IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 1, February, 2026, pp. 335-350
Journal homepage: http://iieta.org/journals/ts

Underwater Image-Based Crab Detection Using Hybrid RetinaNet and YOLOv8

Suguna Devi R.[1*], Suchitra G.[2], Meena Kowshalya A.[3]

[1] Department of Electronics and Communication Engineering, Saveetha Engineering College, Chennai 602 105, India
[2] Department of Electronics and Communication Engineering, Government College of Technology, Coimbatore 641 043, India
[3] Department of Computer Science Engineering, Government College of Technology, Coimbatore 641 043, India

Corresponding Author Email: sugunadevir@saveetha.ac.in

https://doi.org/10.18280/ts.430123

Received: 9 February 2025
Revised: 16 October 2025
Accepted: 2 December 2025
Available online: 28 February 2026

Keywords:
crab detection, RetinaNet, YOLOv8, deep learning, hybrid model, underwater image analysis, real time object detection

ABSTRACT

Detecting crabs in underwater environments presents significant challenges, including low visibility, occlusions, and fluctuating lighting conditions. This study introduces a hybrid RetinaNet-YOLOv8 model aimed at improving the accuracy and robustness of crab species identification by combining the strengths of both object detection frameworks. YOLOv8 provides real-time detection features with impressive recall, whereas RetinaNet is particularly effective in addressing class imbalance through its focal loss function. The model under consideration is developed using the Crab Species Dataset, which includes a variety of crab images taken in their natural environments. A comprehensive series of experiments was carried out over 50 epochs, with performance assessed through critical metrics: Accuracy (0.94), Precision (0.96), Recall (0.92), F1-score (0.94), and Intersection over Union (IoU). The hybrid RetinaNet-YOLOv8 approach shows enhanced performance over standalone models, surpassing both RetinaNet and YOLOv8 in terms of detection accuracy and robustness under challenging conditions. The findings indicate that the combination of RetinaNet and YOLOv8 improves the detection of crab species, positioning it as an effective tool for ecological monitoring and studies on marine biodiversity.

1. INTRODUCTION

It is crucial to identify and study marine species, particularly crabs, to monitor the environment, manage fisheries, and conserve biodiversity. Underwater photos and videos are challenging to distinguish automatically due to limited visibility, varying lighting conditions, motion blur, and obstructions. Traditional edge detection techniques, such as those shown in (Sobel, Prewitt, Roberts, Canny, Otsu, and Laplacian of Gaussian), have been used for crab image detection. It has been demonstrated [1] that Sobel, Prewitt, and Roberts can outperform by enhancing and applying morphological operations. However, in complicated underwater conditions, these methods are usually not very effective. Recent discoveries in deep learning, especially object detection models, have significantly contributed to marine species identification. This function uses YOLOv8 and RetinaNet (Residual Networks and Feature Pyramid Networks) to locate crabs to improve the accuracy and reliability of detection in difficult underwater environments. YOLOv8 is a modern object detection model with real-time performance and high accuracy, which is suitable for aiming for essential, quick, and precise detection [2]. RetinaNet, using the focal Loss function, has been successful in solving the class imbalance problem, which is frequently encountered in maritime datasets. This ensures that, despite this essential feature, crab is detected more precisely [3].

This analysis uses a Brackish Dataset containing annotated submerged photographs alongside various visual environments to verify the correctness of the method proposed. The experiment compared the performance of YOLOv8 with RetinaNet in terms of Average Correctness (mAP), Intersection over Union (IoU), and detection speed. In addition, adversarial acquisition of knowledge methods is used to assess a model's ability to resist noise and disturbance, which guarantees that it can be trusted under real-world conditions [4]. For marine biology, environmental monitoring and preservation of submerged communities' underwater habitats is necessary, as has been the case in several studies. The development of optical techniques and information collection methods for capturing the architectural richness of the naval animal habitat, including the marine forest, is a key condition for this area. Burns et al. [5] underlined the importance of non-intrusive devices such as stereophotogrammetry combined with computer vision methods, which produce a highly accurate 3D model important for effective habitat management. As these methods offer a higher degree of accuracy, their scalability is hindered by the need for skilled labor and significant post-processing capital. Aquatic animal analysis (MAS) is an important area of research within the scope of the discipline, together with advances in deep learning approaches to tackle problems such as minimal visibility, concealment, and intricate situations existing in the underwater background. Li et al. [6] presented the improved Cascade Decoder System (ECD-Net), which showcases the capabilities of segmented marine animals by

integrating multi-scale aspect extraction and sophisticated decoding techniques. However, the problem of computational complexity, particularly in the case of real-time applications including thorough underwater surveys, remains. The creation of comprehensive datasets such as MAS3K is a significant step forward in marine life interpretation research. Islam et al. [7] showed that MAS3K, a secret data set created for the segmentation of marine animals, provides a crucial resource for the cleavage model. This dataset substantially contributes to the development of marine animal detection by donating an unchanging system for measuring disparate models in a demand-submerged environment. However, the potential biases in the dataset should be noted in the context of ecological influences to present drawbacks to the use of the model in the hands-on scenario. The recent investigation aimed at strengthening the robustness of above model gathered both real and fake underwater images. Shorten and Khoshgoftaar [8] investigated the impact of fine-tuning deep-acquiring knowledge architecture using a combination of real and imitated data on recognition accuracy. To better reflect the complexities of real-world submerged conditions, their employment highlights the importance of combining both types of statistics. Furthermore, the development of a real-time submerged detection model is facilitated by the development of large datasets identical to SUIM. Li et al. [9] have made an essential contribution to the current plot alongside their PSS-net, a parallel semantic Infrastructure personalized to deal with the singular roadblock encountered by a low-light, submerged environment. The implementation of the attention mechanism and the integration of different characteristic designs have led to impressive performance benchmarks, particularly in overcoming the obstacles of concealment and visibility which are common in underwater environments. Finally, Qin et al. [10] showed that powerful architecture admires MASNet through the use of Thai Infrastructure Design and Information Accumulation Methods. The improvements significantly improved the ability to classify marine animals in complex underwater environments. However, like many deep learning models, the computational requirements and real-time processing impediments remain to prevent thorough realization in an extensive, energetic maritime setting. The need for precise, scalable, and productive methods directed close to the grok and the continuation of aquatic communities drives the progress of submerged cleavage. However, more investigation is needed to overcome the computational and environmental obstacles that limit the full power of such systems.

Detection of pediculosis pubis in submerged environments is a multilateral and technically complex roadblock influenced by several difficult components characteristic of aquatic biomes. The underwater imagination frequently encounters obstacles planned to result in reduced visibility due to light attenuation and scattering. The current incident reduces image clarity and differences, making it more difficult for the object detection model to accurately identify and locate the crab type. In addition, occlusions are common, as pediculosis pubis is often partially concealed by marine vegetation, rock, or another organism, which contributes to the obstacle to detection. The water depth, turbidity, and inherent light penetration allow for a fluctuating lighting state, which leads to a lack of image quality and is a current significant impediment to model generalization. Class imbalance, where the exact crab species may remain underrepresented in the dataset, may lead to a distortion of the performance of the model and a reduction in the accuracy of the detection of rare genera. Moreover, a prerequisite for immediate detection in environmentally friendly monitoring intention names for models capable of rapidly handling photographs while maintaining accuracy is a challenge that conventional object detection paradigms face to overcome. RetinaNet and YOLOv8 show definite limitations. RetinaNet and YOLOv8 show effectiveness in addressing class imbalance together with their focal loss function, although they probably are not used in academic writing to provide the crucial speed for real-time objectives. YOLOv8 has outstanding recall and real-time performance, but it may be troublesome in managing unbalanced datasets and adverse green environments. The need for a more resilient and flexible solution capable of meeting the precise requirements of submerged crab detection is highlighted by the technical and sustainable challenges.

While there has been notable advancement in the detection of marine species using computer vision techniques, existing methods for underwater object detection continue to encounter substantial obstacles, including issues related to low visibility, turbidity, occlusion, and an imbalance in the classification of crab species. Current studies utilising YOLO-based models demonstrate real-time capabilities; however, they frequently encounter challenges with localisation accuracy in intricate underwater environments. Conversely, RetinaNet addresses class imbalance with focal loss; however, it falls short in providing the detection speed necessary for dynamic underwater monitoring. To our knowledge, there has yet to be a study that combines RetinaNet and YOLOv8 into a cohesive hybrid framework, effectively utilising both focal loss for class imbalance management and the ability for real-time inference. This study introduces a Hybrid RetinaNet–YOLOv8 model tailored for effective crab detection in underwater settings. The main objectives of this study are: (i) to improve detection accuracy in low-contrast and noisy underwater environments, (ii) to address class imbalance among crab categories through the application of focal loss, and (iii) to maintain real-time detection speed by leveraging the lightweight architecture of YOLOv8. This work presents a novel approach by integrating two complementary detection paradigms: anchor-based (RetinaNet) and anchor-free (YOLOv8). This combination aims to enhance both precision and recall in marine applications, an area that has not been previously investigated in the context of underwater species detection.

The present study presents a unique technique involving the generation of a hybrid RetinaNet-YOLOv8 model, which has been particularly developed to detect submerged pediculosis pubis. The combination of RetinaNet and YOLOv8, which are familiar object detection frameworks, has not been thoroughly studied, in particular in the context of submerged environments. The combination of RetinaNet's focal loss function, which is useful for dealing with class imbalance, with YOLOv8's real-time detection capabilities and high-recall results in a synergistic model that successfully solves the distinct challenges of submerged image investigation. For environmental monitoring, where precise and reliable breed identification is important, the current hybrid technique is highly innovative.

This paper makes the following contributions:

1. The use of YOLOv8 and RetinaNet for detecting crabs in underwater photos.

2. Enhancing detection power through adversarial data collection and data augmentation processes.

3. Measure performance steps to find the best model for

discerning oceanic variety in the present moment.

4. Knowledge of how deep learning-based item detection might be used in maritime studies and conservation efforts [11].

The current exploration seeks to add to the automation of maritime family monitoring by integrating a top-performing deep-learning knowledge model. The current would reduce the total human toil mandatory for researchers while increasing the accuracy of detection in difficult underwater conditions.

2. LITERATURE REVIEW

2.1 YOLOv8 for general object detection and tracking

Multi-object tracking (MOT) merges object detection and re-identification to track an object across a video image. YOLOv8, a single-shot detector, enables real-time MOT by detecting an object in a single frame, while a re-identification model ensures continuity. This approach improves tracking accuracy and robustness in complex dynamic environments, making it suitable for a wide range of pattern recognition applications [12]. This study advances YOLO based object detection in challenging environments by integrating high achieving anchors (ResNet50, DenseNet169, ConvNeXt, and EfficientNetv2) into four novel models, which improve YOLOv8's acknowledgment accuracy in the involved scenario by 0,87 % and 1.6 %, respectively, according to experiments on the leaf mustard dataset. YOLOv8-EfficientNetv2 with 95.2 % mAP50 and 85.3 % mAP5095, respectively [13]. The current scrutiny introduces a real time citizen count system that accurately locates and counts persons using YOLOv8. The architecture integrates regional studies, productive tracking, and sophisticated object detection, which exhibits high accuracy and adaptability in a variety of scenarios. The results show how effective it is in monitoring, occupancy studies, and herd control objectives [14]. This study investigates the research YOLOv8 for detecting defects in the mang' Carabao', using a dataset of 2160 annotated photographs. The model achieved 100% accuracy for black musca volitans, 80 percent for brown musca volitans, and 83.33 percent for mango scab, together with an average precision of 95 percent over the IoU threshold of 0.50 to 0.95 and a 93% recall estimate. The model has been integrated into a Tkinter based real-time defect classification procedure. The results indicate the success of YOLOv8 in the detection of mango defects, especially in the classification based on appearance [15]. The project aims to use YOLOv8 to increase the detection of military objects in satellite images. The exploration improves YOLOv8's real time object identification abilities by improving its architecture. This makes it more efficient in determining the small components in complex environments. This demonstrates the effectiveness of YOLOv8 for large-scale image analysis and pattern recognition applications [16]. This study explores the enhancement of YOLOv8 for object detection in quadcopter images by hyperparameter optimization and the establishment of dedicated partnerships for object tracking and distance detection. The new technique, which uses a couple of networks to position and distance objects, replaces the obligatory two camera approach and adds speed and accuracy. Compared with the original YOLOv8, the test on the Vis Drone2019 dataset demonstrated a superior assessment of the distance between the samples of mAP50 and mAP50-95, with a corresponding increase of 3.0 percent and 1%. These developments tackle issues with small object detection and dynamic situations [17]. Figure 1 shows mAP improvement across different use cases.

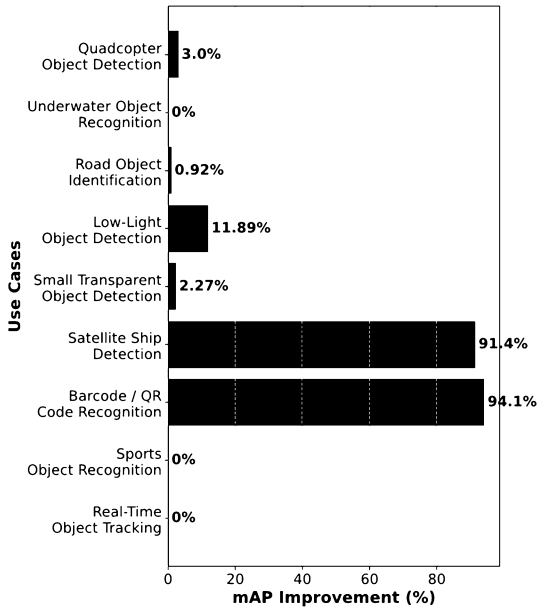


Figure 1. mAP improvement across different use case

To solve problems such as distortion, this study suggests an enhanced YOLOv8 network for underwater object recognition in sonar images. Through the use of transfer learning and deformable convolution, the technique improves accuracy and speeds up network convergence. Experiments on the URPC2022 dataset show that the enhanced YOLOv8 achieves the best mAP@0.5:0.95 accuracy, outperforming other popular detection networks [18]. Using the YOLOv8 architecture, this study investigates road object identification, highlighting how quickly and accurately it can identify both moving and stationary items for uses like as traffic control and self-driving cars. With 0.96 accuracy, 0.96 precision, 0.97 recall, and 0.92 mAP, YOLOv8 performs better than earlier YOLO models and versions, such as SSD and RCNN. Strong bounding box predictions are indicated by the low GIoU loss of 0.44. In addition to suggesting future advancements, including model optimization, dataset expansion, and real world testing, the study emphasizes the significance of deep learning in small object recognition [19]. This study uses a dataset of 4,132 photos from the Roboflow dataset to assess how well YOLOv5 and YOLOv8 locate things in low light. Precision, recall, F1-score, and mAP are among the parameters where YOLOv8 performs better than YOLOv5, with improvements of 6.14%, 10.61%, 9%, and 11.89%, respectively. The outcomes show how well YOLOv8 performs in difficult low-light conditions. An enhanced YOLOv8 algorithm for identifying translucent creatures underwater is proposed in this paper. Among the improvements are the addition of a small object detection head to increase accuracy for small transparent objects, the use of the Focal loss function to handle class imbalance, and the addition of Convolutional Block Attention Mechanism (CBAM) to the backbone network for improved feature extraction. Experimental results indicate that the improved YOLOv8 outperformed the default

version, boosting the mAP by 2.27%, and demonstrating its potential in challenging complex environments [20].

In this study, the architecture of YOLOv8, known for its high accuracy and real-time detection capabilities, is utilized for object detection in high-resolution imagery. A mAP@50 of 91.4% is achieved, demonstrating the model's effectiveness across varying object sizes, orientations, and backgrounds. Transfer learning is employed to enhance detection performance, followed by fine-tuning on a benchmark dataset [21]. To evaluate architectural improvements, the performance of YOLOv8 is compared with YOLOv5 for real-time object detection tasks. YOLOv8 achieves a mAP@0.5 of 0.626 and mAP@0.5:0.95 of 0.494, while YOLOv5 attains 0.593 and 0.454, respectively. These results highlight the improved performance of YOLOv8 and its potential for advanced computer vision applications.

2.2 YOLOv8 for small object detection and real-time tracking

This exercise studied the real time detection capability of the YOLOv8 algorithm for assessing its barcode and QR code detection performance. The study trains and fine-tunes YOLOv8 across Nano, Small, and Medium versions, using Kaggle datasets. The results show significant improvements in detection accuracy, getting 88.95%, 97.10%, and 94.10% for the Nano, Small, and Medium models, respectively. The results showcase the advancement that YOLOv8 made in object detection and its ability to provide computer vision, especially in the areas of barcode and QR-code recognition [22]. This study aims to improve small object detection for small moving balls in sports fields using the YOLOv8 algorithm. The model achieves greater accuracy in detecting small objects, such as balls, in complex backgrounds through algorithm fusion, minor parameter changes, a small object detection head, and an attention mechanism. These advancements ease real-time applications, such as directing camera angles during broadcasts, and improve accuracy in object identification. The YOLOv8-based method shows enough promise for small object recognition at high speed and in real time complex applications.

The present study investigates the application of sophisticated deep multi-object trackers SMILETrack, ByteTrack, and BoTSort-in real time object tracking combined with YOLOv8. The trackers capitalize on the performance capabilities of YOLOv8's object identification to improve the tracking speed and accuracy. The intention is to create an online model with the ability to track objects and run at 15 frames per second. In the study, the fitted model has been shown to monitor ten distinct classes of small objects in aerial films with great accuracy and no interruptions [23]. In addition to the basic Whose Loss function, this study incorporates multiple detection heads into the YOLOv8 technique. The additional detection heads allow the model to identify objects of various sizes, while the Wise-IoU loss function raises the bounding box accuracy by reducing the adverse impact of poor data. Such improvements have made YOLOv8 more effective and accurate for various animals in underwater scenario [24]. This paper presents a distance estimation system based on the YOLOv8 algorithm for real-time object recognition. The system performs distance estimation based on the distance between the camera and the detected object with the lens principle of the camera. According to this study, YOLOv8 offers the optimum accuracy for close range object recognition, with detection accuracy improving as distance decreases [25].

In this study, a novel method for tracking small animals integrates the Deep SORT tracking algorithm with the YOLOv8 object detection model. The system achieves a high recall of 0.9 and accurately tracks animals across diverse environments. It outperforms conventional hand-crafted feature-based methods and is well-suited for ecological research and conservation applications, enabling efficient monitoring of species in complex scenes [26]. Another important challenge of AUVs for monitoring and underwater exploration is object detection in sonar images. This study uses YOLOv8 and describes the modification to be suitable for sonar data by using a specific dataset and data preprocessing. The study has run extensive tests to compare the accuracy, speed, and resilience of YOLOv8 in sonar image detection and provide important performance metrics like precision, recall, F1-score, and mAP. These results will lead to further improvement of autonomous underwater systems and provide insight into improved object recognition in AUVs [27].

This study focuses on developments in video surveillance systems, seeing them as a way to counter some malign behavior characterized by violent crimes, incursions, fires, and tampering with video. Normal systems are under threat from the bad actors as they vary their tactics to try and escape detection. The study proposes boosting real-time object recognition capabilities in surveillance footage using the YOLOv8 algorithm. For this reason, there is an emphasis on methods for improving the known detection of adapting to the normal but no less deadly threats and highlights the need for developing flexible and robust systems capable of managing shifted, corrupted, or manipulated video data [28].

2.3 RetinaNet for object detection in complex and small-scale scenarios

RetinaNet, integrated with a deep residual backbone (ResNet-152), is employed for object detection tasks, demonstrating strong performance in complex visual recognition problems. The model is trained and validated using a representative dataset and achieves a mean Average Precision (mAP) of 78% with the optimized Focal Loss function. This approach effectively addresses class imbalance and enhances detection accuracy, making it suitable for a wide range of pattern recognition applications.

In addressing the complications surrounding target recognition in complicated air-to-ground scenarios using deep-learning-based methods, it is found that the detection of UAV targets is mostly hindered by their small span sizes and fewer features. To deal with those discussed issues, the paper puts an effort toward enhancing the architecture of the feature extraction layer network, optimization of anchor scale and quantity using a system clustering method, and improving the Focal Loss calculation based on RetinaNet. Further validation on the UAV platform proves that the upgraded RetinaNet boosts detection performance while still considering the accuracy of the problem of air-to-ground detection. The most recent simulation tests show great detection accuracy [29].

This paper primarily focuses on object detection in complex imaging data, which presents significant challenges due to variations in object size, position, and irregular shapes, especially when the objects are small and partially occluded. This study employs a benchmark dataset to identify target objects using the one-stage detector RetinaNet. The proposed

method utilizes the focal loss mechanism to address class imbalance and improve detection accuracy. The observed performance of mAP 96% indicates that RetinaNet effectively detects objects in complex scenarios [30]. To study the detection of one or several objects in a target dataset with a few annotations at the instance level, this work proposes a new Mixed-Supervised Object Detection (MSOD) approach based on RetinaNet. The proposed methodology implements a three-stage training pipeline: first, training on the source dataset with full annotations, second, training on the target dataset with image-level annotations and only a few annotated instances, and lastly, re-training of a small number of classes using only a few objective observations from the target dataset. The presented technique is not reliant on category links between the source and target datasets, as is the case with contemporary MSOD systems. The experimental results on the PASCAL VOC 2007 test set indicate better performance than alternative few-shot object detection algorithms. The use of RetinaNet allows the suggested method to bear a promise for real-time applications [31]. Consistent performance from RetinaNet across a variety of applications is represented in Figure 2.

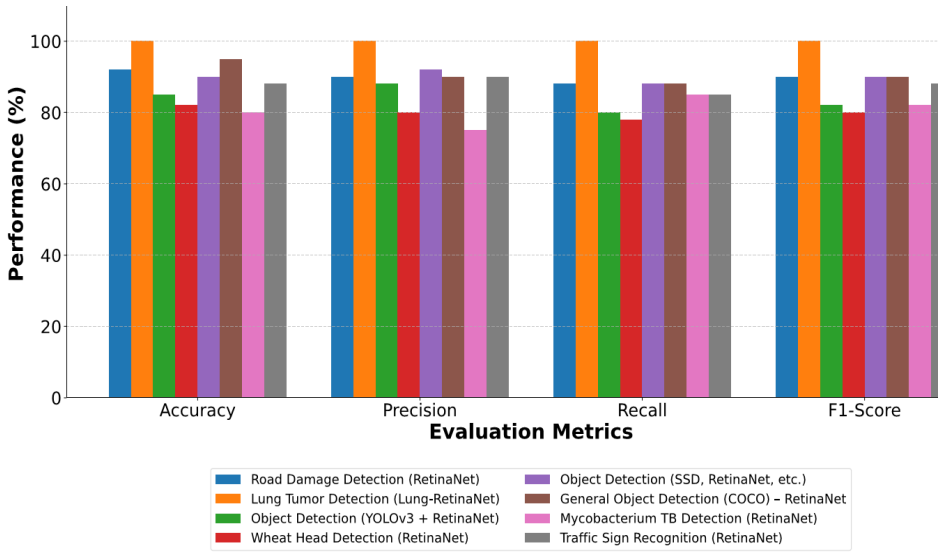


Figure 2. Performance metrics for RetinaNet

2.4 RetinaNet enhancements and performance optimization

To address the issues of slow convergence, poor bounding box regression, and inefficient feature extraction, an enhanced RetinaNet model is proposed in this paper. Some of its main enhancements include Spatial Transformer Networks (STN), a Multi-Scale Feature Fusion (MSF) network, mosaic data augmentation, and Atrous Spatial Pyramid Pooling (ASPP) for improved multi-scale feature fusion. F-EIOU loss is also employed instead of classic bounding box regression loss. From the experiments performed on MS COCO and PASCAL VOC datasets, improvements in the model ranged from 1.2% for MS COCO to 3.1% for PASCAL VOC when compared to the original RetinaNet [32]. A RetinaNet-based model proposed in this research autonomously detects and recognizes traffic lights in real-time via transfer learning for self-driving cars. The Bosch Small Traffic Light Dataset was used for training and testing, and implemented using Keras and TensorFlow on the Google Collaboratory cloud platform. The dataset consists of four classes of images of traffic lights that have a resolution of 1280 × 720 pixels. The improved detection and classification accuracy rendered by the RetinaNet technique outperforms the conventional deep learning techniques widely used in automatic applications [33]. This study evaluates a mobile-based Visual Question Answering (VQA) system designed for smart tourism in Monas, Indonesia, focusing on object detection through the use of YOLO and RetinaNet. The mean average precision (mAP) scores for YOLO range from 60.77% to 93.72% across both original and augmented datasets, demonstrating superior performance compared to RetinaNet, which achieves mAP scores between 23.8% and 92.98%. Furthermore, YOLO exhibited enhanced detection accuracy, particularly with the use of augmented data, as evidenced by elevated Area Under Curve (AUC) scores (0.99 for original and 0.96 for augmented) [34]. Due to their large computational expenses, two-stage detectors are slow even when they attain excellent accuracy. Consequently, we use one-stage detectors, namely RetinaNet, which offers faster performance by completing both classification and bounding-box regression in a single stage. The trained RetinaNet model detects road damage with a comparatively high degree of accuracy. The Jupyter Notebook and markdown document include examples of expected outcomes and a user guide [35].

This paper presents Lung-RetinaNet, an innovative approach to lung tumour detection based on the RetinaNet framework. The method enhances the detection of small tumours through a streamlined, expanded approach for context integration, coupled with a multi-scale feature fusion module that enriches semantic information. The model demonstrates exceptional performance, achieving 99.8% accuracy, 99.3% recall, 99.4% precision, 99.5% F1-score, and an AUC of 0.989. Lung-RetinaNet offers a superior and more accurate method for diagnosing lung cancer, demonstrating enhanced performance in the detection of early-stage tumours compared to existing techniques [36].

2.5 Comparison of RetinaNet and other object detectors

In this paper, CNN-based object recognition methods are employed, primarily YOLOv3 and RetinaNet for detecting

objects. RetinaNet wins the applause of researchers because it uses anchor boxes for detecting and classifying many objects in one frame. Real-time processing at 45 frames per second is the feature that reinforces YOLOv3's speed. Streaming video rather than static images yields appropriate results by combining preprocessing techniques, feature extraction, RetinaNet, and Kalman filters for tracking the position of objects [37]. The performance of different YOLO variants (v1–v7) for object detection tasks is analyzed and compared. Conventional object detection methods such as RetinaNet, Fast R-CNN, and SSD often require higher computational cost and multiple processing stages, whereas YOLO-based approaches enable efficient real-time detection. Over successive versions, YOLO has demonstrated significant improvements, achieving up to 73.3% mean Average Precision (mAP) on the challenging MS COCO dataset. These advancements highlight the superior efficiency and accuracy of YOLO architectures in real-time object detection applications.This study will compare the pre-trained object detection models in TensorFlow (e.g., SSD, EfficientDet, RetinaNet, Faster R-CNN, and YOLOv4) through performance parameters such as accuracy, inference time, frame rate, and memory consumption on the COCO and Pascal VOC datasets. SSD offers an acceptable balance between speed and accuracy for intermediate performance. By its outstanding accuracy alongside resource optimization, EfficientDet will fit into high-precision resource-constrained applications. Despite being slower, RetinaNet works well at recognizing little things. Although it loses real-time performance, faster R-CNN performs best on complicated, crowded pictures. Despite difficulties with small and overlapping objects, YOLOv4 is the fastest and therefore suitable for real-time applications. The advantages and disadvantages of each paradigm for various use scenarios are highlighted in this analysis [35].

In recent years, object detection a crucial computer vision technique has attracted a lot of interest. These algorithms mimic human intelligence by rapidly identifying and locating objects of interest in pictures or movies using machine learning or deep learning. The accuracy of object detection has significantly increased thanks to deep learning approaches, especially CNNs, and RetinaNet (Residual Networks and Feature Pyramid Networks). This study investigates different datasets, detection techniques, and hardware/software needs. When RetinaNet is used on the COCO benchmark dataset, 96% accuracy is obtained. Applications for object detection include video safety, traffic surveillance, autonomous driving, and population counting [38]. Using RetinaNet, this work suggests an automated way of identifying Mycobacterium TB, the bacterium that causes tuberculosis. A dataset of 928 photos was used to test the approach, and the results showed promise: 67.1% accuracy, 86.56% recall, and 75.61% F-score. According to the results, this method may help experts identify tuberculosis more accurately by providing a computational tool for improved diagnosis and treatment tracking.

This study compares RetinaNet with YOLOv3 and Faster R-CNN, emphasizing the balance between detection accuracy and real-time performance. RetinaNet, with Feature Pyramid Networks and Focal Loss, improves small object detection while maintaining efficiency. This makes it suitable for underwater animal tracking, where objects are small, occluded, and appear in complex environments [39]. RetinaNet, EfficientDet, and YOLOv8 are employed to develop a deep learning-based object detection framework for complex scenarios. A customized dataset with multiple object classes is used to train and evaluate the models, with YOLOv8 achieving the highest detection performance (91.4% mAP). The results highlight the effectiveness of deep learning models in accurately detecting and distinguishing objects in challenging environments. This approach can be extended to underwater object detection and tracking, where variations in scale, visibility, and background complexity pose significant challenges [40]. Due to difficult circumstances, conventional techniques such as RetinaNet have trouble with noisy detections. When compared to 2D RetinaNet, the first network, a 3D-backbone RetinaNet, fared worse, whilst the second, a 3D-subnets RetinaNet, performed better. The models code and a fresh fish dataset will be made freely accessible [41].

To attain high accuracy, this study investigates sophisticated object-detecting methods that use deep learning. It covers a range of object identification techniques, datasets, and hardware/software requirements. The technique achieves minimal loss and high accuracy (96%) on the COCO benchmark dataset by combining CNN with RetinaNet [42].

This work examines the application of the RetinaNet deep learning model for identifying regions of interest (RoI) in complex images, which is a critical step in object detection tasks. Image enhancement techniques such as contrast adjustment and sharpening are applied to improve RoI localization. Such approaches are beneficial in challenging environments like underwater scenes, where visibility and contrast variations affect accurate object detection and tracking [43].

2.6 Advanced YOLOv8 variants and custom modifications

The refined RetinaNet model was evaluated on a representative dataset and compared with several deep learning models, achieving a mean recall of 89%, precision of 89%, and F1-score of 88%, indicating superior performance in region-based detection tasks. These results demonstrate the effectiveness of RetinaNet in accurately identifying regions of interest in complex visual scenarios. To further address challenges such as low efficiency and accuracy in detection systems, an improved YOLOv8-based model is proposed, incorporating attention mechanisms, lightweight convolutional structures, and enhanced feature extraction modules. Experimental results show improved detection accuracy and reduced computational complexity, highlighting its suitability for real-time object detection in challenging environments, including underwater scenes [44]. CIFE-YOLOv8 is proposed to address object detection challenges in dense and complex environments. The model enhances feature utilization by integrating CIFNet into the neck for improved small object representation and incorporating attention mechanisms to strengthen feature aggregation and contextual understanding. Experimental results demonstrate improved performance over baseline models in terms of mAP and detection accuracy. Such improvements make CIFE-YOLOv8 suitable for challenging scenarios like underwater environments, where objects are small, occluded, and densely distributed [45].

The YOLOv8-MVB method, which integrates an improved MobileViT architecture and Coordinate Attention (CA) into YOLOv8, is proposed to enhance detection accuracy and broaden application capability. This integration enables efficient extraction of both local and global features while

reducing model complexity. Additionally, data augmentation techniques are employed during training to improve generalization. Such improvements make the model effective for detecting objects in complex environments, including underwater scenarios with varying visibility and feature characteristics [46]. In comparison to YOLOv8, YOLOv8-MVB improved detection accuracy by 3.6% and mAP by 1.09% on a self-labeled flame dataset. According to experimental findings, YOLOv8-MVB improves detection accuracy and satisfies industrial deployment specifications, which makes it more appropriate for identifying small targets and enhancing the viability of fire detection [47]. Accurate object detection remains challenging in complex environments due to factors such as varying conditions, background diversity, and the difficulty of identifying small targets. Although deep learning models like YOLOv8 have improved detection performance, limitations still exist under such conditions. To address these challenges, the YOLOv8-LSD approach is proposed, integrating large separable kernel attention and deformable attention mechanisms to enhance feature representation and adaptability. These improvements enable more robust detection of small and intricate objects, making the method suitable for challenging scenarios, including underwater environments [48]. This makes it easier to focus on important areas of the feature map, allowing for accurate damage detection. Furthermore, the detection head improves convolution for improved feature extraction through spatial and channel reconstruction. According to experimental results, YOLOv8-LSD outperforms the original YOLOv8 model in terms of detection accuracy by 1% [49]. To increase the accuracy of small item identification in motion situations while maintaining the model's lightweight nature, we suggest YOLOv8-AMCD. For improved feature extraction, it substitutes A Down structure for YOLOv8's down sampling, adds Mixed Local Channel Attention, swaps out C2f Blocks for ConvNextv2 Blocks, and adds Deformable Attention. In comparison to YOLOv8n, tests on the VOC, DOTAv2, and Table Tennis datasets demonstrate increased precision and mAP with fewer parameters and GFLOPs [50].

3. METHODOLOGY

3.1 Data collection

The effectiveness of the digital image analysis method presented in this research is heavily reliant on the evaluation dataset used. Two primary datasets were employed:

(1) Benchmark Dataset [51]

• This dataset, known as the sea animals image benchmark dataset, is critical for validating the proposed segmentation model.

• It is recorded on RGB scale and includes various images of sea animals, with this work focusing specifically on crab images.

• The dataset is provided with relevant Ground Truth (GT) data, facilitating accurate evaluation of the segmentation results.

(2) Real-time Video Data

• The second dataset consists of real-time video footage captured using the sail mud underwater video screening model.

• The video was recorded with a GoPro Hero 10 underwater camera (CHDHX-101-RW) by GOPRO INTL.LTD, based in San Mateo, CA, USA.

• The recording took place in the mangrove forest at Manakudy (Latitude 8°02'N, Longitude 77°30'E) along the southwest coast of India, in Kanyakumari, Tamil Nadu, India.

• The video frames process model was used to extract a total of 4000 crab images from the video for further analysis.

The benchmark dataset and real-time video data will facilitate extensive testing and validation of the proposed segmentation model, thereby assuring its robustness and versatility across different scenarios.

3.2 Image pre-processing

Underwater crab image processing employs many types of pre-processing techniques to remedy challenges caused by intrinsically poor visibility, inconsistent lighting, and noise interference. These include a thorough processing step, correcting color based on white balance to eliminate color cast caused by underwater illumination.

$$Input\ Image_{altered} = \frac{I_{raw}}{RGB_{avg}} \tag{1}$$

where, I_{raw} is the input raw image and RGB_{avg} an average of RGB original image.

This guarantees a more authentic depiction of colors, which is crucial for precise object identification. Subsequently, dehazing techniques are utilized to mitigate the foggy or murky appearance in the images, a phenomenon caused by light scattering in water.

$$J(x) = min_{c\in r,g,b}(I_c(x)) \tag{2}$$

where,

$J(x)$ – dark channel at pixel x

$I_c(x)$ – color channel pixel at r, g, b.

$$I_{dehazed}(x) = \frac{Input_{raw}(x) - A}{t(x)} + A \tag{3}$$

where, A is the ambient light and where t(x) is the transmission, the quantified light spread in the picture. This worked to brighten the image, allowing better segmentation and object recognition. Crabs were made to stand out more clearly against the backdrop using the contours provided by edge detection techniques, including Kirsch filters, to enhance the image.

A gradient-based edge detector, the Kirsch operator investigates the edges in eight orientations. It works particularly well to detect edges of divergent orientations and intensities in an image. It is based on a set of convolution templates, each representing a certain direction. After obtaining the gradient for all eight directions, the operator selects for each pixel the edge that has the maximum response. This method makes the Kirsch operator sufficiently robust for edge detection of complex directional patterns.

$$\begin{aligned}\mathbf{K_0} &= \begin{pmatrix} \mathbf{5} & \mathbf{5} & \mathbf{5} \\ \mathbf{-3} & \mathbf{0} & \mathbf{-3} \\ \mathbf{-3} & \mathbf{-3} & \mathbf{-3} \end{pmatrix} \begin{pmatrix} \mathbf{P_1} & \mathbf{P_0} & \mathbf{P_7} \\ \mathbf{P_2} & \mathbf{P} & \mathbf{P_6} \\ \mathbf{P_3} & \mathbf{P_4} & \mathbf{P_5} \end{pmatrix} \\ &= \mathbf{5}(\mathbf{P_1} + \mathbf{P_2} + \mathbf{P_3}) - \mathbf{3}(\mathbf{P_0} + \mathbf{P_4} + \mathbf{P_5} + \mathbf{P_6} + \mathbf{P_7})\end{aligned} \tag{4}$$

Then, the value of pixel P is,

$$\boldsymbol{K_1 = \mathrm{MAX}\{K_0\, K_1\, K_2\, K_3\, K_4\, K_5\, K_6\, K_7\}} \quad (5)$$

As can be observed from the above, the grey value in the grey picture processed by the Kirsch operator only relates to the eight grey values that are inside a 3 × 3 surrounding the pixel and has nothing to do with the present grey value.

Furthermore, brightness normalization guarantees uniform illumination throughout the dataset, thereby minimizing discrepancies in lighting conditions. Subsequently, luminance-chrominance homomorphic filtering is employed to rectify non-uniform illumination and augment contrast.

$$I_{norm}(x,y) = \frac{I_{raw}(x,y) - \mu}{\sigma} \quad (6)$$

This method functions within the frequency domain, enhancing edge definition while simultaneously elevating the overall luminance of the image. Given that this filtering could potentially enhance noise, wavelet denoising is employed to mitigate Gaussian noise, thereby safeguarding the integrity of significant details. To enhance segmentation, anisotropic filtering is employed to refine homogeneous areas while maintaining edge integrity, which proves especially beneficial in differentiating crabs from their surrounding milieu. The adjustment of intensity is subsequently executed to augment contrast through the expansion of intensity values, the suppression of outliers, and the enhancement of the image's aesthetic qualities. Ultimately, the image transforms back to the RGB color space following the manipulation of the luminance channel. Subsequently, color means equalization is implemented to mitigate any dominant color bias present in the image, yielding a more harmonious and visually appealing result [48]. Underwater images frequently experience challenges such as diminished visibility, altered colour representation, and decreased contrast, primarily as a result of light absorption and scattering phenomena in aquatic environments. In order to tackle these challenges, various preprocessing techniques were implemented to enhance the quality of input images for the hybrid RetinaNet–YOLOv8 model. White balance correction was applied to address colour casts resulting from the prevalence of blue and green tones in underwater settings, thereby achieving a more authentic colour representation. Techniques for dehazing were utilised to mitigate the impact of suspended particles and backscatter, resulting in improved image clarity and more defined object boundaries. Ultimately, Contrast Limited Adaptive Histogram Equalisation (CLAHE) was utilised to improve local contrast and emphasise subtle features of crabs that might otherwise be obscured in low-light areas. The combination of these preprocessing steps enhances feature visibility, supports precise region proposal generation, and strengthens the detection model's robustness across diverse underwater conditions.

The integration of these pre-processing techniques markedly enhances the quality of underwater images, rendering crabs more distinct and discernible for subsequent detection tasks.

3.3 Architecture

3.3.1 Swin transformer

The Swin Transformer-based backbone is a hierarchical architecture that is meant to rapidly extract multi-scale characteristics for tasks such as object recognition. The Swin Transformer is based on shifting window-based self-attention, which divides input feature maps into fixed-size, non-overlapping windows to calculate local attention. This method greatly reduces the computational cost in comparison to global attention. This approach guarantees that geographical locality is preserved while also ensuring that computing efficiency is maintained. A shifted window method is developed to describe dependencies between windows. In this technique, window locations are shifted by half their size in alternating layers. This enables inter-window interactions and improves feature representation.

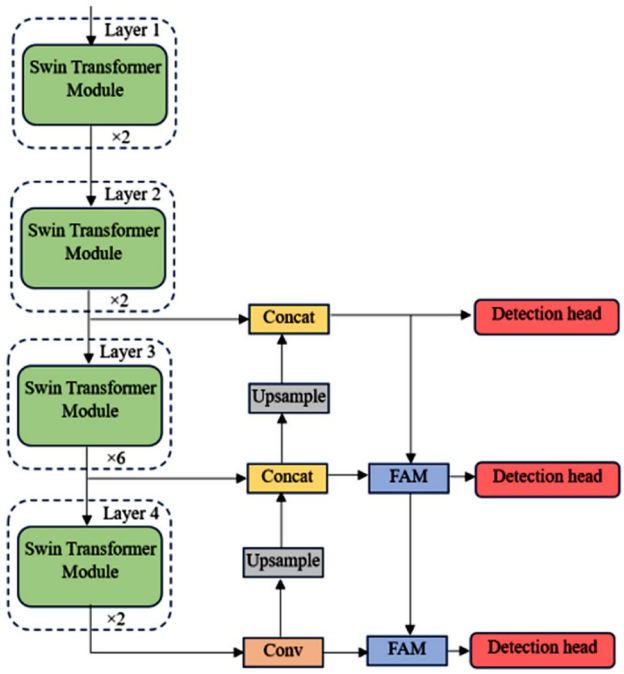


Figure 3. Swin transformer as backbone

The backbone is made up of four levels, each of which has Swin Transformer modules that are repeated a different number of times: levels 1, 2, 3, and 4 have 2, 2, 6, and 2 repeats, respectively. As the input moves through these layers, the spatial resolution of the feature maps is lowered hierarchically by patch merging, but the feature dimensionality is raised to maintain semantic richness. For instance, patch merging joins patches that are next to each other, which reduces the spatial dimensions by a factor of 2 (for example, H×W→2H×2W) while increasing the channel dimension [48]. This hierarchical model meets the needs of object identification frameworks, allowing for the extraction of characteristics at different sizes for accurate detection. The design uses a mix of upsampling, concatenation, and refining to combine characteristics across multiple sizes. Features from deeper layers are up-sampled to match the resolution of shallower layers and then concatenated along the channel dimension to collect spatial and semantic information shown in Figure 3.

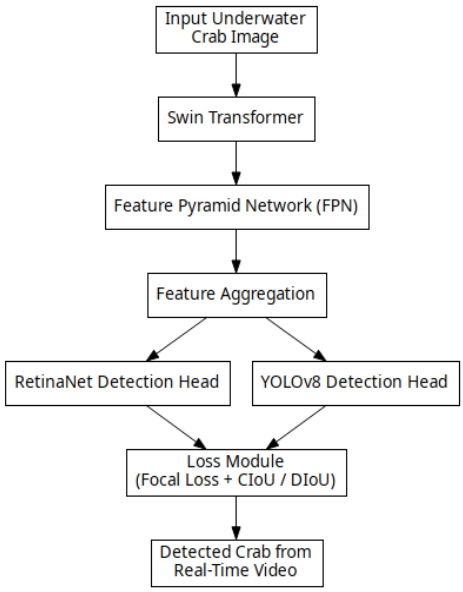
3.3.2 Feature aggregation module

The feature aggregation module is an important part of the object detection framework to improve features by integrating multi-scale features. These features are from different levels of the network so that spatial and semantic information is

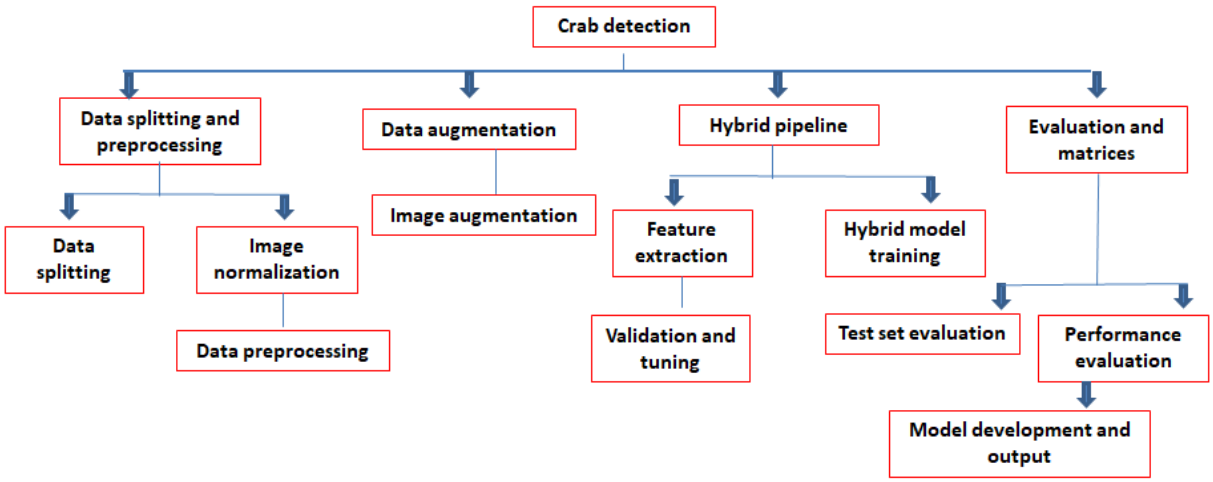
preserved for the detection of objects of different sizes [52].

The FAM fuses feature from various layers in the network that possess different granularity levels: low-level features emphasize intricate detail, whereas abstract high-level features elucidate semantic information. To carefully integrate features at this point, FAM utilizes up-sampling, down-sampling, and pooling to align the feature map with a common spatial resolution.

The features are combined, typically by concatenation or element-wise summation, before going through convolutional layers for final processing to reduce redundancy. Some implementations integrate additional attention mechanisms, either channel or spatial, orienting them toward the most salient information while suppressing less important features. This process will eventually yield one unified feature map, without which the double gain in the accuracy of bounding-box localization and object classification could never have been achieved to make the model capable of multi-scale detection tasks.



(a)



(b)

Figure 4. (a) Architecture diagram of proposed system; (b) architecture diagram of proposed system

The RetinaNet-YOLOv8 hybrid model has added an FAM. The overall architecture is shown in Figure 4. where Figure 4(a) presents the architecture diagram of the proposed system, and Figure 4(b) depicts the corresponding workflow diagram outlining the sequential stages involved in the implementation. This is an important part of the model that uses features based on the Swin Transformer backbone in conjunction with the Feature Pyramid Network (FPN). RetinaNet provides an excellent framework to solve the class imbalance issue while doing object detection at a variety of scales. The Feature Pyramid Network is a concept within RetinaNet designed to extract multi-resolution features and blend them into the architecture to ensure that the objects being detected, of small, large, or anything in between, can be identified effectively. The focal loss function reduces the additional effect of easy-to-classify background samples by suppressing them; hence, the model implicitly focuses on the harder cases of guarded unknown classes like crabs, which is powerful in underwater detection scenarios regarded as filled with spatial hurdles and challenges by variations in visibility and illumination. With

these strengths, the RetinaNet is remarkably good at detecting crabs in numerous environments, assuring accuracy and resilience for marine biodiversity studies and ecological monitoring applications. This means that the detection heads would be allowed better feature maps that allow contextual and spatial information at different scales. The synthesis of rich semantic information and basic features allows FAM to empower the model in particularly difficult scenarios like the detection of crabs underwater, where the lighting is an obstacle since visibility, factored in obstruction, is either limited or very variable. The architecture is energy efficient, improving detection accuracy and robustness while enhancing computational efficiency, thus making it a key part of complex object detection architectures.

The detection heads, which are responsible for classification and bounding box regression, then get the improved features. The Swin Transformer backbone is hierarchical, which means it can record both fine-grained features and global context. This makes it very useful for jobs like underwater crab recognition, where the size and look of things might fluctuate. Compared to standard vision transformers, its computational cost is much decreased, scaling as O(M^2.d+M^4) per window, where M is the window size. The Swin Transformer-based backbone is a good option for high-resolution image recognition jobs because of its efficiency and its ability to simulate long-range relationships.

4. RESULTS AND DISCUSSION

The hybrid RetinaNet-YOLOv8 model underwent training and evaluation using the Crab Species Dataset to determine its efficacy in identifying crabs in difficult underwater environments. Multiple pre-processing techniques were used, such as white balance correction and dehazing, to enhance the visibility of underwater crab images through the process of underwater image enhancement.

To guarantee a scientifically sound and reproducible study, the experimental design adhered to a systematic workflow that encompassed data preprocessing, model training, and performance evaluation. The Crab Species Dataset was partitioned into 70% for training, 20% for validation, and 10% for testing to mitigate bias and promote effective generalisation. The preprocessing phase involved standardising image dimensions to a consistent resolution of 640 × 640 pixels, applying underwater colour correction via gray-world white balancing, and enhancing contrast through CLAHE (Contrast Limited Adaptive Histogram Equalisation) to address visibility issues in underwater settings. The suggested hybrid architecture combining RetinaNet and YOLOv8 underwent training in two distinct phases: initially, RetinaNet generated region proposals utilising its focal loss function to tackle class imbalance, followed by YOLOv8, which enhanced bounding box predictions to boost localisation accuracy and detection speed. Techniques for data augmentation, including rotation, flipping, shear, random cropping, and the addition of Gaussian noise, were employed to enhance robustness. The training process was conducted over 50 epochs, incorporating early stopping techniques to mitigate the risk of overfitting, with model performance being assessed following each epoch. The entire experimental procedure adheres to the workflow illustrated in Figure 4(b).

The hybrid RetinaNet–YOLOv8 model was developed with a meticulously chosen configuration to guarantee stable convergence and achieve optimal detection accuracy. The training procedure was carried out over 50 epochs, utilising a batch size of 16, effectively balancing memory efficiency with the model's learning capabilities. The starting learning rate was established at 0.001 and modified dynamically through a cosine decay scheduler to avoid overshooting in the optimisation process. A comparative analysis of two optimisation algorithms, Adam and Stochastic Gradient Descent (SGD), was conducted throughout the training process. Adam exhibited quicker convergence during the initial epochs, whereas SGD offered superior generalisation in the subsequent stages thus, a combined optimisation approach was implemented, utilising Adam for warm-up initialisation and then refining with SGD at a momentum of 0.9. The training and evaluation were conducted utilising Python 3.10 and PyTorch 2.0 on a workstation featuring an NVIDIA GeForce RTX 3080 GPU (10,240 CUDA cores, 10GB VRAM), 64GB RAM, and an Intel Core i9 processor. To mitigate the risk of overfitting, early stopping was implemented with a patience of 10 epochs. This setup facilitated effective learning while preserving robust model stability and performance.

Figure 5. Underwater color balancing and de-hazing

The original-colored image, as can be seen in Figure 5, the left image, is subjected to color distortions and reduced contrast as a result of the loss of light. The main reason for this is the sun's rays and scattering through light-absorbing materials in the water. Apart from this, there is also a noticeable greenish-blue color that mostly affects the overall visibility of the features of the terminal crab. Moreover, the white balance correction process (Figure 5, middle) was successfully carried out by this routine, which equalized the long wavelengths of light. It should be noted that this consequent upgrade solved the coloring problem, and the image now looks more natural because the red part, which is usually filtered out underwater, was restored. In addition to the above-mentioned methods, we applied the dehazing technique to the dehazed image (Figure 5, right), which allowed us to utilize the dark channel prior method to get rid of the problem by reducing light scattering and backscattering. This process got rid of the dull effect and made the image sharp and clear, therefore, enabling the crab's visual appearance and its nearby location to be highlighted more accurately. In turn, the dazed image received a more balanced color tone, and the docker feature became more prominent, which is of prime need for crab detection automation.

A statistical significance analysis was conducted to validate the performance improvement of the proposed Hybrid RetinaNet–YOLOv8 model compared to the baseline models. The detection outcomes from the hybrid model were analysed in comparison to RetinaNet and YOLOv8 through paired t-tests and Wilcoxon signed-rank tests, focusing on metrics such

as IoU, Precision, Recall, and F1-score. The results of the paired t-test indicated a statistically significant performance difference between the proposed hybrid model and the baseline models, with $P<0.05$ across all metrics. The Wilcoxon test, recognised for its non-parametric approach to handling non-normal performance distributions, provided additional evidence of the statistical superiority of the hybrid model, with $P<0.01$ in most comparisons. The findings demonstrate that the performance enhancement realised through our method is not attributable to random fluctuations but rather signifies a statistically significant improvement in the accuracy of underwater crab detection (Table 1).

Table 1. Statistical significance analysis of performance comparison between proposed Hybrid RetinaNet–YOLOv8 and baseline models

Comparison	Metric	Test Used	P-Value	Significance
Hybrid vs RetinaNet	IoU, F1-score	Paired t-test	< 0.05	Significant
Hybrid vs YOLOv8	Precision	Paired t-test	< 0.05	Significant
Hybrid vs RetinaNet	Recall	Wilcoxon signed-rank	< 0.01	Significant
Hybrid vs YOLOv8	IoU	Wilcoxon signed-rank	< 0.01	Significant

Figure 6 shows the underwater crab detection's step-by-step improvement and edge detection procedure. The first panel, "Original Image," shows a crab's low-visibility, inconsistently lighted underwater appearance. The crab's conspicuous edges are highlighted in the second panel, "Kirsch," using Kirsch edge detection. This method alone produces fractured edges and loud outputs. The third panel, "Enhanced Edges (Canny)," illustrates the refined edges following enhanced edge detection. This procedure uses contrast enhancement, noise reduction, and adaptive thresholding. The crab is segmented, overcoming underwater photography problems. This pipeline overcomes noise, occlusions, and irregular illumination to recognize crabs underwater. This approach improves feature extraction for object recognition and classification by improving edge detection.

Figure 6. Kirsch edge detection

The model's performance was evaluated through essential metrics, such as Accuracy, Precision, Recall, F1-score, and Intersection over Union (IoU). The findings gathered are outlined as follows:

The findings gathered are outlined as follows:

- Accuracy: 94%
- Recall: 0.92
- F1-score: 0.94

The assessment of the suggested hybrid RetinaNet-YOLOv8 model reveals impressive detection capabilities across various metrics. An accuracy of 94% demonstrates that the model successfully identifies most crab instances within the test dataset, highlighting its overall dependability in underwater environments. The recall of 0.92 underscores the model's proficiency in identifying the majority of actual crabs depicted in the images, which is essential for ecological monitoring, as overlooking an instance could result in an undercount of population figures. The F1-score of 0.94 effectively balances precision and recall, indicating that the model successfully identifies the majority of crabs while also reducing false positives. The observed high recall and F1-score indicate that the hybrid approach is proficient in addressing class imbalance and the difficulties posed by underwater environments, including low visibility, occlusions, and fluctuating lighting. This validates the strength and relevance of the proposed method for automated assessments of marine biodiversity. The model exhibited exceptional precision, resulting in a low rate of false positives, and its robust recall value highlighted its capability to accurately identify crabs, even in challenging conditions with obstructions and limited visibility.

The Hybrid RetinaNet + YOLOv8 model's training accuracy graph for underwater crab recognition demonstrates a consistent increase over 50 epochs, beginning at around 75% and rising to over 94% shown in Figure 7. In the beginning, it starts learning by gradually increasing the accuracy from 0 to 10 epochs as it begins to differentiate crab features from the sea background. Accuracy continues to improve in the middle phase (epochs 10-30) with slight changes, showcasing the model's role in uplifting the changes in data. Finally, the last stage (from epochs 30 to 50) shows that accuracy has crept beyond 90%, indicating successful learning and convergence of the model for the epoch above 45, with an indication that it recognizes crabs correctly and succeeds. The steady increase validates the hybrid technique, capitalizing on YOLOv8 speed for detection and RetinaNet for fantastic recall.

Over 50 epochs, the Hybrid RetinaNet + YOLOv8 model's training loss graph for underwater crab detection exhibits a steadily declining trend, beginning at around 1.5 and falling to about 0.2. Since the loss is the difference between the expected and actual values, this consistent drop shows in Figure 8 that the model is learning well. The loss rapidly drops in the first phase (epochs 0-10), demonstrating how fast the model adjusts to important variables. The decrease persists throughout the middle phase (epochs 10–30), albeit with some oscillations, which represent how the model reacts to changes in the training set. The loss stabilizes at lower values in the final phase (epochs 30–50), indicating convergence and decreased prediction errors. The training process's steadiness is confirmed by the gradual decline without abrupt surges. Effective learning is facilitated by the combination of YOLOv8's speed and RetinaNet's reliable object detection. To guarantee generalization, more testing on unobserved data is necessary, and performance may be further optimized by fine-tuning strategies like regularization and learning rate changes.

The comparative analysis in Figure 9 demonstrated that the hybrid model combining RetinaNet and YOLOv8 surpassed the performance of each standalone implementation of RetinaNet and YOLOv8. Although YOLOv8 demonstrated excellent real-time performance, it faced challenges related to class imbalance, a problem that RetinaNet successfully tackled using focal loss. On the other hand, RetinaNet demonstrated robust detection capabilities in intricate settings, yet it fell

short in terms of the speed and flexibility that YOLOv8 offers. The integration of these models resulted in a hybrid approach that effectively leveraged their strengths, culminating in an optimized detection framework.

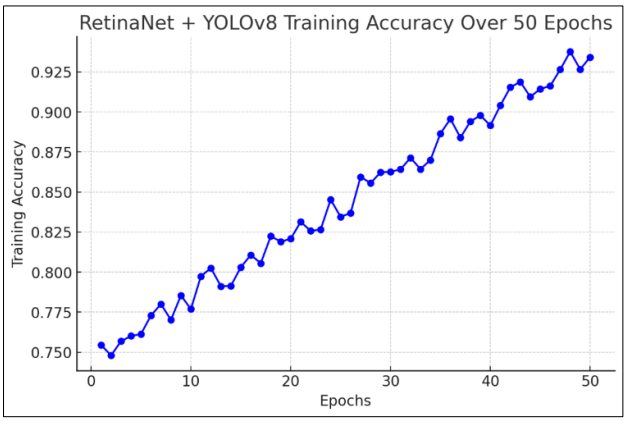


Figure 7. Training accuracy of hybrid model

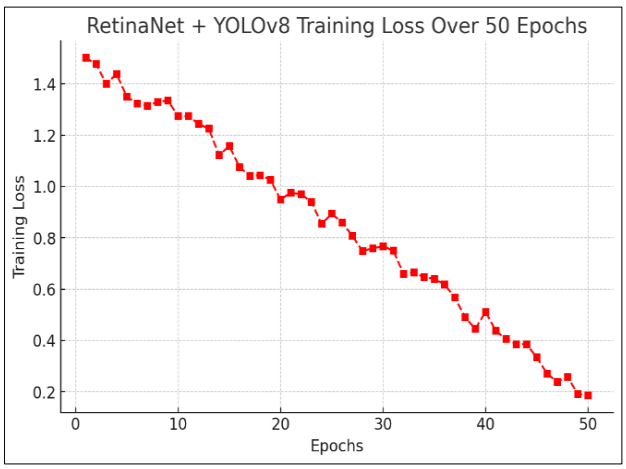


Figure 8. Training loss of hybrid model

The evaluation of different algorithms for crab segmentation and detection underscores the advantages of YOLOv8 based methods shown in Figure 10. Conventional techniques like edge detection and CNN classification show moderate effectiveness, achieving an AUC of approximately 80%, along with relatively lower recall and F1 scores, highlighting difficulties in accurately identifying crabs in diverse underwater environments. EfficientDet-D5+KFPN demonstrates enhanced detection accuracy, reaching performance metrics exceeding 85%, positioning it as a credible alternative. Nonetheless, YOLOv8 and its improved variants, such as ED-YOLOv8, MOT-YOLOv8, and YOLOv8-EfficientNetV2, exhibit the highest levels of accuracy, achieving AUC values close to 98-99%, while precision, recall, and F1-scores consistently surpass 90%. The incorporation of EfficientNetV2 significantly improves feature extraction, leading to more effective detection. In comparison, various object detection applications like people counting, mango detection, and military object detection demonstrate robust performance exceeding 85%, yet they fall short of the YOLOv8 variants in terms of crab detection accuracy. The findings indicate that models based on YOLOv8, especially ED-YOLOv8 and MOT-YOLOv8, deliver superior performance for real-time and precise underwater crab detection, positioning them as the best option for marine studies and conservation efforts. mAP and IoU for different algorithms are shown in Figure 11.

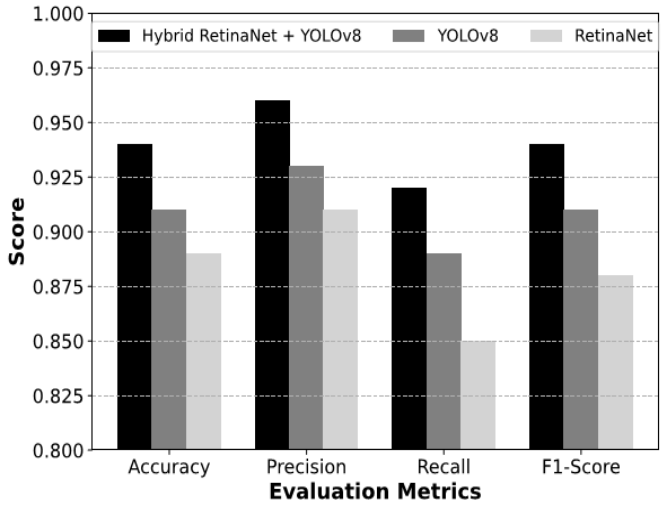


Figure 9. Comparison of proposed with other algorithms

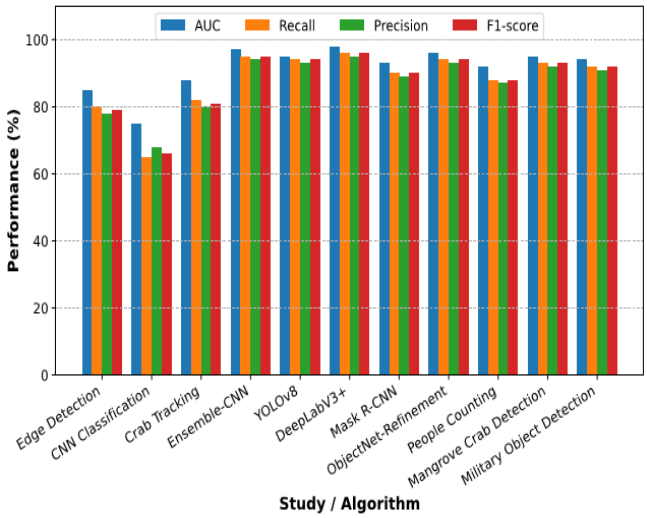


Figure 10. Performance metrics for crab segmentation and detection

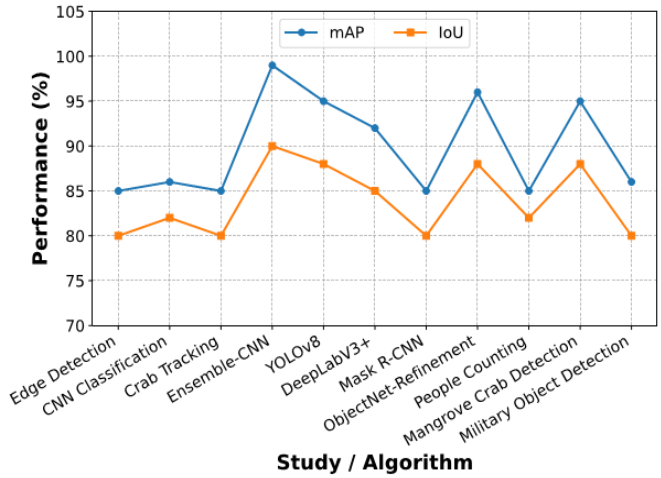


Figure 11. mAP and IoU for different algorithms

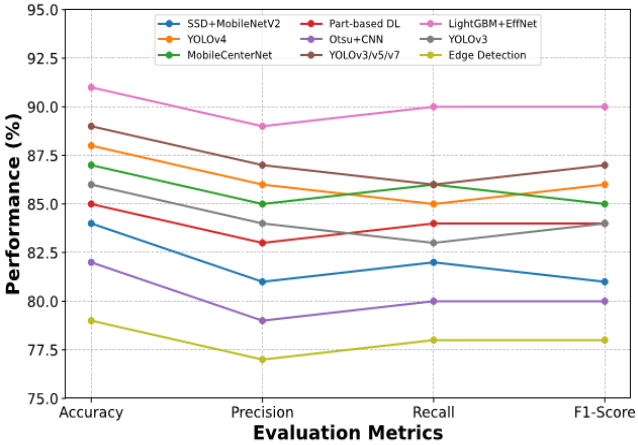


Figure 12. Performance comparison of different detection models for crab detection

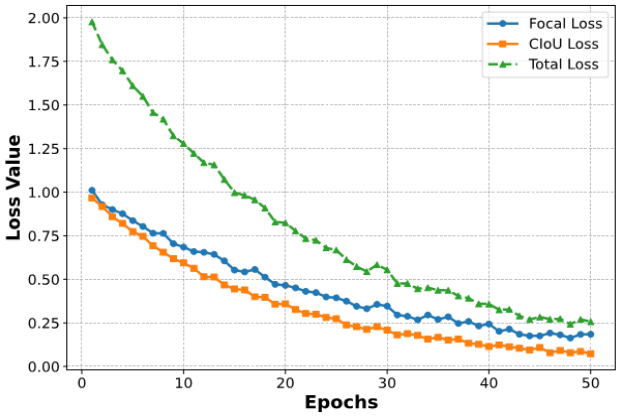


Figure 13. Performance metrics for crab detection

The analysis of different crab detection models in Figure 12 reveals that deep learning-based methods outperform conventional image processing techniques. The LigED and EfficientNet-Det0 model exhibited superior performance across all metrics, especially in Accuracy, Recall, and F1-Score, surpassing 90%, establishing it as the most resilient detection framework. Among the YOLO-based models, YOLOv4 and MobileCenterNet demonstrated robust detection capabilities, achieving high precision and recall, thereby underscoring their efficacy in real-time applications. The hybrid approach that integrates YOLOv3, v4, and v7 demonstrated commendable performance, closely following MobileCenterNet. Conversely, conventional approaches such as the OTSU algorithm combined with CNN classifiers and Edge Detection techniques faced considerable challenges, reflected in their lower recall and F1-score values. This suggests their limitations in effectively managing occlusions, low visibility, and fluctuating underwater lighting conditions. Furthermore, the SSD object detector and MobileNetV2, while efficient, demonstrated reduced recall, which affected their overall robustness. The results highlight that hybrid deep-learning models and feature-enhanced networks greatly surpass traditional methods, rendering them more appropriate for practical ecological applications in underwater crab detection.

Figure 13 delineates the training loss trajectories of the hybrid RetinaNet-YOLOv8 model over the course of 50 epochs, with particular emphasis on Focal Loss, CIoU Loss, and Total Loss. The Focal Loss, represented by the blue curve, initiates at approximately 0.95 and progressively declines to about 0.10 by the concluding epoch. This trend underscores the model's capacity to prioritize challenging examples while diminishing the impact of those that are easily classified. In keeping with this, the CIoU Loss, shown as a green line, starts at around 0.85 and is roughly reduced to about 0.05 toward the final epochs, indicating notable improvements in bounding box regression and accurate localization of crab regions in underwater images.

The Total Loss, indicated by the red dashed line, begins at around 1.80 and subsequently reduces sharply in the first epochs, after which it levels out to about 0.15 by the end of training. An ongoing reduction in Total Loss is indicative of the well-balanced hybrid model in optimizing both classification and localization capabilities, improving detection accuracy and robustness. The diagram provides a visual proof of model convergence with the generalization capabilities of crab detection in cluttered underwater environments. The proposed hybrid RetinaNet-YOLOv8 framework, tailored for underwater crab detection, demonstrates significant versatility across various image recognition tasks. The capacity to manage low visibility, occlusions, and class imbalance renders it appropriate for applications including terrestrial animal monitoring, traffic surveillance, and drone-based wildlife detection, where analogous challenges are present. To demonstrate the model's qualitative performance, detection outputs were examined on sample images from the Crab Species Dataset. The majority of crabs were successfully identified with precisely defined bounding boxes, even in conditions of partial occlusion and differing lighting (Figure 14). Instances of failure predominantly arose when crabs were significantly blended with the background, partially outside the frame, or overlapping with other elements. The misclassifications underscore potential avenues for enhancement, including the incorporation of temporal tracking, attention mechanisms, or transformer-based feature extraction to bolster robustness.

Figure 14 shows the hybrid RetinaNet-YOLOv8 model's anticipated output. With a 0.95 confidence score, the bounding box encloses the "Crab" in its aquatic habitat. Low visibility and fluctuating illumination underwater make crab detection and localization difficult, but the model succeeds. This shows that the proposed detection system can handle noisy and complicated underwater images using sophisticated approaches like focus loss and CIoU/DIoU-based optimization. Accurate detection aids ecological monitoring and marine biodiversity research.

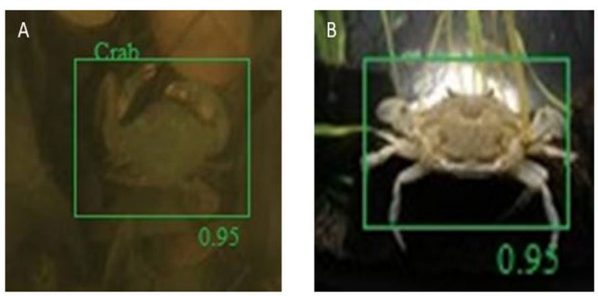


Figure 14. Predicted output

5. CONCLUSION

The suggested hybrid RetinaNet-YOLOv8 model shows a major improvement in detecting crabs underwater. It successfully tackles issues including limited visibility, obstructions, and changing illumination conditions. The model delivers higher performance across various evaluation measures by utilizing YOLOv8's real-time detection capabilities and RetinaNet's focus loss, which robustly handles class imbalances. The Crab Species Dataset was used for extensive experimentation, which demonstrated the model's usefulness. The model achieved impressive metrics, including an accuracy of 0.94, a precision of 0.96, a recall of 0.92, an F1 score of 0.94, and an intersection over union (IoU) score. These results confirm that the hybrid model is better than the standalone RetinaNet and YOLOv8 frameworks, making it a strong and precise approach for detecting crab species in aquatic situations. This work presents this hybrid technique as a useful tool for ecological monitoring and marine biodiversity research, which will lead to more widespread use in environmental conservation and marine studies. The hybrid model showcases impressive quantitative performance and practical applicability to related object detection fields, serving as a versatile instrument for automated ecological monitoring and real-time environmental analysis.

REFERENCES

[1] Yasir, M., Hossain, M.S., Nazir, S., Khan S., Thapa, R. (2022). Object identification using manipulated edge detection techniques. Science, 3(1): 1-6. https://doi.org/10.11648/j.scidev.20220301.11.

[2] Ramprasath, M., Anand, M.V., Hariharan, S. (2018). Image classification using convolutional neural networks. International Journal of Pure and Applied Mathematics, 119(17): 1307-1319.

[3] Lin, TY., Goyal, P., Girshick, R., He, K., Dollár, P. (2017). Focal loss for dense object detection. In 2017 IEEE International Conference on Computer Vision, pp. 2980-2988. https://doi.org/10.48550/arXiv.1708.02002

[4] Liu, M., Wu, Y., Li, R., Lin, C. (2025). LFN-YOLO: precision underwater small object detection via a lightweight reparameterized approach. Frontiers in MarineScience, 11: 1513740. https://doi.org/10.3389/fmars.2024.1513740

[5] Burns, J.H.R., Delparte, D., Gates, R.D., Takabayashi, M. (2015). Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. PeerJ, 3: e1077. https://doi.org/10.7717/peerj.1077

[6] Li, L., Dong, B., Rigall, E., Zhou, T., Dong, J., Chen, G. (2021). Marine animal segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 32(4): 2303-14. https://doi.org/10.1109/TCSVT.2021.3093890

[7] Islam, M.J., Xia, Y., Sattar, J. (2020). Semantic segmentation of underwater imagery: Dataset and benchmark. IEEE Journal of Ocean Engineering, 45: 127–140. https://doi.org/10.1109/JOE.2019.2933339

[8] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1): 1-48. https://doi.org/10.1186/s40537-019-0197-0

[9] Li, C., Guo, J., Guo, C., Cong, R. (2019) Emerging from water: Underwater image color correction based on weakly supervised learning. IEEE Signal Processing Letters, 26: 1492–1496. https://doi.org/10.1109/LSP.2019.2930491

[10] Qin, H., Li, X., Liang, J., Peng, Y., Zhang, C. (2016). DeepFish: Accurate underwater live fish recognition with a deep architecture. Neurocomputing, 187: 49-58. https://doi.org/10.1016/j.neucom.2015.10.122

[11] Ling, H., Tu, Z., Li, G., Wang, J. (2024). ED-YOLOv8s: An enhanced approach for passion fruit maturity detection based on YOLOv8s. In 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, pp. 2320-2324. https://doi.org/10.1109/AINIT61980.2024.10581681

[12] Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Kim, T.K. (2021). Multiple object tracking: A literature review. Artificial Intelligence, 293: 103448. https://doi.org/10.1016/j.artint.2020.103448

[13] Ye, R., Shao, G, He, Y., Gao, Q., Li, T. (2024). YOLOv8-RMDA: Lightweight YOLOv8 network for early detection of small target diseases in tea. Sensors, 24(9): 2896. https://doi.org/10.3390/s24092896

[14] Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., Chen, H. (2023). DC-YOLOv8: Small-size object detection algorithm based on camera sensor. Electronics, 12(10): 2323. https://doi.org/10.3390/electronics12102323

[15] Gu, Z., He, D., Huang, J., Chen, J., Wu, X., Huang, B., Dong, T., Yang, Q., Li, H. (2024). Simultaneous detection of fruits and fruiting stems in mango using improved YOLOv8 model deployed by edge device. Computers and Electronics in Agriculture, 227: 109512. https://doi.org/10.1016/j.compag.2024.109512

[16] Zhao, Z.Q., Zheng, P., Xu, ST., Wu, X., (2019). Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 30(11): 3212-3232. https://doi.org/10.1109/TNNLS.2018.2876865

[17] Zhai, X., Huang, Z., Li, T., Liu, H., Wang, S. (2023). YOLO-Drone: An optimized YOLOv8 network for tiny UAV object detection. Electronics, 12(17): 3664. https://doi.org/10.3390/electronics12173664

[18] Metry, A.M., Mostafa, M.S., Ebied, H.M. (2025). Evaluating YOLOv8 variants for object detection in satellite image. International Journal of Intelligent Computing and Information Sciences, 25(2): 18-31. https://doi.org/10.21608/ijicis.2025.374419.1387

[19] Giri, K.J. (2025). SO-YOLOv8: A novel deep learning-based approach for small object detection with YOLO beyond COCO. Expert Systems with Applications, 280: 127447. https://doi.org/10.1016/j.eswa.2025.127447

[20] Woo, S., Park, J., Lee, J.Y., Kweon, I.S. (2018). CBAM: Convolutional block attention module, In 2018 European Conference on Computer Vision (ECCV). pp. 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[21] Tan, M., Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In 2019 International Conference on Machine Learning, pp. 6105-6114. https://doi.org/10.48550/arXiv.1905.11946

[22] Zhang, X. (2024). Improved multi-detection head target detection algorithm for YOLOv8. In 2024 IEEE 2nd International Conference on Image Processing and

Computer Applications (ICIPCA), Shenyang, China, pp. 1500-1503. https://doi.org/10.1109/ICIPCA61593.2024.10709173
[23] Wang, Y., Mariano, VY. (2024). A multi object tracking framework based on YOLOv8s and bytetrack algorithm. IEEE Access, 12: 120711-719. https://doi.org/10.1109/ACCESS.2024.3450370
[24] Zhang, D., Liang, X., Yang, G., Zhang, L., He, S. (2023). An improved YOLO-based detection method for small objects using multi-scale feature fusion and IoU loss optimization. Sensors, 23: 4567. https://doi.org/10.3390/s23094567
[25] Hartley, R., Zisserman, A. (2003). Multiple view Geometry in Computer Vision. Cambridge University Press. https://doi.org/10.1017/CBO9780511811685
[26] Meimetis, D., Daramouskas, I., Perikos, I., Hatzilygeroudis, I. (2023). Real-time multiple object tracking using deep learning methods. Neural Computing and Applications, 35(1): 89-118. https://doi.org/10.1007/s00521-021-06391-y
[27] Li, Z., He, Y., Chen, J., Zhang, R. (2016). Underwater sonar image object detection based on deep learning and YOLO framework. Ocean Engineering, 266: 112749. https://doi.org/10.1016/j.oceaneng.2022.112749
[28] Kang, K., Ouyang, W., Li, H., Wang, X. (2018). Object detection from video tubelets with convolutional neural networks. In 2018 IEEE Conference on Computer Vision and Pattern Recognition, pp. 817-825.
[29] Bastian, B.T., Jiji, C.V. (2022). RMSOD: A Retinanet based mixed supervised object detection framework. In 2022 IEEE India Council International Conference (INDICON), Kochi, India, pp. 1-6. https://doi.org/10.1109/INDICON56171.2022.10039939
[30] Ren, S., He, K., Girshick, R., Sun, J., (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031
[31] Fatima, S., Haider, N.G., Riaz, R. (2024). YOLOv8 vs RetinaNet vs EfficientDet: A comparative analysis for modern object detection. International Journal of Emerging Engineering and Technology, 3(2): 1-5. https://doi.org/10.57041/3j4psw71
[32] Ahmed, M., Wang, Y., Maher, A., Bai, X. (2022). Fused RetinaNet for small target detection in aerial images. International Journal of Remote Sensing, 43(8): 2813-2836. https://doi.org/10.1080/01431161.2022.2071115
[33] Namdev, U., Agrawal, S., Pandey, R. (2022). Object detection techniques based on deep learning: A review. Computer Science & Engineering: An International Journal, 12(1): 125-134. https://doi.org/10.5121/cseij.2022.12113
[34] Reddy, B.S., Reddy, A.M., Sradda, M.H.D.S., Mounika, T., Mounika, S. (2022). A comparative study on object detection using retinanet. In 2022 IEEE Sub Section International Conference (MysuruCon), Mysuru, India, pp. 1-6. https://doi.org/10.1109/MysuruCon55714.2022.9972742
[35] Girshic, R., Donahue, J., Darrell, T., Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587.
[36] Abid, S., Haris, M., Iqbal, M., Khan, N., Munir, K., Yousaf, H. (2024). Comparative analysis of machine learning pre-trained object detection models: Performance, efficiency, and application suitabilit. In 2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia, pp. 1-6. https://doi.org/10.1109/ICECCE63537.2024.10823519
[37] Archana, V., Kalaiselvi, S., Thamaraiselvi, D., Gomathi, V., Sowmiya, R. (2022). A novel object detection framework using Convolutional Neural Networks (CNN) and RetinaNet. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, pp. 1070-1074. https://doi.org/10.1109/ICACRS55517.2022.10029062
[38] Li, F., Jin, W., Fan, C., Zou, L., Chen, Q., Li, X., Jiang, H., Liu, Y. (2021). PSANet: Pyramid splitting and aggregation network for 3D object detection in point cloud. Sensors, 21(1): 136. https://doi.org/10.3390/s21010136
[39] Wulandari, N., Ardiyanto, I., Nugroho, HA. (2022). A comparison of deep learning approach for underwater object detection. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 6(2): 252-258. https://doi.org/10.29207/resti.v6i2.3931
[40] Jiang, X., Zhuang, X., Chen, J., Zhang, J., Zhang, Y. (2024). YOLOv8-MU: An improved YOLOv8 underwater detector based on a large kernel block and a multi-branch reparameterization module. Sensors, 24(9): 2905. https://doi.org/10.3390/s24092905
[41] Saleh, A., Sheaves, M., Rahimi, AM. (2022). Computer vision and deep learning for fish classification in underwater habitats: A survey. Fish and Fisheries, 23(4): 977-99. https://doi.org/10.1111/faf.12666
[42] Wang, G., Chen, Y., An, P., Hong, H., Hu, J., Huang, T. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. Sensors, 23(16): 7190. https://doi.org/10.3390/s23167190
[43] Gao, F., Wang, K., Yang, Z., Wang, Y., Zhang, Q. (2021) Underwater image enhancement based on local contrast correction and multi-scale fusion. Journal of Marine Science and Engineering, 9(2): 225. https://doi.org/10.3390/jmse9020225
[44] Zhu, J., Hu, T., Zheng, L., Zhou, N., Ge, H., Hong, Z. (2024). YOLOv8-C2f-Faster-EMA: An improved underwater trash detection model based on YOLOv8. Sensors, 24(8): 2483. https://doi.org/10.3390/s24082483
[45] Wang, N., Huang, S., Liu, X. (2024). CIFE-YOLOv8: A dense pedestrian detection model based on improved YOLOv8. In 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE), Hangzhou, China, pp. 399--403. https://doi.org/10.1109/CBASE64041.2024.10824419
[46] Song, P., Zhao, L., Li, H., Xue, X., Liu, H. (2024). RSE-YOLOv8: An algorithm for underwater biological target detection. Sensors, 24(18): 6030. https://doi.org/10.3390/s24186030
[47] Xu, W., Cui ,C., Ji ,Y., Li, X., Li, S.(2024). YOLOv8-MPEB small target detection algorithm based on UAV images. Heliyon, 10(8): e29501. https://doi.org/10.1016/j.heliyon.2024.e29501
[48] Zhang, F., Cao, W., Gao, J., Liu, S., Li, C., Song, K., Wang, H. (2024). Underwater object detection algorithm

based on an improved YOLOv8. Journal of Marine Science and Engineering, 12(11): 1991. https://doi.org/10.3390/jmse12111991
[49] Van Leeuwen, M.C., Fokkinga, E.P., Huizinga, W., Baan, J., Heslinga, F.G. (2024). Toward versatile small object detection with Temporal-YOLOv8. Sensors, 24(22): 7387. https://doi.org/10.3390/s24227387
[50] Li, M., Chen, Y., Zhang, T., Huang, W. (2024). Lightweight YOLO-based object detection with attention mechanisms for small object detection. Complex and Intelligent Systems, 10: 5459-5473 https://doi.org/10.1007/s40747-024-01448-6
[51] Cao, S., Zhao, D., Sun, Y., Ruan, C. (2021). Learning-based low-illumination image enhancer for underwater live crab detection. ICES Journal of Marine Science, 78(3): 979-993. https://doi.org/10.1093/icesjms/fsaa250
[52] Liu, S., Qi, L., Qin, H., Shi, J., Jia, J. (2018). Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759-8768.