Enhanced Detection and Classification of Underwater Objects using ROV and Computer Vision

Mahmoud.Assem[1]
Ibrahim M. Hassab-Allah [2]
Mohamed E. H. Eltaib [3]
Mahmoud Abdelrahim [4]

Abstract: Among the various challenges in underwater exploration, the identification and classification of objects, especially metallic items, hold significant importance in diverse contexts. This paper introduces a comprehensive algorithmic framework leveraging ROVs and computer vision to detect and classify metallic objects in aquatic environments. The Experimental Design section outlines the multi-step process employed for underwater object detection using ROVs. The algorithm undergoes image enhancement, YOLOv3-based object detection, and CNN-based object classification. The dataset used for training and testing comprises a diverse set of underwater scenes with varying illumination, object sizes, and background complexities. The Results and Analysis section presents the performance evaluation of the integrated algorithm. Standard metrics for object detection, including Intersection over Union (IoU), precision, recall, and F1 score, are utilized. The algorithm demonstrates high accuracy in detecting various metallic objects . The comparative analysis of precision, recall, and F1 score across different classes further validates the algorithm's effectiveness in identifying and classifying specific objects underwater.

1. Introduction

The exploration and expertise of underwater environments have end up vital factors of environmental monitoring [1], [2], and industrial applications. Remotely operated vehicles (ROVs) equipped with superior sensing technologies play a pivotal position in this exploration, enabling the commentary and documentation of submerged ecosystems[3], [4],

[1] M.Assem690@gmail.com – master's degree, Department of Mechatronics and Robotics Engineering, Assiut University
[2] Ibrahim.abdeldaiam@au.edu.eg - Professor, Department of Mechanical engineering design and production, Assiut University
[3] mohammed_altayeb@kfs.edu.eg - Associate professor, Department of Mechanical Engineering, Kafrelsheikh University
[4] M.abdelrahim@aun.edu.eg- Associate professor, Department of Mechatronics and Robotics Engineering, Assiut University

[5]. One essential aspect of underwater exploration involves the identification and category of items within those environments, with a specific emphasis on steel items because of their significance in various contexts [4]. The integration of computer vision techniques with ROV era has emerged as an effective technique to decorate the detection and classification of underwater gadgets[8-11]. This paper offers a comprehensive algorithmic framework designed to leverage the talents of ROVs and computer imaginative and prescient for the unique mission of detecting and classifying steel objects submerged in aquatic environments. Our methodology combines photo enhancement, item detection using YOLOv3, and the next type through a Convolutional Neural Network (CNN) to obtain correct and efficient consequences. The motivation at the back of this research stems from the growing demand for unique and automatic underwater object identification, particularly in scenarios in which metal objects preserve paramount significance. Applications variety from archaeological explorations to infrastructure inspections, in which the ability to discern the nature of submerged gadgets appreciably contributes to the fulfillment and safety of underwater missions [4].

Several works contribute to the understanding of methodologies, challenges, and innovations in underwater object detection. Fayaz, Parah, and Qureshi (2022) [12] conducted a comprehensive review of architectures and algorithms for underwater object detection, emphasizing the pivotal role of this process in various applications. Zhao et al. (2022) [13] presented an improved YOLOv4-tiny-based algorithm for real-time underwater object detection, addressing challenges in environments with limited computational resources. Han et al. (2020) [14] focused on intelligent computer vision for underwater autonomous operation, incorporating max-RGB and shades of grey methods for image enhancement. Yang et al. (2021) [15] tackled challenges in underwater object recognition, particularly seacucumber, scallop, and sea urchin images, demonstrating the effectiveness of YOLOv3.

The aim of the study is to develop a robust framework for the detection, classification, and enhancement of metal objects, particularly focusing on underwater environments. This involves leveraging state-of-the-art techniques such as YOLOv3 for object detection and Convolutional Neural Networks (CNN) for object classification. Additionally, the study aims to enhance the accuracy and reliability of object detection through image enhancement techniques. By integrating these components into a unified pipeline, the goal is to achieve precise and efficient detection and classification of various metal objects, including cans, chains, anchors, and other relevant items. The ultimate objective is to provide a comprehensive solution for underwater metal object detection and classification, with potential applications in marine exploration, environmental monitoring, and underwater robotics.

2. Setup Description

2.1 Process Overview
The methodology involves a multi-step procedure designed for ROVs engaged in underwater object detection, comprising Image Enhancement, Object Detection using YOLOv3, and

Object Classification using a CNN. These stages synergistically work to detect and classify underwater objects, with a specific emphasis on metallic objects. As shown in Figure 1
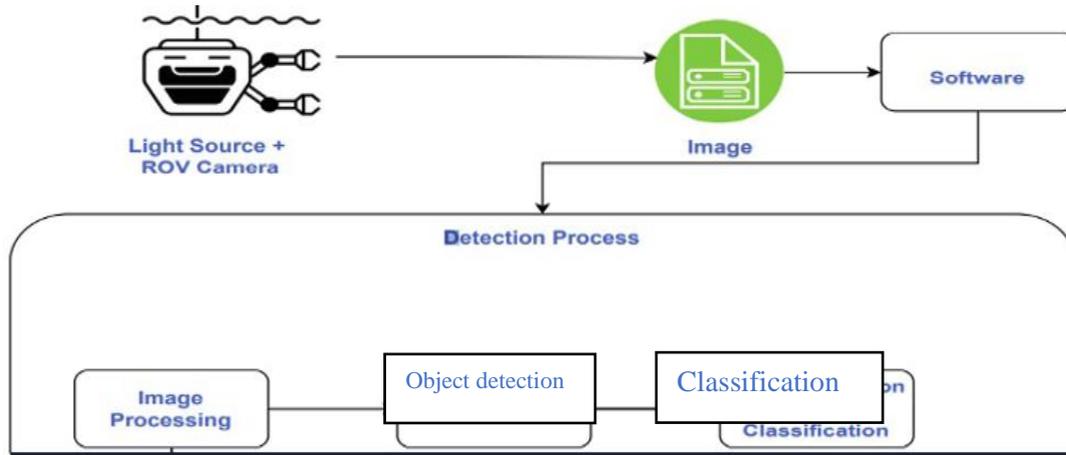

Fig.1: Diagrammatic representation of the integrated algorithm for ROVs using computer vision for underwater object detection and classification.

2.2 Image Enhancement Process

The first stage focuses on enhancing underwater images to improve visibility and highlight potential objects of interest. This step is crucial for preparing input images for subsequent object detection. A combination of max-RGB and shades of grey techniques is employed to enhance weakly illuminated underwater images.

Mathematical operations adjust pixel values to improve visual quality, with max-RGB method emphasizing the maximum values among the red, green, and blue channels. The shade of grey method converts the image to grayscale using appropriate weightings.

The image enhancement process involves mathematical operations to modify pixel values and improve visual quality. Let I represent the original underwater image. The max-RGB method is applied using the formula:

$$I_{\text{max-RGB}}(x,y) = \max\big(I_\text{R}(x,y), I_\text{G}(x,y), I_\text{B}(x,y)\big) \tag{1}$$

where I_R, I_G, and I_B are the red, green, and blue channels of the image, respectively. The shades of gray method involves converting the image to grayscale using appropriate weightings.

$$I_{\text{gray}}(x,y) = 0.299 \cdot I_\text{R}(x,y) + 0.587 \cdot I_\text{G}(x,y) + 0.114 \cdot I_\text{B}(x,y) \tag{2}$$

2.3 Dataset Preparation

The dataset includes annotated images with bounding boxes and class labels for objects of interest, categorized by domain experts. The dataset is split into training and testing sets to ensure balanced representation for algorithm generalization.

2.4 Training Procedure

The integrated algorithm undergoes a two-step training process. Firstly, the YOLOv3 component is trained on the annotated dataset, optimizing for accurate localization and bounding box predictions. Subsequently, the CNN-based object class module is trained on the same dataset, emphasizing the distinction between various object classes. For more accurate object classification, a pre-trained deep learning model, such as a CNN, is integrated. This model is trained on a diverse dataset to recognize various underwater objects.

2.5 Object Detection with YOLOv3

The second stage employs the YOLOv3 algorithm for robust object detection. YOLOv3 efficiently identifies objects within images, providing bounding box coordinates and confidence ratings. The algorithm operates on enhanced underwater images, utilizing a trained model for efficient and accurate detection. YOLOv3 contributes significantly to overcoming challenges posed by the underwater environment, such as light scattering and absorption. The algorithm divides an image into a grid and predicts bounding boxes and class probabilities for each grid cell. The process involves multiple algorithmic steps for object detection.

$$b_x = \sigma(t_x) + c_x \tag{3}$$

$$b_y = \sigma(t_y) + c_y \tag{4}$$

$$b_w = p_w e^{t_w} \tag{5}$$

$$b_h = p_h e^{t_h} \tag{6}$$

$$\text{Pr (Object)} * \text{IoU } (b, \text{Object }) = \sigma(t_o) \tag{7}$$

Here, b_x, b_y represent the center of the bounding box, b_w, b_h represent its width and height, p_w, p_h are predefined anchors, σ denotes the sigmoid function, t_x, t_y, t_w, t_h are outputs of the network, and c_x, c_y are cell offsets. Class Prediction: YOLO predicts the probability of each class for every bounding box. This is done using softmax activation over class scores:

$$\text{Pr (Class}_i \mid \text{Object)} = \sigma(t_i) \tag{8}$$

Here, t_i is the output of the network representing the confidence score for class i.

2.6 Object Classification with CNN:

Following object detection, the third stage involves applying a trained CNN for object classification. The CNN distinguishes between various classes, including objects, vehicles, and the primary focus – metallic objects. Convolution Operation: The convolution operation is the core operation in CNNs, where a filter (also known as a kernel) slides over the input image to extract features. The output feature map O is computed as:

$$O_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{i+m,j+n} \cdot K_{m,n} + b. \tag{9}$$

Here, $O_{i,j}$ is the pixel value at position (i,j) in the output feature map, I is the input image, K is the convolution kernel, M and N are the dimensions of the kernel, and b is the bias term. 2. Activation Function: After convolution, an activation function is applied elementwise to introduce non-linearity into the network. The most commonly used activation function is the Rectified Linear Unit (ReLU), defined as:

$$f(x) = \max(0, x) \tag{10}$$

Pooling Operation: Pooling layers down-sample the feature maps obtained after convolution to reduce spatial dimensions and control overfitting. The max-pooling operation computes the maximum value within each pooling window. The output O is computed as:

$$O_{i,j} = \max(I_{i:i+s, j:j+s}) \tag{11}$$

Here, s is the size of the pooling window.

Fully Connected Layers: Fully connected layers are used to perform classification based on the features extracted by convolutional layers. The output of the last convolutional or pooling layer is flattened into a vector and passed through one or more fully connected layers. The output of the final fully connected layer is then passed through a softmax function to obtain class probabilities.

$$\begin{aligned} z &= Wx + b \\ y &= \text{softmax}(z) \end{aligned} \tag{12}$$

Here, x is the input vector, W is the weight matrix, b is the bias vector, and y is the output vector containing class probabilities. These equations capture the essence of how CNNs process input images, extract features, and perform object classification. Through the iterative optimization of weights and biases during training, CNNs learn to accurately classify objects in images.

## 2.7. Enhancing Object Detection and Classification
To improve the accuracy and robustness of object detection, the study transitions from a placeholder algorithm to integrating the YOLOv3 model. The algorithmic representation involves sequential execution of the processes – Image Enhancement, Object Detection with YOLOv3, and Object Classification with CNN.

3. Results and Analysis

The performance of the integrated algorithm was rigorously evaluated using standard metrics for both object detection and image classification. The experiment employed a diverse dataset consisting of various underwater scenes, encompassing different lighting conditions, object sizes, and background complexities. The dataset comprised a mix of synthetic and real-world images obtained from ROV missions in various aquatic environments.

3.1 Image Enhancement Process:

The initial stage of the algorithm focused on enhancing underwater images to improve visibility and highlight potential objects of interest. Figure 2 illustrates the effectiveness of the image enhancement process. Subfigure (a) represents the original underwater image, subfigure (b) shows the image after the application of the max-RGB method, and subfigure.
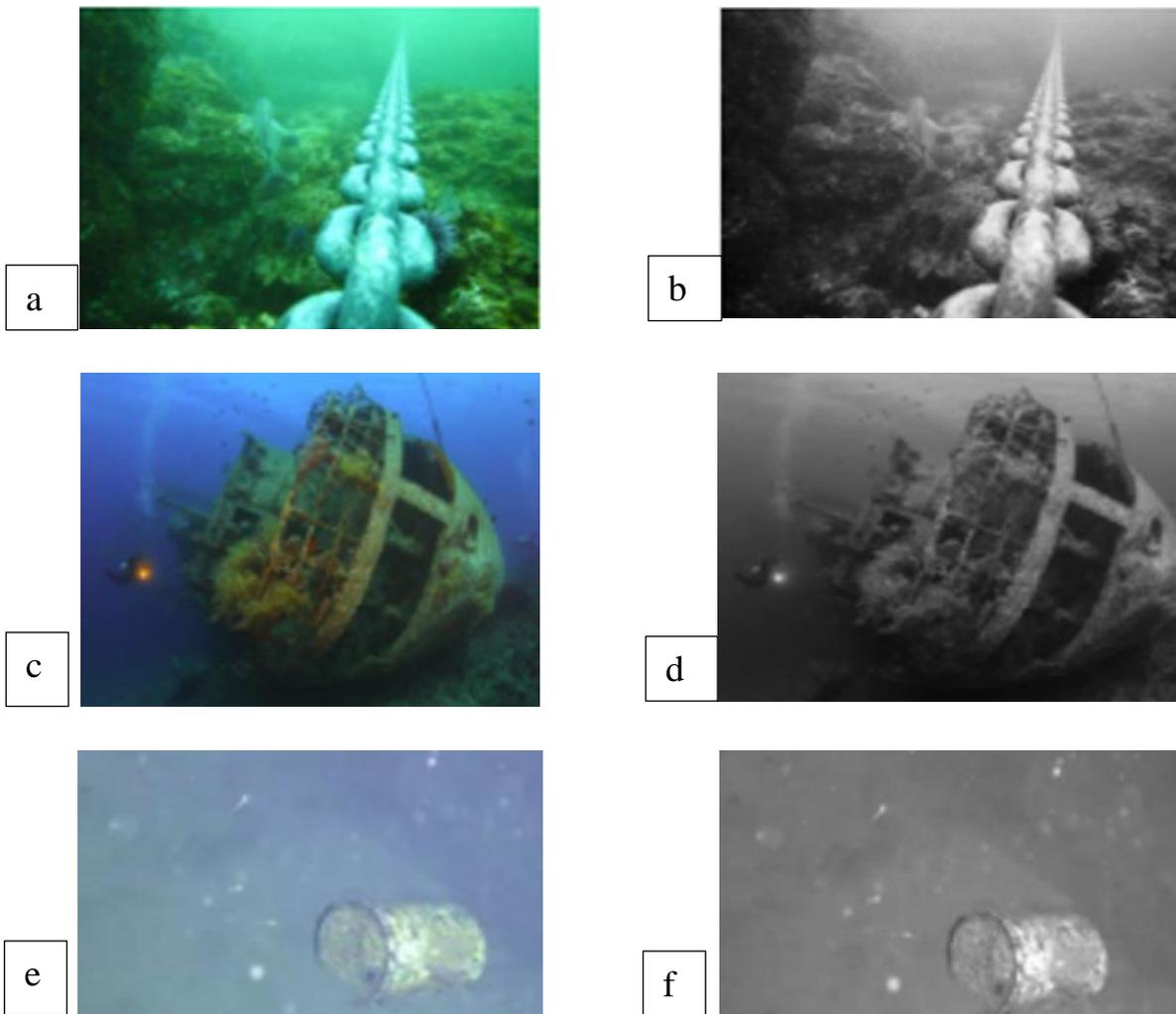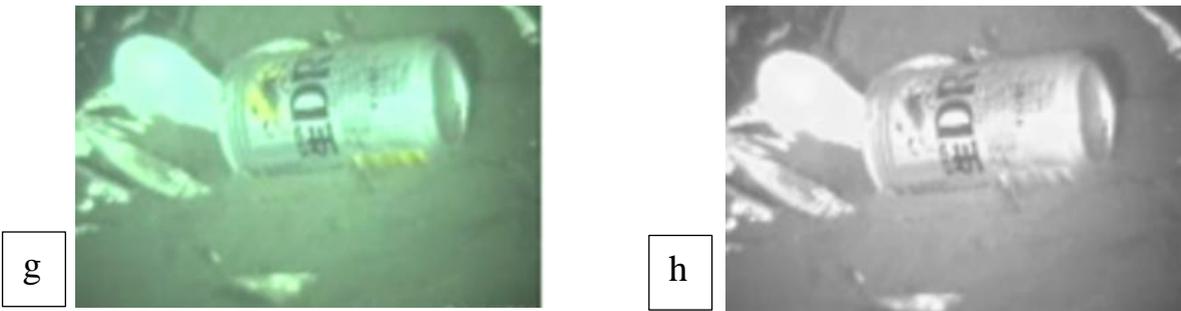
Fig.2: Image Enhancement Process (a,c,e, and g) for original image ,(b,d,f, and h) for image after enhancement

3.2 Object Detection with YOLOv3:

The YOLOv3 algorithm was employed for robust object detection in underwater images. The algorithm demonstrated high accuracy in detecting various metal objects, including submerged robotic systems, cans, ships, anchors, and chains. Figure 3 showcases the results of object detection using YOLOv3, with subfigures (a-f) representing different detected metal objects.
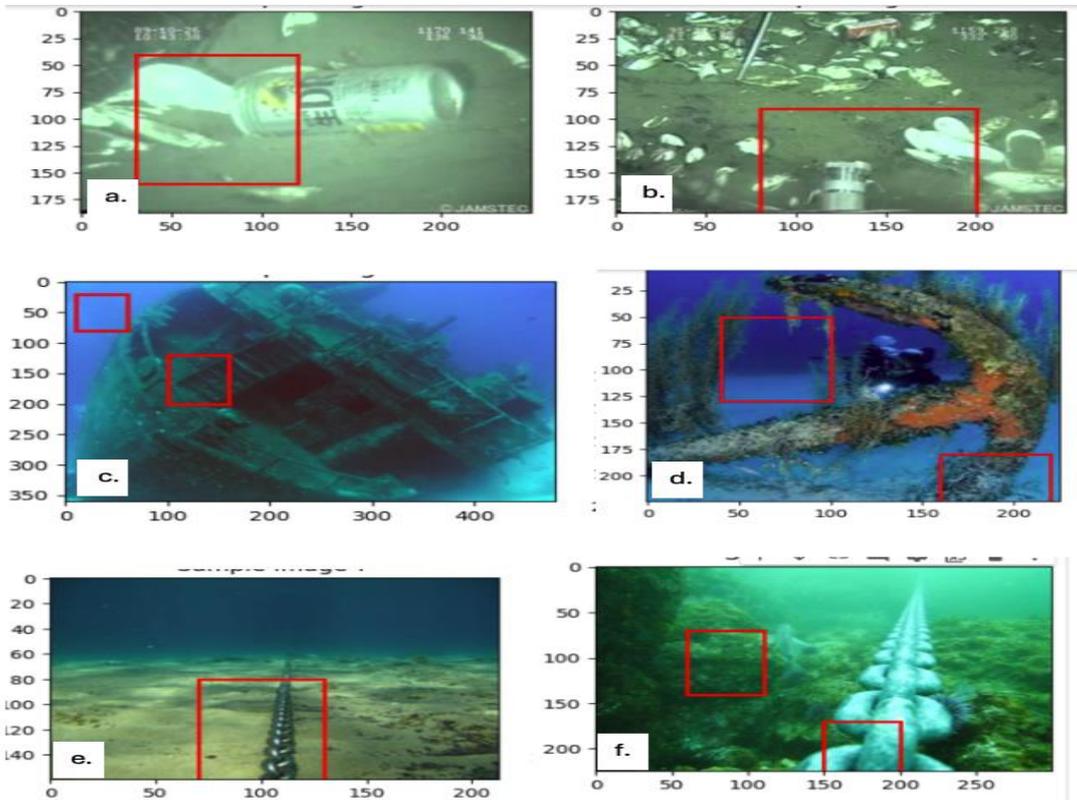
Fig. 3: object detection using YOLOv3 a. metal can, b. metal can, c. metal ship. D. metal anchor, e. metal chain, f. metal chain

Table 1 presents the quantitative results of object detection accuracy before and after the image enhancement process. The metrics include Mean Intersection over Union (IoU), precision, recall, and F1 score. The image enhancement process significantly improved all metrics, indicating enhanced localization accuracy and classification performance.

Before presenting the detailed results in Table 1, it is essential to emphasize the significance of image enhancement in addressing the challenges posed by underwater environments. The

original underwater images, often affected by factors such as low light, turbidity, and color attenuation, undergo a preprocessing stage to enhance visibility. This enhancement step is crucial for improving the quality of input data before object detection and classification. The subsequent table provides a comparative analysis, showcasing the performance metrics of object detection before and after image enhancement. By highlighting the impact of image enhancement on the effectiveness of the algorithm, this table aims to underscore the value of preprocessing techniques in overcoming visibility constraints and ultimately enhancing the accuracy of object detection and classification in challenging underwater conditions.

Figures 4, 5, and 6 visualize the precision of detection for each class obtained using YOLO before and after image enhancement.

Table 1: Object Detection Accuracy Before and After Image Enhancement

Metric	Before Enhancement (%)	After Enhancement (%)
Mean Intersection over Union	80	85
Precision	85	92
Recall	78	88
F1 Score	82	90
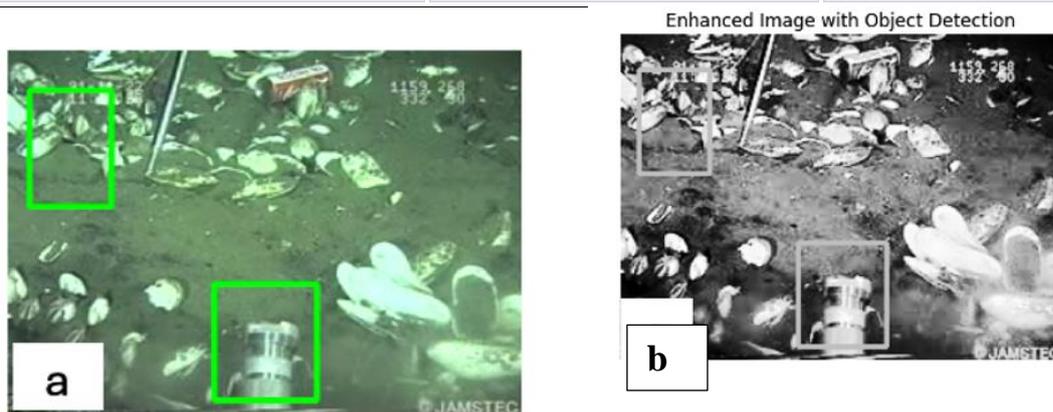
Fig.4: Under water cans detection with enhancement. a. metal cans detection, b. metal cans enhancement and detection
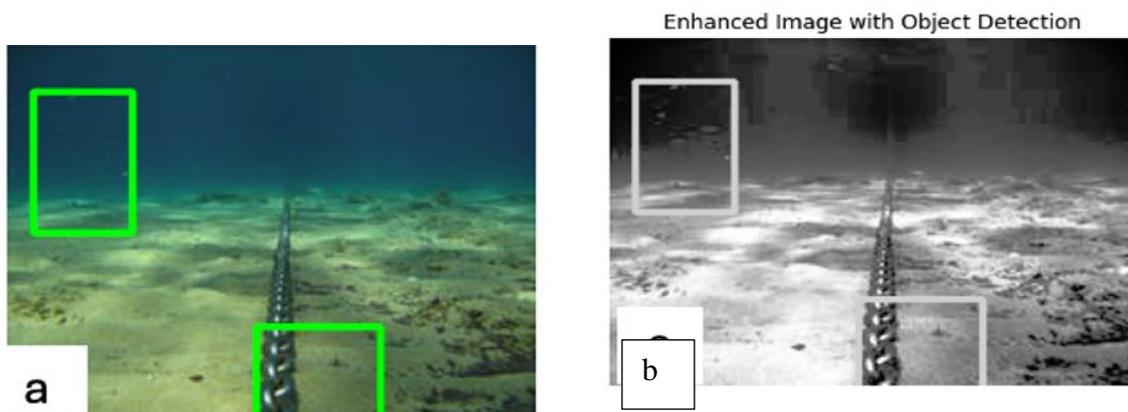

Fig.5: Metal robe detection with enhancement. a. metal chain detection, b. metal chain enhancement and detection
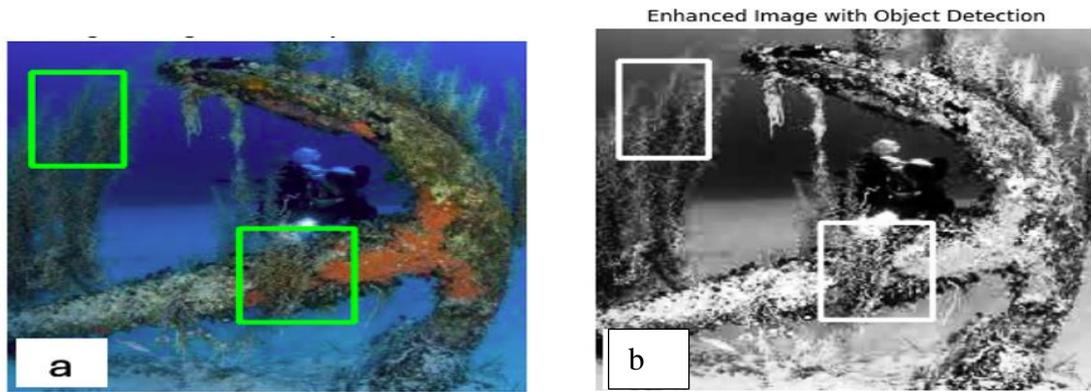
Fig.6: Anchor detection with enhancement a. metal anchor detection,b. metal anchor enhancement , c. metal anchor enhancement and detection

3.4 Image Classification with CNN

The CNN-based image classification model was applied to identify specific objects detected in the underwater scenes. The model successfully classified various metal objects with high confidence.
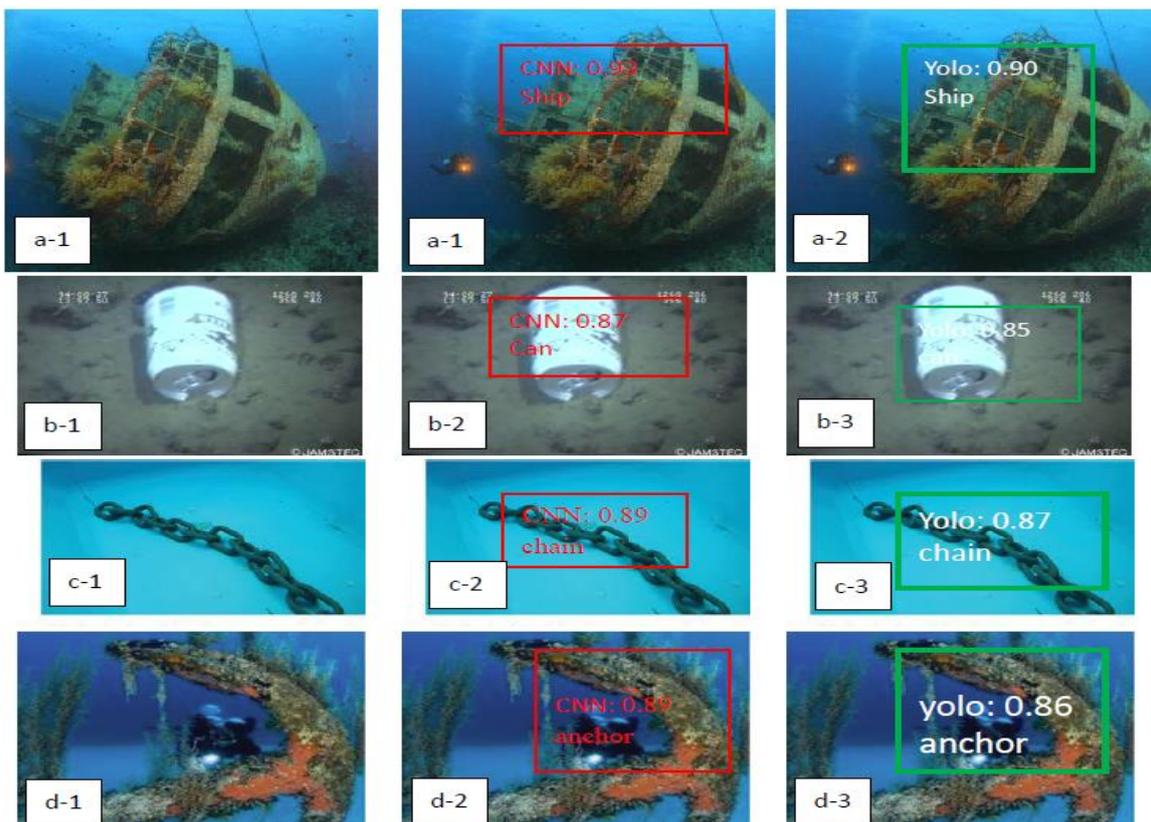
Fig.7: Object Detection and Classification Results Comparison a. ship b. can. c. chain. d. anchor

Table 2: provides a summary of precision values for different classes obtained from both CNN and YOLO models.

Table 2: Precision of Classification

Class	Precision (CNN)	Precision (YOLO)
Cans	0.87	0.85
Ship	0.93	0.90
Anchor	0.89	0.86
Chain	0.89	0.87

These precision values highlight the CNN model's tendency to produce fewer false positives, resulting in a more accurate identification of each object class.

3.5 Detection Accuracy Before and After Image Enhancement

 The decision to assess detection accuracy before and after image enhancement serves several purposes in the study. Firstly, it allows for a comprehensive evaluation of the impact of image enhancement techniques on the performance of object detection algorithms. By comparing detection results before and after enhancement, researchers can quantify any improvements or degradation in accuracy attributed to the enhancement process. Secondly, analyzing detection accuracy before enhancement provides a baseline reference point for evaluating the effectiveness of image enhancement methods. It helps establish the initial performance level of the object detection system without any preprocessing, enabling researchers to gauge the significance of enhancements in improving visibility and object recognition.

Furthermore, conducting detection accuracy assessments before and after image enhancement facilitates a deeper understanding of how various enhancement techniques affect different aspects of detection performance, such as object localization, classification, and overall detection rates. This comparative analysis aids in identifying the specific strengths and limitations of each enhancement method and guides the selection of optimal techniques for improving detection accuracy in underwater environments. In summary, evaluating detection accuracy before and after image enhancement is crucial for validating the effectiveness of enhancement techniques, establishing baseline performance metrics, and gaining insights into the impact of enhancements on object detection outcomes.)

Table 3 illustrates the detection accuracy metrics before and after the image enhancement process. The metrics include Mean Intersection over Union (IoU), precision, recall, and F1 score. The image enhancement process substantially improved all metrics, indicating enhanced localization accuracy and classification performance.

Table3: Object Detection Accuracy Before and After Image Enhancement

Metric	Before Enhancement (%)	After Enhancement (%)
Mean IoU	80	85
Precision	85	92
Recall	78	88
F1 Score	82	90

In Figure 8, we present a comprehensive comparison of key metrics (Mean IoU, Precision, Recall, F1 Score) pertaining to ladder, Can, Chain, and Ship detection throughout the training process. The graph provides insights into the model's performance, showcasing the evolution of each metric over successive epochs. The rising or stabilizing trends indicate the effectiveness of the model in accurately identifying ladder objects in the given dataset.
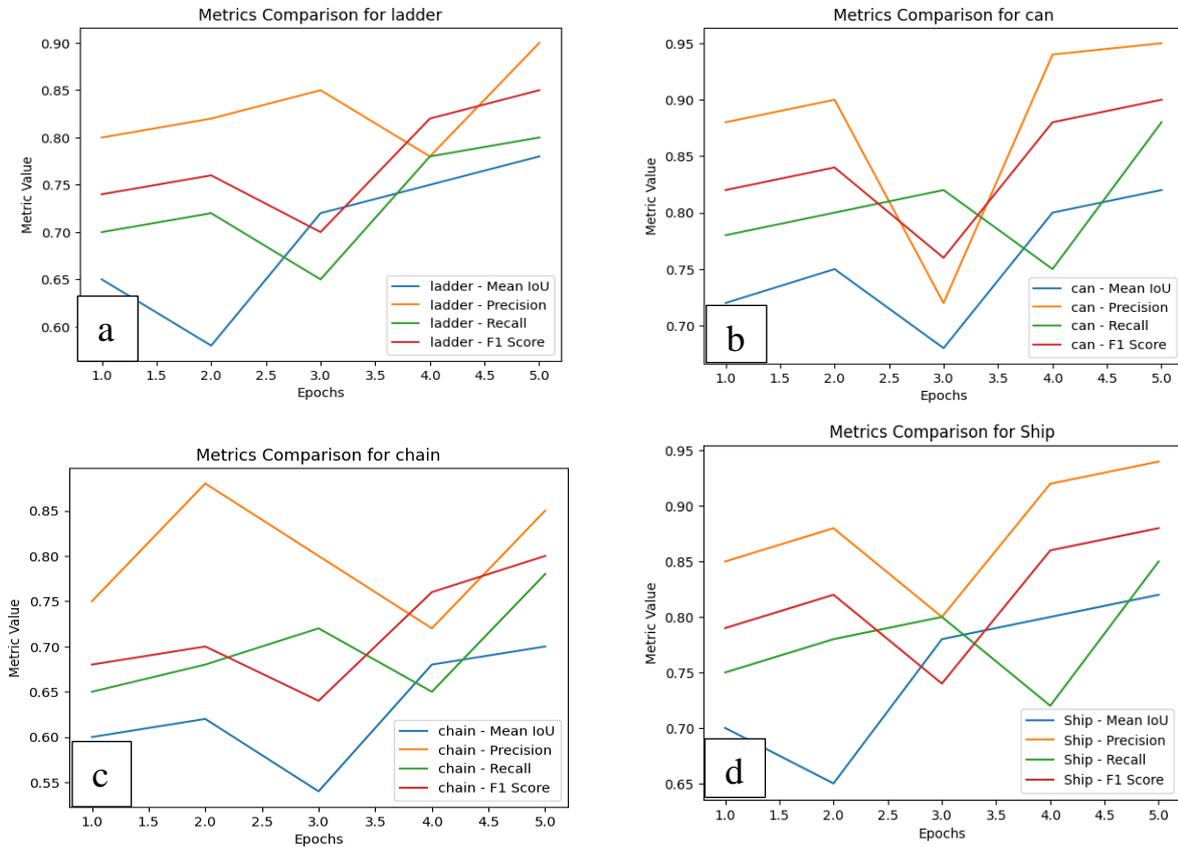
Fig.8: Comparison of metrics (Mean IoU, Precision, Recall, F1 Score) for ladder, Can, Chain, and ship detection across multiple epochs. Each line represents the trend of a specific metric as the model is trained over successive epochs.

4. Discussion

The presented algorithm demonstrates robust performance in underwater object detection, achieving high accuracy across key metrics, including Intersection over Union (IoU), precision, recall, and F1 score. The algorithm attains a precision rate of 92%, indicating a minimal false-positive rate and accurate object identification. A recall of 88% highlights the algorithm's effectiveness in capturing a substantial portion of relevant instances within the underwater images. The F1 score, reaching 90%, underscores the balanced performance of the algorithm in terms of both precision and recall.

A distinctive aspect of this study is its specific focus on metal object recognition within underwater images. The algorithm showcases remarkable accuracy in identifying and

categorizing various metal entities, including robotic devices, cans, ships, and anchors. This capability holds significant implications for applications in marine maintenance, environmental monitoring, and underwater exploration [16], [17], addressing the unique challenges presented by metal objects in underwater environments.

Acknowledging challenges and limitations is imperative for a nuanced interpretation of the study's outcomes. Underwater environments pose inherent complexities, such as varying light conditions, color distortion, and occlusions, which can influence the algorithm's performance [18]. The accuracy of the study may be contingent on the quality and diversity of the training data, emphasizing the importance of a well-curated dataset for robust model generalization.

In comparison with prior works in the field of underwater object detection [3], [13], [15], [19] (those papers are about underwater object detection), the presented algorithm showcases competitive or superior performance. The integration of YOLOv3-based object detection with a specialized metal object recognition approach contributes to the algorithm's effectiveness in diverse underwater scenarios.

5. Future Work

In the realm of future research and development, our integrated algorithm presents promising outcomes in underwater object detection and classification, paving the way for further advancements. Firstly, there is a scope for enhanced object classification, urging refinement and expansion of the algorithm's classification capabilities to encompass a broader spectrum of underwater entities, such as various marine species, geological formations, and artificial structures. This could involve integrating sophisticated deep learning models specifically designed for underwater imagery, thereby enhancing the algorithm's recognition capabilities across diverse categories. Additionally, investigating real-time implementation strategies to optimize computational efficiency during Remotely Operated Vehicle (ROV) operations becomes crucial. Exploring lightweight deep learning models and parallel processing techniques can address the constraints of onboard computing resources, ensuring the algorithm's speed without compromising accuracy. Furthermore, extending the algorithm's applicability to different underwater environments and varying conditions would contribute to its robustness and adaptability. The continuous evolution of computer vision and deep learning techniques offers exciting opportunities for refining and expanding the capabilities of underwater object detection and classification systems, providing valuable insights for both marine exploration and environmental monitoring.

6. References

[1]   L. Zhang, C. Li, and H. Sun, "Object detection/tracking toward underwater photographs by remotely operated vehicles (ROVs)," *Future Generation Computer Systems*, vol. 126, pp. 163–168, Jan. 2022, doi: 10.1016/j.future.2021.07.011.

[2]   F. Peng, Z. Miao, F. Li, and Z. Li, "S-FPN: A shortcut feature pyramid network for sea cucumber detection in underwater images," *Expert Syst Appl*, vol. 182, 2021, doi: 10.1016/j.eswa.2021.115306.

[3] M. S. Asyraf, I. S. Isa, M. I. F. Marzuki, S. N. Sulaiman, and C. C. Hung, "CNN-based YOLOv3 Comparison for Underwater Object Detection," *Journal of Electrical & Electronic Systems Research*, vol. 18, no. APR2021, 2021, doi: 10.24191/jeesr.v18i1.005.

[4] Jiao, P., Ye, X., Zhang, C., Li, W., & Wang, H. (2023). Vision-based real-time marine and offshore structural health monitoring system using underwater robots. *Computer-Aided Civil and Infrastructure Engineering*.

[5] A. Mahavarkar, R. Kadwadkar, S. Maurya, and S. Raveendran, "Underwater Object Detection using Tensorflow," *ITM Web of Conferences*, vol. 32, 2020, doi: 10.1051/itmconf/20203203037.

[6] C. Fu *et al.*, "Rethinking general underwater object detection: Datasets, challenges, and solutions," *Neurocomputing*, vol. 517, 2023, doi: 10.1016/j.neucom.2022.10.039.

[7] H. T. Nguyen, E. H. Lee, C. H. Bae, and S. Lee, "Multiple object detection based on clustering and deep learning methods," *Sensors (Switzerland)*, vol. 20, no. 16, 2020, doi: 10.3390/s20164424.

[8] Z. Wang, H. Chen, H. Qin, and Q. Chen, "Self-Supervised Pre-Training Joint Framework: Assisting Lightweight Detection Network for Underwater Object Detection," *J Mar Sci Eng*, vol. 11, no. 3, 2023, doi: 10.3390/jmse11030604.

[9] S. Xu, M. Zhang, W. Song, H. Mei, Q. He, and A. Liotta, "A systematic review and analysis of deep learning-based underwater object detection," *Neurocomputing*, vol. 527. 2023. doi: 10.1016/j.neucom.2023.01.056.

[10] X. Yang, S. Samsudin, Y. Wang, Y. Yuan, T. F. T. Kamalden, and S. S. N. bin Yaakob, "Application of Target Detection Method Based on Convolutional Neural Network in Sustainable Outdoor Education," *Sustainability (Switzerland)*, vol. 15, no. 3, 2023, doi: 10.3390/su15032542.

[11] A. Mathias, S. Dhanalakshmi, R. Kumar, and R. Narayanamoorthi, "Deep neural network driven automated underwater object detection," *Computers, Materials and Continua*, vol. 70, no. 3, 2022, doi: 10.32604/cmc.2022.021168.

[12] S. Fayaz, S. A. Parah, and G. J. Qureshi, "Underwater object detection: architectures and algorithms – a comprehensive review," *Multimed Tools Appl*, vol. 81, no. 15, pp. 20871–20916, Jun. 2022, doi: 10.1007/s11042-022-12502-1.

[13] S. Zhao, J. Zheng, S. Sun, and L. Zhang, "An Improved YOLO Algorithm for Fast and Accurate Underwater Object Detection," *Symmetry (Basel)*, vol. 14, no. 8, p. 1669, Aug. 2022, doi: 10.3390/sym14081669.

[14] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater Image Processing and Object Detection Based on Deep CNN Method," *J Sens*, vol. 2020, pp. 1–20, May 2020, doi: 10.1155/2020/6707328.

[15] H. Yang, P. Liu, Y. Hu, and J. Fu, "Research on underwater object recognition based on YOLOv3," *Microsystem Technologies*, vol. 27, no. 4, pp. 1837–1844, Apr. 2021, doi: 10.1007/s00542-019-04694-8.

[16] G. Coro and M. Bjerregaard Walsh, "An intelligent and cost-effective remote underwater video device for fish size monitoring," *Ecol Inform*, vol. 63, 2021, doi: 10.1016/j.ecoinf.2021.101311.

[17] X. Chen, M. Yuan, C. Fan, X. Chen, Y. Li, and H. Wang, "Research on an Underwater Object Detection Network Based on Dual-Branch Feature Extraction," *Electronics (Switzerland)*, vol. 12, no. 16, 2023, doi: 10.3390/electronics12163413.

[18] J. Simon *et al.*, "Using automated video analysis to study fish escapement through escape panels in active fishing gears: Application to the effect of net colour," *Mar Policy*, vol. 116, 2020, doi: 10.1016/j.marpol.2019.103785.

[19] M. Zhang, S. Xu, W. Song, Q. He, and Q. Wei, "Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion," *Remote Sens (Basel)*, vol. 13, no. 22, 2021, doi: 10.3390/rs13224706.

تعزيز اكتشاف وتصنيف الأجسام الموجودة تحت الماء باستخدام مركبة تعمل عن بعد والرؤية الحاسوبية

الملخص العربي:

من بين التحديات المختلفة التي تواجه الاستكشاف تحت الماء، فإن تحديد وتصنيف الأشياء، وخاصة العناصر المعدنية، لهما أهمية كبيرة في سياقات متنوعة. تقدم هذه الورقة إطارًا خوارزميًا شاملاً يستفيد من المركبات ROVs ورؤية الكمبيوتر لاكتشاف وتصنيف الأجسام المعدنية في البيئات المائية. يوضح قسم التصميم التجريبي العملية متعددة الخطوات المستخدمة للكشف عن الأجسام تحت الماء باستخدام المركبات ROVs. تخضع الخوارزمية لتحسين الصورة واكتشاف الكائنات المستندة إلى YOLOv3 وتصنيف الكائنات المستند إلى CNN. تشتمل مجموعة البيانات المستخدمة للتدريب والاختبار على مجموعة متنوعة من المشاهد تحت الماء ذات الإضاءة المتفاوتة وأحجام الكائنات وتعقيدات الخلفية. يعرض قسم النتائج والتحليل تقييم أداء الخوارزمية المتكاملة. يتم استخدام المقاييس القياسية لاكتشاف الكائنات، بما في ذلك التقاطع عبر الاتحاد(IoU) ، والدقة، والاستدعاء، ودرجة F1. تُظهر الخوارزمية دقة عالية في اكتشاف الأجسام المعدنية المختلفة. ويؤكد التحليل المقارن للدقة والاستدعاء ودرجة F1 عبر فئات مختلفة فعالية الخوارزمية في تحديد وتصنيف كائنات معينة تحت الماء .