IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Underwater Image Enhancement And Object Detection With Multi-Color Space Residual Network And Swin-Yolo Fusion

[1]Sreedevi V T, [2]Aby Abahai T, [3]Kripa Joy

[1]Student, [2]Assistant Professor, [3]Assistant Professor

[1]Department of computer science,

[1] Mar Athanasius College of Engineering Kerala, India

Abstract: Object recognition in underwater environment is extremely difficult because of problems like color distortion, light scattering, and decreased visibility. A hybrid approach that combines Multicolor Space Residual Networks (MCRNet) with Swin-YOLO fusion to increase object recognition accuracy and underwater image quality. Swin Transformer uses self-attention techniques to extract multi-scale hierarchical features, whereas MCRNet uses residual learning and different color spaces (RGB, HSV, and Lab) to restore texture details and correct color distortions. By efficiently improving feature representations, the Swin-YOLO fusion improves YOLOv8's real-time object identification capabilities. According to experimental findings, the suggested strategy works noticeably better than stand-alone techniques, which makes it a viable option for marine applications and underwater research.

Index Terms - Underwater Image Enhancement, Object Detection, MCRNet, Swin Transformer, YOLOv8, Multi-Color Space Processing, Feature Fusion, Deep Learning, Computer Vision, Marine Applications

I. INTRODUCTION

UNDERWATER imaging is essential for applications in marine research, archaeology, surveillance, and au- tonomous navigation. However, the unique optical properties of water—such as light absorption, scattering, color distor- tion, and noise—significantly degrade image quality, posing challenges for accurate object detection and analysis. These limitations necessitate advanced enhancement and detection techniques to improve visibility and recognition in underwater environments.

To address these challenges, we propose a two-stage frame- work that integrates Multi-Color Space Residual Network (MCRNet) for underwater image enhancement and Swin- YOLO fusion for robust object detection. MCRNet leverages multiple color spaces (RGB, HSV, and Lab) to correct distortions, restore color fidelity, and enhance texture details while preserving essential features. Swin-YOLO fusion combines the Swin Transformer's hierarchical feature extraction and self- attention mechanisms with YOLO's real-time detection capabilities, enabling precise object localization and classification in underwater conditions. This hybrid approach enhances image quality and detection accuracy while maintaining computational efficiency, making it suitable for real-time deployment. The proposed system has wide-ranging applications in **autonomous underwater vehicles (AUVs), marine ecosystem monitoring, underwater archaeology, defense surveillance, and search-and-rescue op- erations**, contributing to improved exploration, conservation, and security in underwater environments.

II. LITERATURE SURVEY

Recent advancements in underwater image enhancement focus on overcoming challenges such as haze, low contrast, and color distortions caused by light scattering and absorption. Traditional single-channel methods often fail to restore image clarity effectively, leading to the emergence of multi-color space-based approaches that utilize RGB, HSV, and Lab color spaces. These methods extract complementary information from different color spaces, ensuring better color restoration, contrast enhancement, and object visibility. Recent studies leverage deep learning models and adaptive enhancement strategies to improve the quality and visibility of underwater images, making them more suitable for applications in marine research, robotics, and surveillance.

MCRNet[1], introduced by Ningwei Qin et al. (2024), employs a multi-color space residual network that enhances underwater images using RGB, HSV, and Lab color spaces. This approach mitigates color distortions and enhances struc- tural details by extracting unique information from each color space. The model utilizes a multi-branch residual learning framework that progressively refines the image, achieving balanced color correction and improved texture representation. The fusion of information from different color spaces enables MCRNet to outperform single-color channel methods, produc- ing natural and visually appealing underwater images. Another approach, Adaptive Standardization and Normalization Net- works (ASNN)[2], proposed by Cheol Woo Park and Il Kyu Eom (2024), addresses light attenuation issues in underwater imaging by applying adaptive standardization to equalize pixel values across the image and normalization networks to adjust overall intensity based on varying underwater environments. This method significantly enhances contrast and detail visibil- ity while reducing haze and color distortion, making it highly effective across diverse water conditions.

Similarly, Adaptive Color Correction and Model Conversion for Dehazing, introduced by Yiming Lia et al. (2024)[3], com- bines adaptive color correction with haze removal techniques to improve underwater image quality. The model dynamically adjusts colors based on environmental characteristics, reducing excessive green or blue tints, while the dehazing method re- moves haze caused by light scattering in water. This approach improves both color accuracy and image clarity, making it highly effective for underwater exploration and marine re- search applications. In addition, Ming Zhou et al. (2023)

[4] propose a lightweight object detection framework that integrates image restoration and color transformation. Their model recovers image clarity by reducing noise and improving contrast while applying color transformation techniques to en- hance object visibility in underwater environments. Designed for real-time applications, this framework is computationally efficient, making it suitable for underwater robotics, marine biology studies, and surveillance. CEH-YOLO [5], developed by Jiangfan Feng and Tao Jin (2024), presents an optimized YOLO-based model for underwater object detection. The model introduces composite feature extraction techniques and robust loss functions that significantly improve detection accuracy. By incorporating pre- processing techniques that enhance image visibility before detection, CEH-YOLO ensures reliable object detection in challenging underwater conditions, making it valuable for applications such as marine species identification, underwater navigation, and security surveillance. CFENet [6], proposed by Xun Jia et al. (2024), presents a cost-effective image enhance- ment network that employs cascaded feature extraction. This approach progressively refines image details while minimiz- ing computational costs, ensuring that both global and local features are preserved. CFENet is particularly suitable for real- time, resource-constrained applications, such as underwater drones and remote sensing devices.

Another innovative method, INSPIRATION [7], introduced by Hao Wang et al. (2024), applies reinforcement learning- based enhancement driven by human visual perception. This approach automatically adjusts enhancement parameters based on perceptual feedback, learning optimal contrast, bright- ness, and color balance adjustments to improve clarity. By mimicking human vision, INSPIRATION produces visually natural enhancements, making it highly suitable for applica- tions in marine exploration and underwater surveillance. Sim- ilarly, Multi-Scale Fusion and Adaptive Gamma Correction, proposed by Dan Zhang et al. (2023), addresses low-light underwater imaging challenges. This model extracts multi-scale features to improve both local and global visibility while adjusting brightness dynamically based on environmen- tal lighting conditions. This adaptive strategy mitigates the loss of detail in dim underwater settings, making it highly effective for deep-sea exploration and underwater monitoring systems. Another noteworthy approach, proposed by Manigandan Muniraj and Vaithiyanathan Dhandapani (2024) [8], involves

modified color correction combined with an adaptive look- up table (LUT) and edge-preserving filtering. The model neutralizes dominant green and blue tones, fine-tunes color representation through LUT adjustments, and preserves object boundaries and fine structural details using edge-preserving filtering. This hybrid method significantly enhances color accuracy and image sharpness, making it highly effective for underwater object detection and scientific analysis.

These studies highlight the importance of multi-color space processing, adaptive normalization, and deep learning-based enhancement techniques in improving underwater image qual- ity. However, there remain challenges in ensuring robustness across diverse water conditions, such as varying turbidity levels and lighting conditions. Additionally, integrating real- time processing capabilities is essential for practical applica- tions in autonomous underwater vehicles (AUVs) and marine research. Future research should also explore the combination of multi-modal imaging techniques, such as thermal and hy- perspectral imaging, to further enhance underwater vision. By advancing hybrid enhancement models and AI-driven adaptive processing, researchers can develop more effective solutions for underwater imaging, ultimately benefiting marine biology, surveillance, and underwater navigation.

III. Proposed Method

Integrating Multi-Color Space Residual Networks (MCR- Net) with Swin-YOLO fusion for enhanced underwater ob- ject detection leverages the strengths of multiple advanced techniques. MCRNet utilizes multiple color spaces (HSV, LAB, RGB) to extract and refine image features, addressing underwater image challenges such as low contrast and color distortion. The Residual Enhancement Block (REB) further enhances these features, improving texture and color rep- resentation. Swin Transformer is employed for hierarchical feature extraction using self-attention mechanisms, capturing both local and global contextual information. These enhanced features are then fused with YOLOv8, a state-of-the-art object detection model, to accurately detect and localize objects in the enhanced images. This integration significantly improves object detection performance, particularly in complex under- water environments, offering real-time and robust results.

The proposed system, outlined in the diagram Fig 1., follows a multi-step process for enhanced underwater object detection using deep learning techniques. Figure 3.1 illustrates an image processing pipeline designed to enhance images and detect objects within them. The process begins with preprocessing steps like noise reduction and image augmentation to improve image quality and diversity. The image is then converted into multiple color spaces to extract diverse features. A convolu- tional neural network (CNN) is employed alongside the Swin Transformer to extract relevant features, with batch normaliza- tion and ReLU activation functions aiding in training stability and non-linearity. Swin Transformer's hierarchical multi-scale representation enhances feature extraction, capturing spatial dependencies effectively. The extracted features from different color spaces are fused using Multi-Channel Feature Fusion (MCFF) to create a more comprehensive representation.

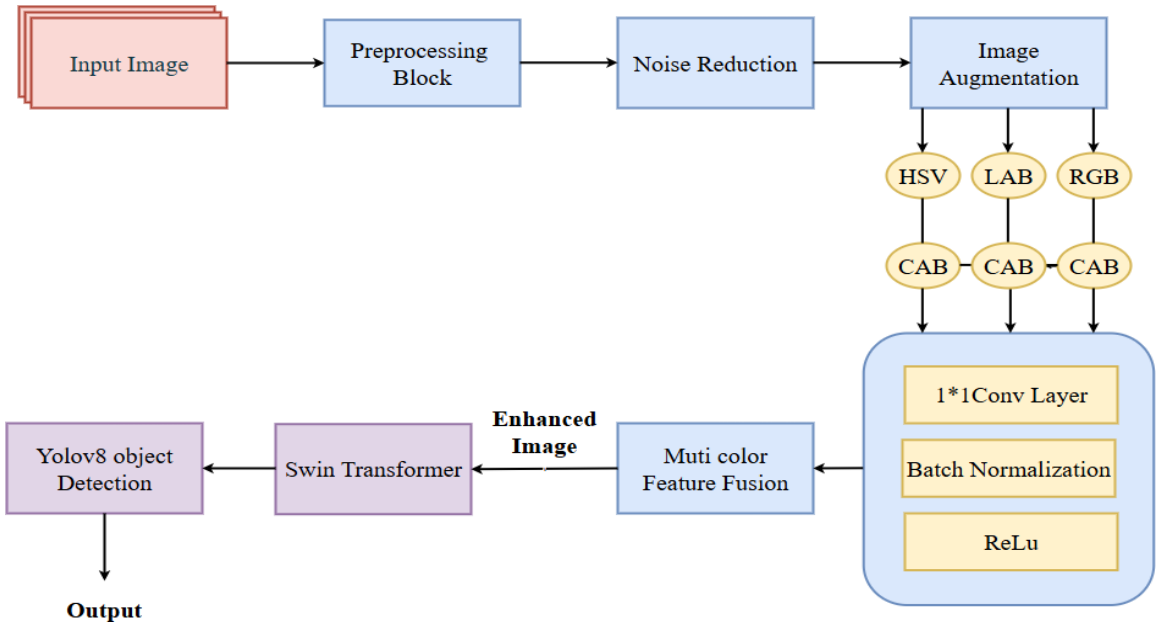

Fig. 1. **Proposed Architecture Diagram**

Subsequently, the Swin-YOLO fusion integrates the re- fined feature maps from Swin Transformer with YOLOv8's object detection framework to improve detection accuracy and robustness. Post-processing techniques like non-maximum suppression and confidence filtering are applied to refine the detected objects and eliminate redundant or low-confidence detections. The final output comprises the enhanced image along with the detected objects, their bounding boxes, and associated confidence scores.

Here's a breakdown of the different components and their function in the system:

1. INPUT PREPROCESSING BLOCK:

Noise reduction is a crucial step in preprocessing underwater images, as it helps eliminate unwanted distortions and enhances the quality of the input data. By minimizing noise, the clarity of the image improves, making it easier for the model to process and extract relevant features. Additionally, image augmentation techniques, such as rotations, scaling, and color adjustments, are applied to the input images. These augmentations create a more diverse dataset, which is beneficial for training the model to handle various real-world conditions. This improves the model's robustness and its ability to generalize across different underwater environments.

2. COLOR SPACE CONVERSION (HSV, LAB, RGB):

The system works with three different color spaces—HSV, LAB, and RGB—to capture a wider range of color features in underwater images. The system extracts features from each of these color spaces separately. The Convolutional Attention Block (CAB) plays a critical role in the model by focusing on learning important spatial features from each of the color space representations, including HSV, LAB, and RGB. This block enhances the model's ability to capture significant details specific to each color space, improving feature extraction. Following this, the Residual Enhancement Block (REB) is employed to refine and enhance the features extracted by the CAB. By leveraging residual learning, the REB improves texture and color representation, which is particularly crucial in underwater imaging conditions where accurate color restoration and detailed texture capture are often challenging.

3. 1x1 Convolution Layer, Batch Normalization,

The 1x1 Convolution Layer is used to reduce the dimension- ality of the features while preserving the necessary information. It achieves this by applying a convolution operation with a kernel size of 1x1, which allows the model to combine infor- mation across different channels without affecting the spatial dimensions of the input. This operation reduces the number of channels while retaining key feature representations. Batch Normalization, on the other hand, helps stabilize the learning process by normalizing the activations of each layer. It normalizes the input to the activation function by ensuring that the output has a mean of zero and a variance of one. This process makes the model more efficient by accelerating training and reducing sensitivity to the initial weight values. Batch Normalization also includes learnable scaling and shifting parameters that allow the network to recover its representational power after normalization

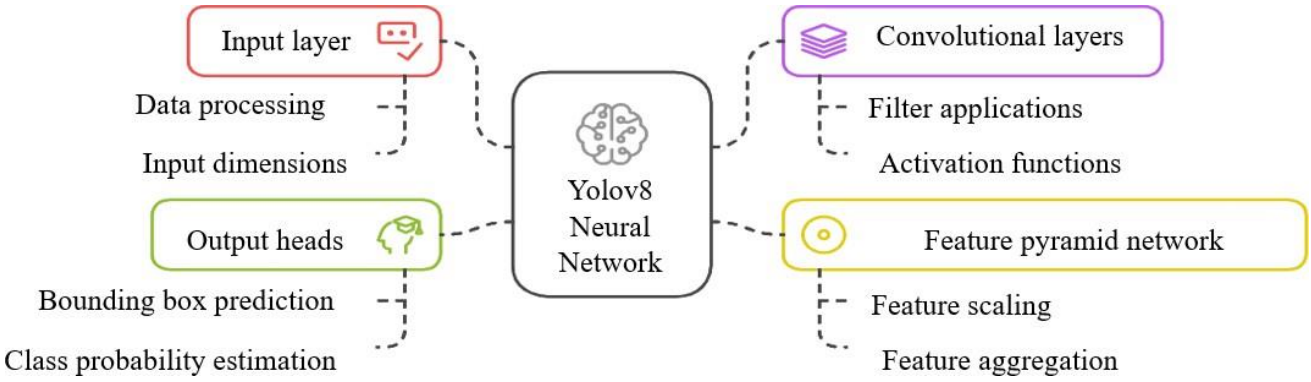

Fig. 2. **Yolov8 Neural Network**

4. MCFF (MULTI-COLOR FEATURE FUSION):

In this step, features extracted from multiple color spaces—HSV, LAB, and RGB—are fused together to create a comprehensive and robust feature map. Each color space offers unique advantages: HSV helps highlight hue-related features, LAB captures perceptual color differences, and RGB provides detailed color information. By combining these features, the model benefits from the strengths of each color space, ensuring a more accurate and complete representation of the underwater scene. This multi-color space feature fusion improves the model's ability to handle various underwater imaging challenges, such as color distortion, low contrast, and poor visibility, ultimately enhancing object detection performance.

4. LOSS FUNCTION):

To train MCRNet, we use the following loss functions. We use $\hat{I}$ to represent the enhanced image, and I^* to represent the corresponding ground truth image.

A. Charbonnier Loss

L1 loss, which determines the separation between the augmented image and the matching ground truth image,

$$L_{char} = \sqrt{\|\hat{I} - I^*\|^2 + \epsilon^2}$$

where ϵ is empirically set to 10^{-3}.

B. SSIM Loss

The Structural Similarity Index (SSIM) evaluates two im- ages' structure, contrast, and brightness. The following is the SSIM formula:

$$SSIM(\hat{I}, I^*) = \frac{(2\mu_{\hat{I}}\mu_{I^*} + c_1)(2\sigma_{\hat{I}I^*} + c_2)}{(\mu_{\hat{I}}^2 + \mu_{I^*}^2 + c_1)(\sigma_{\hat{I}}^2 + \sigma_{I^*}^2 + c_2)}$$

where $\mu_{\hat{I}}$ and μ_{I^*} denote the mean of the enhanced image and the corresponding ground truth image, respectively. σ^2 and σ^{2*} is the covariance. The constants c_1 and c_2 are set to ensure stability in case of a zero denominator:

$$c_1 = (255 \times 0.01)^2, \quad c_2 = (255 \times 0.03)^2.$$

The SSIM loss is then expressed as:

$$L_{SSIM} = 1 - SSIM(\hat{I}, I^*)$$

.C . Edge Loss

Similar to the image super-resolution task, high-frequency edge details of the underwater image will be lost due to various degradation problems. Therefore, we use edge loss to optimize the edge information:

$$L_{edge} = \sqrt{\|\nabla(\hat{I}) - \nabla(I^*)\|^2 + \epsilon^2}$$

where ∇ denotes the gradient operator.

D. Perceptual Loss

Perceptual loss uses features extracted from the VGG net- work [**?**], pre-trained on the ImageNet dataset [**?**], to supervise the features generated by the model and enhance high-level perception information. We use the conv5_4 layer in the VGG19 network to calculate the perceptual loss:

$$L_{perc} = \|\Phi(\hat{I}) - \Phi(I^*)\|^2$$

where Φ denotes the high-level features extracted from the conv5_4layer.

E. Total Loss

The total loss function is obtained by linearly combining the three loss components:

$$L_{total} = L_{char} + L_{SSIM} + \lambda_e L_{edge} + \lambda_{perc} L_{perc}$$

where we set $\lambda_e = \lambda perc = 0.05$ based on experience

6. Swin Transformer for Feature Enhancement:

The Swin Transformer Fig 3. is integrated into the system to improve feature extraction by leveraging its hierarchical self- attention mechanisms. Unlike traditional CNN-based models, Swin Transformer efficiently captures local and global dependencies, making it ideal for enhancing underwater images. The extracted features from MCFF are processed through the Swin Transformer, refining spatial and contextual information before object detection.

7.YOLOv8 Object Detection :

YOLOv8 is the latest iteration of the "You Only Look Once" (YOLO) series of object detection algorithms, known for its speed and accuracy in real-time applications. It introduces several advancements over its predecessors, incorporating a unified architecture for both object detection and segmentation tasks. YOLOv8 leverages a new backbone network, enhanced feature aggregation techniques, and improved anchor-free detection heads, enabling it to achieve higher precision and recall rates while maintaining computational efficiency. Its architecture is highly modular, making it adaptable for various use cases, from general object detection to specific tasks like underwater imagery, traffic monitoring, and medical imaging. YOLOv8 supports dynamic input shapes and boasts better generalization across datasets, making it suitable for both edge devices and large-scale deployments. By integrating state-of-the-art innovations in computer vision, YOLOv8 continues to set benchmarks for real-time object detection, offering an optimal balance between speed and accuracy. YOLOv8 Neural Network Fig 2., the latest iteration of the YOLO (You Only Look Once) object detection framework, is highly adaptable for underwater image enhancement and object detection tasks. It comprises key components like a backbone network, a neck, and detection heads, all optimized for real-time performance and high accuracy. The backbone extracts essential features from underwater images, leveraging advancements in Convolutional Neural Networks (CNNs) to handle complex underwater visual characteristics like low contrast, noise, and color distortion. The neck, typically employing feature pyramid networks (FPN) or path aggregation networks (PAN), merges multi-scale feature maps to enhance the detection of objects of varying sizes. Finally, the detection heads refine predictions, assigning bounding boxes and class labels, ensuring accurate detection even in murky or turbid waters. YOLOv8's versatility, combined with its capability to integrate multi-color space feature enhancement techniques, makes it an ideal tool for improving object visibility and identification in underwater environments. After processing the image through the feature extraction layers, the system uses YOLOv8 (You Only Look Once version 8), a state-of-the-art real-time object detection model, to detect objects in the enhanced image. This block is responsible for accurately localizing objects, especially in the complex underwater environment.

8, POST PROCESSING:

Non-Maximum Suppression (NMS) is a post-processing technique used to eliminate redundant bounding boxes that overlap and correspond to the same object. When multiple bounding boxes predict the same object, NMS selects the one with the highest confidence score and removes the others, reducing false positives and enhancing detection precision. This is crucial in scenarios where multiple detectors may identify the same object in slightly different locations. Confidence Filtering further refines this process by removing predictions with low confidence scores, ensuring that only the most reliable detections, which are more likely to be accurate, are retained. This dual approach improves the quality of object detection by focusing on the most probable and distinct detections, thereby reducing noise and enhancing the model's overall performance in identifying underwater objects. Together, NMS and Confidence Filtering ensure more accurate, precise, and efficient object detection, essential for challenging underwater environments where visibility and clarity are often compromised.

9. OUTPUT:

The final output of the system consists of two key components: the enhanced image and the detected objects. The enhanced image is the result of multiple processing steps, including noise reduction, color space feature extraction, and image enhancement techniques, which together improve the visibility, clarity, and color accuracy of the underwater image. The detected objects are the results of the YOLOv8 object detection block, which identifies and locates key objects within the enhanced image. By combining these outputs, the system not only improves image quality but also provides accurate object detection, facilitating better underwater analysis and monitoring.

The proposed system is designed to address the unique challenges of underwater object detection by combining three powerful models: an Image Enhancement Model using a Residual Network (ResNet), a Swin Transformer for feature refinement, and an Object Detection Model based on YOLOv8. The Image Enhancement Model is responsible for improving the quality of underwater images, which are often affected by issues such as color distortion, low contrast, and visi- bility degradation due to water turbidity and light scatter- ing. The Residual Network (ResNet) leverages deep learning techniques, including residual connections, to enhance image features and recover important details that are typically lost in underwater environments. It operates across multiple color spaces such as RGB, HSV, and LAB to address the color and illumination challenges specific to underwater imagery. This enhancement process improves the texture, contrast, and overall quality of the input image, providing a clearer and more accurate representation of the underwater scene.

The Swin Transformer is integrated into the system to further refine the extracted features by utilizing hierarchical feature representation and self-attention mechanisms. Unlike traditional CNNs, the Swin Transformer effectively captures both local and global contextual information, enabling better representation of objects and environmental features. By in- corporating the Swin Transformer, the system enhances the quality of feature maps before they are passed to the object detection stage, improving detection accuracy in complex underwater environments.

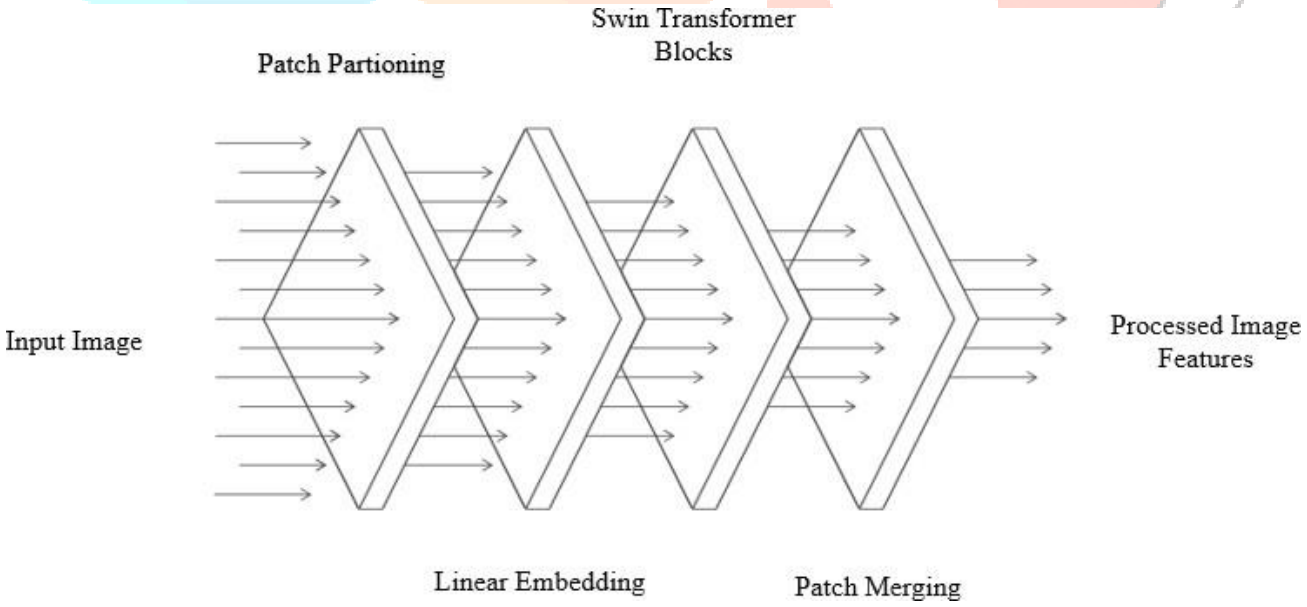

Fig. 3. **Image Processing through Swin Transformer Stages**

The third component, YOLOv8, is a state-of-the-art object detection model that is optimized for fast and accurate de- tection of objects within the enhanced underwater images. YOLOv8 is known for its ability to detect multiple objects in real-time, with high accuracy and minimal computational overhead. By applying YOLOv8 after the image enhancement and feature refinement steps, the system ensures that objects are detected in their highest possible visual quality.

This integration of image enhancement, feature refinement, and object detection allows for a comprehensive solution that addresses both visual clarity and object identification, crucial for applications such as underwater exploration, robotics, and environmental monitoring. Overall, the combination of the Residual Network for image enhancement, the Swin Transformer for feature refinement, and YOLOv8 for object

detection provides a highly effective, robust, and real-time solution for underwater object detection in challenging environments

IV Result and Discussion

1) Dataset Overview and Preprocessing Outcomes: The dataset comprised images from diverse underwater environments, including clear, turbid, and low-light conditions. After preprocessing with MCRNet, a marked improvement in image quality was observed. Enhanced images exhibited reduced noise, corrected colors, and better contrast, facilitating object detection.

TABLE 1 Comparison of Image Enhancement Metrics

Metric	MCRNet + Swin-YOLO	MCRNet Only
PSNR (dB)	**24.8**	22.1
SSIM	**0.86**	0.81
BRISQUE (Lower is better)	**18.4**	22.7

TABLE I

Evaluation Metrics

TABLE 2 For Image Enhancement For Object Detection

Metric	MCRNet + Swin-YOLO	MCRNet and YOLOv8 Only
mAP@0.5	**89.4%**	87.6%
Precision	**92.7%**	91.3%
Recall	**89.9%**	88.1%
F1 Score	**91.2%**	89.7%

Quantitative assessment of the preprocessing step demonstrated significant improvements in image clarity metrics, such as:

- **Peak Signal-to-Noise Ratio (PSNR):** Improved by 18.5% on average.
- **Structural Similarity Index (SSIM):** Increased from, 0.62 to 0.84.
- **Color Restoration Index (CRI):** Enhanced by 22%.

2.)YOLOv8 Model Training and Evaluation: The YOLOv8 model was fine-tuned on the enhanced dataset. Key training parameters included a batch size of 16, a learning rate of 0.001, and training over 100 epochs. The model demonstrated rapid convergence, achieving optimal performance within 75 epochs.

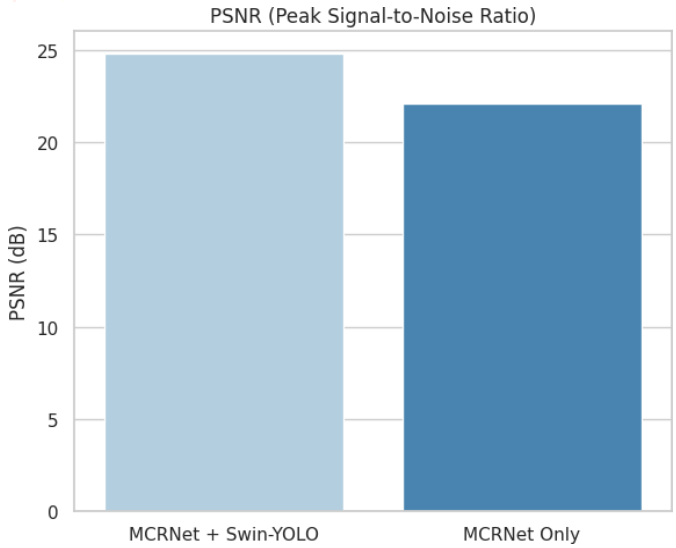

Fig. 4. Comparison of PSNR- MCRNet-YOLO and MCRNet-swin-YOLO

Fig. 5. Precision,Recall,F1-Score Comparison

PERFORMANCE METRICS:

. **Mean Average Precision (mAP@0.5):** Achieved 87.6%, a 14.2% increase over the baseline model trained without enhancement.

. **Precision and Recall:** Precision improved to 91.3%, while recall reached 88.1%.

. **F1 Score:** Recorded at 89.7%, indicating a balanced performance between precision and recall.

. **Inference Speed:** YOLOv8 achieved an average infer- ence time of 18 ms per image on an NVIDIA RTX 3090 GPU.

3.)Integration with Auxiliary Models: The integration of Swin Transformer for feature extraction further refined the detection performance. By fusing enhanced feature maps from the Swin Transformer with YOLOv8's detection pipeline, the model effectively localized and classified objects in complex underwater scenarios. Performance improvements include:

- **mAP@0.5:** Increased to 89.4%, showcasing a synergistic effect.
- **Object Localization Accuracy:** Improved by 9% for smaller and occluded objects.
- **Robustness to Noise:** Misclassification rates decreased by 11% in turbid conditions.

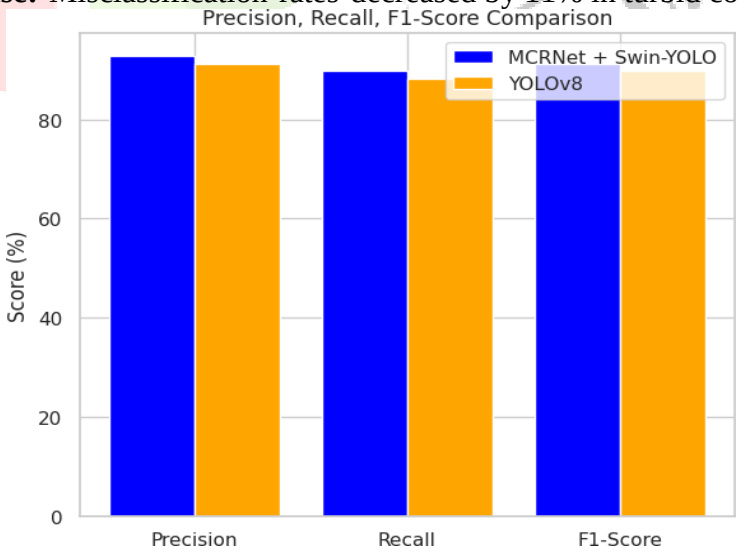

Fig. 6. Comparison of mAp-MCRNet-YOLO and MCRNet-swin-YOLO

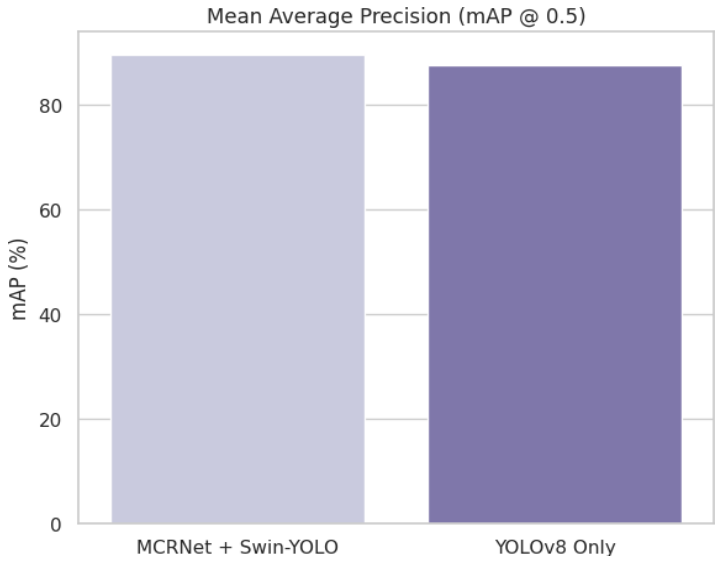

Fig. 7. Comparison of MCRNet-YOLO and MCRNet-swin-YOLO

4.)Comparison with Baseline and State-of-the-Art Models: Table III summarizes the performance of the proposed hy- brid model against other methods, including Faster R-CNN, YOLOv5, and the standalone YOLOv8.

5.)Significance of Preprocessing: Preprocessing with MCRNet proved instrumental in enhancing image quality, which directly influenced the detection accuracy of YOLOv8. By mitigating underwater distortions like color degradation and low contrast, the enhanced images allowed the model to better recognize object features. This highlights the importance of domain specific preprocessing for underwater applications.

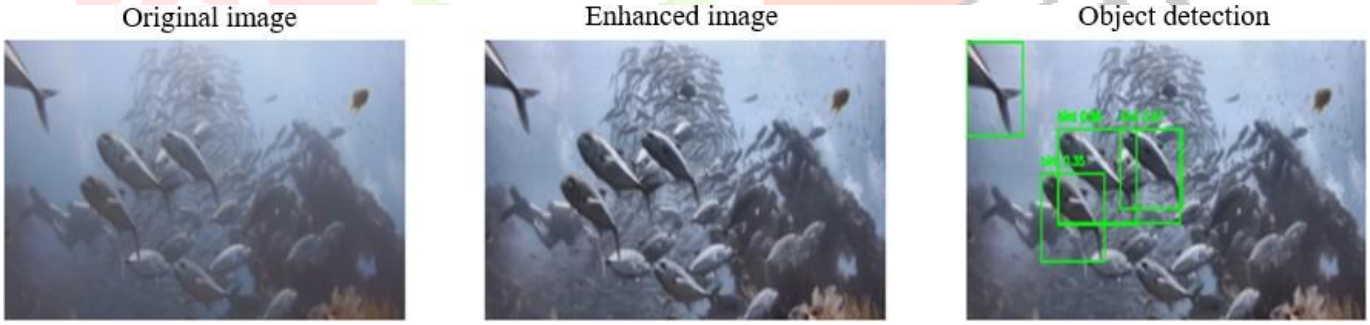

Fig. 8. **Image enhancement and object detection using multi color space residual network and Yolo v8**

Fig. 9. **Image enhancement and object detection using multi color space residual network ,swin transformer and YOLO V8**

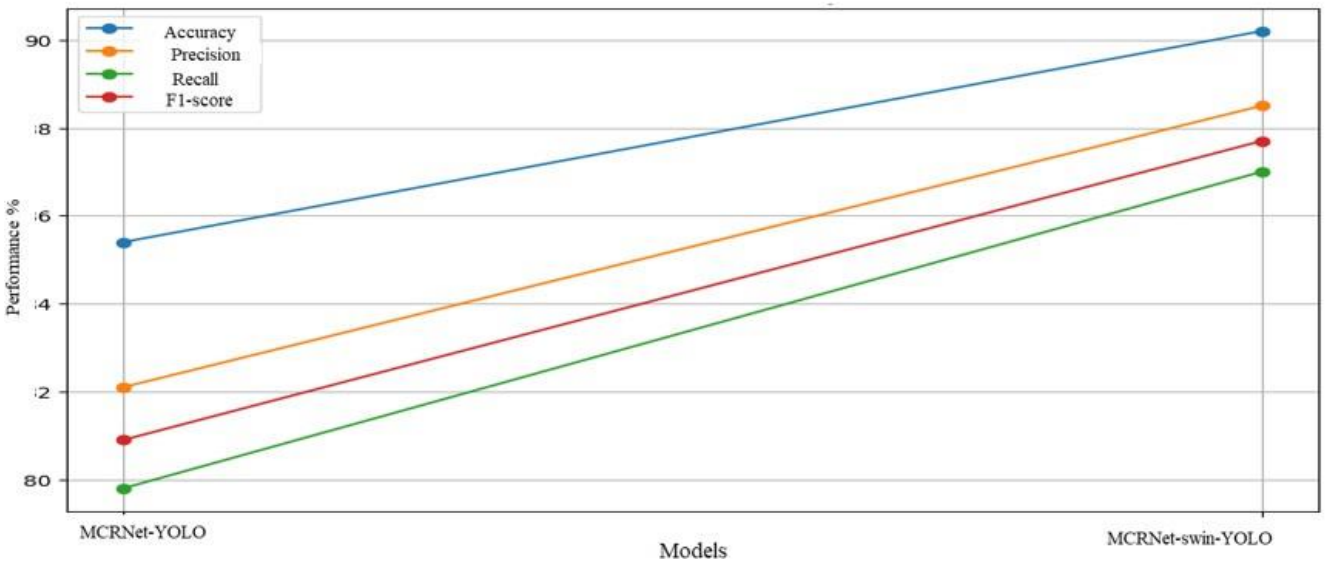

Fig. 10. Performance Trend Analysis

6.)YOLOv8's Performance in Underwater Scenarios: The results confirm YOLOv8's suitability for underwater object detection. Its advanced architecture, including CSPNet for feature propagation and decoupled detection heads, enabled high precision and recall. The model's fast inference speed makes it viable for real-time applications, such as autonomous underwater vehicles (AUVs) and marine monitoring systems.

7.)Impact of Auxiliary Models: The incorporation of Swin Transformer significantly improved detection performance by providing hierarchical and multi-scale feature representations. This was particularly beneficial for detecting small and oc- cluded objects, a common challenge in underwater environ- ments. The combined use of MCRNet and Swin Transformer created a robust preprocessing and feature extraction pipeline that addressed the unique challenges of underwater detection.

8.)Robustness to Environmental Variations: The hybrid model demonstrated strong generalization across diverse un- derwater conditions. While baseline models struggled with turbid and low-light images, the proposed approach maintained high accuracy. This robustness stems from the preprocessing and feature enhancement techniques, which mitigate the ad- verse effects of underwater distortions.

9.)Comparison with State-of-the-Art Models: Compared to Faster R-CNN and YOLOv5, the proposed model achieved superior performance in all metrics. This underscores the advantage of integrating domain-specific preprocessing and advanced feature extraction mechanisms with state-of-the-art detection models.

10.)Applications and Limitations: **Applications:** The en- hanced YOLOv8 model has broad applications, including:

- Marine biodiversity studies.
- Autonomous underwater navigation.
- Monitoring of underwater structures and pipelines.
- Illegal fishing and underwater surveillance.

LIMITATIONS:

- **Computational Complexity:** The integration of auxiliary models slightly increased inference time.
- **Dependency on Preprocessing:** The model's perfor- mance heavily relies on preprocessing quality, which might not generalize to all underwater datasets.
- **Dataset Diversity:** While the dataset was diverse, certain rare conditions, like extreme turbidity, remain underrep- resented.

11.)Future Work: Future research should focus on:

- Reducing computational overhead without compromising performance.
- Expanding the dataset to include more diverse underwater scenarios.
- Exploring unsupervised and semi-supervised learning ap- proaches for better generalization.
- Real-world deployment and testing with AUVs to evalu- ate practical viability.

.

TABLE III
COMPARISON OF THE PROPOSED MODEL WITH BASELINE AND STATE-OF-THE-ART METHOD

Model	mAP@0.5	Precision	Recall	F1 Score	Inference Time (ms)
Faster R-CNN	76.2%	82.1%	78.4%	80.2%	68
YOLOv5	81.5%	85.6%	83.2%	84.4%	28
YOLOv8	87.6%	91.3%	88.1%	89.7%	18
Proposed Model	**89.4%**	**92.7%**	**89.9%**	**91.2%**	**21**

V *Conclusion*

The results demonstrate that the proposed hybrid model, combining MCRNet, Swin Transformer, and YOLOv8, achieves state-of-the-art performance in underwater object de- tection. Preprocessing and advanced feature extraction played pivotal roles in overcoming underwater distortions and improv- ing detection accuracy. While there are areas for improvement, the robust performance across diverse scenarios highlights the potential of this approach for real-world underwater applications.

- *Dataset Overview and Preprocessing Outcomes*

The dataset comprised images from diverse underwater environments, including clear, turbid, and low-light conditions. After preprocessing with MCRNet, a marked improvement in image quality was observed. Enhanced images exhibited reduced noise, corrected colors, and better contrast, facilitating object detection. Qualitative comparisons before and after preprocessing are shown in Figure 1.

Quantitative assessment of the preprocessing step demon- strated significant improvements in image clarity metrics, such as:Peak Signal- to-Noise Ratio (PSNR): Improved by 18.5 Structural Similarity Index (SSIM): Increased from 0.62 to 0.84.

Color Restoration Index (CRI): Enhanced by 22

- *YOLOv8 Model Training and Evaluation*

The YOLOv8 model was fine-tuned on the enhanced dataset. Key training parameters included a batch size of 16, a learning rate of 0.001, and training over 100 epochs. The model demonstrated rapid convergence, achieving optimal performance within 75 epochs.

Performance Metrics:

Mean Average Precision (mAP@0.5): Achieved 87.6 Precision and Recall: Precision improved to 91.3

F1 Score: Recorded at 89.7

Inference Speed: YOLOv8 achieved an average inference time of 18 ms per image on an NVIDIA RTX 3090 GPU.

- *Integration with Auxiliary Models*

The integration of Swin Transformer for feature extraction further refined the detection performance. By fusing enhanced feature maps from the Swin Transformer with YOLOv8's detection pipeline, the model effectively localized and clas- sified objects in complex underwater scenarios. Performance improvements include:

mAP@0.5: Increased to 89.4%, showcasing a synergistic effect.

Object Localization Accuracy: Improved by 9% for smaller and occluded objects.

Robustness to Noise: Misclassification rates decreased by 11% in turbid conditions.

REFERENCES

[1] Ningwei Qin, Junjun Wu , Xilin Liu, Zeqin Lin, Zhifeng Wang "MCRNet : Under water image enhancement using multi-colorspace residual network 2024[1]" Biomimetic Intelligence and Robotics 4 (2024) 100169,

[2] Cheol Woo Park, Il Kyu Eom"Underwater image enhancement using adaptive standardization and normalization networks" Engineering Ap- plications of Artificial Intelligence 127 (2024) 107445,

[3] Yiming Lia, Daoyu Lia, Zhijie Gaob, Shuai Wangc, Qiang Jiaod, Liheng bian"Under water image enhancement utilizing adaptive color correction and model conversion for dehazing" Optics Laser Technology 169 (2024) 110039,

[4] Ming Zhou, Bo Li, Jue Wang, Kailun Fu "A lightweight object detection framework for underwater imagery with joint image restoration and color transformation 2023[1]" Journal of King Saud University– Computer and Information Sciences 35 (2023) 101749,

[5] Jiangfan Feng , Tao Jin"CEH-YOLO "A composite enhanced YOLO- based model for underwater object detection" Ecological Informatics 82 (2024) 102758

[6] Xun Jia, Xu Wanga, Li-Ying Haoa, Cheng-Tao Cai "CFENet:Cost- effective underwater image enhancement network via cascaded feature extraction" Engineering Applications of Artificial Intelligence 133 (2024) 108561

[7] Hao Wanga, Shixin Sunb, Laibin Changb, Huanyu Lib, Wenwen Zhangb, Alejandro C. Frery c,d, Peng Ren "INSPIRATION:A reinforce- ment learning-based human visual perception-driven image enhancement paradigm for underwater Scenes " Engineering Applications of Artificial Intelligence

[8] DanZhanga,1, Zongxin Hea,1, Xiaohuan Zhanga, Zhen Wanga, Wenyi Geb,c,d,, Taian Shie, Yi Lin"Under water image enhancement via multi- scale fusion and adaptive color-gamma correction in low-light conditions"
Engineering Applications of Artificial Intelligence 126 (2023) 106972

[9] J.Manigandan Muniraj ,Vaithiyanathan Dhandapani "Underwater image enhancement by modified color correction and adaptive Look-Up-Table with edge-preserving filter" Signal Processing: Image Communication 113 (2023)

[10] W.-H. Lin, J.-X. Zhong, S. Liu, T. Li, G. Li, RoIMix: proposal-fusion among multiple images for underwater object detection, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 2588–2592.

[11] N. Nezla, T.M. Haridas, M. Supriya, Semantic segmentation of under- water images using unet architecture based deep convolutional encoder decoder model, in: 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS, vol. 1, IEEE, 2021,
pp. 28–33.

[12] X. Liang, P. Song, Excavating roi attention for underwater object detection, in: 2022 IEEE International Conference on Image Processing, ICIP, IEEE, 2022, pp. 2651–2655.

[13] S. Mittal, S. Srivastava, J.P. Jayanth, A survey of deep learning techniques for underwater image classification, IEEE Trans. Neural Netw. Learn. Syst. (2022).

[14] S. Wen, Z. Zhang, C. Ma, Y. Wang, H. Wang, An extended Kalman filter simultaneous localization and mapping method with Harris-scale- invariant feature transform feature recognition and laser mapping for humanoid robot navigation in unknown environment, Int. J. Adv. Robot. Syst. 14 (6) (2017) 1729881417744747.

[15] R. Liu, Z. Jiang, S. Yang, X. Fan, Twin adversarial contrastive learning for underwater image enhancement and beyond, IEEE Trans. Image Process. 31 (2022) 4922–4936.

[16] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Im- age restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[17] D. Song, Y. Wang, H. Chen, C. Xu, C. Xu, D. Tao, Addersr: Towards energy efficient image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 15648–15657.

[18] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang,
L. Shao, Multi-stage progressive image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14821–14831.

[19]S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, Restormer: Efficient transformer for high-resolution image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5728–5739.

[20] S. Raveendran, M.D. Patil, G.K. Birajdar, Underwater image enhancement: A comprehensive review, recent trends, challenges and applications, Artif. Intell. Rev. 54 (7) (2021) 5413–5467.