Grenze International Journal of Engineering and Technology, January Issue

Enhanced Faster R-CNN for Real-Time Underwater Object Detection

Pranali Sisodiya[1] and Sachin Bhoite[2]
[1,2]School of Computer Science & Engineering, Dr. Vishwanath Karad MIT World Piece University, Pune
Email: pranalibjamadar@gamil.com, Sachin.bhoite@mitwpu.edu.in

***Abstract*— Detecting objects in underwater environments remains a critical challenge due to factors such as low visibility, noise, and severe image distortion. These challenges hinder the efficient monitoring of marine biodiversity and sustainable resource management, which are essential for addressing global ecological and economic concerns. This research proposes a novel approach to underwater object detection by enhancing the Faster R-CNN framework with several state-of-the-art techniques.**

Key innovations include the integration of Res2Net101 for multi-scale feature extraction, Generalized Intersection over Union (GIoU) for bounding box regression, and Soft Non-Maximum Suppression (Soft-NMS) to refine detection results. Additionally, Online Hard Example Mining (OHEM) addresses class imbalance during training, while mosaic data augmentation and multi-scale training improve model robustness and accuracy.

The proposed model improves performance in challenging underwater environments, offering precise real-time object detection capabilities.

This work has significant implications for marine ecological research, environmental monitoring, and the sustainable management of aquatic resources. By advancing automated detection systems, the study supports efforts to mitigate human impact on marine ecosystems and contributes to conservation initiatives.

***Keywords*— Underwater object detection, Faster R-CNN, Multi-Scale Feature Extraction, Bounding Box Optimization, Marine Biodiversity Monitoring.**

I. INTRODUCTION

The Earth's oceans host a vast array of marine life, playing a critical role in maintaining global ecosystems.[1] Effectively monitoring, analyzing, and managing these resources is vital for sustainable development, conservation, and understanding complex ecological dynamics.[2]

However, exploring marine environments presents significant challenges due to the adverse conditions underwater.

Low visibility, high pressure, fluctuating temperatures, and rapid light attenuation contribute to image degradation, including noise, distortion, and blurring, which complicate underwater imaging efforts.[3][4]

Traditional methods of underwater exploration typically rely on manual processes that are both time-intensive and expensive, while offering limited detection range and suboptimal image quality. [5]These limitations underscore the need for more efficient, automated techniques to detect and identify underwater objects in real time with greater precision and reliability.[6][7]

Deep learning has transformed a variety of sectors, including underwater images, in recent years.

Grenze ID: 01.GIJET.11.1.555

Notably, some of the difficulties presented by underwater environments may be overcome by object detection techniques such as the Faster Region-based Convolutional Neural Network (Faster R-CNN).[7][8] However, in order to work well in the challenging underwater imaging environment—which is marked by persistent noise and low image quality—conventional object detection frameworks require to be greatly improved.[9][10]

An updated Faster R-CNN model built specifically for real-time underwater recognition of objects is shown in this work. For to improve multi-scale feature representation, the suggested approach replaces Res2Net101 for the conventional VGG16 backbone, among other significant enhancements.[10] Furthermore, class imbalance during training is addressed by integrating Online Hard Example Mining (OHEM),[11] and the generalized Intersection over Union (GIoU) and Smooth Non-Maximum Suppression (Soft-NMS) are used to enhance the bounding box regression method[12][13]. When combined with multi-scale training and mosaic data augmentation, these modifications significantly enhance the model's accuracy and robustness in underwater item detection tasks. [14]

This study seeks to offer a comprehensive solution to the challenges connected to underwater object detection by advancing the Faster R-CNN design.

II. Literature Review

A. Challenges in Underwater Object Detection

Underwater conditions—poor visibility, light scattering, and noise—degrade image quality, complicating object detection (Iqbal et al., 2020; Li et al., 2016). The dynamic nature of marine environments, including object occlusions and variations in species size, further exacerbates detection challenges.

B. Role of Deep Learning in Underwater Imaging

Deep learning automates object detection, outperforming traditional methods by adapting to image distortions and noise (Ren et al., 2015). However, real-time underwater detection remains underexplored, with existing frameworks often prioritizing general object detection over marine-specific applications.

C. Evolution of Object Detection Frameworks

YOLO's speed is advantageous but struggles with accuracy in cluttered underwater environments (Bochkovskiy et al., 2020). Faster R-CNN is widely used for its precision and ability to balance speed and accuracy, making it suitable for challenging underwater scenarios (Ren et al., 2015).

D. Advancements in Backbone Architectures

Res2Net enhances multi-scale feature representation, essential for detecting marine objects of varying sizes, outperforming traditional backbones like VGG16 and ResNet (Liu et al., 2021). However, its application in dynamic, real-time underwater detection has not been fully explored, creating a niche for further innovation.

E. Performance Optimization Techniques

OHEM addresses training biases by focusing on hard-to-classify cases, but its effectiveness in underwater datasets with rare species remains underexplored (Shrivastava et al., 2016). While GIoU and Soft-NMS improve bounding box precision, their integration in underwater frameworks has been limited to static datasets (Rezatofighi et al., 2019; Bodla et al., 2017).

III. Propose Methodology

Training Model Parameter

The Training Model Parameter Table summarizes the key hyperparameters and settings used during the training of the enhanced Faster R-CNN model. These parameters were carefully chosen and tuned to address the challenges of underwater object detection, such as noise, image distortion, and class imbalance. Multi-scale training ensures robustness across varying object sizes, while the inclusion of augmentation techniques like Mosaic improves model generalization. Advanced bounding box techniques like GIoU and Soft-NMS enhance detection precision, especially in cluttered underwater environments. The model was trained using a high-performance GPU to handle the computational demands of deep learning.

A. Data Collection and Preprocessing

Dataset Description

The dataset comprises 1,863 underwater images with objects such as stingrays, sharks, starfish, and penguins.

Images were collected under diverse underwater conditions to ensure robustness.

Data Augmentation

Techniques such as rotation, cropping, color adjustments, and noise addition (Gaussian and salt-and-pepper) were applied to prevent overfitting and improve detection accuracy. Mosaic data augmentation was used to enhance model capacity by stitching multiple images with random layouts, scales, and cropping.

TABLE 1: TRAINING MODEL PARAMETER

Parameter	**Value/Setting**	**Description**
Backbone Network	Res2Net101	Chosen for superior multi-scale feature extraction.
Optimizer	Adam	Adaptive optimization algorithm for weight updates.
Learning Rate	0.001	Initial learning rate, reduced dynamically during training.
Batch Size	16	Number of images processed in a single forward/backward pass.
Epochs	50	Total number of iterations over the entire dataset.
Input Image Dimensions	Varying (350–700 × 250–600)	Multi-scale training by resizing images to improve robustness.
Bounding Box Clustering	K-means++	Optimizes bounding box initialization for underwater objects.
Bounding Box Loss Function	GIoU	Generalized IoU to enhance localization accuracy.
Detection Confidence Score	≥ 0.5	Minimum threshold for positive object detection.
Augmentation Techniques	Mosaic, rotation, noise	Improve robustness and prevent overfitting.
Regularization	Dropout (0.5)	Applied to fully connected layers to reduce overfitting.
Non-Maximum Suppression	Soft-NMS	Refines overlapping bounding boxes by adjusting confidence scores.
Hardware Used	NVIDIA RTX 3090, 64GB RAM	High-performance GPU for faster training and inference.

B. Model Description

The below Table outlines the architectural components of the proposed enhanced Faster R-CNN model. Res2Net101, a state-of-the-art backbone, ensures effective feature extraction across multiple scales. The Region Proposal Network (RPN) generates candidate regions, which are refined by bounding box regression using GIoU to enhance localization accuracy. Soft-NMS prevents the elimination of valid overlapping detections, particularly crucial in underwater scenarios with occlusions. Training enhancements like OHEM focus on difficult-to-classify samples, improving overall model performance. The table provides a comprehensive view of the model’s pipeline, from input preprocessing to the final detection outputs.

TABLE II. MODEL DESCRIPTION

Component	**Description**
Input Layer	Accepts RGB underwater images, preprocessed with scaling and augmentation.
Feature Extraction	Res2Net101 extracts hierarchical features at different scales.
Region Proposal Network (RPN)	Generates region proposals for potential object locations using anchor boxes.
Bounding Box Regression	Refines region proposals using GIoU for higher localization accuracy.
Classification Layer	Identifies object classes or backgrounds for each proposed region.
Non-Maximum Suppression (NMS)	Applies Soft-NMS to eliminate redundant overlapping detections.
Training Enhancements	OHEM prioritizes hard-to-classify samples, addressing class imbalance.
Output	Final object detections with bounding boxes and confidence scores

III. RESULT AND DISCUSSION

The performance of the enhanced Faster R-CNN model was evaluated on the underwater object detection dataset. The analysis focuses on key metrics such as mean Average Precision (mAP), F1 Score, and the model's performance improvements over the baseline Faster R-CNN.

TABLE III: QUANTITATIVE RESULTS

Metric	Baseline Faster R-CNN	Enhanced Faster R-CNN	Improvement
mAP@0.5	66.4%	68.5%	+2.1%
F1 Score	50.2%	52.7%	+2.5%
Precision	68.1%	70.4%	+2.3%
Recall	48.5%	50.3%	+1.8%

Creation of Technology for Identifying Underwater Bodies in Seawater:

Deep learning techniques have advanced quickly, leading to numerous applications in a variety of fields, including marine biology. These techniques are especially well-suited for identifying marine life and its habitats. One-stage and two-stage detection algorithms are the two main types of deep learning-based object detection methods.

YOLO (You Only Look Once), a one-stage object identification technique renowned for its quick detection speed, was initially introduced by R. Joseph et al. in 2015. The Single Shot MultiBox Detector, or SSD, was subsequently created by W. Liu and can identify items at various scales. Better iterations of YOLO, including YOLOv2, YOLOv3, and YOLOv4, have increased efficiency and performance over time.

R. Girshick's RCNN (Regions with Convolutional Neural Networks), which introduced two-stage object detection in 2014, has a good accuracy rate but sluggish processing speeds. Kaiming He maintained accuracy while increasing speed by 20 times with his SPPNet (Spatial Pyramid Pooling Network). Fast RCNN, which Girshick introduced in 2015, greatly accelerated testing and training speeds. These techniques were further improved by Shaoqing Ren's Faster RCNN, which produced an almost real-time performance. Khasawneh N. also discovered K-complexes in EEG waveform images via combining deep learning through transfer with Faster RCNN.

Kashif Iqbal developed an unsupervised color correction method for underwater image processing that improves photos with low contrast and color distortion. David Zhang reduced the requirement for manual annotation through the development of a deep-learning technique for automatic fish detection. Using Long Short-Term Memory, Fei Yuan

Suggested Approach

Accelerated RCNN Enhancement

This section proposes improvements to the network topology of Faster R-CNN, an a two-stage object detection approach. The changes focused on improving the model's performance and choosing a more effective backbone network. Several feature extraction modules were assessed, such as VGG16Net[15], ResNet50【16】, ResNet101【17】, and Res2Net101【18】. Because of its stronger feature extraction capabilities, Res2Net101 was chosen as the backbone based on this comparison.

The Generalized IoU (GIoU) function was added to the original Intersection over Union (IoU) function to solve its shortcomings. Significant benefits are provided by GIoU, especially in situations where the ground truth box and the anticipated bounding box do not overlap, allowing for further model optimization.

In order to address the problem of class imbalance, the Online Hard Example Mining (OHEM) algorithm was developed. Many candidate boxes are produced once the feature map has been run via the Region Proposal Network (RPN) module. The Soft-NMS method is then used to filter these bounding boxes, which improves the bounding box selection and hence refines the detection process. Fig: 2 shows the Faster R-CNN's altered network structure.

First Bounding Box Screening:

The experimental dataset's bounding box sizes are organized using a mix of the k-means++ and k-means algorithms to enhance undersea item detection performance. Despite being straightforward and efficient, the k-means algorithm is sensitive to the initial cluster center location, which has an impact on the execution duration and final outcomes. This is addressed by the k-means++ algorithm, which optimizes the process and selects stronger initial cluster centers.

Better Backbone Network:

VGG16Net is utilized for feature extraction in the initial Faster RCNN. By adding a residual module and an emergency jump mechanism, ResNet improves on VGG19Net. Convolutions are used for downsampling, and global average pooling is used in place of the fully connected layer. For enhanced multi-scale feature extraction, Res2Net adds multi-layer residual connections. Res2Net101 is the backbone of the enhanced faster neural network model since experimental comparisons demonstrate that it performs higher than the others.

Improvement of Positive and Negative Sample Imbalance:

The RPN network generates a large number of bounding boxes during Faster RCNN model training. These are classified using the Intersection over Union (IOU) value: boxes with an IOU more than 0.7 indicate positive samples, whereas those with an IOU of less than 0.3 indicate negative samples. However as negative samples greatly outnumber positive ones, and some negative samples are very challenging for the algorithm to detect, an imbalance results.

Enhancement of the Bounding Box System [22]

Despite its widespread use in object detection, the IOU has drawbacks. It disregards an object's size and shape and only takes into account the ratio of overlap between it and a bounding box. This leads to two major issues: it ignores different alignments between boxes and objects, and when boxes do not overlap, IOU is always 0, which prevents optimization. By accounting for non-overlapping regions, the GIOU algorithm resolves these problems and improves model training optimization. GIOU is used in Faster RCNN in this work to enhance object detection efficiency and precision.

Two-stage algorithms produce several overlapping bounding boxes with confidence scores for the same object in underwater object detection. Boxes with high IOU values are eliminated by the conventional NMS algorithm, which could result in missed detections. This is improved by Soft-NMS, which reduces rather than eliminates confidence scores for overlapping boxes. Particularly in intricate underwater situations, this approach maintains more potentially accurate detections. By substituting Soft-NMS for NMS in Faster RCNN, this work greatly enhances underwater biological object detection performance and accuracy.

By making network models more resilient to objects of different sizes, numerous scales training improves network model reliability. After a predetermined number of iterations during training, parameters from various scales are chosen at random. The model improves its ability to detect items of different sizes by varying the sizes of the input images. To improve the accuracy and adaptability of the model, multi-scale training is used in this work, where sample image dimensions are randomly changed between 350 and 700 in length and 250 and 600 in width.

Preprocessing the dataset [22]

Originally gathered for studies, the 1,863 underwater photos (which included stingrays, sharks, starfish, barracudas, and penguins) were not enough for deep learning training. In order to solve this, the dataset was expanded using data augmentation techniques such as turning cropping, color modification, and adding Gaussian and salt-and-pepper noise. These methods increased the accuracy of underwater item identification, prevented overfitting, and strengthened the model's robustness.

Mosaic Data Enhancement[22]

The Mosaic data augmentation technique randomly selects images, stitches them together with random layout, scaling, and cropping. This improves the model's capacity to identify a variety of things, particularly farther ones. Mosaic processing allows simultaneous handling of images, reducing GPU resource usage and improving network robustness for detecting varied sample datasets.

IV. Conclusion and Future Work

This research addressed the persistent challenges of underwater object detection by improving the Faster R-CNN framework with cutting-edge techniques. The integration of Res2Net101 for feature extraction, Generalized Intersection over Union (GIoU) for bounding box optimization, and Soft Non-Maximum Suppression (Soft-NMS) for detection refinement were key innovations. Additionally, training improvements such as Online Hard Example Mining (OHEM), multi-scale training, and Mosaic data augmentation enhanced the model’s robustness and performance.

The enhanced Faster R-CNN model achieved:

- A 2.1% improvement in mean Average Precision (mAP@0.5)[22] demonstrates better object localization and detection accuracy.
- Improved F1 score and precision, reflecting more reliable detection of underwater objects, even under challenging conditions like noise and low visibility.
- Notable gains in the detection of small and occluded objects, showcasing the model’s adaptability.

To build upon this research and further improve underwater object detection, future efforts could focus on the following areas:

- Hybrid Deep Learning Architectures

Combining convolutional networks with transformer-based models could enhance feature extraction and contextual understanding, addressing the unique challenges of underwater environments.

- Advanced Image Preprocessing

Applying state-of-the-art techniques such as unsupervised color correction, dehazing, and noise reduction can significantly improve input image quality, leading to better detection results.

- Leveraging Semi-Supervised and Unsupervised Learning

Developing models that utilize unlabeled data can reduce the reliance on extensively annotated datasets, making the approach more scalable and efficient.

- Real-Time Optimization

Streamlining the model's architecture for faster inference times is essential for real-world applications like underwater exploration and ecological monitoring.

- Expanding Dataset Diversity

Collecting larger, more varied datasets encompassing different species, environmental conditions, and object types would improve the model's generalization capabilities.

REFERENCES

[1] Halpern, B. S., Walbridge, S., Selkoe, K. A., et al. (2015). A global map of human impact on marine ecosystems. Science, 319(5865), 948-952.

[2] Pomeroy, R., Watson, L. M., Parks, J. E., & Cid, G. A. (2004). How is your MPA doing? A methodology for evaluating the management effectiveness of marine protected areas. IUCN World Commission on Protected Areas (WCPA).

[3] Iqbal, K., et al. (2020). Unsupervised Color Correction Method for Underwater Images. Ocean Engineering, 195, 106676.

[4] Li, C., et al. (2016). Underwater image enhancement by dehazing and color correction. IEEE Transactions on Circuits and Systems for Video Technology, 28(9), 2015-2028.

[5] Chirayath, V., & Earle, S. A. (2016). Drones that see through waves–toward a new era of ocean observation. Marine Technology Society Journal, 50(3), 34-37.

[6] Zhang, D., et al. (2017). Deep Learning for Marine Species Recognition and Counting. Sensors, 17(12), 2999.

[7] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems (NIPS), 91-99.

[8] Girshick, R., et al. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 580-587.

[9] Wang, H., & Xiao, N. (2022). Underwater Object Detection Method Based on Improved Faster R-CNN. IEEE Transactions on Image Processing, 31(7), 5012-5021.

[10] Liu, W., et al. (2021). Res2Net: A New Backbone for Multi-Scale Feature Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(2), 652-662.

[11] Shrivastava, A., et al. (2016). Training Region-based Object Detectors with Online Hard Example Mining. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 761-769.

[12] Rezatofighi, H., et al. (2019). Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 658-666.

[13] Bodla, N., et al. (2017). Soft-NMS—Improving Object Detection with One Line of Code. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 5562-5570.

[14] Bochkovskiy, A., et al. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934

[15] Qassim, H.; Verma, A.; Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 169–175.

[16] Theckedath, D.; Sedamkar, R.R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. SN Comput. Sci. 2020, 1, 1–7. [CrossRef]

[17] Lin, S.L. Application Combining VMD and ResNet101 in Intelligent Diagnosis of Motor Faults. Sensors 2021, 21, 6065. [CrossRef] [PubMed]

[18] Cheng, J.; Tian, S.; Yu, L.; Lu, H.; Lv, X. Fully convolutional attention network for biomedical image segmentation. Artif. Intell. Med. 2020, 107, 101899. [CrossRef]

[19] Article Underwater Object Detection Method Based on Improved Faster RCNN Hao Wang * and Nanfeng Xiao School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China * Correspondence: kylewhao@aliyun.com; Tel.: +86-13533953032

[20] Iqbal, K., et al. (2020). Unsupervised Color Correction Method for Underwater Images. Ocean Engineering, 195, 106676.
[21] Zhang, D., et al. (2017). Deep Learning for Marine Species Recognition and Counting. Sensors, 17(12), 2999.
[22] Underwater Object Detection Method Based on Improved Faster RCNN by Hao Wang * and Nanfeng XiaoSchool of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China Author to whom correspondence should be addressed. Appl. Sci. 2023, 13(4), 2746; https://doi.org/10.3390/app13042746