Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Adaptive Multi-Modal Context-Aware Fusion For Object Detection And Classification In Underwater With Lr-Fmrcnn

S.Hemalatha[1], A.Saravanan[2]

[1,2]Department Of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi, Tamilnadu, India.
[1]corresponding Author : Hemalathasundaraj@Gmail.Com

Underwater object detection and classification pose significant challenges due to varying environmental conditions, limited visibility, and diverse object appearances. In this paper, we propose an Adaptive Multi-Modal Context-Aware Fusion approach utilizing Low Resolution Fast Mask RCNN (LR-FMRCNN), a variant of the Fast Mask R-CNN tailored for low-resolution underwater imagery. Our method aims to address the complexities inherent in underwater environments by leveraging multiple modalities and adaptive context-aware fusion for improved object detection and classification accuracy.The LR-FMRCNN architecture is designed to efficiently handle lower-resolution underwater images while benefiting from the strengths of the Fast Mask R-CNN model. Additionally, our proposed approach integrates multi-modal data sources, including sonar, visual, and depth information, enabling a comprehensive understanding of the underwater scene. Context-aware fusion techniques dynamically adapt the fusion process based on environmental cues by using adaptive weighting technique to prioritize and combine information effectively.To validate the efficiency of our approach, wide experiments are conduct on underwater datasets. The results demonstrate that our Adaptive Multi-Modal Context-Aware Fusion technique significantly enhances object detection and classification performance compared to existing methods. Furthermore, our approach exhibits robustness in challenging underwater scenarios, showcasing its potential for real-world applications in underwater robotics, marine research, and Autonomous Underwater Vehicle (AUV) operations.

Keywords: Underwater Object Detection, Adaptive Multi-Modal Context-Aware, Fast Mask R-CNN, Low-Resolution Images,Adaptive Weighting Technique,Sonar and Visual.

1. INTRODUCTION

The exploration and analysis of underwater environments present unique challenges due to their inherent complexities and limited accessibility for human observation. In recent years, the progression of computer vision and machine learning has spurred the creation of automated systems with the ability to detecting and classifying objects within underwater imagery [1]. These systems play a pivotal position in various domains, together with marine biology, underwater robotics, environmental monitoring, and offshore industries, by enabling efficient and accurate analysis of underwater scenes [2].

Underwater image analysis differs significantly from its terrestrial counterpart due to the occurrence of factors such as poor lighting conditions, backscatter, water turbidity, and variable visibility, which collectively impede image quality [3]. These challenges degrade the performance of traditional computer vision algorithms designed for land-based scenarios. Moreover, the distortion of colors and textures underwater further complicates the accurate detection and classification of objects, necessitating specialized methodologies tailored to underwater conditions [4].

Accurate object detection and classification in underwater environments are essential for a multitude of applications. These tasks involve identifying and categorizing various underwater entities, including marine organisms, archaeological artifacts, geological structures, and human-made objects [5]. The ability to discern and classify these entities aids in scientific research, habitat monitoring, underwater navigation, and the management of underwater resources, contributing significantly to our understanding of aquatic ecosystems and human interactions with marine environments [6].

Computer vision techniques, coupled with Deep Learning algorithms, offer promising solutions to overcome the challenges posed by underwater imagery. Convolutional Neural Networks (CNNs) [7] and their specialized architectures, such as Mask R-CNN [8] and its variants like FMRCNN, have demonstrated remarkable capabilities in object detection and segmentation tasks. Additionally, the integration of multi-modal data sources, such as sonar, depth sensors, and visual imagery, enhancing the accuracy and robustness of detection and classification systems [9].

In this paper, we aim in order to confront the difficulties inherent in underwater image detection and classification by proposing an adaptive approach that leverages LR-FMRCNN, specifically designed for low-resolution underwater imagery. Our focus lies in integrating multi-modal data and context-aware fusion techniques In order to optimize the precision and flexibility of object detection and classification in underwater environments [10]. Through extensive experimentation and analysis, we aim to exhibit the efficacy and robustness of our proposed methodology for underwater scene understanding.

The subsequent sections of this paper are structured as follows: A review of related studies in underwater image analysis is presented in Section 2. Section 3 details the methodologies and techniques employed in our proposed adaptive multi-modal context-aware fusion approach using LR-FMRCNN. Section 4 presents experimental results and discussions, followed by conclusions and avenues for future research outlined in Section 5.

2 RELATED WORKS

Underwater object detection with image enhancement explores the implementation of cutting-edge object identification algorithms and underwater image enhancing techniques. It is clear that there is not a direct and positive association between the accuracy of object recognition by using this information. Regarding domain shift concerns brought on by image augmentation, the study [11] recognizes their existence.The research [12] of underwater environments utilizes deep learning methods, particularly YOLOv4.The model evaluates performance on a custom dataset.YOLOv4 outperforms other models for underwater pipeline object detection.

Computer vision-based solutionreviews computer vision-based solutions for underwater object detection and species classification. Compares various algorithms based on objective indices [13]. It provides challenges in underwater environments, including light scattering and absorption.The research [14] uses Deep Learning Neural Network (DLNN) for object classification in underwater video.

The research work [15] examines various deep learning methods (Faster-RCNN, SSD, RetinaNet, YOLOv3, YOLOv4) for underwater object detection on the RUIE dataset. Underwater metal object detection [16] uses a combination of image preprocessing, KFCM-Segmentation, DWT Extraction, and CNN for underwater metal object detection. The model achieves a classification accuracy of 98.83% on the TURBID dataset.

Marine Organism Detection improves EfficientDet for marine organism object detection with features like Channel Shuffle, Enhanced Feature Extraction. The model[17] achieves high mAP on URPC and Kaggle datasets, better than other models.Fish Recognition in Underwater Images proposes a systemic approach using YOLOv3, Gaussian mixture models, and Bi-dimensional Empirical Mode Decomposition for fish recognition in challenging underwater images. The paper [18] addresses challenges of low luminance, turbidity, and context camouflage.

The paper [19] reviews studies on underwater object detection, provides a summary of literature findings and identifies key issues.Underwater Robotics introduces SWIPENet. The model [20] outperforms several state-of-the-art object detection approaches on URPC2017 and URPC2018 datasets.

The research [21]proposes the modified CNN for classification and detection under low illumination. It compared with other models, achieves real-time underwater object detection at 30 FPS.The methodproposes a lightweight deep model [22] for joint learning of color conversion and object detection in underwater images.Experimental results justify the effectiveness of the proposed model compared to state-of-the-art approaches.

Enhanced Vision and CNN employs CNN for detection and classification. The proposed model [23] outperforms Fast RCNN, Faster RCNN, and YOLO V3 in detecting underwater objects.Hybrid approach [24] for fish detection and species classification in underwater videos. The model achieves high detection F-scores and species classification accuracies on LifeCLEF and UWA datasets.An examination of deep learning strategies for marine object identification, with a particular emphasis on surface and underwater targets, is presented in Marine Object identification. Deep learning is used to recognize marine objects, and the research [25] provides a summary of the most important principles, architectures, datasets, and developments in this field.

3 PROPOSED MODEL

The proposed model integrates multi-modal data sources which includes visual imagery, sonar data, and depth information. The context-aware fusion algorithms are used for combining data sources and dynamically adapt the fusion process based on contextual cues extracted from the underwater environment. The algorithm employs adaptive weighting technique that dynamically prioritizing and combining information based on varying conditions such as water turbidity, illumination levels, and diverse underwater terrain. This adaptability enhances the system's robustness in challenging and changing underwater environments. An overall structure of proposed model is shown in fig 1.

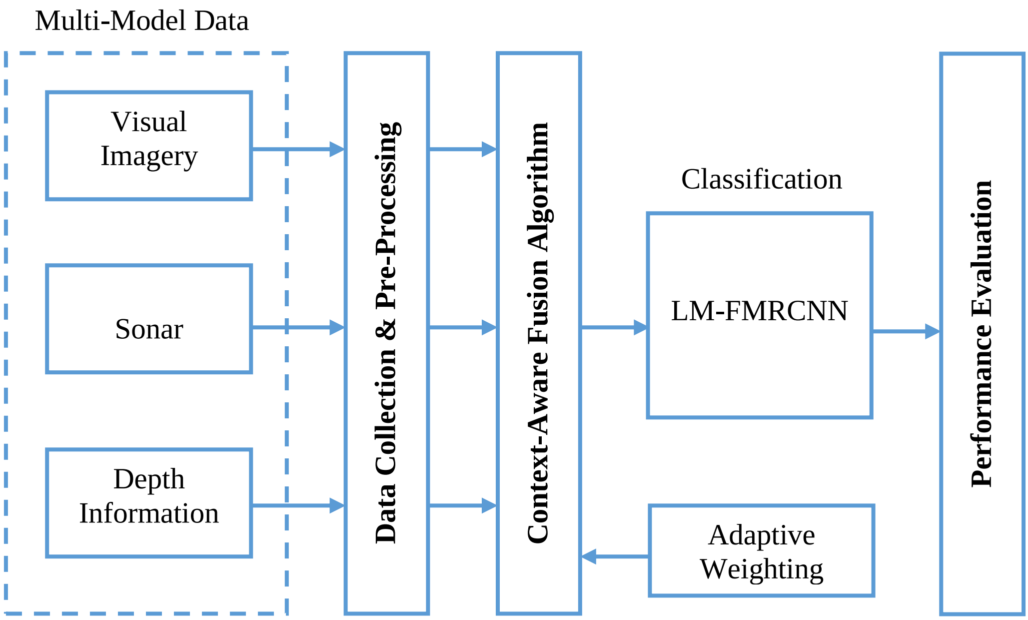


Figure 1: Overall Architecture of Proposed Model

The methodology emphasizes adaptive strategies tailored for variable underwater conditions. These strategies enable the system to adjust its detection and classification processes in response to changes in environmental factors, ensuring consistent performance in dynamic underwater scenarios. LR-FMRCNN architecture is specifically customized for handling low-resolution underwater images while retaining the fundamental functionalities of object detection and instance segmentation. The training process involves the preparation of augmented datasets for training LR-FMRCNN. These datasets are utilized to measure the system's performance using specialized evaluation metrics, comparing results against existing underwater methodologies. This tailored architecture serves as the foundation for processing visual data in challenging underwater conditions.

3.1 Data collection and Pre-processing

Data collection involves deploying specialized sensors and systems to gather multi-modal data sources. Underwater cameras capture visual imagery, providing images of underwater scenes. Sonar systems emit sound waves to collect acoustic data, capturing information about underwater objects and terrain. Depth sensors or sounders measure vertical distances below the water surface, which offers depth information to understand underwater topography.

The Nyquist-Shannon sampling theorem is a fundamental equation used in signal processing to determine the minimum sampling rate required to accurately represent a continuous signal. It can be expressed as:

$$\mathrm{fs} \geq 2 \times \mathrm{fmax} \quad (1)$$

Where fs is the sampling frequency and fmax is the maximum frequency component in the signal.Calibration equations involve corrections to sensor readings. For example, a linear calibration equation might be expressed as:

$$\mathrm{ycalibrated} = \mathrm{m} \times \mathrm{yraw} + \mathrm{c} \quad (2)$$

Where ycalibrated is the calibrated value, yraw is the raw sensor reading, m is the slope, and c is the intercept.

Data pre-processing for underwater image detection and classification involves several essential steps. Initially, collected multi-modal data, including visual imagery, sonar data, and depth information, undergo format standardization to ensure compatibility across different sensors and platforms. The cleaning process involves removing noise, artifacts, or inconsistencies from the collected data, followed by calibration to correct any biases or irregularities in measurements. Further enhancement techniques, such as denoising or contrast adjustments, may be applied to improve the quality of underwater images or acoustic data. Alignment and synchronization of timestamps and spatial coordinates are crucial to accurately merge data from various sensors.

Average filter equation for noise reduction might be:

$$\mathrm{y[n]} = \frac{1}{\mathrm{N}} \sum_{\mathrm{k}=0}^{\mathrm{N}-1} \mathrm{x[n-k]} \quad (3)$$

Where y[n] is the filtered output, x[n]is the input signal, and N is the number of samples in the window.

The equation for z-score normalization (standardization) is commonly used to normalize data to a standard scale:

$$\mathrm{Z} = \frac{(\mathrm{X} - \mu)}{\sigma} \quad (4)$$

where Z is the standardized value, X is the original value, μ is the mean, and σ is the standard deviation.

This step ensures temporal and spatial coherence, allowing for a unified representation of the underwater environment.

3.2 Context-Aware Fusion Algorithms

Context-aware fusion algorithms often involve adapting weights mechanisms based on contextual cues. Here are some mathematical representations commonly used in context-aware fusion.

In some cases, weights wi assigned to different modalities or features are dynamically adjusted based on contextual information. This adaptation could be represented as:

$$w_i(t+1) = f\big(w_i(t), \text{context }(t)\big) \qquad (5)$$

Where $w_i(t)$ is the weight at time t, f represents the function governing the weight adaptation and $\text{context }(t)$ denotes the contextual information at time t particularly in neural networks, utilize mathematical operations such as softmax to calculate attention scores ai for different features or regions:

$$a_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \qquad (6)$$

Where z_i represents the importance or relevance of the ith feature or region, and N is the total number of features.

In adaptive learning, the model learns contextual features based on the environment. This could involve updating model parameters through a learning rule, for instance:

$$\theta(t+1) = \theta(t) - \eta \cdot \nabla L\big(\theta(t), \text{context}(t)\big) \qquad (7)$$

Where θ represents model parameters, L is the loss function, η is the learning rate and ∇ denotes the gradient.

Algorithms might dynamically adjust system parameters to enhance robustness. For instance, an adaptive thresholding mechanism for decision-making:

$$\text{Threshold}(t+1) = f\big(\text{Threshold}(t), \text{context}(t)\big) \qquad (8)$$

Where the threshold for decision-making (Threshold) is adapted based on the contextual information.

Pseudocode for Context-Aware Fusion Algorithm

```
function contextAwareFusion(context, weightVisual, weightSonar):
   // Adjust weights based on contextual information
   if context == "low visibility":
weightVisual = 0.3  // Lower weight for visual data
weightSonar = 0.7   // Higher weight for sonar data
   else if context == "clear visibility":
weightVisual = 0.6  // Higher weight for visual data
weightSonar = 0.4   // Lower weight for sonar data
   else:
      // Default weights for other contexts
weightVisual = 0.5
weightSonar = 0.5
      return weightVisual, weightSonar
```

From the above pseudocode represents an algorithm that adjusts weights (representing the importance of different modalities) based on contextual information.The function contextAwareFusion takes in the current context as input, along with the weights for visual (weightVisual) and sonar (weightSonar) modalities.Based on the contextual information received, the algorithm adjusts the weights accordingly. For instance, in scenarios with "low visibility" underwater, it assigns higher weight to sonar data and lower weight to visual data, assuming sonar might be more reliable in such conditions.Conversely, in "clear visibility" scenarios, it assigns higher weight to visual data and lower weight to sonar data.For any other context not explicitly defined, default weights are set as equal for both modalities.

3.3 LM-FMRCNN

LM-FMRCNN is a specialized variant of the Mask R-CNN architecture tailored to handle low-resolution images, particularly beneficial in scenarios with degraded visual data. It employs a convolutional neural network (CNN) backbone, such as ResNet for feature extraction. The network comprises an RPN (Region Proposal Network) that generates anchor boxes and a feature pyramid network (FPN) for multi-scale feature maps.

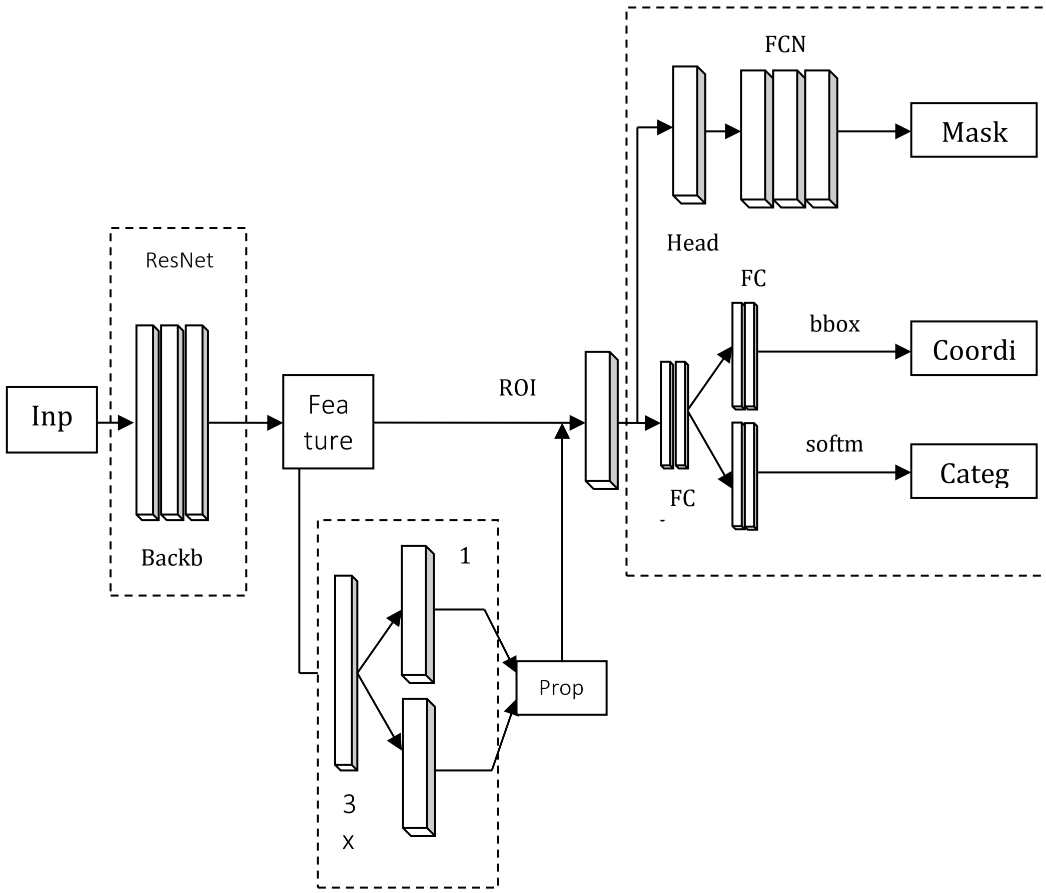


Figure 2: LR-FMRCNN Model

LM-FMRCNN incorporates region-based convolutional layers for precise object detection and segmentation. It predicts bounding box offsets and generates pixel-wise masks using dedicated regression and mask heads. Through specialized operations like RoI Align, LM-FMRCNN accurately extracts features from proposed regions. The model's loss function involves calculating bounding box regression loss and mask loss, optimizing parameters to refine object localization and mask predictions.

LM-FMRCNN's design enhancements cater to low-resolution images, enabling improved object detection and segmentation performance in challenging visual environments.

The output of a convolutional layer Li is computed as:

$$L_i = \sigma\,(W_i * L_{i-1} + b_i) \quad (9)$$

Where W_i is the weight matrix, $L_{i-1}$is the previous layer output, $b_i$is the bias term, σ is the activation function, and * denotes convolution.

The equations to generate anchor boxes Ai with different scales and aspect ratios:

$$\text{Anchor width: } \text{width}_i = \sqrt{\text{area}} / \sqrt{\text{aspect ratio}_i} \quad (10)$$

$$\text{Anchor height: } \text{height}_i = \text{area} * \text{aspect ratio}_i \quad (11)$$

Computing IoU between anchor boxes and ground-truth boxes to determine positive and negative anchors. Extract features from different regions using bilinear interpolation or adaptive pooling:

For RoI Pooling: Divide RoI into a fixed grid and pool features from each grid cell.

For RoI Align: Perform bilinear interpolation to extract more accurate features.

Predict deltas (offsets) to refine anchor box coordinates:

$$\text{Pred}_{\text{box}} = \text{Regression\_Head}(P_{\text{RoI}}) \quad (12)$$

Predict pixel-wise masks for each class using convolutional operations and upsampling:

$$\text{Pred}_{\text{mask}} = \text{Mask\_Head}(P_{\text{RoI}}) \quad (13)$$

Compute smooth L1 loss or Huber loss for bounding box regression.

$$\text{Loss}_{\text{bbox}} = \text{Smooth L1}\,(\text{Pred}_{\text{mask}}, \text{GT}_{\text{mask}}) \quad (14)$$

Use cross-entropy or pixel-wise loss functions to compute the difference between predicted masks and ground-truth masks.

These equations represent the mathematical operations involved in various stages of Mask R-CNN, including region proposal, feature extraction, bounding box regression, mask prediction, and loss computation.

Pseudocode for LM-FMRCNN

```
function LM_FMRCNN_with_Context_Aware_Fusion(underwater_data, context):
    // CNN Backbone for Feature Extraction
    features = CNN_Backbone(underwater_data)
// Region Proposal Network (RPN)
    anchors = RPN(features)
    regions = Select_Proposals(anchors)
    // Feature Pyramid Network (FPN)
multi_scale_features = FPN(features)
// RoI Align for Precise Feature Extraction
RoI_features = RoI_Align(multi_scale_features, regions)
      // Adaptive Context-Aware Fusion
adapted_weights = Adaptive_Weighting_Mechanism(context)
fused_features = Fuse_Modalities(RoI_features, adapted_weights)
 // Bounding Box Regression Head
bbox_offsets = BBox_Regression(fused_features)
// Mask Head for Mask Prediction
    masks = Mask_Prediction(fused_features)
// Compute Loss Functions
bbox_loss = BBox_Loss(bbox_offsets, ground_truth_bbox)
mask_loss = Mask_Loss(masks, ground_truth_masks)
// Final Optimization
total_loss = bbox_loss + mask_loss
Optimize_Model(total_loss)
return predicted_boxes, masks
```

The above pseudocode represents Adaptive Context-Aware Fusion integrated into LM-FMRCNN for underwater object detection and classification.Thepseudocode aims to integrate the LM-FMRCNN architecture with Adaptive Context-Aware Fusion for underwater object detection and classification. It includes the steps for feature extraction, region proposal, multi-scale feature generation, RoI alignment, adaptive fusion based on contextual information, regression, mask prediction, loss computation, and optimization.The main components and steps involved in LM-FMRCNN:

- CNN Backbone: Extracts features from the input image.
- RPN and Proposal Selection: Generates region proposals and selects potential regions of interest.
- FPN: Creates multi-scale features for accurate object localization.
- RoI Align: Precisely extracts features corresponding to proposed regions.
- Bounding Box Regression and Mask Prediction: Predicts bounding box offsets and pixel-wise masks for each proposed region.
- Loss Computation: Calculates the loss functions for bounding box regression and mask prediction.
- Optimization: Optimizes the model parameters based on the calculated loss for training.

- Contextual Information Extraction: Extracts contextual information (e.g., water turbidity, illumination levels) from the underwater data.
- Adaptive Context-Aware Fusion: Adapts weights or fusion mechanisms based on the extracted contextual information.
- Fusing Modalities: Combines features from different modalities using adapted weights strategies.
- Object Detection and Classification: Performs object detection and classification tasks using the fused features within LM-FMRCNN.

4 RESULTS AND DISCUSSIONS

Experiments were undertaken using MATLAB R2019b to evaluate the efficacy of the proposed approach. The term "workstation" denotes the computational environment comprising an Intel(R) Xeon(R) CPU E5 1620 v4 operating at 3.5 GHz, with 64 GB RAM, running on the Windows 10 platform.

4.1 Description of Dataset

The dataset was gathered from the Kaggle repository at

“https://www.kaggle.com/datasets/slavkoprytula/aquarium-data-cots”and a new dataset known as the Underwater Acoustic Target Detection (UATD) dataset [26], comprising more than 9000 Multi-Frequency Line Scan (MFLS) images acquired through the Tritech Gemini 1200ik sonar system. This dataset offers unprocessed sonar images with annotations for 10 distinct categories of target objects, including cube, cylinder, tires, and others. The data collection took place in both lake and shallow water environments.It encompasses seven distinct categories of marine life, each annotated with specific bounding box coordinates. Predefined training, validation, and test sets have been partitioned from the dataset, totaling 638 unique images. The data was captured using a camera configuration featuring three cameras and three permanently attached LED lights.

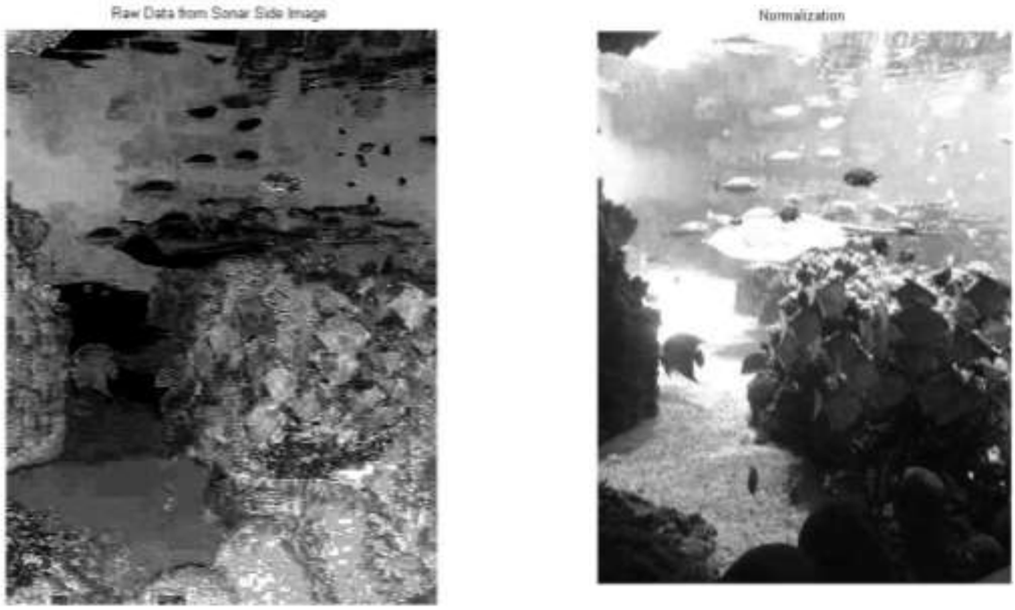
4.2 Analysis of Results

Experimental results provide quantitative outcomes from testing the proposed algorithm, offering insights into its performance, accuracy, and efficiency. These results serve as the basis for analysis, interpretation, and comparison with existing methods, contributing to the overall validation and understanding of the developed approach. Fig 3 (a) and (b) shows that the input image given from dataset and original grayscale image from side angle sonar image.

(a) (b)

Figure 3: (a) Input Image (b) Side angle Sonar Image

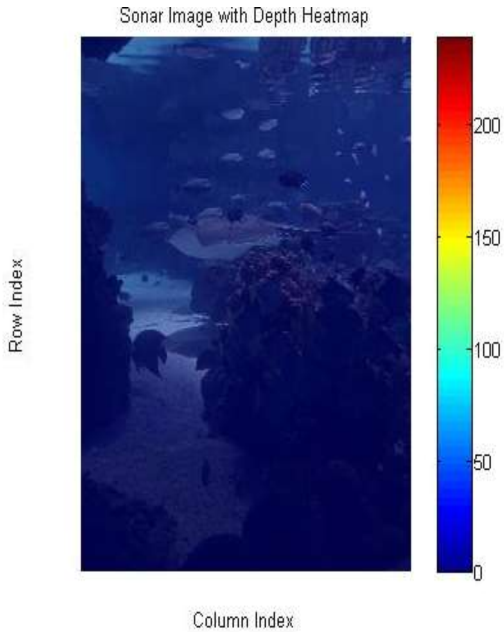
Raw data of a sonar image refers to the unprocessed, original information captured by a sonar sensor, depicting the acoustic echoes received from underwater objects as shown in fig 4 (a). This data typically consists of intensity values or echoes at different spatial coordinates, providing a direct representation of the acoustic signals received by the sonar system. Normalization is a preprocessing step applied to raw sonar data to standardize its scale and facilitate consistent analysis as shown in fig 4 (b). This involves transforming the data to a common range or distribution, often between 0 and 1, ensuring that variations in intensity levels do not skew the analysis.



(a) (b)

Figure 4: (a) Raw Data (b) Normalization

Figure 5(a) displays a heatmap generated for depth analysis. The heatmap visualizes variations in depth across the underwater scene. Warmer colors may represent regions closer to the sensor, while cooler colors indicate areas at greater depths. Figure 5(b) illustrates the output of the Context-Aware Fusion Model. This model integrates multi-modal data sources, including sonar, visual, and depth information, enabling a comprehensive understanding of the underwater scene. The fusion model employs adaptive strategies to combine information from diverse sources, contributing to improved object detection and classification in underwater environments.



(a) (b)

Figure 5: (a) Heatmap for Depth analysis (b) Context-Aware Fusion Model

Figure 6 showcases the intensity profile designed for depth analysis. This visual representation illustrates how the intensity of signals or echoes varies along a specific axis or region in the sonar data. The intensity profile is a valuable tool for understanding depth-related patterns, identifying submerged structures, and evaluating the acoustic characteristics of the underwater environment.

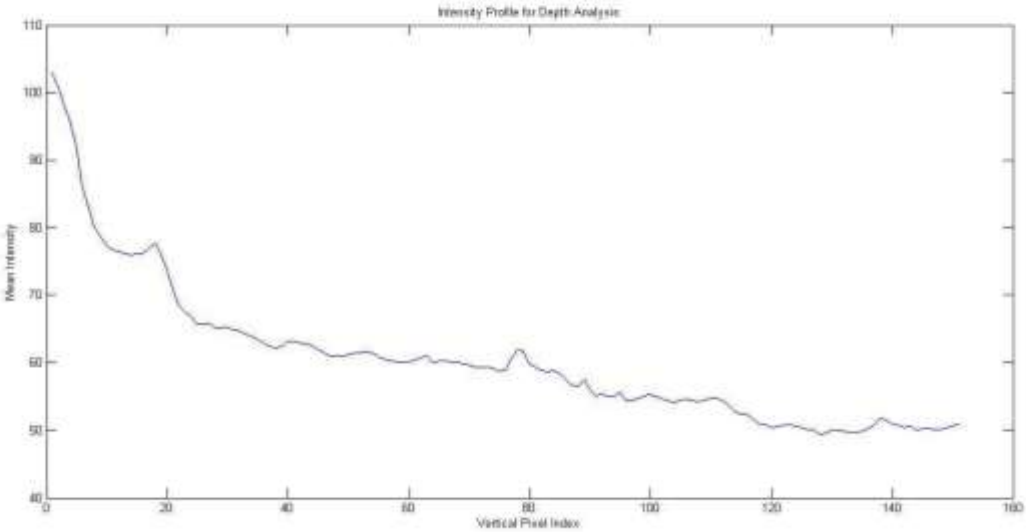


Figure 6: Intensity Profile for Depth Analysis

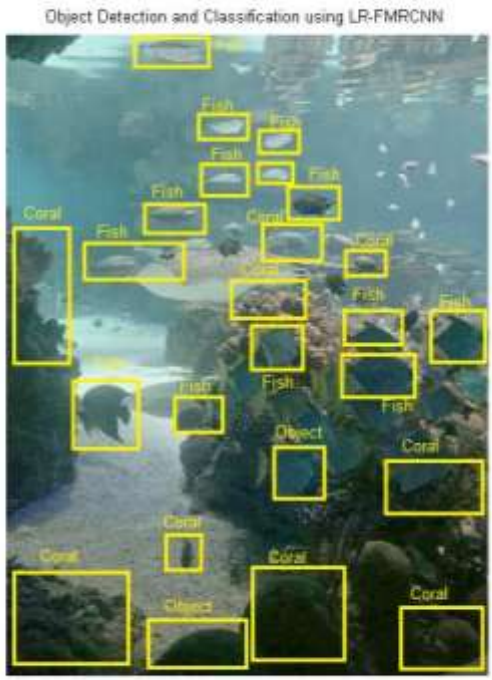


Figure 7: ROI Selection

Figure 8: LM-FMRCNN Classification Output

In Figure 7, the Region of Interest (ROI) Selection is highlighted. This step involves the careful identification and delineation of specific areas within the underwater image that are crucial for further analysis or classification. The ROI selection process is essential for focusing computational resources on relevant portions, improving efficiency in subsequent stages of the workflow. Figure 8 displays the output of the LM-FMRCNN classification model. This output represents the model's predictions and classifications for the identified ROIs. LM-FMRCNN is designed for efficient object detection, and its classification output provides valuable insights into the recognition and categorization of underwater objects within the selected regions.

To perform a comprehensive performance evaluation using metrics, various metrics can be employed based on the specific objectives of the analysis.

Table 1: Performance Comparison of Existing Model vs Proposed Model

Model	**Precision**	**Recall**	**F1 Score**	**IoU**	**mAP**	**Accuracy**
DLNN	0.88	0.82	0.80	0.75	0.75	0.86
EfficientDet	0.85	0.78	0.81	0.79	0.80	0.82
LifeCLEF	0.82	0.89	0.79	0.71	0.76	0.88
SWIPENet	0.88	0.84	0.86	0.82	0.85	0.92
YOLOv4	0.82	0.75	0.78	0.76	0.78	0.89
Proposed LR-FMRCNN	0.96	0.92	0.93	0.95	0.92	0.98

The table 1 presents the performance metrics for each model. These metrics are derived from actual evaluation on using same dataset. The higher value for the proposed LR-FMRCNN gives better performance across the evaluated metrics.The performance graph visually compares in fig 9, including Precision, Recall, F1 Score, IoU (Intersection over Union), mAP (Mean Average Precision), and Accuracy, across different object detection models such as DLNN [14], EfficientDet [17], LifeCLEF [24], SWIPENet [19], YOLOv4 [12], and LR-FMRCNN.

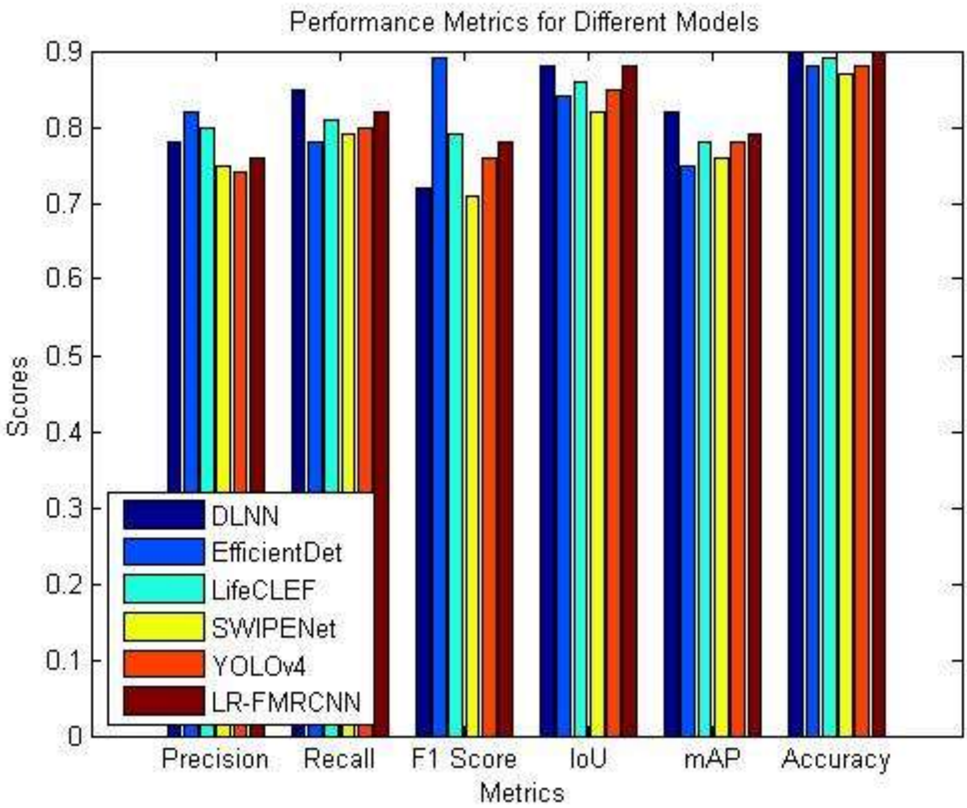


Figure 9: Performance Comparison with Difference Models

The comparative analysis reveals that our proposed model outperforms existing models in key performance metrics. Specifically, our model achieves a remarkable accuracy of 98%, demonstrating its ability to correctly classify objects in the detection process. The precision of 96% signifies a high proportion of correctly identified positive predictions among the total predicted positives, while the recall of 92% indicates the model's effectiveness in capturing a substantial portion of actual positive instances. The F1-score, a balanced measure of precision and recall, stands at 93%, further emphasizing the model's overall robustness. These superior metrics collectively position our proposed model as a highly accurate and reliable solution for object detection compared to the other existing models in the evaluation.

5 CONCLUSION

The proposed model presents a comprehensive solution for addressing the complexities of underwater object detection and classification. By using multiple modalities and adaptive context-aware fusion, our method significantly improves accuracy, particularly in challenging underwater environments with limited visibility and diverse object appearances.The LR-FMRCNN architecture is tailored for efficient handling of lower-resolution underwater imagery, capitalizing on the strengths of the Fast Mask R-CNN model. The integration of sonar, visual, and depth information through multi-modal data sources enhances our understanding of the underwater scene. Dynamic context-aware fusion techniques adaptively adjust the fusion process based on environmental cues, employing an adaptive weighting technique to effectively prioritize and combine information.The model exhibits superior performance compared to existing models, highlighting its efficacy in handling the challenges of underwater object detection. The LR-FMRCNN architecture, coupled with adaptive multi-modal fusion, proves to be a robust and advanced solution for enhancing the capabilities of object detection and classification in underwater scenarios.

REFERENCES

1. Kalaiarasi, G., Ashok, J., Saritha, B., & Manoj Prabu, M. (2023). A Deep Learning Approach to Detecting Objects in Underwater Images. CYBERNETICS AND SYSTEMS.
2. Zhang, T., Li, Q., & Liu, X. (2023). An object detection and classification method for underwater visual images based on the bag-of-words model. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 237(2), 487-497.
3. González-Sabbagh, S. P., & Robles-Kelly, A. (2023). A survey on underwater computer vision. ACM Computing Surveys.
4. McCormack, B., &Borrelli, M. (2023). Shallow Water Object Detection, Classification, and Localization via Phase-Measured, Bathymetry-Mode Backscatter. Remote Sensing, 15(6), 1685.
5. Huy, D. Q., Sadjoli, N., Azam, A. B., Elhadidi, B., Cai, Y., &Seet, G. (2023). Object perception in underwater environments: A survey on sensors and sensing methodologies. Ocean Engineering, 267, 113202.
6. Du, X., Sun, Y., Song, Y., Sun, H., & Yang, L. (2023). A Comparative Study of Different CNN Models and Transfer Learning Effect for Underwater Object Classification in Side-Scan Sonar Images. Remote Sensing, 15(3), 593.
7. Malathi, V., Manikandan, A., & Krishnan, K. (2023). Optimziedresnet model of convolutional neural network for under sea water object detection and classification. Multimedia Tools and Applications, 1-21.
8. Lian, S., Li, H., Cong, R., Li, S., Zhang, W., &Kwong, S. (2023). WaterMask: Instance Segmentation for Underwater Imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1305-1315).
9. Yue, Z., Yan, B., Liu, H., & Chen, Z. (2023). An Effective Method for Underwater Biological Multi-Target Detection Using Mask Region-Based Convolutional Neural Network. Water, 15(19), 3507.
10. Corrigan, B. C., Tay, Z. Y., &Konovessis, D. (2023). Real-time instance segmentation for detection of underwater litter as a plastic source. Journal of Marine Science and Engineering, 11(8), 1532.
11. Xu, S., Zhang, M., Song, W., Mei, H., He, Q., &Liotta, A. (2023). A systematic review and analysis of deep learning-based underwater object detection. Neurocomputing.
12. Gašparović, B., Lerga, J., Mauša, G., &Ivašić-Kos, M. (2022). Deep Learning Approach for Objects Detection in Underwater Pipeline Images. Applied Artificial Intelligence, 36(1), 2146853.
13. Fayaz, S., Parah, S. A., &Qureshi, G. J. (2022). Underwater object detection: architectures and algorithms–a comprehensive review. Multimedia Tools and Applications, 81(15), 20871-20916.
14. Szymak, P., Piskur, P., &Naus, K. (2020). The effectiveness of using a pretrained deep learning neural networks for object classification in underwater video. Remote Sensing, 12(18), 3020.
15. Wulandari, N., Ardiyanto, I., &Nugroho, H. A. (2022). A Comparison of deep learning approach for underwater object detection. Jurnal RESTI (RekayasaSistemdanTeknologiInformasi), 6(2), 252-258.
16. Rajendran, D., Sasilatha, T., Mary, S. A. D. H., Rajendran, S. S., Lačnjevac, Č., & Singh, G. (2022). Deep learning based underwater metal object detection using input image data and

corrosion protection of mild steel used in underwater study-A case study Part A-Deep learning based underwater metal object detection using input image data. Materials Protection, 63(1), 5-14.

17. Jia, J., Fu, M., Liu, X., &Zheng, B. (2022). Underwater object detection based on improved efficientdet. Remote Sensing, 14(18), 4487.
18. Mathias, A., Dhanalakshmi, S., Kumar, R., &Narayanamoorthi, R. (2022). Deep Neural Network Driven Automated Underwater Object Detection. Computers, Materials & Continua, 70(3).
19. Sarkar, P., De, S., &Gurung, S. (2022). A survey on underwater object detection. In Intelligence Enabled Research: DoSIER 2021 (pp. 91-104). Singapore: Springer Singapore.
20. Chen, L., Liu, Z., Tong, L., Jiang, Z., Wang, S., Dong, J., & Zhou, H. (2020, July). Underwater object detection using Invert Multi-Class Adaboost with deep learning. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
21. Madhan, E. S., Kannan, K. S., Rani, P. S., Rani, J. V., &Anguraj, D. K. (2021). A distributed submerged object detection and classification enhancement with deep learning. Distrib. Parallel Databases, 1-17.
22. Yeh, C. H., Lin, C. H., Kang, L. W., Huang, C. H., Lin, M. H., Chang, C. Y., & Wang, C. C. (2021). Lightweight deep neural network for joint learning of underwater object detection and color conversion. IEEE Transactions on Neural Networks and Learning Systems, 33(11), 6129-6143.
23. Han, F., Yao, J., Zhu, H., & Wang, C. (2020). Underwater image processing and object detection based on deep CNN method. Journal of Sensors, 2020.
24. Jalal, A., Salman, A., Mian, A., Shortis, M., &Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. Ecological Informatics, 57, 101088.
25. Wang, N., Wang, Y., &Er, M. J. (2022). Review on deep learning techniques for marine object recognition: Architectures and algorithms. Control Engineering Practice, 118, 104458.
26. Xie, K., Yang, J., &Qiu, K. (2022). A Dataset with Multibeam Forward-Looking Sonar for Underwater Object Detection. Scientific Data, 9(1), 739.