Iraqi Journal of Intelligent Computing and Informatics (IJICI)
Vol. 4, 2. December 2025, pp. 114~123
ISSN: 2791-2868, DOI: 10.52940/ijici.v4i2.106

Review of Underwater Object Detection Using YOLO: Advances, Challenges, and Future Directions

Shaymaa K. Hussein[1], Zainab N. Nemer[2]
[1,2]Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Iraq

Article Info

Article history:

Received May 4, 2025
Revised Jun 29, 2025
Accepted July 7, 2025

Keywords:

Underwater Object Detection
YOLO (You Only Lock Once)
Attention Mechanism
Underwater Datasets
Deep learning

ABSTRACT

Underwater object detection is critical for environmental monitoring, maritime security, and rescue operations, yet it faces challenges such as light scattering, color distortion, and low visibility. This paper presents a comprehensive review of YOLO (You Only Look Once) algorithms and their integration with attention mechanisms to address these challenges. We systematically analyze the evolution of YOLO models—from YOLOv1 to YOLOv11—highlighting key architectural advancements, including anchor-free detection, multi-scale feature fusion, and attention modules like CBAM and SimAM. These innovations enhance detection accuracy in underwater environments, where small, occluded objects and dynamic backgrounds degrade performance.

We evaluate YOLO variants on underwater datasets (e.g., URPC, SUIM, RUIE), comparing metrics such as mean Average Precision (mAP), inference speed (FPS), and computational complexity. Attention mechanisms, including spatial, channel, and self-attention, are shown to improve feature discrimination, achieving up to a 25% reduction in false positives. Challenges such as limited annotated data and real-time processing constraints are discussed, along with solutions like semi-supervised learning and synthetic data augmentation.

Based on our findings, YOLOv8 and YOLOv9 models integrating attention mechanisms provide the best trade-offs between accuracy and efficiency for underwater detection. These suggest other directions for future research such as novel lightweight attention designs and multi-sensor fusion to give even more robustness in complex aquatic environments. This study provides a useful reference for researchers and practitioners who contribute to the development of underwater object detection techniques.

Corresponding Author:

Shaymaa K. Hussein
Department of Computer Science, Basra University, Basra, Iraq
Email: shaimaa.khudher@uobasrah.edu.iq
Email: zainab.nemer@uobasrah.edu.iq

1. INTRODUCTION

Underwater object detection is a critical field with growing significance in applications such as environmental monitoring, maritime security, and rescue operations. This technology is essential for monitoring environmental changes, including water pollution and the degradation of marine habitats. It also plays a crucial role in maritime security by detecting risks like shipwrecks and underwater mines[1] The primary obstacles stem from inherent optical properties of water, including light scattering and absorption, which lead to low contrast, color distortion, and reduced visibility. The presence of suspended particles further complicates object recognition [2, 3].

To address these challenges, deep learning techniques have emerged as a state-of-the-art solution. Among these, one-stage detectors like the YOLO framework have proven to be particularly effective, offering a powerful balance of real-time speed and high accuracy suitable for the dynamic nature of underwater scenes [4, 5].

Objective of the Study

The primary objective of this study is to provide a comprehensive and systematic review of the YOLO framework's application to underwater object detection. This research aims to:

1. **Systematically Document the Evolution of YOLO:** To analyze the key architectural and methodological advancements from YOLOv1 to its latest versions, focusing on features relevant to underwater challenges (e.g., multi-scale fusion, anchor-free design).
2. **Evaluate the Role of Attention Mechanisms:** To investigate and synthesize literature on how integrating attention mechanisms (e.g., CBAM, self-attention) with YOLO models enhances feature discrimination for detecting small or occluded underwater targets.
3. **Analyze and Compare Model Performance on Underwater Datasets:** To review and compare the reported performance of various YOLO models on established underwater datasets, using standard metrics such as mAP and FPS.
4. **Identify Key Challenges and Future Research Directions:** To consolidate the primary challenges that remain in this specific field and identify promising pathways for future research, such as lightweight model design and multi-sensor fusion.

This review is structured to guide the reader from foundational concepts to specific analyses and future outlooks. It begins with a comprehensive review of related literature, followed by a detailed examination of the YOLO framework's architectural and methodological progression.

The paper then contextualizes this evolution by analyzing the specific challenges that YOLO faces in underwater environments. Building on this, it evaluates the role of attention mechanisms as a key technological solution. To support the analysis, the review provides an overview of relevant underwater datasets and a comparative performance evaluation of various YOLO models.

Finally, the paper consolidates its findings to identify key challenges and propose promising future research directions, thereby addressing all of the study's primary objectives.

2. Related work

Research in underwater object detection has been approached from two primary angles: the development of novel technical solutions for specific challenges, and comprehensive reviews that survey the field. Before delving into specific contemporary studies, it is useful to contextualize the technological evolution of object detection methodologies. This progression from traditional, feature-based methods to modern deep learning paradigms highlights the critical trade-offs between accuracy, speed, and robustness in the challenging underwater environment.

To appreciate why one-stage deep learning detectors like YOLO have become the dominant paradigm, it is essential to compare their capabilities against other computational approaches[6], as summarized in Table 1.

Table 1: Comparison of Different Object Detection Paradigms for Underwater Environments

Methodology	Key Principle	Strengths	Weaknesses in Underwater Context
Traditional Computer Vision	Rule-based image processing using hand-crafted features (e.g., color thresholds, edge detection).	Fast; requires no training data.	Extremely Brittle: Fails completely with changes in lighting, turbidity, and color distortion. Cannot generalize across different underwater scenes.
Classic Machine Learning (ANN, SVM)	Learns decision boundaries from hand-crafted features (e.g., HOG, SIFT) which are then fed into a classifier like an ANN.	More adaptable than rule-based methods.	Feature-Dependent: Performance is entirely limited by the quality of the hand-crafted features, which are not robust to severe underwater image degradation. ANNs on their own are not well-suited for complex spatial detection tasks from raw pixels.
Deep Learning (Two-Stage Detectors) e.g., R-CNN, Faster R-CNN	Region Proposal + Classification. Proposes potential object regions first, then classifies each one.	Highest Accuracy (mAP): Excellent at precise localization and classification.	Slow and Computationally Heavy: The multi-stage pipeline makes them unsuitable for real-time detection on resource-constrained platforms like AUVs.
Deep Learning (One-Stage	Unified End-to-End Regression. Predicts	Excellent Speed-Accuracy Trade-off: Achieves real-	Historically struggled with very small objects, though recent versions (YOLOv5

Methodology	Key Principle	Strengths	Weaknesses in Underwater Context
Detectors) e.g., YOLO, SSD	bounding boxes and class probabilities in a single pass.	time performance (high FPS) while maintaining very competitive accuracy.	and later) have significantly improved in this area.

As the table illustrates, traditional and classic machine learning methods are ill-suited for the dynamic nature of underwater environments. Deep learning overcomes this by learning robust features directly from data. Within this paradigm, one-stage detectors like YOLO provide the most effective solution by balancing the need for high-speed processing for real-time monitoring with the high accuracy required for reliable detection.

This evolution provides the context for evaluating recent literature. For instance, some technical solutions still employ multi-stage pipelines, as seen in the work of Rajendran et al. [7], who proposed a pipeline involving sequential pre-processing, segmentation, DWT feature extraction, and CNN classification[1]. While reporting high accuracy[2], this approach does not leverage a dedicated end-to-end framework like YOLO, nor does it use standard object detection benchmarks (mAP, FPS), making it difficult to compare against contemporary methods.

On the other hand, several comprehensive reviews have sought to map the research landscape. The works by Er et al[8], Jian et al. [9], and Khan et al.[10] provide valuable overviews of the field. However, due to their broad scope, they tend to treat the YOLO framework as a general sub-category without a deep, version-by-version analysis of its architectural lineage for underwater tasks, and they lack quantitative, controlled performance benchmarks.

Even reviews focused specifically on the YOLO family, such as the comprehensive work by Terven et al. [11], are limited by their focus on general-purpose computer vision. That review does not analyze how YOLO's innovations address unique underwater challenges and contains no performance analysis on underwater datasets.

Beyond broad surveys, specific technical papers demonstrate the practical application of the concepts central to our review. For example, the work by Zhang et al.[12] directly addresses the need for efficient underwater detection by proposing a lightweight model based on YOLOv4. They replaced the standard backbone with MobileNetv2 and integrated a modified attentional feature fusion module (AFFM) to enhance the feature pyramid. This approach achieved a high mAP of 92.65% on the brackish dataset while reducing the model size to just 19.53% of the original YOLOv4, demonstrating a successful trade-off between accuracy and efficiency.The second path is input enhancement through pre-processing. A very recent example by Roy & Talukder [13] demonstrates this by applying a MaxRGB filter to images before feeding them to a standard YOLOv8n model, achieving an impressive 98.6% mAP_{50} on The Brackish Dataset.

Therefore, a comprehensive survey of the literature reveals a clear and specific research gap. While general-purpose reviews map the field broadly, and individual technical papers like those of Zhang et al. [12]and Roy & Talukder [13]provide excellent specific solutions, a systematic review that synthesizes and compares these advanced, specialized models is missing. This paper is motivated by the need to provide a systematic analysis of the YOLO framework's evolution specifically for the underwater domain, to offer a focused synthesis on the integration of attention mechanisms with YOLO for this niche, and, most critically, to present a comparative performance analysis on relevant underwater datasets—a contribution currently absents from the literature.

3. The Yolo principle

The YOLO framework marks a fundamental shift in object detection by reformulating the task as a single, end-to-end regression problem. In contrast to multi-stage methods, YOLO's core principle is to make predictions based on a global view of the input image. As visualized in Figure 1, this is achieved by dividing the image into an S×S grid, where each grid cell is responsible for detecting any object whose center falls within its boundaries[14] .

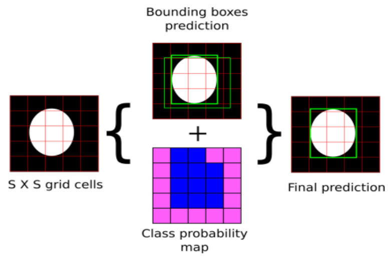

Figure 1: Principle of yolo algorithm[15]

For each grid cell, the model simultaneously predicts two distinct sets of information: a predetermined number of bounding boxes and a vector of conditional class probabilities. Each bounding box is parameterized by five values:

its center coordinates (x, y), its dimensions (width w and height h), and a confidence score. This score represents both the model's determination that a box contains an object and the accuracy of its fitness[16]. This process is formalized into step-by-step algorithm 1.

Algorithm 1: The Foundational YOLO Framework [17]
Input: • I_square: A preprocessed input image of fixed square size S×S. • Ground Truth: True bound boxes and classes (during training). • Thresholds: Confidence and NMS thresholds. Output: • Final Detections: A set of refined bounding boxes with class labels and scores. • Trained Model Weights (during training). Steps: 1. Grid Partitioning: Divide the input image I_square into an S x S grid. 2. Unified Prediction: For each grid cell, pass the image through the network once to predict: o A set of B bounding boxes (BBox_pred). o A vector of class probabilities (P_pred). 3. Post-Processing: a. Generate final confidence scores for each box by combining BBox_pred confidence with P_pred. b. Apply Non-Maximum Suppression (NMS) using the defined thresholds to filter and refine Final Detections. 4. Model Learning (During Training Only): a. Calculate the Total Loss between Predictions and Ground Truth. b. Update model weights via backpropagation based on Total Loss.

This algorithm details the pipeline from initial image processing to the final application, which eliminates redundant detections and utilizes Non-Maximum Suppression (NMS). It is this “single shot” processing that, within the one unified architecture, allows the YOLO framework to have its characteristic speed of real-time detection[18].

3.1 The Evolution of YOLO Models

The evolution of the YOLO framework has significantly transformed real-time object detection, with each iteration introducing key architectural and methodological enhancements to improve the trade-off between speed and accuracy. This progression reflects a continuous effort to address the limitations of previous versions and adapt to new challenges in computer vision.

- YOLOv1 and YOLOv2: Foundational Concepts and Early Refinements The original YOLOv1 pioneered real-time detection by framing the task as a single regression problem. While revolutionary for its speed, it struggled with significant localization errors and had difficulty detecting small or clustered objects due to its coarse grid structure. To address these shortcomings[10], YOLOv2 introduced crucial improvements such as Batch Normalization for more stable training and anchor boxes to achieve more precise localization. Furthermore, its variant YOLO9000 extended detection capabilities to over 9000 categories through hierarchical classification[16].
- YOLOv3 and YOLOv4: Multi-Scale Detection and Optimization A major leap came with YOLOv3, which substantially improved the detection of small objects by incorporating a more powerful residual backbone (Darknet-53) and introducing predictions at three different scales[4]. Subsequently, YOLOv4 focused on optimizing the training process and architecture without a significant increase in computational cost. It integrated a "bag-of-freebies" for training (e.g., mosaic augmentation, CIoU loss) and a "bag-of-specials" for the architecture itself (e.g., SPP module), achieving a state-of-the-art balance between speed and accuracy at the time[19].
- YOLOv5, YOLOv6, and YOLOv7: Usability and Industrial Advancements The development of YOLOv5 marked a significant shift by moving the framework to the more accessible PyTorch library, which greatly improved its usability and deployment flexibility. It also introduced enhancements like automated anchor box optimization and a computationally efficient SPPF layer[20, 21]. Following this, YOLOv6 and YOLOv7 continued the trend of optimization for industrial applications. YOLOv6 introduced an efficient anchor-free design and a reparametrized backbone[22], while YOLOv7 further

pushed efficiency with E-ELAN networks and advanced model scaling techniques, achieving exceptional real-time performance[23].

- YOLOv8 and Beyond: The Modern Anchor-Free Era Finally, YOLOv8, the baseline model for this research, represents a culmination of these advancements. It features a completely anchor-free design and a decoupled head, which simplifies the detection pipeline and resolves the conflict between classification and regression tasks[24]. Furthermore, it incorporates a redesigned C2f module for richer feature extraction and advanced loss functions like Distribution Focal Loss (DFL) to enhance localization accuracy[25]. Subsequent versions like YOLOv9 and YOLOv10 have continued to innovate by focusing on optimal gradient flow and lightweight architecture[26].

4. Challenges in Underwater Object Detection

The underwater environment introduces unique challenges that degrade the performance of conventional computer vision models. The three primary obstacles are: (1) Image quality degradation caused by light absorption, scattering, and color distortion [27]; (2) Detection challenges including small objects, occlusion, and dynamic backgrounds[28]; and (3) A severe lack of annotated data [29].

4.1 Image quality degradation

Underwater images suffer from severe degradation due to three primary physical phenomena: light attenuation, scattering, and color distortion. Light attenuation causes exponential visibility reduction with depth, particularly affecting red wavelengths (absorption coefficient of ~0.3 m^{-1} vs. 0.01 m^{-1} for blue in clear water) [30]. Scattering from suspended particles creates backscattering effects that reduce contrast by up to 80% in turbid waters [31], while wavelength-dependent absorption creates strong blue-green color casts that distort object appearances [32]. To address these challenges, researchers have developed two complementary approaches: physical model-based enhancement and data-driven methods. Physics-based techniques like histogram equalization [33] can recover up to 60% of lost contrast in shallow waters, while deep learning approaches [34]achieve PSNR improvements of 5-8 dB by learning degradation patterns. For training data augmentation, GAN-based synthetic datasets [35]have demonstrated 15-20% mAP improvement on detection tasks by generating physically accurate underwater variations. These solutions collectively address the fundamental optical challenges that degrade underwater imaging systems.

4.2 YOLO-Specific Challenges in the Underwater Environment and Their Technical Solutions

YOLO algorithms face unique challenges when applied to underwater environments, as the harsh visual conditions not only degrade image quality but also directly impact the model's architectural components, thus hindering its effectiveness. The most prominent of these challenges are the loss of small object features and the network's weakened ability to discriminate features due to image degradation.

A primary challenge is the loss of features for small or occluded objects. YOLO's hierarchical architecture, which uses successive down-sampling layers, can cause the spatial information from objects smaller than 50 pixels to be completely lost as features pass through the network's backbone and neck [36].To address this, technical solutions focus on modifying the network's structure. These include enhancing multi-scale architectures like Feature Pyramid Networks (FPN) and Path Aggregation Networks (PANet) to improve the flow of fine-grained details, which has shown to improve mean Average Precision (mAP) by 12-18% in some modified YOLOv5 implementations. Another solution is to add extra detection heads at earlier, higher-resolution stages to specialize in detecting small targets[37].

A second challenge is weakened feature discrimination due to poor image quality. Low contrast, color cast[37], and dynamic backgrounds can make visual patterns ambiguous for YOLO's convolutional filters[38], resulting in weak feature maps and a higher rate of false positives. Researchers address this through two main strategies:

1. **Pre-processing Enhancement:** This approach improves image quality before it enters the YOLO model. For instance, a recent study applied a MaxRGB filter to underwater images, enabling a YOLOv8n model to achieve a high detection accuracy of 98.6% mAP_{50} [13].
2. **Integrating Attention Mechanisms:** This method modifies the YOLO architecture internally. Attention modules adaptively learn to amplify features relevant to the target while suppressing noise and background

elements. The work by Zhang et al. is a key example, where integrating a modified attentional feature fusion module (AFFM) into the YOLOv4 architecture improved performance in underwater environments [12].

While these advancements show significant potential for applications in marine archaeology and ecological monitoring, challenges remain in achieving consistent performance across diverse underwater environments.

4.3 A severe lack of annotated data

The scarcity of annotated underwater datasets, due to high annotation costs and limited public availability[39], hinders robust object detection model development. While datasets like URPC, SUIM, Seaclear, and FishTrack23[40] [41] [42] provide some training data, their scope remains restricted. Semi-supervised learning [40]and transfer learning [39]offer promising solutions by leveraging unlabeled data and pretrained terrestrial models. However, underwater-specific challenges—such as visibility variations and occlusion—still limit performance. Further research is needed to improve dataset availability and algorithmic adaptability.

5. Attention Mechanisms in Underwater Object Detection

Attention mechanisms are computational techniques that enable neural networks to dynamically focus on the most relevant features of input data, thereby improving model performance [43]. In underwater object detection, where challenges such as low visibility, light scattering, and occlusions are prevalent, attention mechanisms enhance feature discrimination by suppressing noise and emphasizing critical regions [44]. By adaptively weighting feature maps, these mechanisms improve detection accuracy in complex underwater environments[45].

5.1 Types of Attention Mechanisms

1. **Self-Attention (Intra-Attention):**
   Self-attention computes relationships between all positions in a feature map, allowing the model to capture long-range dependencies [43]. In underwater object detection, it helps in identifying distant or partially obscured objects by modeling contextual interactions [46].

2. **Channel Attention:**
   Channel attention mechanisms, such as Squeeze-and-Excitation (SE) networks [47], recalibrate channel-wise feature responses by learning adaptive weights. This is particularly useful in underwater scenarios where certain spectral channels may be more informative due to varying light absorption[48].

3. **Spatial Attention:**
   Spatial attention focuses on salient regions within feature maps, enhancing object localization[49]. In underwater imaging, where objects may be partially occluded or blurred, spatial attention helps in precisely detecting object boundaries [50].

These mechanisms are often integrated into convolutional neural networks (CNNs) or transformers to optimize underwater object detection, addressing challenges such as low contrast and background clutter [50].

5.2 Attention Mechanisms Integrated into various YOLO architectures

Attention mechanisms have been effectively integrated into various YOLO architectures to enhance object detection performance across diverse applications. In YOLOv5, the incorporation of Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation (SE), and Channel Attention (CA) improved weapon detection accuracy, achieving a 95.6% mean Average Precision (mAP)—a 3.1% increase over the baseline[51]. Similarly, YOLOv7 combined attention mechanisms with recursive gated convolutions, maintaining high speed while improving detection accuracy for autonomous driving scenarios[52]. YOLOv8 further advanced small object detection in UAV imagery using spatial and channel attention, reaching a 50.2% mAP[53]. For license plate recognition, YOLO-SLD integrated into the SimAM attention module, boosting accuracy from 98.44% to 98.91% without altering the base architecture[54]. Additionally, YOLOv4's cascade attention mechanism, fusing channel and spatial features, increased mAP by 4.77% for small object detection[55]. While these enhancements demonstrate significant accuracy improvements, they also introduce computational trade-offs, requiring careful optimization for real-time applications. Future research may focus on lightweight attention designs to further balance speed and performance in next-generation YOLO models.

6. Datasets for Underwater Object Detection

The datasets utilized for underwater ecological monitoring and robotics present unique applications and challenges. Each dataset serves a specific purpose, from tracking marine life to identifying debris, while also facing environmental difficulties that impact data quality and analysis. Below is a summary of the datasets, their uses, and their associated challenges.

Underwater object detection relies on several specialized datasets shown in Table 2, each designed to address specific challenges in aquatic environments. The Brackish Dataset is primarily used for detecting fish, crabs, and debris in brackish water, supporting ecological monitoring and robotics applications, though it faces challenges such as low contrast, turbidity, and regional bias, particularly in Northern Europe[56]. Similarly, the URPC Dataset focuses on tracking sea cucumbers and scallops in aquaculture settings but struggles with class imbalance and variability in deep-sea environments. The SUIM Dataset, which segments divers and reefs in coral environments, encounters difficulties with dynamic lighting conditions and occlusions that hinder visibility[57]. For salvage operations, the DUO Dataset identifies underwater mines and debris, though sparse annotations in murky water complicate detection efforts. The RUIE Dataset enhances low-visibility images before detection tasks; however, it is affected by color distortion and light scattering, which degrade image quality. Lastly, the HabCam Dataset maps marine habitats and species distributions, though challenges such as environmental variability and data integration issues are often encountered[56]. While these datasets are indispensable for advancing underwater monitoring technologies, they underscore the need for improved methodologies to overcome the inherent difficulties posed by underwater environments, including poor visibility, limited annotated data, and real-time processing constraints. Addressing these challenges will require innovations in synthetic data generation, multi-sensor fusion, and lightweight model development to enhance detection accuracy and operational efficiency.

Table 2: Underwater Object Detection Datasets

Dataset	Environment	Key Objects	Primary Use Cases	Main Challenges	Reference
Brackish	Brackish water	Fish, crabs, debris	Ecological monitoring, underwater robotics	Low contrast, turbidity, regional bias	[56]
URPC	Deep sea	Sea cucumbers, scallops	Aquaculture, underwater robotics	Class imbalance, deep-sea variability	[58]
SUIM	Coral reefs	Divers, wrecks, reefs	Marine research, diver safety	Dynamic lighting, occlusions	[57]
DUO	Murky water	Mines, debris	Military, salvage operations	Sparse annotations, low visibility	[59]
RUIE	Low-visibility water	Enhanced images (pre-processing)	Image enhancement for detection tasks	Color distortion, light scattering	[32]

7. Comparison of YOLO Versions on Underwater Datasets

The comparison of YOLO versions in underwater object detection reveals significant advancements in performance across different iterations. Recent studies indicate that while YOLOv5 generally excels in mean Average Precision (mAP), newer versions like YOLOv8 and YOLOv7 also demonstrate notable improvements in accuracy and speed, making them suitable for challenging underwater environments, as shown in Table 3.

Performance Metrics

- **YOLOv5**: Achieved the highest mAP score in underwater conditions, showcasing superior precision and recall compared to its predecessors[60].
- **YOLOv7**: While slightly slower, it outperformed YOLOv5 and YOLOv3 in detection accuracy, with an mAP of 0.82[61].
- **YOLOv8**: Demonstrated high accuracy (mAP of 0.84) and rapid detection capabilities, making it effective for real-time applications[62].

Table 3. Comparison of YOLO versions in underwater object detection

Version	mAP	Speed (FPS)	FLOPs (G)	Key Strengths	Limitations & Challenges
YOLOv5 [60]	0.82	60-85	16.5	Best balance of accuracy/speed; superior precision/recall in turbid water	Struggles with extreme occlusion
YOLOv7 [61]	0.82	50-70	104.7	Higher accuracy than v5/v3; robust to shape distortions	Computationa lly heavy (high FLOPs)
YOLOv8 [62]	0.84	90-140	28.6	Anchor-free design; fastest real-time performance; handles light variations	Smaller object detection needs refinement

8. Future Work

To advance underwater object detection, future research should prioritize:

1. Enhancing Data Quality – Generative Adversarial Networks (GANs) can synthesize realistic underwater imagery to address dataset limitations, improving model generalization.
2. Optimizing Lightweight Models – Developing efficient YOLO variants for deployment on low-power edge devices (e.g., AUVs, ROVs) without sacrificing accuracy.
3. Boosting Small/Object Detection – Refining multi-scale attention mechanisms and high-resolution feature fusion to better identify occluded or minute marine objects.

Conclusion

This review highlights YOLO's transformative role in underwater object detection, driven by architectural innovations like attention mechanisms and anchor-free designs. Despite progress, challenges persist—particularly data scarcity and small-object detection in turbid environments. Future efforts must focus on data augmentation, computational efficiency, and adaptive detection frameworks to unlock robust, real-time underwater vision systems. Collaborative advancements in AI and marine technology will be pivotal in overcoming these barriers.

References

1. Hoehner, F., et al., Object Detection and Tracking in Maritime Environments in Case of Person-Overboard Scenarios: An Overview. Journal of Marine Science and Engineering, 2024. 12(11): p. 2038.
2. Hu, Z., et al., Underwater Target Detection with High Accuracy and Speed Based on YOLOv10. Journal of Marine Science and Engineering, 2025. 13(1): p. 135.
3. Zhai, L., Underwater Object Detection Using Deep Learning. 2017, Prairie View A & M University.

4. Redmon, J. and A. Farhadi, Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
5. Liu, M., et al., LFN-YOLO: precision underwater small object detection via a lightweight reparameterized approach. Frontiers in Marine Science, 2025.
6. Xu, S., et al., A systematic review and analysis of deep learning-based underwater object detection. Neurocomputing, 2023. 527: p. 204-232.
7. Sasilatha, T., et al., Deep learning-based underwater metal object detection using input image data and corrosion protection of mild steel used in underwater study: A case study: Part A: Deep learning-based underwater metal object detection using input image data. Zastita Materijala, 2022. 63(1): p. 5-14.
8. Er, M.J., et al., Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. Sensors, 2023. 23(4): p. 1990.
9. Jian, M., et al., Underwater object detection and datasets: a survey. Intelligent Marine Technology and Systems, 2024. 2(1): p. 9.
10. Khan, A., et al., Underwater target detection using deep learning: methodologies, challenges, applications, and future evolution. IEEE Access, 2024. 12: p. 12618-12635.
11. Terven, J., D.-M. Córdova-Esparza, and J.-A. Romero-González, A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. Machine learning and knowledge extraction, 2023. 5(4): p. 1680-1716.
12. Zhang, M., et al., Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion. Remote Sensing, 2021. 13(22): p. 4706.
13. Roy, J. and K.H. Talukder. Under water objects detection and classification using deep learning technique. in 2024 international conference on advances in computing, communication, electrical, and smart systems (iCACCESS). 2024. IEEE.
14. Varghese, R. and M. Sambath. Yolov8: A novel object detection algorithm with enhanced performance and robustness. in 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). 2024. IEEE.
15. Durve, M., et al., Tracking droplets in soft granular flows with deep learning techniques. The European Physical Journal Plus, 2021. 136(8): p. 864.
16. Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
17. Sunitha, G., et al. Road Object Detection Using Yolov8. in 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC). 2024. IEEE.
18. Cantelmo, C., A novel approach for detecting signatures: YOLO neural network. 2022-2023: Luiss Guido Carli
19. Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
20. Maheswaran, S., et al. Yolov5 based a real time automatic number plate and helmet recognition system. in 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2022. IEEE.
21. Hussain, M., Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo. IEEE access, 2024. 12: p. 42816-42833.
22. Li, C., et al., YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976, 2022.
23. Wang, C.-Y., A. Bochkovskiy, and H.-Y.M. Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
24. Hussain, M., Yolov5, yolov8 and yolov10: The go-to detectors for real-time vision. arXiv preprint arXiv:2407.02988, 2024.
25. Sohan, M., T. Sai Ram, and C.V. Rami Reddy. A review on yolov8 and its advancements. in International Conference on Data Intelligence and Cognitive Informatics. 2024. Springer.
26. Wang, C.-Y., I.-H. Yeh, and H.-Y. Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. in European conference on computer vision. 2024. Springer.
27. Akkaynak, D. and T. Treibitz. A revised underwater image formation model. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
28. Jaffe, J.S., Underwater optical imaging: the past, the present, and the prospects. IEEE Journal of Oceanic Engineering, 2014. 40(3): p. 683-700.
29. Elmezain, M., et al., Advancing Underwater Vision: A Survey of Deep Learning Models for Underwater Object Recognition and Tracking. IEEE Access, 2025.
30. Schettini, R. and S. Corchs, Underwater image processing: state of the art of restoration and image enhancement methods. EURASIP journal on advances in signal processing, 2010. 2010: p. 1-14.
31. Drews, P.L., et al., Underwater depth estimation and image restoration based on single images. IEEE computer graphics and applications, 2016. 36(2): p. 24-35.
32. Li, C., et al., An underwater image enhancement benchmark dataset and beyond. IEEE transactions on image processing, 2019. 29: p. 4376-4389.
33. Akkaynak, D. and T. Treibitz. Sea-thru: A method for removing water from underwater images. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
34. Fabbri, C., M.J. Islam, and J. Sattar. Enhancing underwater imagery using generative adversarial networks. in 2018 IEEE international conference on robotics and automation (ICRA). 2018. IEEE.
35. Islam, M.J., Y. Xia, and J. Sattar, Fast underwater image enhancement for improved visual perception. IEEE Robotics and Automation Letters, 2020. 5(2): p. 3227-3234.
36. Chi, J., L. Zheng, and J. Miao. Underwater Object Detection Algorithm Based on Improved YOLOv5. in International Conference on Autonomous Unmanned Systems. 2023. Springer.
37. Guan, X., X. Luo, and D. Wang. YOLOv5-UOD: Underwater Object Detection Method Based on Improved YOLOv5. in 2023 13th International Conference on Information Science and Technology (ICIST). 2023. IEEE.
38. Wang, N., et al. Marine object detection based on improved YOLOv5. in 2022 5th International Conference on Intelligent Autonomous Systems (ICoIAS). 2022. IEEE.
39. Levy, D., et al. Automated analysis of marine video with limited data. in Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018.
40. Chen, G., et al., A cooperative training framework for underwater object detection on a clearer view. IEEE Transactions on Geoscience and Remote Sensing, 2024.
41. Dawkins, M., et al. FishTrack23: An Ensemble Underwater Dataset for Multi-Object Tracking. in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024.
42. Đuraš, A., et al., A Dataset for Detection and Segmentation of Underwater Marine Debris in Shallow Waters. Scientific data, 2024. 11(1): p. 921.

43. Vaswani, A., et al., Attention is all you need. Advances in neural information processing systems, 2017. 30.
44. Zhang, R., et al., Survey on Deep Learning-Based Marine Object Detection. Journal of Advanced Transportation, 2021. 2021(1): p. 5808206.
45. Zhang, Z. and D. Zhang. Research on underwater object detection based on frequency domain attention mechanism. in Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024). 2025. SPIE.
46. Gao, Z., Y. Shi, and S. Li, Self-attention and long-range relationship capture network for underwater object detection. Journal of King Saud University-Computer and Information Sciences, 2024. 36(2): p. 101971.
47. Hu, J., L. Shen, and G. Sun. Squeeze-and-excitation networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
48. Zhong, Y., J. Wang, and Q. Lu, SCAUIE-Net: Underwater image enhancement method based on spatial and channel attention. IEEE Access, 2023. 11: p. 72172-72185.
49. Woo, S., et al. Cbam: Convolutional block attention module. in Proceedings of the European conference on computer vision (ECCV). 2018.
50. Wang, X., et al., Underwater object detection algorithm based on adding channel and spatial fusion attention mechanism. Journal of Marine Science and Engineering, 2023. 11(6): p. 1116.
51. Khalfaoui, A., A. Badri, and I. El Mourabit. A Deep Learning Framwork with Attention-Guided Feature Learning for Robust Object Detection in Surveillance Systems. in 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM). 2024. IEEE.
52. Shi, Y. and A. Hidaka. Improvement in On-road Object Detection Using YOLOv7 with High-order Spatial Interaction and Attention Mechanism. in 2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). 2023. IEEE.
53. Rupak, N., et al. An Investigation into YOLO-v8 Model Optimization for Small Object Detection in UAVs using Attention Mechanism. in 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2024. IEEE.
54. Chung, M.-A., Y.-J. Lin, and C.-W. Lin, Yolo-sld: An attention mechanism-improved yolo for license plate detection. IEEE Access, 2024.
55. Mao, G., G. Liao, and H. Zhu. Object detection algorithm based on channel-spatial fusion cascade attention. in 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS). 2022. IEEE.
56. Pedersen, M., et al. Brackishmot: The brackish multi-object tracking dataset. in Scandinavian Conference on Image Analysis. 2023. Springer.
57. Islam, M.J., et al. Semantic segmentation of underwater imagery: Dataset and benchmark. in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2020. IEEE.
58. Liu, C., et al. A dataset and benchmark of underwater object detection for robot picking. in 2021 IEEE international conference on multimedia & expo workshops (ICMEW). 2021. IEEE.
59. Wei, Y., et al., Underwater Detection: A Brief Survey and A New Multitask Dataset. International Journal of Network Dynamics and Intelligence, 2024: p. 100025-100025.
60. Gašparović, B., et al. Evaluating Yolov5, Yolov6, Yolov7, and Yolov8 in underwater environment: Is there real improvement? in 2023 8th International Conference on Smart and Sustainable Technologies (SpliTech). 2023. IEEE.
61. Shankar, R. and M. Muthulakshmi. Comparing YOLOV3, YOLOV5 & YOLOV7 architectures for underwater marine creatures detection. in 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). 2023. IEEE.
62. Pavani, D., A.N.N. Reddy, and N. Saw. YOLO object detection method for underwater trash collecting robot: Octacleaner. in 2023 IEEE Technology & Engineering Management Conference-Asia Pacific (TEMSCON-ASPAC). 2023. IEEE.