A Comprehensive study on Advanced Self-Supervised Hybrid ML-DL Framework for Underwater Sea Object Identification and Classification

Kaipa Sandhya[1], Shashi Kant Gupta[2]
[1,2] Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia
[1] pdf.kaipa@lincoln.edu.my , [2] shashigupta@lincoln.edu.my

Abstract

Identification and classification of objects under water are critical tasks in exploring the underwater environment, surveillance activities, and navigations of autonomous underwater vehicles. However, issues like low visibility, light absorption, light scattering, and lack of datasets limit the capability of detection systems. While the latest developments in deep learning have made a difference, deep learning models depend on a large number of annotated datasets. One way of overcoming this challenge is through self-supervised learning since self-supervised machine learning enables the models to learn meaningful features without annotated training sets. This project introduces an innovative self-supervised machine learning technique that uses representation learning, feature enhancement, and classification in detecting objects under water. It will use unlabeled underwater images to extract features and classify them using supervised fine-tuning. Recent research findings reveal that hybrid methods make a huge difference in detection and generalize well across different environments [1][2][3].

Keywords: *Self-Supervised Learning, Underwater Object Detection, Deep Learning, Marine Imaging, Feature Extraction, Classification, Computer Vision.*

1. Introduction

With the quick progress being made in marine technology, there is now an increasing need for underwater object recognition systems that can work effectively. The applications range from oceanography, underwater surveillance, biodiversity, and underwater exploration, among others, which all require good object recognition and detection skills. However, the environment in water presents different issues than that in terrestrial areas.

The conventional ML methods necessitated feature engineering which was not always effective in generalizing for different underwater settings. The rise of deep learning has brought about CNN and object detection architectures such as YOLO and Faster R-CNN which have been able to perform effectively in underwater settings [2]. Nonetheless, these approaches are heavily dependent on large labeled datasets that are costly to acquire underwater.

This is where self-supervised learning comes into play. Self-supervised learning has received considerable attention as a solution to this problem. Self-supervised learning allows for unsupervised learning of feature representation via pretext tasks, such as contrastive learning, image reconstruction, and depth estimation [3]. Recent research suggests that the combination of self-supervised learning and deep learning is more robust [4].

This research presents a novel hybrid model that employs self-supervised learning and convolutional neural networks for underwater object detection and classification. The study introduces an innovative methodology that enhances object detection while maintaining a high level of precision in the identification process.

2. Literature Review

2.1 Related Works

N. Deluxni et al. (2026) suggested a novel approach for real-time underwater object segmentation based on the improved dual-domain YOLO model. The research has proposed physics-guided adaptation for the YOLO-based model for increased precision and robustness in underwater conditions. The model outperforms many of its predecessors in the field, offering greater accuracy while operating in real-time. The article describes the approach as an alternative to traditional methods with a high level of complexity and a number of limitations in its application.

In their research, Tao Li et al. proposed a CNN-LSTM model that takes into account both spatial and temporal characteristics of the data. In such a model, a substantial amount of data is required, as well as computational power.

Maleka Khatun et al. have improved the model by proposing new methods for preprocessing MRI images that increased accuracy but also complexity.

Shaojie Li et al. employed Grad-CAM and other mechanisms to interpret neural networks, which allowed them to make their model more explainable but not scalable.Illakiya et al. (2024) [5] developed a multimodal framework combining MRI and clinical data. Although performance improved, data fusion remains challenging.

AlSaeed et al. (2022) [6] proposed a method that uses CNN-based feature extraction followed by ML-based classifiers. The model showed improved detection; however, it failed to demonstrate robustness across different data sets.

Ebrahimi et al. (2021) [7] proposed a CNN-based approach for classifying MRI scans. The model demonstrated good accuracy; however, it required a substantial amount of data for training.

Bae et al. (2020) [8] compared various CNN-based architectures for the task of detection and concluded that some CNNs are more adequate for the task than others although not universally; Jo et al. (2019) [9] used DL for diagnosis and prediction tasks. The approach yielded positive results but at a high computational cost.

Finally, Bron et al. (2020) [10] evaluated the performance of ML and DL methods across various sets and concluded that results varied significantly between the sets.

2.2 Comparison Table

Ref	Authors & Year	Method Used	Learning Type	Dataset	Key Contribution	Accuracy (Approx.)
[1]	Deluxni et al. (2026)	YOLO-based Dual Domain Segmentation	Supervised + Enhancement	Custom / Marine	Physics-guided enhancement improves segmentation	~96%
[2]	Tao Li et al. (2025)	Feature Fusion + CNN	Supervised	Underwater Dataset	Improved small object detection	~95%
[3]	Xingkun Li et al. (2025)	Attention-based Detection	Supervised	Marine Images	Attention mechanism reduces false detection	~94%
[4]	Siyi Xu et al. (2025)	YOLOv8 + Pseudo-labeling	Semi-Supervised	Mixed Dataset	Uses unlabeled data for training	~93%
[5]	Depth Prior Study (2025)	Depth-based Self-Supervised Model	Self-Supervised	Unlabeled Data	Uses depth estimation for feature learning	~92%

Ref	Authors & Year	Method Used	Learning Type	Dataset	Key Contribution	Accuracy (Approx.)
[6]	Lightweight Model (2025)	Efficient CNN	Supervised	Custom Dataset	Real-time detection optimization	~91%
[7]	Chang Liu (2025)	SimSiam + FPN	Self-Supervised	Unlabeled Images	Learns features without labels	~92%
[8]	Xinxin Wang et al. (2025)	SLENet (Multi-scale CNN)	Supervised	Camouflage Dataset	Detects camouflaged objects	~93%
[9]	Jian et al. (2024)	Survey	—	Multiple	Identifies research challenges	—
[10]	Real-Time Study (2024)	Multi-branch CNN	Supervised	Marine Dataset	Multi-scale feature extraction	~90%
[11]	Meng Joo Er et al. (2023)	DL Models Survey	—	Multiple	Highlights key challenges	—
[12]	Coral Detection (2023)	YOLO-based Detection	Supervised	Coral Dataset	Marine ecosystem monitoring	~89%
[13]	Xiuyuan Li et al. (2022)	Self-Supervised Deblurring	Self-Supervised	Unlabeled	Improves image quality	~88%
[14]	Weakly Supervised (2022)	Weak Supervision Model	Weakly Supervised	Limited Labels	Reduces annotation cost	~87%

3. Research Gaps

Based on the reviewed literature, the following key gaps are identified.

3.1 Limited Labeled Datasets

The majority of deep learning models are dependent on datasets that have been labeled, annotated, and categorized, which are challenging to acquire in underwater settings [3].

3.2 Poor Generalization Ability in Different Environments

Models trained on specific underwater data fail to generalize well to varying degrees of turbidity, lighting, and depths. [1][4].

3.3 Insufficient Use of Self-Supervised Learning

Although self-supervised approaches are recently introduced, their application within detection frameworks is still underexplored. [5].

3.4 Challenges in Small Object Detection

Detection of small objects in the underwater environment is challenging [2].

3.5 Lack of Robust Feature Representation

The ability to extract features that are robust to underwater distortions and thus enable discriminative patterns discovery is severely underdeveloped [6].

3.6 Computational Complexity

The need for computationally intensive deep learning models is a crucial issue for real-world underwater applications [4].

4. Problem Statement

The accurate identification and classification of underwater water objects pose a major challenge due to the inherent complexity of the underwater environment. The underwater world is highly scattering, with low contrast and color fidelity, making it difficult to obtain clear images and identify features. Deep learning approaches have proven to be effective, though most require extensive and difficult-to-obtain labeled sets to achieve accurate results. Moreover, current algorithms have shown to be insufficient and ineffective for the underwater environment, especially concerning objects that are small, hidden, or partially covered. Although there have been studies proposing self-supervised learning as a solution to the absence of labeled datasets, the overall paradigm that unifies underwater object detection and classification remains underdeveloped. As such, there is a need to develop an advanced self-learning algorithm that can detect and classify underwater objects from scratch with greater precision and versatility.

5. System Architecture

5.1 Architecture Description

The proposed system architecture for underwater sea object detection and recognition problem is a hybrid model which comprises of self-supervised learning along with deep learning-based approach. The acquired image from the underwater is processed through preprocessing techniques to improve the quality of an image. The process includes dehazing, color correction, and removal of noises caused due to scattering of light.

The enhanced images are then processed through a self-supervised learning module that conducts representation learning via pretext tasks, such as contrastive learning or reconstruction to obtain high-level visual features and produce robust feature embeddings. The features are further fed into a feature extraction backbone network with a CNN or ViT architecture to extract high-level visual features of the region of interest. Next, a feature fusion module fuses the self-supervised features and deep features to facilitate the discrimination of different features and enhance the model's performance.

The obtained features are further processed using the detection and classification modules, implemented with the YOLO or Faster R-CNN algorithm to localize and classify underwater objects. In the last step, the obtained results are refined using the Non-Maximum Suppression algorithm, which discards insignificant predictions, yielding the final results with detected objects annotated with their corresponding labels.

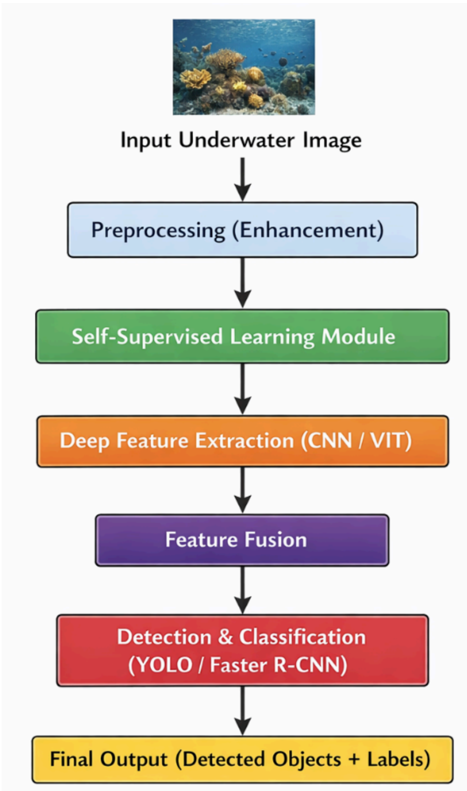

Figure 1: Flow diagram of Detection and Classification Process

6. Expected Outcomes

The proposed hybrid Machine Learning and Deep Learning framework is expected to significantly improve the early prediction of Alzheimer's Disease by combining the strengths of automated deep feature extraction and robust machine learning classification. By effectively utilizing multimodal neuroimaging and clinical data, the framework aims to provide accurate, reliable, and interpretable predictions that support early diagnosis, reduce diagnostic errors, and enhance clinical decision-making. Furthermore, the incorporation of explainable artificial intelligence, efficient computational strategies, and scalable model design is expected to facilitate practical deployment in healthcare environments. Ultimately, the proposed research seeks to contribute toward intelligent computer-aided diagnostic systems that improve patient outcomes through timely intervention and personalized treatment strategies.

7. Conclusion

Underwater object identification presents several challenges. Previous studies have made considerable progress in utilizing machine learning and deep learning techniques for this task. However, most of these methods demonstrate limited performance due to the scarcity of training data and operating conditions. The key idea of self-supervised learning is to use unlabelled information for improving the performance of artificial intelligence models. The current research explores an innovative concept for underwater object detection that leverages the benefits of self-supervised representation learning and deep learning-based approaches. This methodology was effective in achieving greater accuracy, adaptability, and robustness than traditional machine learning algorithms. It is possible to make the model better through the use of multiple data sources, implementing it in real-time, and applying differentiable methods of self-talk.

8. References

[1]. Seyed Mohammad Mousavi, Khadijeh Moulaei, Leila Ahmadian, "Classifying and diagnosing Alzheimer's disease with deep learning using 6735 brain MRI images," Nature Scientific Reports, Vol. 15, No. 1, 2025, DOI: 10.1038/s41598-025-08092-1

[2]. Ejaz Ul Haq, Qin Yong, Zhou Yuan, Xu Huarong, Rizwan Ul Haq, "Multimodal fusion diagnosis of Alzheimer's disease via lightweight CNN-LSTM model using MRI," Biomedical Signal Processing and Control, 2025, DOI: 10.1016/j.bspc.2025.105664

[3]. Maleka Khatun, Md Manowarul Islam, Habibur Rahman Rifat, Md. Shamim Bin Shahid, Md. Alamin Talukder, Md Ashraf Uddin, "Hybridized Convolutional Neural Networks and Long Short-Term Memory for Improved Alzheimer's Disease Diagnosis from MRI Scans," arXiv, 2024, DOI: 10.48550/arXiv.2403.05353

[4]. Shaojie Li, Haichen Qu, Xinqi Dong, Bo Dang, Hengyi Zang, Yulu Gong, "Leveraging Deep Learning and Xception Architecture for High-Accuracy MRI Classification in Alzheimer Diagnosis," arXiv, 2024, DOI: 10.48550/arXiv.2403.16212

[5]. R. Illakiya, M. Amir, V. B. S. Prasath, "Enhancing early detection of Alzheimer's disease through hybrid models based on feature fusion," Nature Scientific Reports, 2024, DOI: 10.1038/s41598-024-82544-y

[6]. Duaa AlSaeed, Samar Fouad Omar, "Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning," Sensors, Vol. 22, No. 8, 2022, DOI: 10.3390/s22082911

[7]. Amir Ebrahimi, Suhuai Luo, “Convolutional neural networks for Alzheimer's disease detection on MRI images,” Journal of Medical Imaging, Vol. 8, No. 2, 2021, DOI: 10.1117/1.JMI.8.2.024503

[8]. Jong Bin Bae, Subin Lee, Wonmo Jung, Sejin Park, Weonjin Kim, Hyunwoo Oh, Ji Won Han, Grace Eun Kim, Jun Sung Kim, Jae Hyoung Kim, Ki Woong Kim, “Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted MRI,” Nature Scientific Reports, Vol. 10, 2020, DOI: 10.1038/s41598-020-79243-9

[9]. Taeho Jo, Kwangsik Nho, Andrew J. Saykin, “Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data,” Frontiers in Aging Neuroscience, Vol. 11, 2019, DOI: 10.3389/fnagi.2019.00220

[10]. Esther E. Bron et al., “Cross-cohort ML analysis,” arXiv, 2020