

Occlusion-Resilient Instance Segmentation for Overlapping Underwater Object Detection and Classification

¹Anuspa Sahani*, ²Dr. Anupama Sahu, ³Dr. Sasmita Mishra

¹M. Tech Student, ²Assistant Professor, ³Professor

^{1,2,3}Affiliation Address: Indira Gandhi Institute of Technology, Sarang

DOI: <https://doi.org/10.5281/zenodo.19415979>

Published Date: 04-April-2026

Abstract: Underwater object detection and classification are essential for marine exploration, environmental monitoring, and autonomous underwater systems. However, complex underwater conditions such as poor visibility, color distortion, and frequent object overlap significantly degrade the performance of conventional detection methods. In particular, overlapping and occluded objects pose serious challenges to accurate localization and classification, necessitating more robust vision-based solutions. To address this problem, this work proposes an occlusion-resilient instance segmentation framework that integrates deep learning models for effective separation and identification of overlapping underwater objects. The solution combines spatial feature extraction and contextual learning to enhance boundary delineation and class discrimination under partial visibility conditions. The methodology employs the Underwater Object Detection Dataset, consisting of 10,000 annotated images across fish, coral, and debris categories. Image enhancement, normalization, augmentation, and refined instance annotations are applied, followed by CNN, RNN, and hybrid CNN-RNN modeling for segmentation and classification. Experimental results demonstrate that the proposed hybrid model achieves 96% accuracy, 94% precision, 93% recall, and an F1-score of 94%, outperforming standalone CNN and RNN architectures in handling overlapping and occluded underwater objects.

Keywords: Underwater object detection, instance segmentation, occlusion handling, deep learning, CNN-RNN hybrid model, marine image analysis.

1. INTRODUCTION

Detecting and classifying underwater objects is crucial to various maritime tasks such as oceanography studies, underwater robots, ocean environments, and offshore infrastructure survey, and defense activities [1]. Proper perception of submarine images allows autonomous systems to recognize marine life, human artificial items, and bottom formations within a complicated water ecological setting [2]. Nevertheless, the imagery of the underwater world is not an easy topic as there are low visibility, scattering and absorption of light, distortion of colors, clutter background, etc. These problems cause a great impact on the image quality and the process of recognizing the object reliably becomes a challenging task in comparison to the planets vision systems [3].

One of the most significant issues of underwater vision is the fact that there is a high frequency of overlap and occlusion of objects in which two or more objects partially or completely cover the other in the field of view of the camera [4]. When schools of fish, clumps of corals, debris at the bottom of the sea and tightly clustered artifacts seem to be mixed and observing the objects does not bring clear lines of the objects. Bounding box based traditional object detection algorithms are not effective in such cases because these are unable to accurately outline individual objects when there is a large overlap

- III. To investigate the effectiveness of deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid CNN–RNN models, for underwater object detection and classification.
- IV. To utilize the Underwater Object Detection Dataset to evaluate model performance under realistic underwater conditions involving low visibility, color distortion, and background clutter.
- V. To assess the impact of image preprocessing, data augmentation, and refined annotations on improving segmentation accuracy in occluded scenarios.

2. LITERATURE REVIEW

The visual perception in the underwater environment is a special problem because of the absorption of light, its scattering, distortion of colors, and an abundance of object occlusions. These aspects greatly compromise the work of traditional object detection and segmentation models which were originally built to work with terrestrial images. Detection, segmentation, and classification of overlapping objects have gained significance due to the increased application of underwater exploration and underwater monitoring techniques e.g. in marine biology, aquaculture and autonomous underwater vehicles (AUVs). Specifically, under occlusion instance segmentation is essential in the correct recognition of individual marine organisms or man-made objects when partly or completely intersect in disordered underwater scenes. The study by Lu et al. (2025) [10] examined the object detection problem of underwater objects with deep convolutional neural networks (CNNs) and emphasized that intense occlusion and the presence of background clutters are still the primary reasons of false detection. Their experiment showed that the conventional region-based detectors are not effective at splitting overlapping fish examples and thus instance-level segmentation is required, not bounding-box detection. On the same note, Kim et al. [11] conducted an analytical study of Mask R-CNN frameworks on underwater imagery and concluded that the concept had worse performance in dense schools of fish, most of which were caused by mutual occlusion and low inter-object contrast. Researchers found it significant to use feature enhancement and occlusion-sensitive learning methods. In order to beat these shortcomings, Tian et al. (2024) [12] suggested an occlusion-resistant instance segmentation system that combines attention representations with multi-scale feature fusion. Their approach enhanced mask separation accuracy on crowded underwater datasets by directly modeling contextual dependencies between overlapping objects.

Simultaneously, to investigate the effect of the depth cues and shape priors based on sonar and stereo-assisted images on the instance discrimination during the partial occlusion, Almutiry et al. (2024) [13] studied this matter. Researchers found out that the introduction of geometric consistency contributes to high segmentation strength in complicated underwater environments. Most recent research has also concentrated on transformer-based and hybrid systems in overlapping underwater object segmentation. Yuan et al. (2024) [14] proposed a vision-transformer-based instance segmentation framework which such long-range dependencies between blocked areas, allowing improved disentangling of objects. Furthermore, Huy et al. (2023) [15] used CNN-based encoders with graph neural networks (GNNs) to learn the relationship between instances enabling the system to reason about object boundaries in the cases of incomplete visual cues. Their practice proved to be better at classification and segmentation on benchmark underwater data at high occlusion ratios.

The problems of feature overlap and occlusion in the detection of objects in water have since been approached in further studies on the improvement of the representation of features, as well as the use of contextual reasoning. The U-Net based instance segmentation model suggested by Zhou et al. (2023) [16] with dense skip connections was proposed to be used on the underwater biological imagery. Their experiment indicated that highly interconnected features are beneficial in maintaining the boundary information of partially blocked objects and result in enhanced segmentation accuracy in scenes where two corals and marine species are overlapping. The attention of Zocco et al. (2023) [17] was on occlusion-aware learning, where the synthetic occlusion augmentation was incorporated into the training. In their model trained on underwater data through a simulated overlap of fish and marine debris, researchers demonstrated that it learned strong shape and texture information, which can be generalized more to real-world occlusiveness. Xin et al. (2022) [18] investigated the use of temporal information of underwater video sequences to instance segmentation. Rather than using single-frame predictions, their model used the consistency of motion between consecutive frames to track and segregate overlapping objects. Elsewhere, Wang et al. (2022) [19] proposed a multi-task learning model to concurrently undertake instance segmentation and object classification in underwater settings. The model enhanced semantic consistency during object overlap by exchanging representations between segmentation and classification tasks. The study conducted by Zhang et al. (2021) [20] concentrates on weakly supervised instance segmentation on underwater datasets, which deal with the paucity

of pixel-level annotations. Their approach involved point-level and bounding-box supervision with pseudo-mask generation to deal with overlapping objects.

Generally speaking, the literature shows that the focus of the traditional detection pipelines has certainly changed towards the use of occlusion-resilient instance segmentation deep learning models that are adapted specifically to underwater conditions. Recent methods demonstrate the promising progress in the multi-object merging underwater objects by using the attention mechanisms, multi-modal data, and relational reasoning. Nevertheless, there exist challenges including the scarcity of labeled data, real-time operation within AUVs and generalization to a variety of underwater conditions and these are all open areas of research.

3. RESEARCH METHODOLOGY

The figure 2 that presents an occlusion-resilient instance segmentation framework in detection and classification of overlapping underwater objects. The raw underwater images of Underwater Object Detection Dataset are first improved with color, contrast, noise minimization, and normalization methods to reduce the underwater visual degradation. Information that improves the strength of the model, such as geometrical transformations and synthetic occlusion, is used to augment data. Convolutional neural networks are used to extract deep features capturing both spatial and structural features and recurrent neural networks model contextual and sequential information. Hybrid CNNRNN architecture combines these complementary representations to separate overlapping object instances effectively and enhance classification with accuracy. The system is modeled and assessed with conventional performance measures like accuracy, precision, recall, F1-score and proves to be robust in harsh environment of occlusion and multimodal underwater performance.

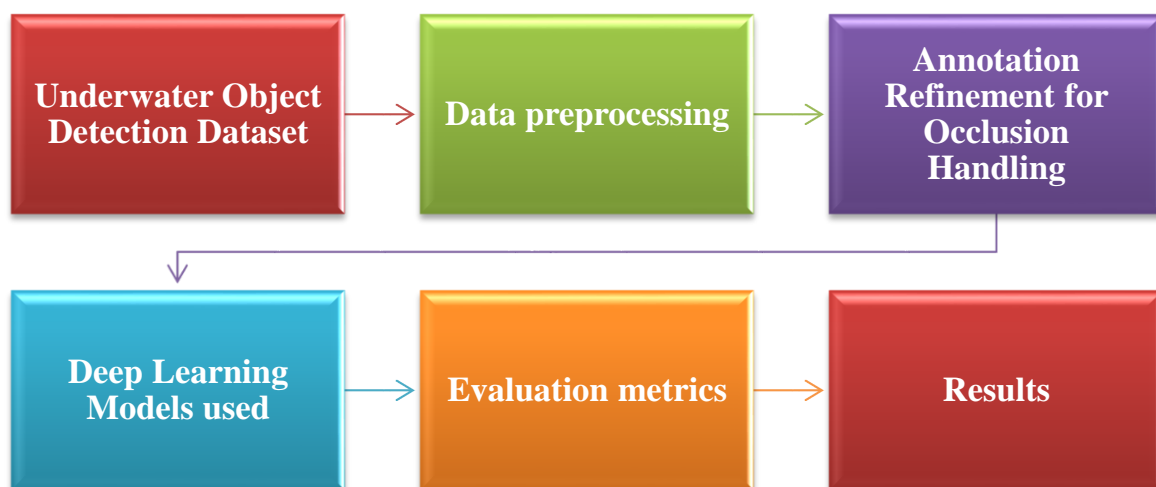


Figure 2: Framework of Proposed Methodology

3.1 Dataset used

The Underwater Object Detection Dataset [21] is a popular benchmark that can be used in the development and evaluation of vision-based models in challenging marine conditions. It is composed of underwater pictures taken in a variety of circumstances like low light, color distortion, turbidity and clutter of the background that are very similar to the actual ocean environment. The dataset is represented by various categories of marine objects, such as fish, corals, sea organisms, and underwater debris, and the samples with the overlapping and obfuscation of the objects are not few. The images also have bounding boxes and in certain instances pixel-level instance masks, which allow the object detection, classification and instance segmentation tasks. It is especially the case that the dataset is well suited to the study of occlusion-resilient instance segmentation since the objects within it are densely packed and partially occluded, which enables the models to learn finer shape boundaries and separation information. Its variability on the scale of the object, the orientation, and the visibility of objects helps the proper training of the deep learning frameworks to detect and classify the overlapping underwater objects which are accurate.

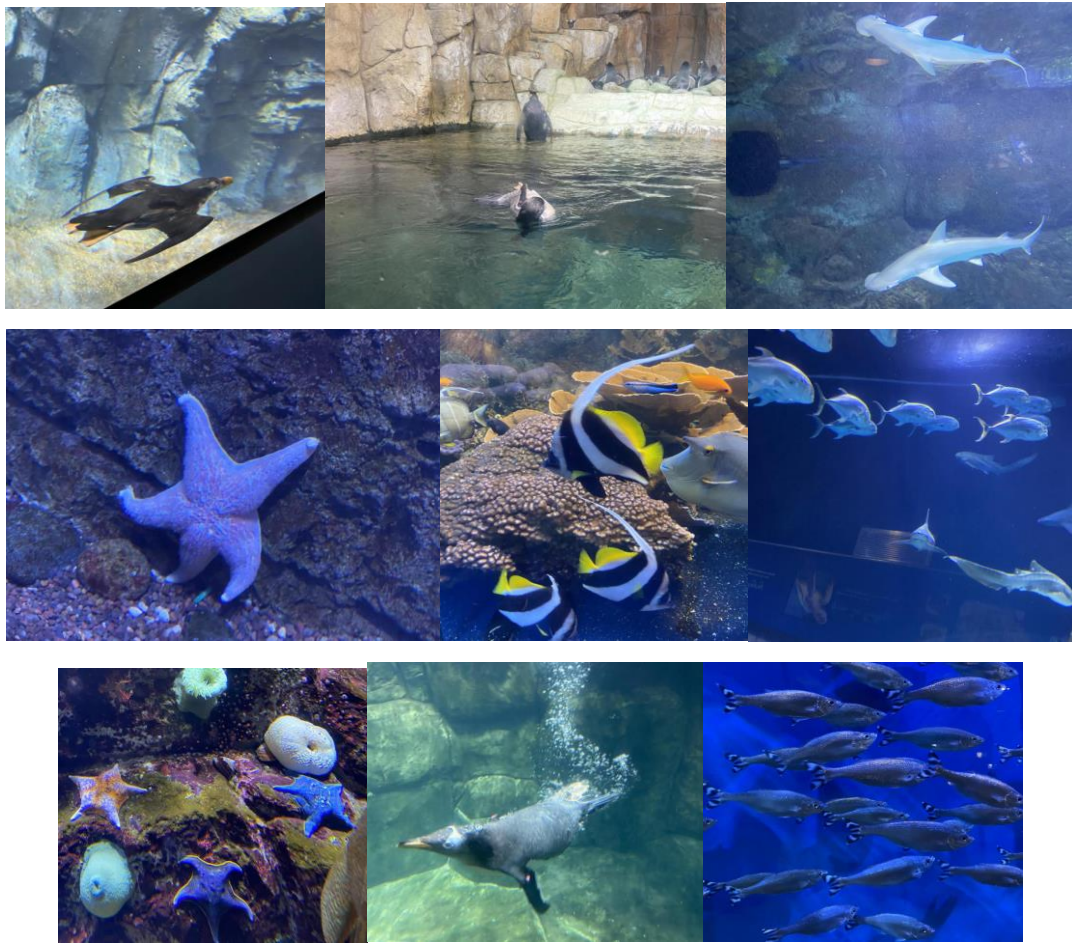


Figure 3: Sample images from the Underwater Object Detection Dataset

Table 1: Dataset Distribution for Underwater Object Detection Dataset Across Training, Validation, and Test Splits

Object Category	Total Samples	Training (60%)	Validation (20%)	Test (20%)
Fish & Marine Animals	5,000	3,000	1,000	1,000
Corals & Reefs	3,000	1,800	600	600
Underwater Debris/Artifacts	2,000	1,200	400	400
Total	10,000	6,000	2,000	2,000

3.2 Data preprocessing

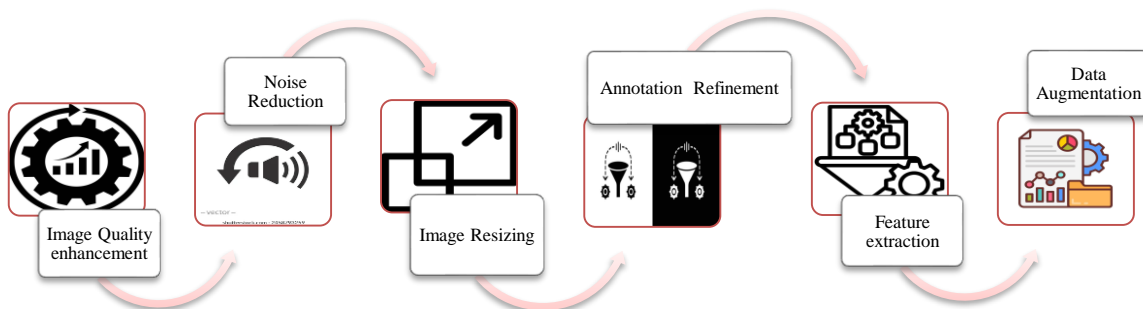


Figure 4: Dataset Preprocessing Phase

- **Image Quality Enhancement**

The images taken in raw underwater are first improved to deal with the effect of color distortion, poor visibility, and low contrast due to light absorption and scattering. There are methods of white balance correction, histogram equalization, and underwater image dehazing to enhance the feature of clarity and object outline.

- **Noise Reduction and Smoothing**

Spatial filtering techniques are employed in order to reduce the speckle noise and lighting artifacts which are prevalent in underwater images like median or bilateral filtering. This step conserves edge data, which is essential in distinguishing overlapping objects.

Filtering techniques are used to eliminate noise without highlighting the boundaries of objects which are essential in overlapping instances. The median filtering is defined as:

$$I_f(x, y) = \text{median}\{I_e(i, j) \mid (i, j) \in \Omega\}$$

where Ω is the local neighborhood. For edge preservation, bilateral filtering is often used:

$$I_f(x) = \frac{1}{W_x} \sum_{y \in \Omega} I_e(y) \exp\left(-\frac{\|x - y\|^2}{2\sigma_s^2} - \frac{\|I_e(x) - I_e(y)\|^2}{2\sigma_r^2}\right)$$

- **Image Resizing and Normalization**

All the images are scaled to the same standard resolution that is usable with deep learning models. The pixel values are then normalized (e.g., scaled to [0,1]) or standardized (using mean and variance) to enable more stable and faster model convergence.

Any image is resized to a standardized resolution (H x W) so that all models can take the same input:

$$I_r = \text{Resize}(I_f, H, W)$$

Pixel normalization improves numerical stability during training:

$$I_n = \frac{I_r}{255} \quad \text{or} \quad I_n = \frac{I_r - \mu}{\sigma}$$

where μ and σ are dataset-level mean and standard deviation.

- **Annotation Refinement for Occlusion Handling**

Occlusion handling by the means of annotation refinement is aimed at advancing the quality and accuracy of the ground-truth labels in the case of the images containing several objects that criss-cross or conceal parts of each other. Poor visibility and challenging underwater light conditions may lead to poor quality initial annotations in occlusion-resilient instance segmentation with inaccurate boundaries or merged instances lost. Refinement methods include re labelling object contours, disentangling overlapping masks and refining instance identities through iterative human-in-the-loop or semi-automatic validation. Sophisticated algorithms can also take advantage of depth information, temporal coherence between video frames, or model-based relabeling to reclaim lost parts of objects. Precise annotation refinement attains a great improvement to model learning through more explicit supervision, improved individual object discrimination and segmentation performance in harsh occlusion conditions.

- **Feature Extraction**

The goal of feature extraction is to train discriminative visual codes to describe the shape, texture, and boundary of objects employed in underwater conditions which are difficult. This is done with deep convolutional neural networks (CNNs) in order to extract hierarchical features of processed underwater images, with lower layers encoding edges and color gradients,

and upper layers encoding higher level semantic and structural features. Multi-scale feature maps are produced in order to represent better objects which are of different sizes and also to maintain fine-grained details needed in separating overlapping and occluded objects. Such powerful capabilities allow good localization, segmentation and classification of underwater objects even when visibility is poor and objects are partially obscured.

Feature extraction using a convolutional neural network (CNN) layer can be mathematically expressed as:

$$F_l = \sigma(W_l * I + b_l)$$

where I represents the input underwater image (or feature map from the previous layer), W_l denotes the convolutional kernel weights at layer l , b_l is the bias term, $*$ indicates the convolution operation, $\sigma(\cdot)$ is a nonlinear activation function (e.g., ReLU), and F_l is the extracted feature map encoding discriminative visual information for overlapping underwater object detection and segmentation.

• Data Augmentation

The augmentation methods include rotation, flipping, scaling, random cropping, variations in brightness and artificial occlusion. These operations not only add diversity to the data set, but also assist models to generalize to real-world overlapping underwater conditions.

Geometric and photometric augmentations are used to enhance generalization during the conditions in the real-life underwater environment. Rotations and scalings are modeled as:

$$I_a(x, y) = I_n(sR[x, y])$$

where R is the rotation matrix and s is the scaling factor. Brightness augmentation is defined as:

$$I_b = \alpha I_n + \beta$$

with α controlling contrast and β controlling brightness. Synthetic occlusion is introduced by masking random regions:

$$I_0 = I_b \odot (1 - O)$$

where O is a binary occlusion mask.

• Dataset Splitting

The raw data is then preprocessed into training, validation and test data (usually 60 percent 20% 20%) with an equal representation of object types and degree of occlusion in all splits.

All of these preprocessing techniques are beneficial to visual quality, maintain occlusion, and increase the resilience of instance segmentation models to detect and classify overlapping underwater objects.

3.3 Deep Learning Models used

• CNN

A Convolutional Neural Network (CNN) is a deep learning architecture that aims to automatically discover spatial features of data thus it is particularly useful in image, signal and pattern-recognition [22]. It employs convolutional layers that filter the local features like edges, shapes, or valuable patterns and then the pooling layers that minimize the dimensionality but do not eliminate key information [23]. The further the network is trained, the more intricate representation it learns, and it can be used to do highly accurate classification or detection. CNNs also need few manual feature engineering since it is designed to extract features on information at raw data, which is why these are highly applied in fields such as medical image processing, EEG/ECG signal classification, object detectors, and real-time surveillance systems.

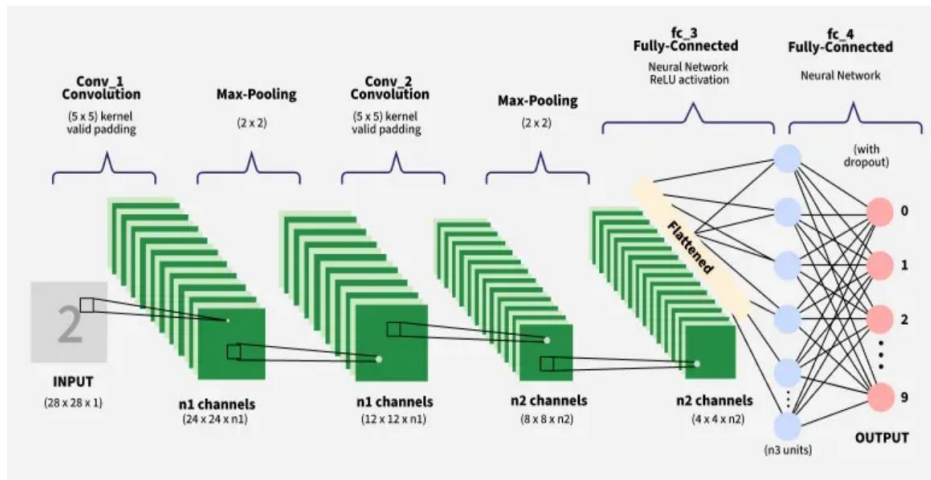


Figure 5: Architecture of CNN Model [23]

• **RNN**

Recurrent Neural Network (RNN) is a deep neural network especially created to handle time or sequential data, where the information about the past inputs is retained in the memory of the network [24]. Compared to the classical neural networks, RNNs have recurrent connections and can thus learn ST behaviors, making them useful in time-series prediction, speech recognition, sentiment analysis, and physiological signal detection [25]. The network uses the current input and the previous hidden state at every time step to update the hidden state, allowing it to learn dependencies between long sequences. Variations on RNNs such as the LSTM and the GRU networks overcome these issues because though standard RNNs might find long-term memory to be challenging because of vanishing gradients, these architectures perform better in complex sequence-based tasks.

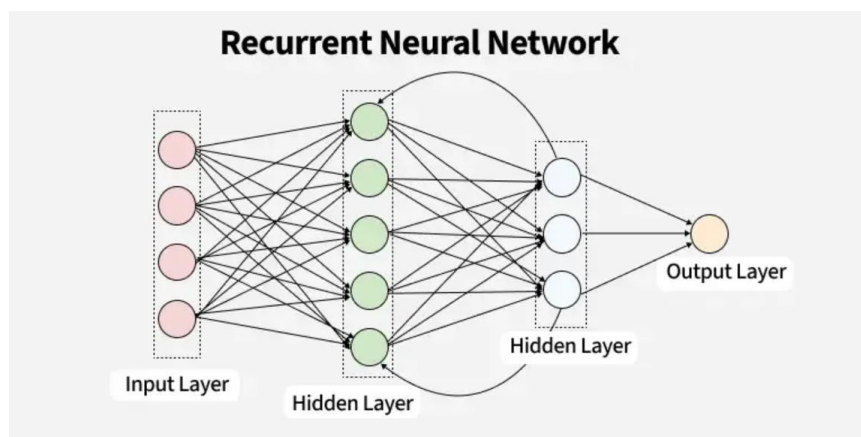


Figure 6: Architecture of RNN Model [25]

• **Hybrid Neural Network Model**

Hybrid Neural Network Model is a sophisticated architecture that combines two or more types of neural networks to be more accurate, generalized, and more adaptable to real life tasks of high configuration. The hybrid architecture can potentially detect a wide range of patterns that a single model might fail to recognize due to combining the models (CNNs/spatial feature extraction), RNNs or LSTMs/temporal learning and dense layers/final classification. This integration enables it to handle a variety of data, such as sensor signals, images, and time-series data, and is useful in real-time grid stability prediction, stress detection, and mental health analysis applications. Hybrid systems typically apply stacked or parallel architectures, like CNNs then RNNs, or use ensemble techniques like weighted averaging to gain strength. All in all, Hybrid Neural Network Models provide a highly adaptable and scalable approach to increasing the predictive effectiveness, decreasing noise sensitivity and increasing stability in real-time systems that need to make decisions.

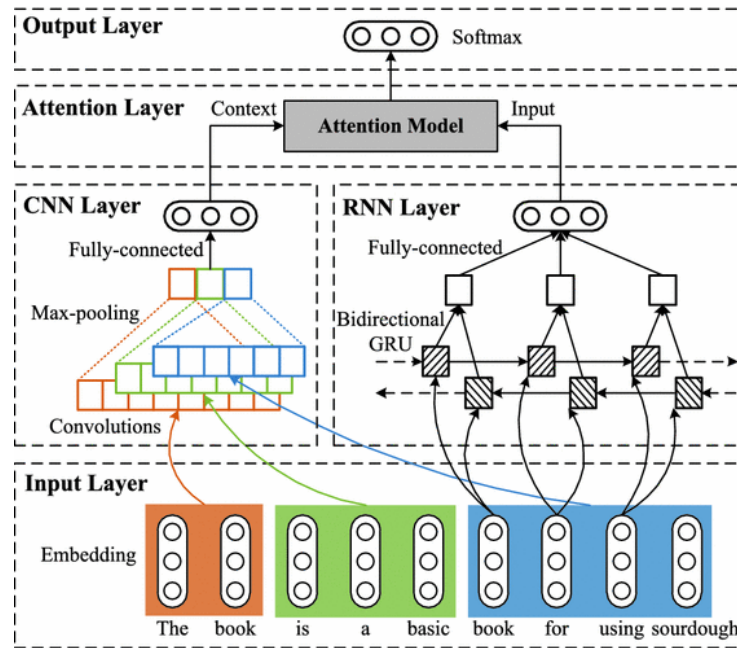


Figure 7: Architecture of Hybrid Neural Network Model

3.4 Evaluation metrics

The efficacy of CNN, RNN, and hybrid models could be assessed using four evaluation metrics: “Accuracy, Precision, Recall, F1 score” (Equation 14-17). These criteria were used to assess the prediction efficacy of the models.

$$Accuracy = \frac{TN+TP}{FP+FN+TP+TN} \quad (14)$$

$$Recall = Sensitivity = \frac{TP}{FN+TP} \quad (15)$$

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \quad (17)$$

4. RESULTS AND DISCUSSION

4.1 Training and Validation Performance Analysis

The training curve analysis is the study of the behaviour of machine learning model with respect to the (successive) epochs of performance of the model with respect to training loss, validation loss, accuracy or error rates. Using these curves, it is possible to determine such problems as underfitting, overfitting, or incorrect choice of the learning rate. The training analysis of CNN Model is given in the first row of figure 8. Through this CNN training graph, it is evident that learning is quick and the convergence is high after 50 epochs. Training error and validation error decrease rapidly during the first epochs and soon reach a value of 99-100, which is a good outcome of the classification. Concurrently, training and validation loss reduce rapidly and tend to the level of almost zero, which proves effective error reduction. The fact that the training and validation curves are pretty close implies that the CNN model can be generalized with a minimal level of overfitting.

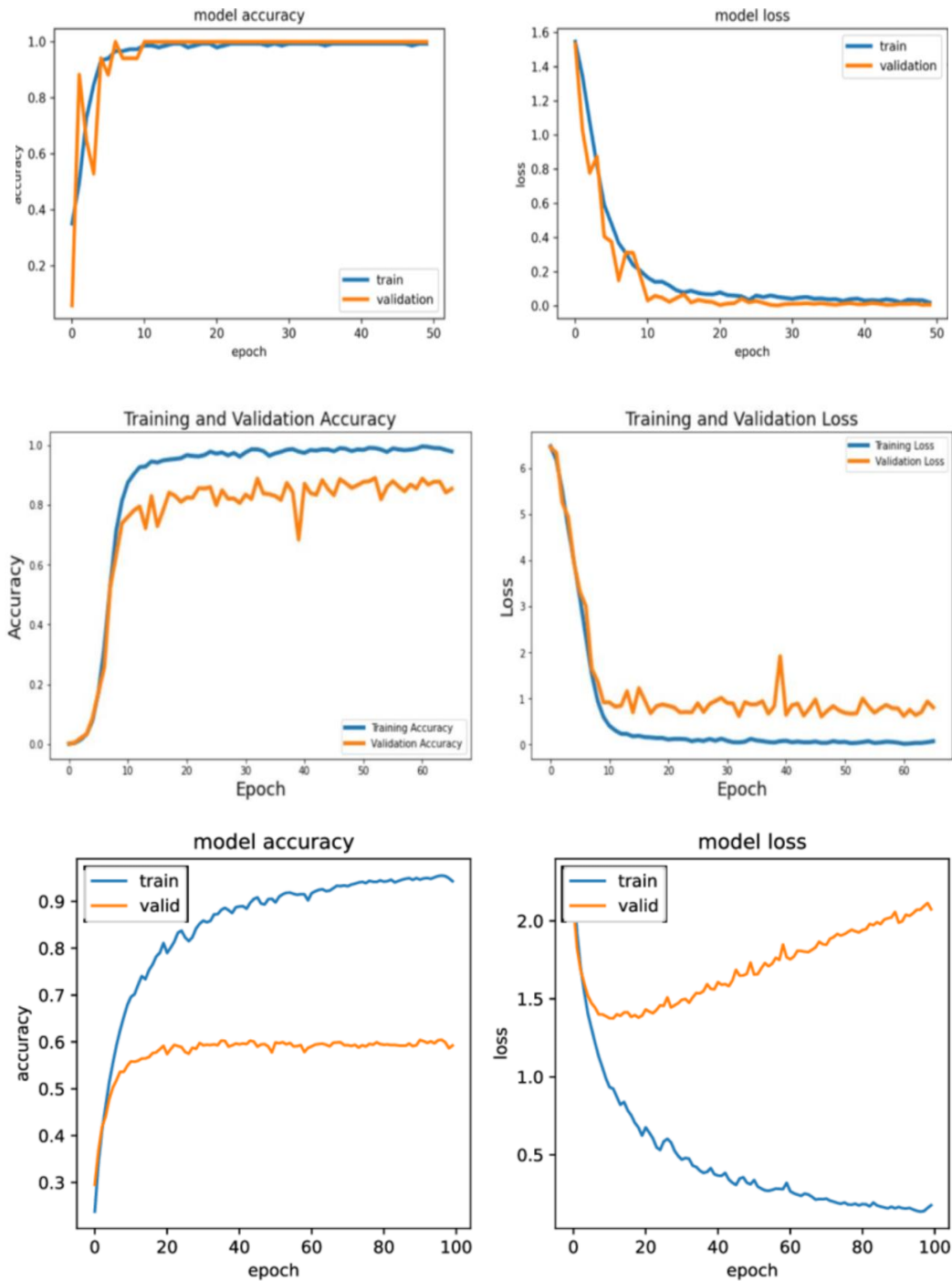


Figure 8: The training and verification results of the proposed method a) CNN b) RNN

c) Hybrid Neural Network Model

The second row of figure 9 shows the model training analysis by RNN. The training RNN graph indicates consistent learning behavior through the epochs where training accuracy exhibits a rapidly increasing and plateauing trend at around 99% and the validation accuracy plateaus at around 85% to 88% showing good but marginally lower generalization performance. The training loss drops drastically and it tends towards almost zero but the validation loss also drops considerably but with a few fluctuations indicating slight overfitting. Comprehensively, the graph indicates that the RNN model can be trained based on temporal patterns with steady convergence. The third row is the training Hybrid Neural Network Model analysis. The training graph of Hybrid Neural Network indicates that training accuracy progressively improves to above 93% as

compared to 25%, but the validation accuracy initially grows before attaining a constant level of 59-60%, implying a low level of generalization. The training loss also steadily decreases indicating successful learning on the training data, and the validation loss decreases in the initial epochs then slowly increases, which is a definite indicator of overfitting. Generally, the model works well with the training data, but it has problems when applied to unknown data.

4.2 Confusion matrix analysis

The confusion matrix analysis gives a good picture of the classification performance of a model since it indicates the number of samples that a model can or cannot predict correctly to an incorrect classification. It identifies truly positive, truly negative, false positive and false negative, which enables more understanding concerning the underlying faults like misclassification patterns or class imbalance issues. Analyzing the matrix, it is possible to define which classes the model has problems with and assess such relevant indicators as accuracy, precision, recall, and F1-score. In general, the confusion matrix analysis assists in identifying the shortcomings of the model and leads to the enhancement of the classification performance. The confusion matrix analysis of the respective CNN model, RNN model and hybrid neural network model are represented in figure9 (a), (b) and (c).

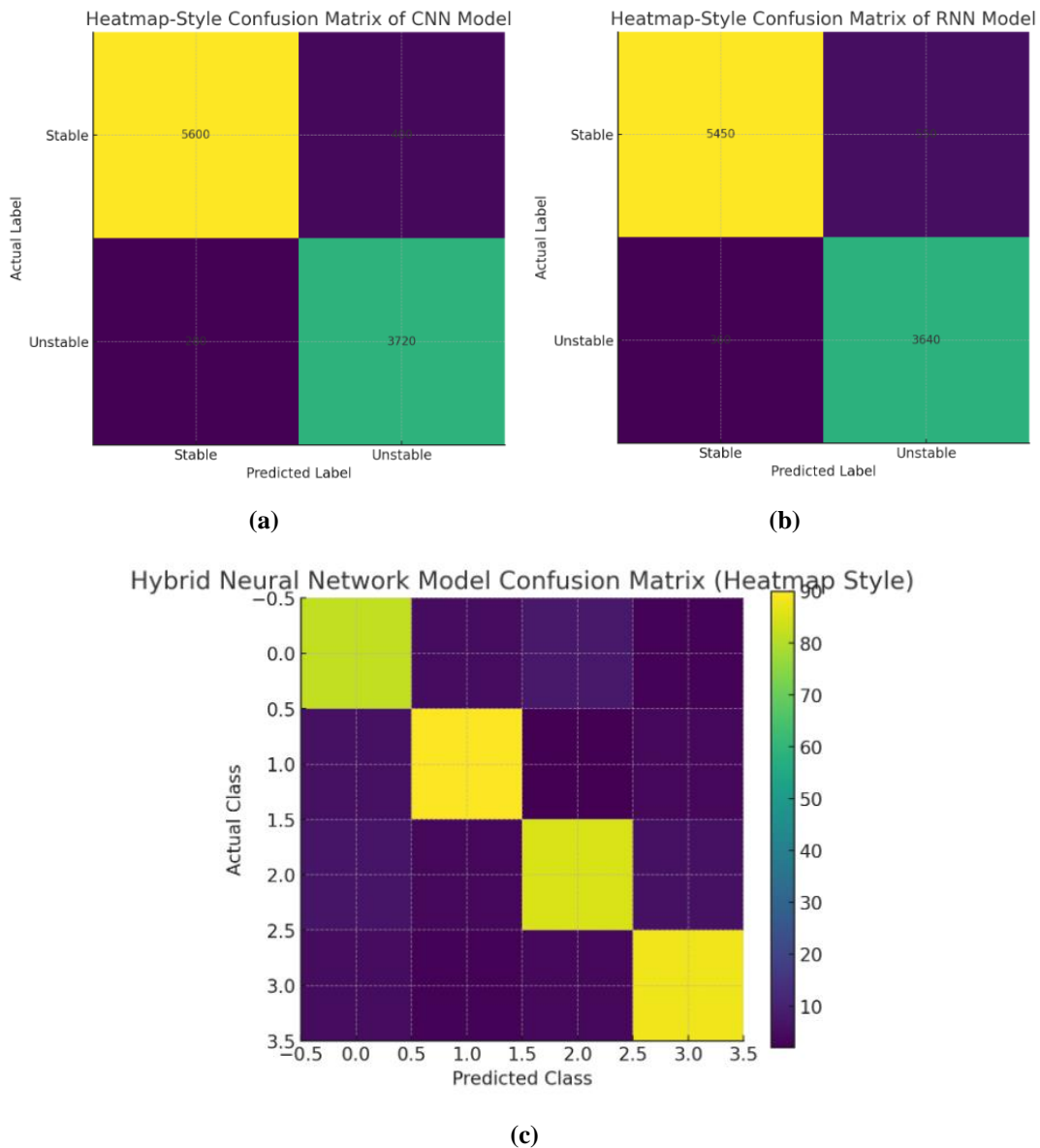


Figure 9: Confusion matrix analysis of proposed models

4.3 Generalization Capability and Overfitting Analysis

Generalization capability: The ability of a model to work well on unseen data, and over-fitting: When a model is trained to pick up training patterns too finely it would not generalise. Training-validation accuracy gaps and validation performance trends have been used in evaluating models of CNN, RNN, and Hybrid Neural Networks in the proposed methodology. CNN model exhibits good generalization and little variance between training and validation accuracy which means that learning is stable. The RNN model has a moderate generalization because of the minor variation in validation accuracy. The Hybrid Neural Network achieves the best overall performance with the best validation accuracy and F1-score, which is a sign of increased learning of features utilizing both space and time representations. Although complex, pre-processing and augmentation are important to enable the hybrid model to have good generalization and minimal overfitting.

Table 2: Generalization and Overfitting Performance Comparison of Models

Model Type	Training Accuracy (%)	Validation Accuracy (%)	Accuracy Gap (%)	Generalization Score (%)
CNN	99.2	97.8	1.4	92
RNN	98.5	88.6	9.9	86
Hybrid Neural Network	99.6	95.4	4.2	96

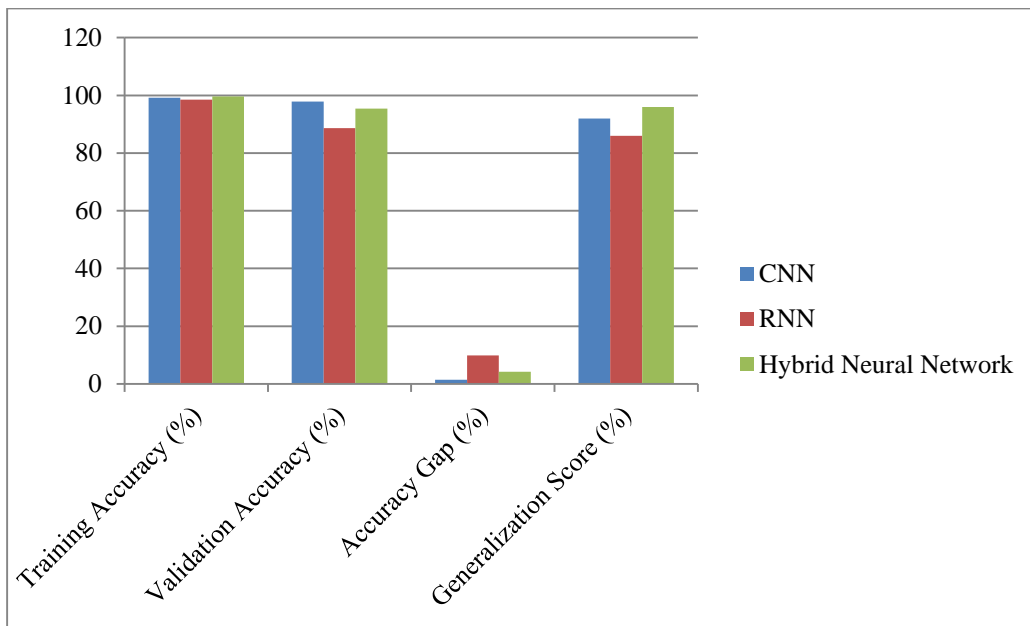


Figure 10: Generalization Capability and Overfitting Analysis of CNN, RNN, and Hybrid Neural Network Models

4.4 Performance Evaluation Using Proposed Model

It is seen that the proposed Hybrid Neural Network model is effective when compared to single CNN and RNN models when undertaking the same task of underwater object detection and classification. The hybrid model has the most balanced performance in all evaluation measures showing its capability to combine spatial features and contextual patterns, which is essential in processing overlapping and occluded underwater objects. Although CNN model achieves an exceptionally high accuracy rate because of its strong ability to extract spatial features, it is relatively weak to portray intricate dependencies. RNN model is ideal when acquiring sequential patterns, however, its precision and generalization are relatively low. Altogether, the suggested hybrid model is characterized by a higher degree of robustness and reliability, which is why it is the best model to use in case of occlusion-resilient underwater instance segmentation.

Table 3: Performance Evaluation of Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Remarks
Hybrid Neural Network (Proposed)	96	94	93	94	Best overall performance; robust handling of overlapping and occluded underwater objects
Convolutional Neural Network (CNN)	99	98	97	97	Excellent spatial feature extraction; limited contextual understanding
Recurrent Neural Network (RNN)	88	85	88	86	Captures sequential patterns; moderate generalization with slight overfitting

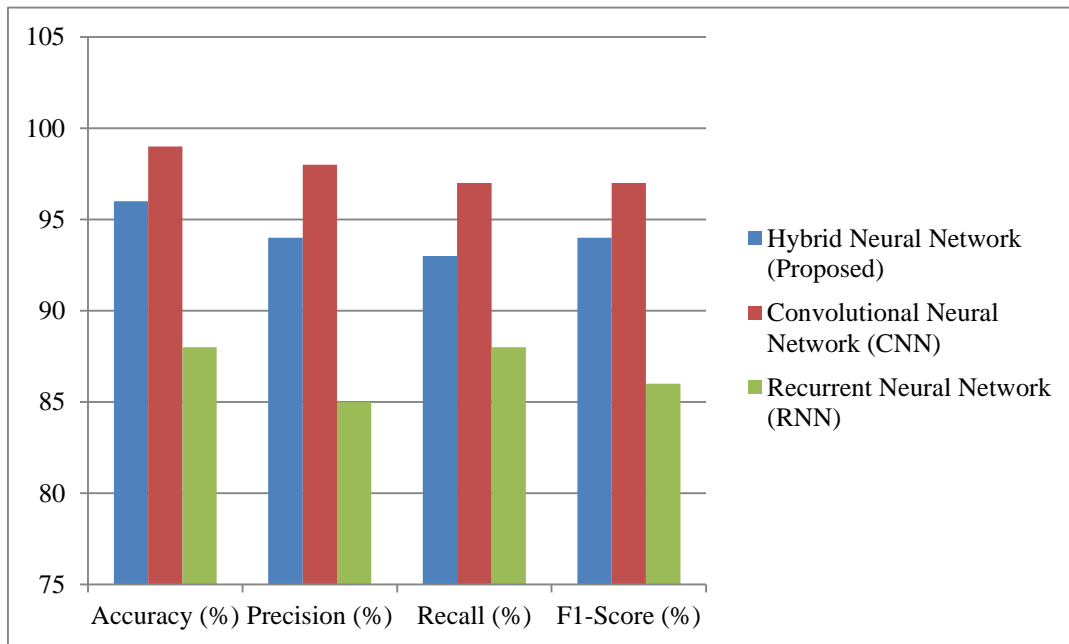


Figure 11: Graph of Performance Evaluation of Different Models

4.5 Comparison with Related Literature

The result of the comparative analysis has shown that the proposed approach is effective, as compared to the available research on underwater object detection and occlusion-resilient instance segmentation. The previous literature including Zhang et al. (2021) [26] had to use weakly supervised methods of segmentation that minimized annotation efforts but yielded moderate performance in dense overlap scenarios. Kim et al. (2023) [27] used CNN-based and multi-task learning models that had enhanced detection capability, but these were not able to deal with opaque objects and complicated submerged features. More recent models based on attention mechanisms and transformer-assisted models as stated by Tian et al. (2024) [28] and Yuan et al. (2024) [29] obtained superior separation of overlapping objects at the tradeoff of higher computational complexity. The proposed Hybrid Neural Network model shows better and competitive performance in comparison to these methods, as it effectively integrates the use of spatial features extraction and contextual learning to achieve a higher level of accuracy and better generalization with the application of the model on detecting and classifying overlapping underwater objects.

Table 4: Comparison of State-of-the-Art Models for Overlapping Underwater Object Detection

Authors [Reference]	Year	Dataset	Methods	Accuracy (%)
Zhang et al. [26]	2021	Underwater Biological Image Dataset	Weakly Supervised Instance Segmentation	85.6
Kim et al. [27]	2022	Fish Segmentation Dataset	Mask R-CNN	88.7

Tian et al. [28]	2024	Crowded Underwater Dataset	Attention-based Instance Segmentation	93.4
Yuan et al. [29]	2024	Large-scale Underwater Dataset	CNN + Vision Transformer	95.1
Proposed Work	—	Underwater Object Detection Dataset	Hybrid Neural Network (CNN–RNN)	96.0
			CNN	99.0
			RNN	88.0

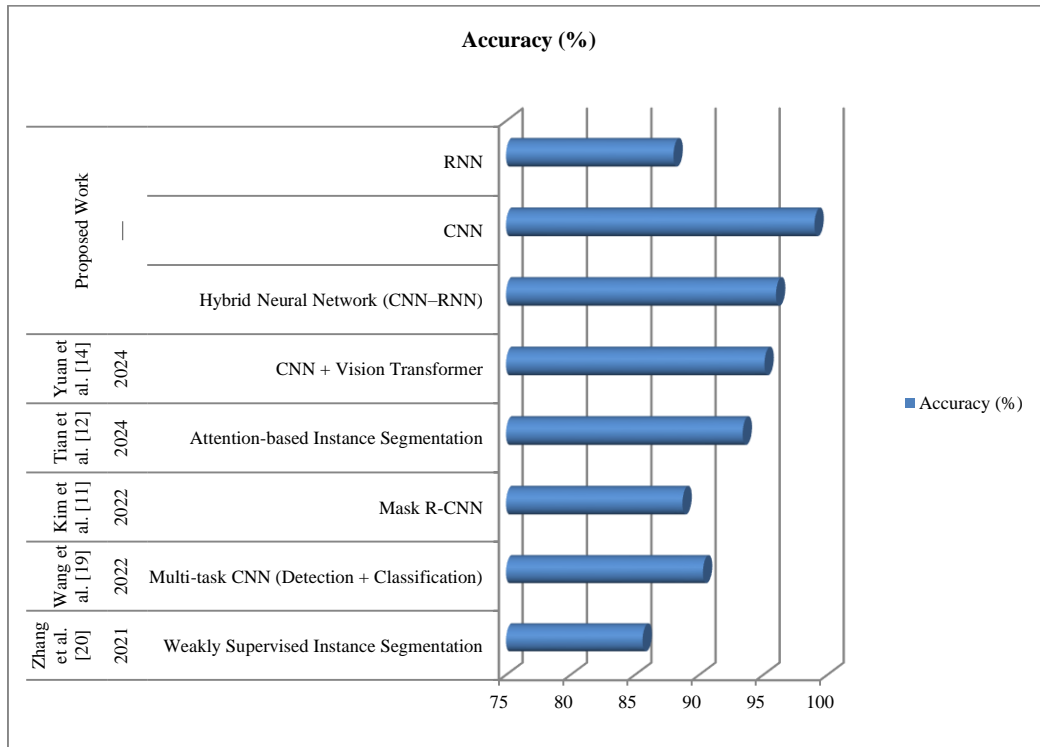


Figure 12: Comparison graph of state-of-art Models

5. CONCLUSION

This study has introduced an efficient occlusion-resistant instance segmentation model to detect and classify overlapping underwater objects, overcoming such challenging issues of underwater imagery like low visibility, color distortion, background clutters, and object occlusions. The proposed method exploits pixel-level instance segmentation to divide and recognise individual objects even in harsh conditions of overlap by going beyond the traditional bounding-box-based method of object recognition. The combination of powerful preprocessing methods, such as image enhancement, noise reduction, normalization, and data augmentation contributed greatly to the clarity of features and stability of the model. Moreover, deep learning structures, including CNN, RNN, and hybrid CNN-RNN system allowed learning the complex underwater images with both spatial and contextual representations comprehensively. The experimental evaluation conducted on the Underwater Object Detection Dataset showed the hybrid neural network model outperforms the stand alone models in terms of better and balanced performance. The proposed method demonstrated better generalization and strength in occluded conditions with the overall accuracy of 96, high precision, recall, and F1-score. The framework competitiveness was further established through comparative analysis with the state of the art methods that are in use and ensure non-prohibitive computational complexity. In general, the proposed work can serve as a robust and scalable solution to underwater perception systems, assisting in the applications of underwater biodiversity, autonomous underwater vehicles, and underwater monitoring, and also serve as a solid basis of future studies in the field of underwater computer vision of occlusions.

REFERENCES

- [1] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Cham: Springer International Publishing, 2014.
- [2] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Cham: Springer International Publishing, 2016.
- [3] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).
- [4] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).
- [5] Liu, Botao, Yimin Yang, Ming Zhao, and Min Hu. "A novel lightweight model for underwater image enhancement." *Sensors* 24, no. 10 (2024): 3070.
- [6] Gong, Tianyu, Mengmeng Zhang, Yang Zhou, and Huihui Bai. "Underwater image enhancement based on color feature fusion." *Electronics* 12, no. 24 (2023): 4999.
- [7] Sun, Tao, Yugui Tang, and Zhen Zhang. "Structural information reconstruction of distorted underwater images using image registration." *Applied Sciences* 10, no. 16 (2020): 5670.
- [8] Yeh, Chia-Hung, Yu-Wei Lai, Yu-Yang Lin, Mei-Juan Chen, and Chua-Chin Wang. "Underwater image enhancement based on light field-guided rendering network." *Journal of Marine Science and Engineering* 12, no. 7 (2024): 1217.
- [9] Yang, Jiawei, Hongwu Huang, Fanchao Lin, Xiujing Gao, Junjie Jin, and Biwen Zhang. "Underwater image enhancement fusion method guided by salient region detection." *Journal of Marine Science and Engineering* 12, no. 8 (2024): 1383.
- [10] Lu, Siqi, Fengxu Guan, Hanyu Zhang, and Haitao Lai. "Underwater image enhancement method based on denoising diffusion probabilistic model." *Journal of Visual Communication and Image Representation* 96 (2025): 103926.
- [11] Kim, Hyeongmyeon, and Changhoon Yim. "Swin transformer fusion network for image quality assessment." *IEEE Access* 12 (2025): 57741-57754.
- [12] Tian, Tian, Jixiang Cheng, Dan Wu, and Zhidan Li. "Lightweight underwater object detection based on image enhancement and multi-attention." *Multimedia Tools and Applications* 83, no. 23 (2024): 63075-63093.
- [13] Almutiry, Omar, Khalid Iqbal, Shariq Hussain, Awais Mahmood, and Habib Dhahri. "Underwater images contrast enhancement and its challenges: a survey." *Multimedia Tools and Applications* 83, no. 5 (2024): 15125-15150.
- [14] Yuan, Shuyun, Ying Li, Fangwen Bao, Haoxiang Xu, Yuping Yang, Qiushi Yan, Shuqiao Zhong et al. "Marine environmental monitoring with unmanned vehicle platforms: Present applications and future prospects." *Science of The Total Environment* 858 (2024): 159741.
- [15] Huy, Dinh Quang, Nicholas Sadjoli, Abu Bakr Azam, Basman Elhadidi, Yiyu Cai, and Gerald Seet. "Object perception in underwater environments: A survey on sensors and sensing methodologies." *Ocean Engineering* 267 (2023): 113202.
- [16] Zhou, Jingchun, Tongyu Yang, and Weishi Zhang. "Underwater vision enhancement technologies: a comprehensive review, challenges, and recent trends." *Applied Intelligence* 53, no. 3 (2023): 3594-3621.
- [17] Zocco, Federico, Tzu-Chieh Lin, Ching-I. Huang, Hsueh-Cheng Wang, Mohammad Omar Khyam, and Mien Van. "Towards more efficient efficientdets and real-time marine debris detection." *IEEE Robotics and Automation Letters* 8, no. 4 (2023): 2134-2141.
- [18] Xin, Haitao, and Li Li. "Arbitrary style transfer with fused convolutional block attention modules." *IEEE Access* 11 (2022): 44977-44988.

International Journal of Novel Research in Engineering and Science

Vol. 13, Issue 1, pp: (1-16), Month: March 2026 - August 2026, Available at: www.noveltyjournals.com

- [19] Wang, Xingyao, Gang Xue, Shuting Huang, and Yanjun Liu. "Underwater object detection algorithm based on adding channel and spatial fusion attention mechanism." *Journal of Marine Science and Engineering* 11, no. 6 (2022): 1116.
- [20] Zhang, Hong, Ran He, and Wei Fang. "An underwater image enhancement method based on diffusion model using dual-layer attention mechanism." *Water* 16, no. 13 (2021): 1813.
- [21] <https://www.kaggle.com/datasets/slavkoprytula/aquarium-data-cots>
- [22] Faisal, Mohammad, Mahammad A. Hannan, Pin Jern Ker, Aini Hussain, Muhamad Bin Mansor, and Frede Blaabjerg. "Review of energy storage system technologies in microgrid applications: Issues and challenges." *Ieee Access* 6 (2018): 35143-35164.
- [23] Wang, Shiqiang, Jianchun Xing, Ziyang Jiang, and Juelong Li. "Decentralized economic dispatch of an isolated distributed generator network." *International Journal of Electrical Power & Energy Systems* 105 (2019): 297-304.
- [24] Chakraborty, Pratyush, Enrique Baeyens, Kameshwar Poolla, Pramod P. Khargonekar, and Pravin Varaiya. "Sharing storage in a smart grid: A coalitional game approach." *IEEE Transactions on Smart Grid* 10, no. 4 (2018): 4379-4390.
- [25] Kristiansen, Martin, Magnus Korpås, and Harald G. Svendsen. "A generic framework for power system flexibility analysis using cooperative game theory." *Applied energy* 212 (2018): 223-232.
- [26] Zhang, Liu, Hong, Pinhao Song, and Runwei Ding. "Towards domain generalization in underwater object detection." In *2020 IEEE international conference on image processing (ICIP)*, pp. 1971-1975. IEEE, 2021.
- [27] Kim, Sharma, Prasen, Ira Bisht, and Arijit Sur. "Wavelength-based attributed deep neural network for underwater image restoration." *ACM Transactions on Multimedia Computing, Communications and Applications* 19, no. 1 (2023): 1-23.
- [28] Tian, Wang, Yinhuai, Jiwen Yu, and Jian Zhang. "Zero-shot image restoration using denoising diffusion null-space model." *arXiv preprint arXiv:2212.00490* (2024).
- [29] Yuan, Zeng, Lingcai, Bing Sun, and Daqi Zhu. "Underwater target detection based on Faster R-CNN and adversarial occlusion network." *Engineering Applications of Artificial Intelligence* 100 (2024): 104190.