.2021.05.653>
- [18] Liao Z, Hu K, Meng Y, et al. An advanced three stage lightweight model for underwater human detection. *Scientific Reports*. 2025;15:18137. <https://doi.org/10.1038/s41598-025-03677-2>
- [19] Hanumanthappa S, Guruprakash CD. A novel technique for brain tumor detection and classification using T1-weighted MR image. *International Journal of Online and Biomedical Engineering (ijOE)*. 2023;19(17):51–65. <https://doi.org/10.3991/ijoe.v19i17.44309>
- [20] Hanumanthappa S, Guruprakash CD. Fusion of handcrafted and deep-learning features for brain tumor detection and classification using T1-weighted magnetic resonance images. *SN Computer Science*. 2024;5:1107. <https://doi.org/10.1007/s42979-024-03470-4>