A Novel Approach to Underwater Object Detection via Bottleneck Layers and Morphological Operations

Sumanth Kumar Panguluri [a], Ponduri Vasanthi [b], Shaik Basheera [b], Laavanya Mohan [c], Veeravalli Venkata rao [d], Madhukumar Patnala[a]

[a] Department of Electronics and Communication Engineering, Bapatla Engineering College, Bapatla, Andhra Pradesh, 522102, India

[b] Department of Electronics and Communication Engineering, Eswar College of Engineering, Narasaraopet, Guntur, Andhra Pradesh, 522601, India

[c] Department of Electronics and Communication Engineering, Vignan's Foundation for Science, Technology, and Research, Vadlamudi, Guntur, Andhra Pradesh, 522213, India

[d] Department of Electronics and Communication Engineering, Narasaraopeta Engineering College, Narasaraopet, Guntur, Andhra Pradesh, 522601, India

vasanthi457@gmail.com, skp6472@gmail.com, shaikbphd@gmail.com, laavanvijay@gmail.com, vvraoece2023@gmail.com

ABSTRACT

Underwater object detection poses significant challenges due to the inherent distortion and light attenuation experienced in aquatic environments. This paper outlines a comprehensive approach to enhance object detection in such challenging conditions. The proposed methodology combines morphological edge enhancement techniques with an efficient detection model featuring bottleneck layers. The initial phase involves the acquisition of a meticulously labeled dataset comprising underwater images containing objects of interest. Prior to model training, a critical preprocessing step is undertaken to rectify underwater distortions, encompassing tasks like color correction and contrast enhancement. To further fortify the model's adaptability to diverse underwater scenarios, the dataset is enriched through augmentation, introducing variations in lighting conditions, water clarity, and object poses. Bottleneck layers act as information bottlenecks, reducing the spatial dimensions of feature maps while simultaneously increasing their depth. This transformation not only compresses information but also mitigates computational overhead, thereby facilitating efficient object detection. This proposed model undergoes experimental validation on the underwater dataset, achieving significantly higher metrics such as a mean average precision (mAP) of 85.1%, precision of 84.4%, and recall of 79.9%. These experimental findings strongly indicate that the proposed model surpasses existing models in its ability to detect exceedingly underwater objects effectively.

KEYWORDS

Edge enhancement, Morphological transformations, BottleNeck layers, Small Object Detection

1. INTRODUCTION

Underwater object detection is a crucial field of study and technology used in various underwater applications, including marine research, offshore industries, naval operations, and environmental monitoring. It involves the use of specialized equipment and algorithms to identify and locate objects or features submerged beneath the water's surface. Detecting various objects in water, including fish, jellyfish, penguins, puffins, sharks, starfish, and stingrays, is a complex and important task in underwater image processing and computer vision. Each of these objects serves different ecological and scientific purposes, and their detection can aid in conservation efforts, research, and monitoring of aquatic environments.

Here's an introduction to object detection for these aquatic creatures: 1. Fish Detection: Fish detection is crucial for marine biology, fisheries management, and ecological studies. Techniques include image processing and computer vision algorithms that identify fish based on color, shape, and size. Convolutional Neural Networks (CNNs) are commonly used for fish detection in underwater images and videos. 2. Jellyfish Detection: Detecting jellyfish is essential to assess their population, which can have significant ecological and economic impacts. Detection methods involve identifying the unique translucent and gelatinous structure of jellyfish in images. Machine learning models can be trained to distinguish jellyfish from other objects in the water. 3. Penguin and Puffin Detection: Penguins and puffins are seabirds that are often studied in their natural habitats. Detection methods may involve identifying the distinctive shapes and colors of these birds, even when partially submerged. Deep learning models can be trained to recognize these species in images and videos. 4. Shark Detection: Shark detection is vital for public safety at beaches and for studying these apex predators' behaviors. Detection methods can include identifying shark silhouettes, dorsal fins, and body markings. Real-time detection systems using drones or underwater cameras are used for early shark warnings. 5. Starfish Detection: Studying starfish can provide insights into reef health and marine ecosystems. Detection techniques focus on the radial symmetry and distinct shapes of starfish. Image processing algorithms can segment and identify starfish in underwater imagery. 6. Stingray Detection: Stingray detection is essential for beach safety and understanding the distribution of these species.

Detection methods may involve identifying the unique flattened body shape and long tail of stingrays. Machine learning models can be trained to distinguish stingrays from other underwater objects. In many cases, the detection of these aquatic objects involves using underwater cameras, sonar, or other specialized sensors to capture images or acoustic data. Machine learning and computer vision techniques, including deep learning, play a significant role in automating the detection process. These systems can be integrated into autonomous underwater vehicles (AUVs) or remotely operated vehicles (ROVs) for efficient data collection and object detection in underwater environments. Accurate detection and monitoring of these aquatic species contribute to the understanding and preservation of marine ecosystems.

The significance of biological detection in underwater environments has garnered widespread attention. This application has found utility in diverse areas, including underwater robotics [1], underwater surveillance [2,3], and marine research [4,5]. Achieving autonomous and highly accurate detection of marine organisms is now a pressing need. However, when it comes to underwater detection, there are unique challenges. Unlike land-based detection scenarios [6,7,8], underwater imaging faces severe natural conditions on the seabed. Consequently, the majority of images are afflicted by issues like color casts [9,10,11,12,13], low contrast [14,15,16,17], blurriness [18,19], and noise [20,21]. As a result, accurate identification of target marine organisms in complex and dynamic underwater environments remains an immense challenge, one that only a handful of researchers have taken on. Hence, the development of a high-performance OD framework tailored for marine OD tasks is of paramount importance.

Underwater target detection methods are most effective when they account for specific underwater conditions like clear visibility, moderate currents, and suitable lighting. Traditional detection approaches primarily rely on extracting features such as “color, texture, and geometry”. However, as DL technology has advanced, neural networks have been introduced as frameworks for underwater OD, capable of identifying and locating objects within images, thereby improving detection accuracy. Nevertheless, real-world underwater conditions are often far from ideal, leading to a decline in image quality that adversely impacts detection performance. To address these challenges, a comprehensive underwater OD framework tailored for complex underwater environments is proposed. This framework incorporates an image enhancement module with efficient detection network structure.

2. LITERATURE SURVEY

Initially, a series of image enhancement preprocessing steps is applied to the original underwater image. These steps include operations like CLAHE [22], dynamic thresholding [23], multi-scale color adaptive correction [24,25], and other conventional techniques. Their purpose is to enhance image clarity, contrast, and detailed texture features, ultimately improving image quality. This, in turn, enhances the accuracy and generalization capability of the model. With the continuous evolution of attention mechanisms, recent scholarly work has introduced these mechanisms into OD frameworks. A notable example is the introduction of a self-attention mechanism for action classification within a pyramid network [26]. However, it's important to note that the computational burden of the self-attention layer increases quadratically as image resolution grows. To tackle this challenge, BoTNet [27] has replaced the 3 × 3 convolution typically found in the middle of the Bottleneck used by ResNet50 [28] with Multi-Head Self-Attention (MHSA), yielding promising results in object detection tasks.

Given the intricacy and variability of underwater environments, traditional data enhancement techniques are often ineffective due to low contrast and pronounced color differences in existing datasets. Notably, previous research [29] has also highlighted that the inclusion of data augmentation operations can enhance model detection accuracy. The study of underwater target detection holds substantial significance in domains like ocean current observation [30], marine biology, and security [31]. In efforts to enhance the performance of the improved YOLOv5, some researchers have focused on refining the loss function and detection layer of the backbone module, primarily relying on conventional data augmentation techniques. In our paper, we propose an efficient OD framework aimed at achieving high-precision detection. To achieve this, we employ the morphological image enhancement method for data augmentation, processing the original dataset's images as inputs for the improved network.

Xu, F., Wang at.el., proposed a novel approach for improving object detection in images. They introduce a specialized network called SA-SPPN (Spatial Pyramid Pooling Network with Attention) designed to enhance relevant information and expand the receptive field in the original features obtained from the backbone network. This strategy combines features from different levels, creating robust feature maps for object detection [32]. Wang, J., et.al., introduced Poisson fusion is applied to augment the data at the input, ensuring a balanced representation of detected targets. Following this, wavelet transform is employed to execute Style Transfer on the improved images, leading to image restoration. This fusion method merges the effective feature layer derived from the Backbone layer, resulting in heightened detection precision and accelerated model detection speed [33].

Jia, J. et. Al., presents an enhancement to the EfficientDet detector and introduces a novel model called EfficientDet-Revised (EDR), tailored for marine organism object detection. Firstly, the MBConvBlock has been restructured by integrating the Channel Shuffle module. Secondly, the attention module's fully connected layer has been eliminated, and convolutional layers have been employed instead. This change aims to reduce the overall number of network parameters while maintaining or improving performance. Lastly, an Enhanced Feature Extraction module has been devised to support multi-scale feature fusion. This module enhances the network's capability to extract features from objects of varying sizes, contributing to its adaptability in detecting different objects effectively [34].

Ouyang, W et. Al., proposed two lightweight modules, namely the Attention-GB and Bottom-Enhancement modules. The Attention-GB module is designed to incorporate prior knowledge related to the differing attenuation coefficients of red light, green light, and blue light in water. This knowledge is essential for accurate object detection in underwater settings. The Bottom-Enhancement module addresses the need to enhance semantic information in the shallow layers of the model. This enhancement is particularly important for improving the accuracy of detecting small objects, a common challenge in underwater object detection [35]. Zhang, J. et. Al., presenting the CSPLayer incorporates a sizable convolution kernel, enabling the detection network to capture contextual information and with greater precision [36].

The contributions of the proposed methodology can be summarized as follows:

- The methodology integrates advanced morphological edge enhancement techniques; it enhances the quality of input images by emphasizing important object boundaries and features, which in turn aids the detection model in making more accurate predictions.

- To enhance the model's adaptability to diverse underwater scenarios, the dataset is enriched through data augmentation, this process introduces variations in lighting conditions, water clarity, and object poses, making the model more robust in underwater environments.

- The inclusion of bottleneck layers in the detection model reduces the spatial dimensions of feature maps while increasing their depth. This dual function effectively compresses information, enabling the model to operate more efficiently.

3. PROPOSED MODEL

Underwater OD is a challenging task due to the distortion and attenuation of light in water. To improve object detection in underwater environments, combine morphological edge enhancement techniques with the efficient detection model with bottleneck layers.

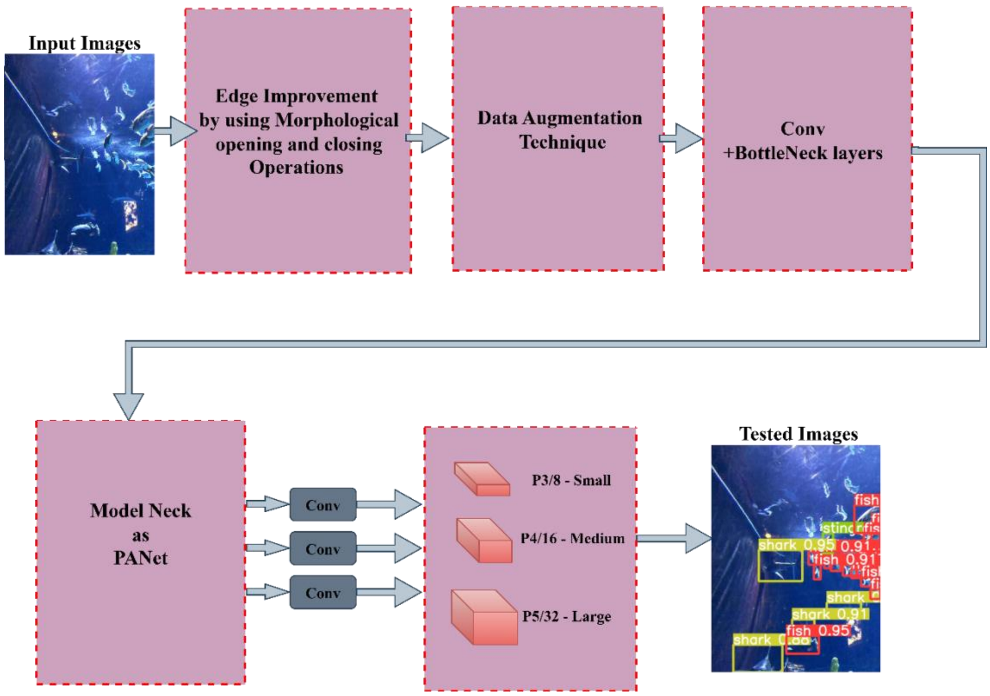


Fig.1. The proposed model block diagram.

The proposed model block diagram, and layer architecture is demonstrated in **Fig.1** and **Fig.2**. At first, gather a labelled dataset of underwater images with objects of interest. Pre-process the images to correct for underwater distortions, such as color correction and contrast enhancement. This step is essential to improve the quality of input data for the model. Morphological Edge Enhancement operations, such as erosion, dilation, and gradient operations, to enhance edges and object boundaries in the pre-processed images. This helps in making object boundaries more distinguishable. After that, augment the dataset with variations in lighting conditions, water clarity, and object poses to make the model more robust to different underwater scenarios. The primary function of bottleneck layers is to act as a bottleneck within the network, reducing the spatial dimensions (e.g., width and height) of the feature maps while increasing the number of channels (depth). This transformation helps in compressing information and reducing the computational load while preserving valuable information for detection. Finally, evaluate the model's performance using metrics like mean Average Precision (mAP) on a validation dataset.

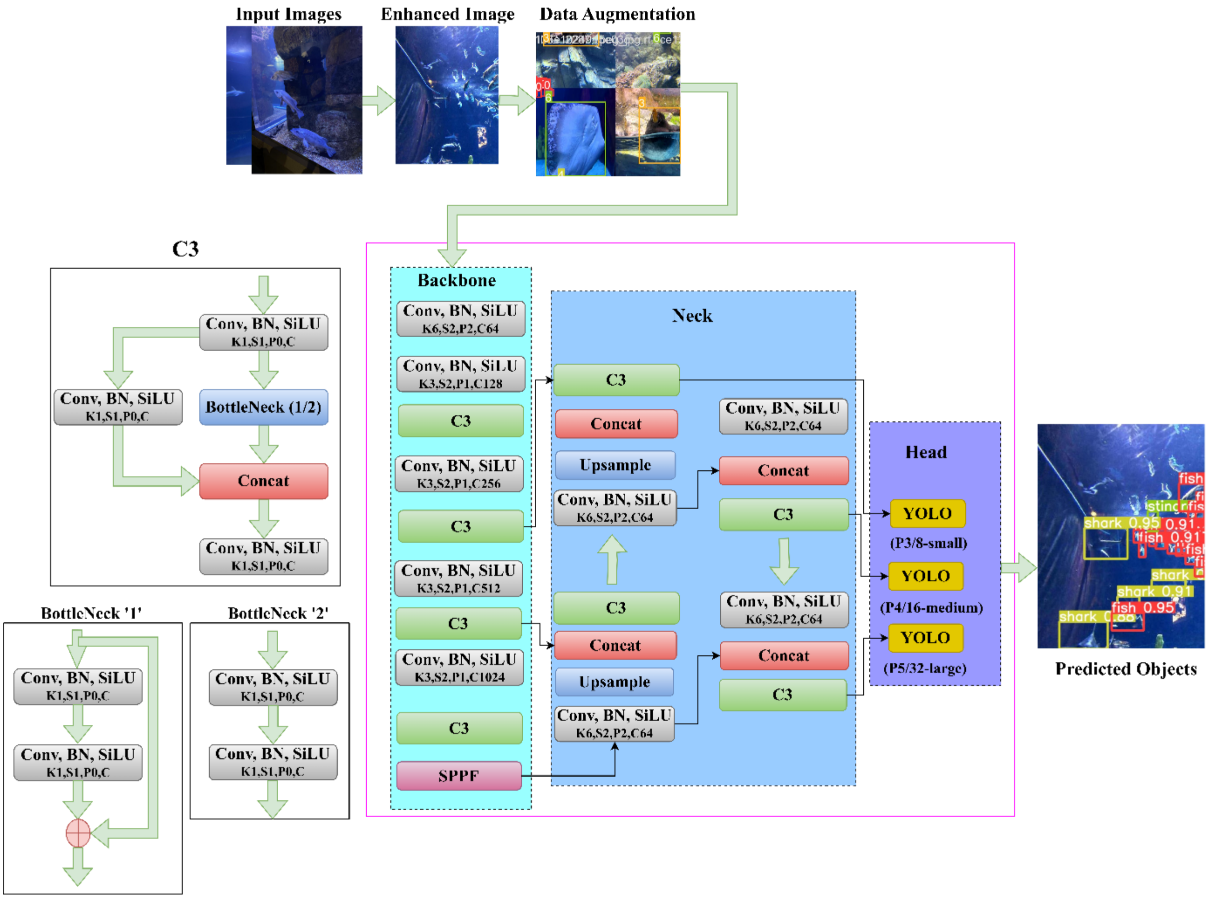


Fig.2. Layer architecture of the Proposed Model.

3.1. EDGE IMPROVEMENT TECHNIQUE

A new edge improvement technique developed by using morphological opening and closing operations is displayed in **Fig.3**.

Generation of output image using above technique involves following steps:

Step1: First we have to take input image. Let consider it as $i(x, y)$.

Step2: Apply morphological opening operation on input image $i(x, y)$. Mathematically it can be expressed as

$$f_{opening}(x,y) = i(x,y) \circ S_e \tag{1}$$

Where S_e represents structuring element.

Step3: Apply morphological closing operation on output of morphological opening operation. Mathematically it can be expressed as

$$f_{closing}(x,y) = f_{opening}(x,y) \bullet S_e \tag{2}$$

The output of morphological closing operation is the blurred image.

Step4: Subtract output of morphological closing operation from input image $i(x,y)$. Mathematically it can be expressed as

$$f_e(x,y) = i(x,y) - f_{closing}(x,y) \tag{3}$$

Where $f_e(x,y)$ represents edge imge.

Step5: Add input image $i(x,y)$ with $f_e(x,y)$. Mathematically it can be expressed as

$$E(x,y) = i(x,y) + f_e(x,y) \tag{4}$$

Where $E(x,y)$ represents output image or edge improved image.

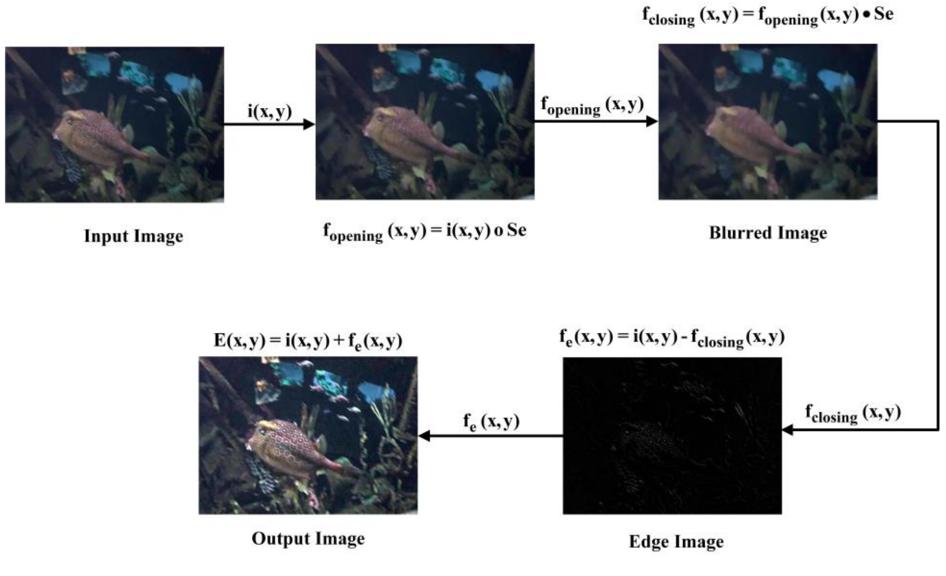


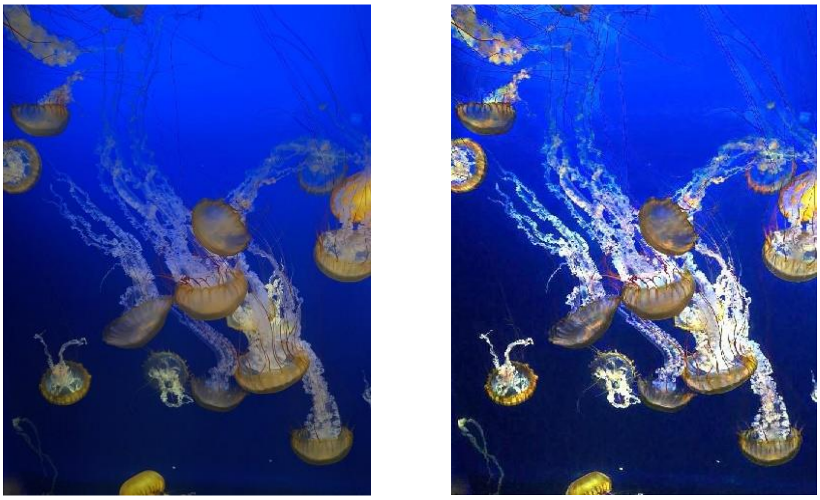
Fig.3. Edge improvement technique using morphological opening and closing operations.

The input images and edge enhanced output images are displayed in **Fig.4**.

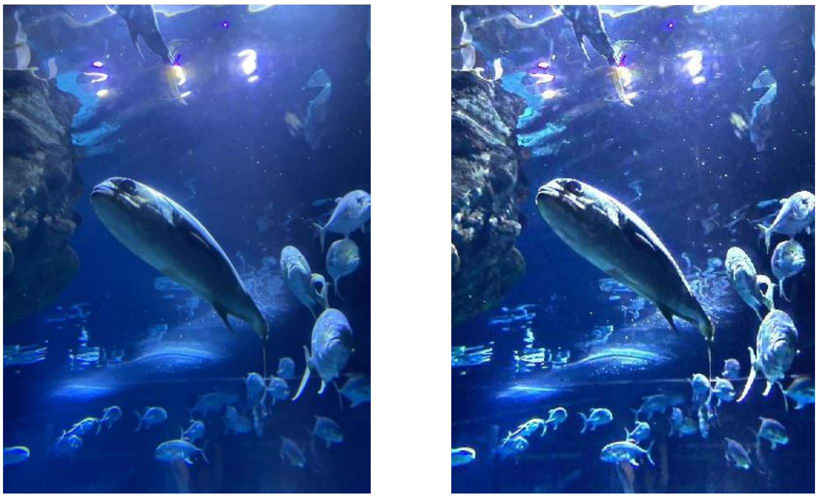
(a)

(b)

(c)



(d)

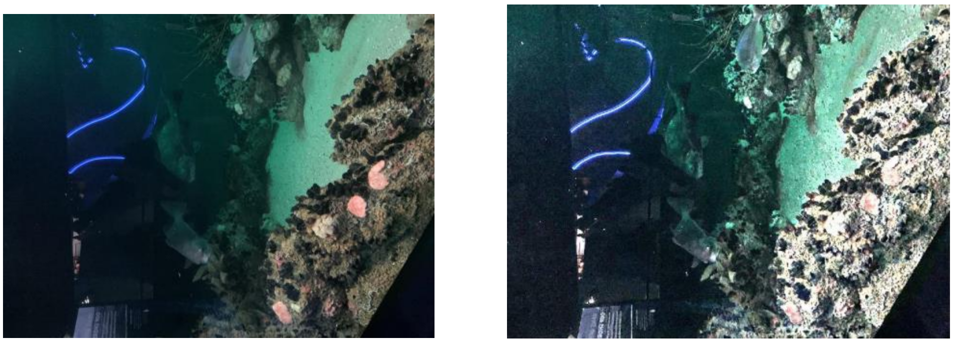


(e)

(f)

(g)

(h)



(i)

Fig.4. Enhancement output results.

3.2. DATA AUGMENTATION

Data augmentation is a crucial technique in training deep learning models like YOLOv5. It involves applying various transformations to the training data to increase the diversity of the dataset, which, in turn, improves the model's ability to generalize to different scenarios and conditions. The optimizer used in this context is Stochastic Gradient Descent (SGD) with a learning rate (lr) of 0.01. This SGD optimizer is configured with different parameter groups: The first parameter group (123 weight) has no weight decay (decay=0.0). The second parameter group (126 weight) has weight decay set to 0.0005. The third parameter group (126 bias) does not have weight decay specified, which implies it may use the default value.

Additionally, data augmentation techniques are applied using the Albumentations library. These augmentations include: Randomly applying blurring with a probability of 1% and a blur limit between 3 and 7. andomly applying median blurring with a probability of 1% and a blur limit between 3 and 7. Randomly converting the image to grayscale with a probability of 1%. Applying Contrast Limited Adaptive Histogram Equalization (CLAHE) with a probability of 1%. The clip limit is between 1 and 4.0, and the tile grid size is 8x8. These data augmentation techniques are applied to the training data to increase its diversity and help the model generalize better to various conditions and inputs during training. The optimizer configuration and data augmentation strategies work together to train the model effectively, controlling the learning process and enhancing the model's ability to handle different scenarios and data variations.

3.3. BACKBONE NETWORK

Proposed model utilizes a modified CSPDarknet backbone network as its feature extractor. The key features of the modified CSPDarknet backbone network include Cross Stage Module (CSP). The CSP module is the core element of the CSPDarknet backbone network. It facilitates cross-stage feature fusion, allowing feature maps from different stages of the network to be combined effectively. This fusion enhances the network's ability to capture features at various scales and resolutions. The backbone is built upon the Darknet-53 architecture, which is a deep convolutional neural network. Darknet-53 is designed to extract hierarchical features from the input image. It consists of various convolutional layers and residual blocks. Proposed model incorporates feature pyramids within the CSPDarknet backbone. Feature pyramids enable the network to capture and utilize multi-scale features, which is crucial for detecting objects of different sizes in an image. Before performing feature fusion, the CSP module often concatenates a portion of the feature maps from the main path with those from the shortcut path. This step ensures that relevant information from both paths is combined during feature fusion.

In proposed model, Convolution, Batch Normalization, and the Silu (Sigmoid Linear Unit) activation function play crucial roles in the model architecture. Convolutional layers are used to extract features from the input image, enabling the network to learn hierarchical representations of the data. Batch Normalization is a technique used to stabilize and accelerate the training of deep neural networks. It normalizes the activations of each layer across a batch of data, reducing internal covariate shift. BatchNorm layers are often placed after convolutional layers. They help improve the model's convergence speed and generalization by maintaining stable activations. The Silu activation function, also known as the Swish activation, is a smooth approximation of the Rectified Linear Unit (ReLU) function. Silu has been found to perform well in deep neural networks due to its non-linearity and smoothness, which can help mitigate vanishing gradient problems.

In proposed architecture's C3 module, the Conv layer with a bottleneck layer is a crucial component that contributes to the network's ability to extract hierarchical features effectively. This module is designed to improve the network's capacity to capture and represent information at different scales, which is important for object detection tasks. The term "bottleneck layer" is often associated with ResNet-style architectures, and it typically refers to a sequence of convolutional layers with a specific structure.

SPPF LAYER

The SPPF layer is a type of pooling layer used to handle objects of different sizes and scales within an image. It helps the model capture multi-scale features effectively. The primary objective of the SPPF layer is to divide the input feature maps into a grid of fixed-size regions, perform pooling operations on each region, and then concatenate the results. This enables the network to capture context and information at multiple scales. SPPF layers are particularly useful in object detection tasks, where objects can vary significantly in size and location within an image. Proposed backbone uses SPPF layer to improve its ability to detect objects of different sizes and scales, contributing to better under water object detection performance.

3.4.NECK NETWORK

The PAN (Path Aggregation Network) is a type of feature fusion network that can be used in the "neck" part of an underwater OD model. The neck is the part of the architecture that follows the backbone and is responsible for aggregating features from different scales before feeding them into the detection head. The PAN network is designed to improve the handling of multi-scale information and enhance object detection performance. In underwater OD, objects can vary in size, and it's essential to capture information at multiple scales to detect both small and large objects accurately.

The PAN network typically takes feature maps from different levels of the backbone network, creating a feature pyramid. Each level of the feature pyramid contains features at a specific spatial scale. In the top-down pathway, feature maps from higher-resolution levels of the feature pyramid are down sampled to match the spatial resolution of lower-resolution levels. This helps align feature maps from different scales. In the bottom-up pathway, features are propagated from lower-resolution levels to higher-resolution levels. This allows high-level semantics to be combined with detailed information from lower scales. Lateral connections connect feature maps from the top-down and bottom-up pathways. These connections allow information to flow bidirectional between different scales, facilitating feature fusion. It combines feature maps from different levels, ensuring that the final feature maps have consistent spatial resolutions across scales. The PAN network is designed to handle the challenges of underwater OD effectively. By aggregating and fusing features from various scales, it enhances the model's ability to detect objects of different sizes within an underwater image.

3.5. BOTTLENECK LAYER

A bottleneck layer typically consists of three main components: 1x1 Convolution: The first component reduces the number of input channels (dimensionality reduction) using a 1x1 convolutional layer. This step helps reduce the computational cost. 3x3 Convolution: After the 1x1 convolution, a 3x3 convolutional layer is applied. This layer captures spatial features and relationships within the feature maps. 1x1 Convolution (Expansion): Finally, another 1x1 convolutional layer is used to expand the number of channels back to the desired level. This expansion can introduce non-linearity and expressive power. Bottleneck layers serve dimensionality reduction: The initial 1x1 convolution reduces the number of channels, which reduces the computational cost and memory requirements. Increased Non-Linearity: The two 1x1 convolutional layers sandwiching the 3x3 convolution introduce additional non-linearity to the network, enhancing its ability to capture complex relationships in the data. Bottle neck layers are displayed in **Fig.5**.

The use of bottleneck layers has several advantages:

- Improved training efficiency: The reduced number of parameters and computations speeds up training.
- Increased depth: Deeper networks can capture more intricate patterns and features.
- Better generalization: Deeper networks often generalize better to unseen data.

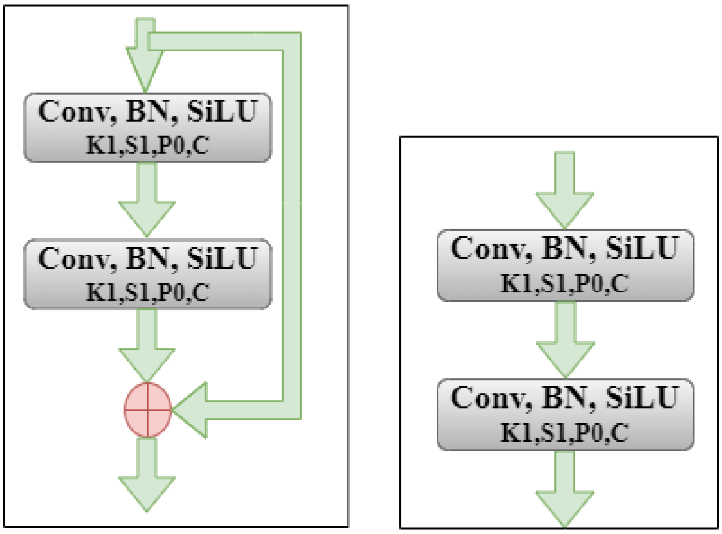


Fig.5. Bottle Neck layers.

3.6. MODEL HEAD

In proposed model, the "model head" refers to the portion of the neural network architecture responsible for producing the final output, which includes underwater OD predictions. The model head processes the feature maps extracted by the backbone network and generates bounding box coordinates, object class predictions, and objectness scores for detected objects within the underwater image. The model head often starts with a series of convolutional layers. These layers are responsible for further processing the feature maps extracted by the backbone network. It includes detection layers that predict the presence of objects within specific anchor boxes at different spatial scales. These layers are responsible for predicting bounding box coordinates (x, y, width, height), objectness scores (probability that an object is present within a given anchor box), and class probabilities (probability of the detected object belonging to each class). Proposed model employs specific loss functions for different components of the output, including the bounding box regression loss, objectness loss, and classification loss. These loss functions are used during training to optimize the model's predictions. The final output format of the model head typically consists of a grid of predictions across different spatial scales.

4. EXPERIMENTS

4.1.DATASET

In our study, we conducted evaluations to assess the performance of proposed framework using the publicly available Underwater Robot Picking Contest (URPC) dataset. This dataset encompasses images containing seven distinct categories of underwater objects, namely 'fish,' 'jellyfish,' 'penguin,' 'puffin,' 'shark,' 'starfish,' and 'stingray.' You can observe instances of these images in **Fig.6**. As part of our experimental setup, we standardized the image sizes to 640x640 pixels. For the verification process, we utilized the raw underwater images as our test set. This test set served as the basis for evaluating the overall framework's performance in terms of object detection. This approach allowed us to rigorously evaluate the effectiveness of proposed OD framework in identifying and localizing underwater objects from the URPC dataset, even when dealing with the challenges posed by raw underwater imagery.

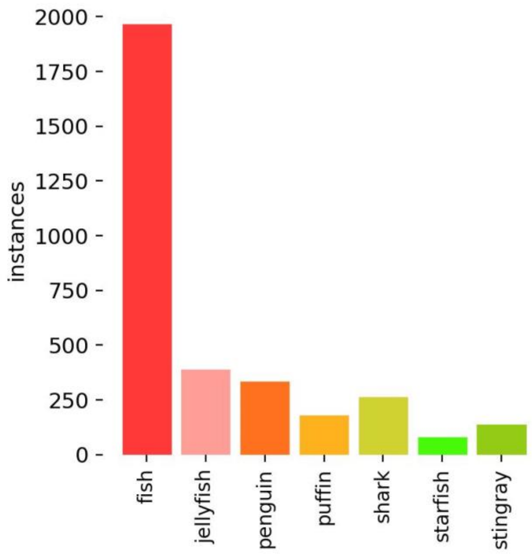


Fig.6. Instances of URPC Dataset

4.2. EVALUATION CRITERIA

In our research, we conducted a thorough evaluation of the model's performance using a comprehensive set of evaluation criteria. These criteria included essential metrics such as mean Average Precision (mAP), precision (P), recall (R), F1-score, and Frames Per Second (FPS). A pivotal aspect of our evaluation process was the utilization of a Confusion Matrix. This matrix played a crucial role in providing a detailed and insightful assessment of the classifier's performance by comparing its predictions to the actual ground truth. The Confusion Matrix consists of four fundamental elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These components offered us a fine-grained understanding of the model's classification and object detection capabilities. This comprehensive evaluation approach allowed us to thoroughly analyze and quantify the model's performance across various aspects of its functionality.

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{7}$$

$$F1Score = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{8}$$

4.3. EXPERIMENTAL OUTCOMES

During the initial stages of training, our model leverages the URPC dataset and fine-tunes its hyper-parameters for optimal performance. The training process incorporates advanced data augmentation techniques, as demonstrated in **Fig.7**. One such augmentation technique is Mosaic augmentation, a powerful approach that combines four distinct images into a single composite image. During training, the network processes these augmented images using a batch size of 16. In **Fig.7**, you can observe bounding boxes, each associated with a label ranging from 0 to 9, representing the presence of seven different objects within the URPC dataset. These bounding boxes vividly illustrate the model's capability to detect objects effectively. Furthermore, in **Fig.8**, we provide visual representations of the model's object detection prowess. These visuals showcase bounding boxes, associated labels, and detection probabilities, highlighting the proposed model's remarkable accuracy in identifying objects, including those that are small and densely packed. Importantly, this accuracy is achieved even when working with input images resized to 640x640 pixels, underscoring the model's robust performance. The figure in **Fig. 9** illustrates a sequence of input images, enhanced images, and detected images.

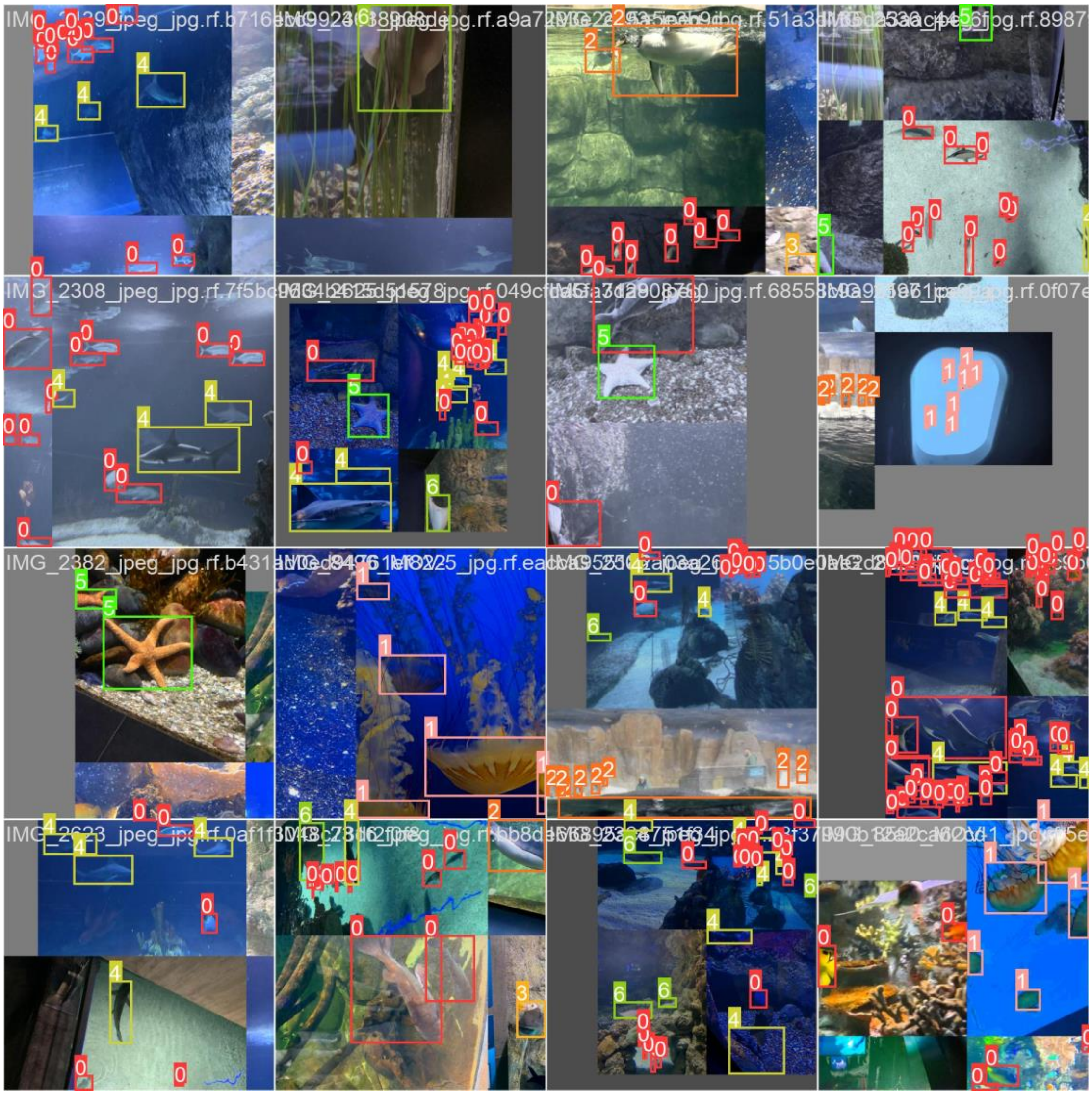
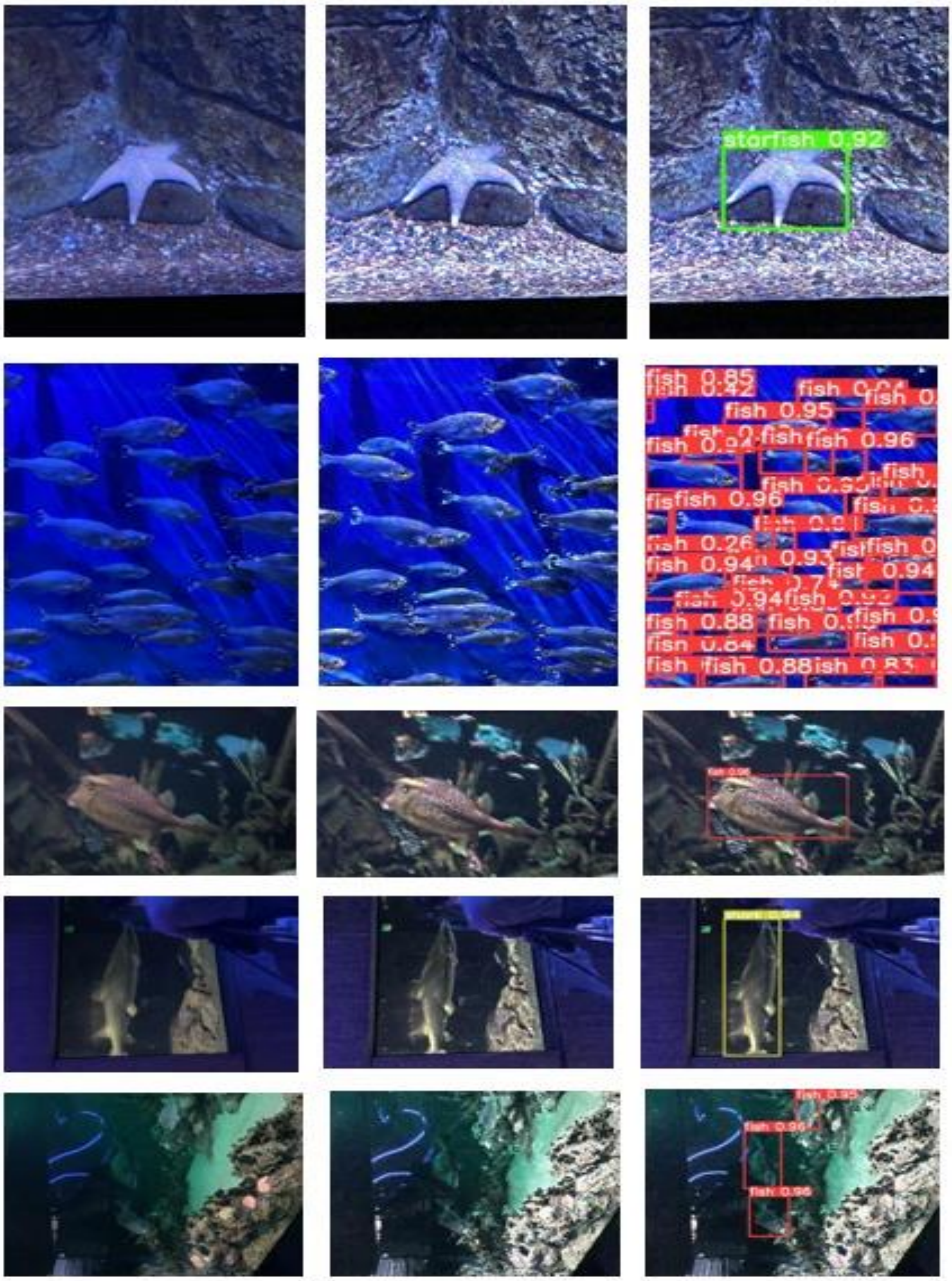


Fig.7. Augmented Images

Fig.8. Validation Images.

In **Fig.10**, we provide a visual representation of the confusion matrix. This visualization plots the true values against the model's predicted values for the dataset, offering a clear and intuitive depiction of the model's classification performance. This comprehensive evaluation process ensures that we have a thorough understanding of the model's capabilities in classifying objects within the dataset. The proposed model undergoes experimental validation on the underwater dataset, achieving significantly higher metrics such as a mean average precision (mAP) of 85.1%, precision of 84.4%, and recall of 79.9%. They are displayed in **Fig.11.**



(a) (b) (c)

Fig.9. (a) Input Images, (b) Enhanced Images, and (c) Test Images.

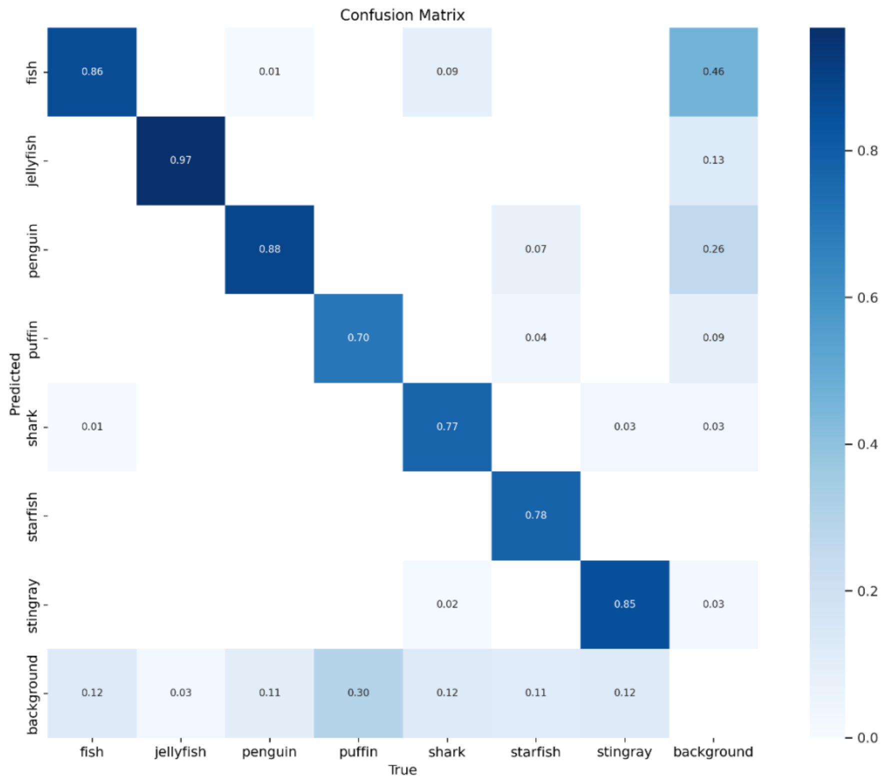


Fig.10. Confusion Matrix

Table 1: The detection performance analysis on the underwater object detection dataset.

'Framework'	**'Year'**	**'Backbone'**	**'Test Pixels'**	**'AP50'**
SA-SPPN [32]	2022	DarkNet-53	608x608	79.23
B-YOLOX-S [33]	2022	DarkNet-53	640x640	82.68
SWIPENET-CMA [37]	2022	SWIPENET	512x512	68
SWIPENET-IMA [38]	2020	SWIPENET	512x512	64.5
RoIAttn [39]	2022	ResNet-50	1000x600	82
RFTM-XT101 [40]	2023	ResNetXT101	640x640	84.7
YOLOX-U [41]	2023	Cascade Mask R-CNN with Swin-T	640x640	83.65
Proposed Model		**Modified CSP Darkent**	**512x512**	**85.62**

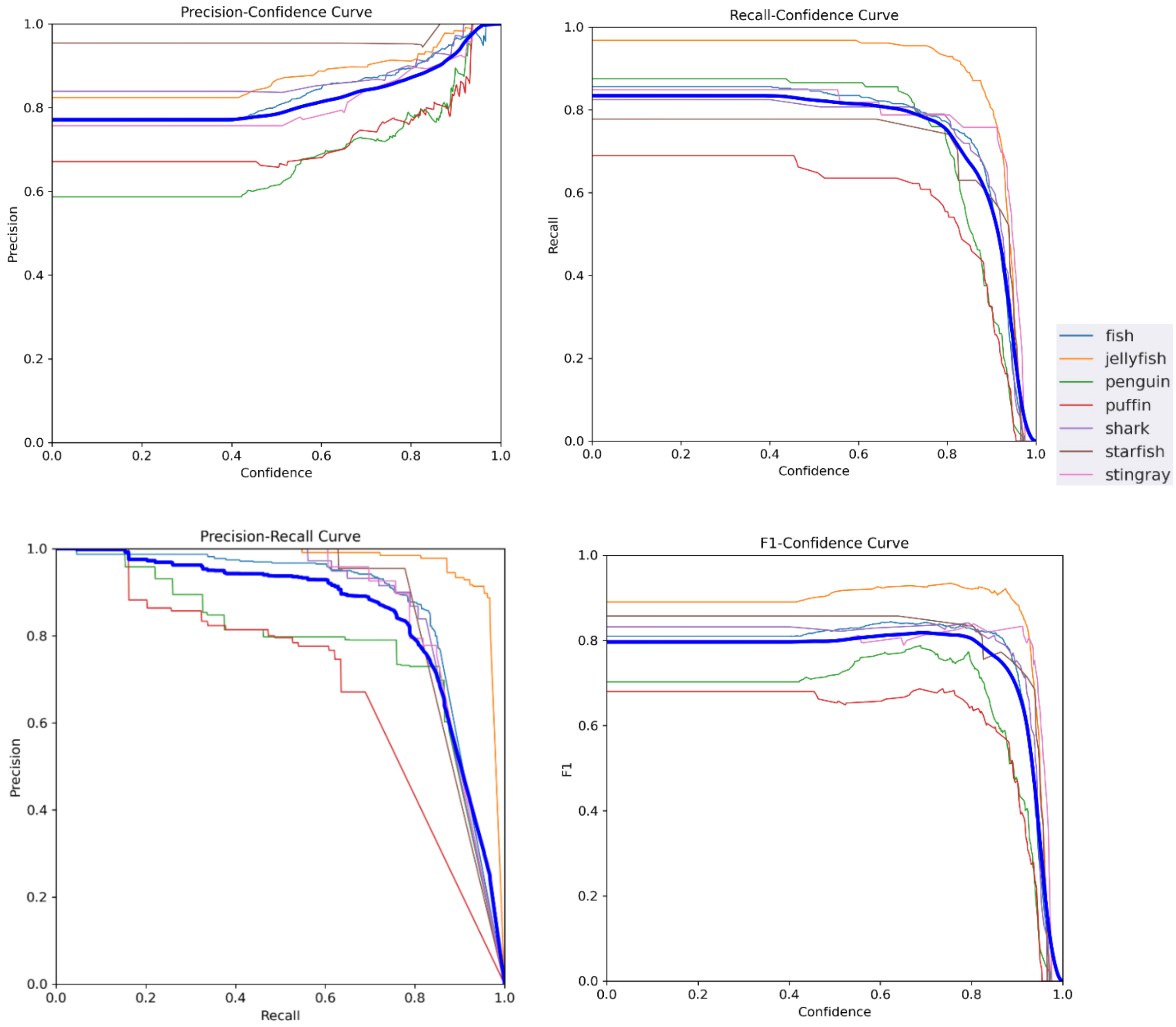


Fig.11. Performance metrics

Table 2: The detection results for 7 class categories of the underwater object detection dataset images

Class	P	R	mAP@50	mAP50-95
Fish	87.2	81.3	87.7	56.4
Jellyfish	90.3	95.5	96.9	63.4
Penguin	72.6	83.9	78.1	45.2
Puffin	74.4	62.8	69.2	41.8
Shark	86.6	80.7	87.8	65
Starfish	95.4	76.3	87.7	74.9
Stingray	84.3	78.8	88.1	70.6

4.4. DISCUSSION

In this training configuration, we outline the parameters and settings used for training the proposed model model. The training process spans 100 epochs, with each batch comprising 16 images resized to 640x640 pixels. The training process utilizes the following specifications: lr0=0.01, lrf=0.01, momentum=0.937, weight_decay=0.0005, warmup_epochs=3.0, warmup_momentum=0.8, warmup_bias_lr=0.1, box=0.05, cls=0.5, cls_pw=1.0, obj=1.0, obj_pw=1.0, iou_t=0.2, anchor_t=4.0, fl_gamma=0.0, hsv_h=0.015, hsv_s=0.7, hsv_v=0.4, degrees=0.0, translate=0.1, scale=0.5, shear=0.0, perspective=0.0, flipud=0.0, fliplr=0.5, mosaic=1.0, mixup=0.0, copy_paste=0.0.

The **table 1** presents a comparative analysis of various models in terms of their performance metrics, year of publication, backbone architecture, test image resolution (in pixels), and average precision at 50% overlap (AP50) scores. These models have been evaluated on a common dataset, and their respective results are summarized as: SA-SPPN [32], published in 2022, utilizes the DarkNet-53 backbone and operates on test images with a resolution of 608x608 pixels. It achieves an AP50 score of 79.23, indicating its ability to accurately detect objects with a 50% overlap. B-YOLOX-S [33], also from 2022, is based on the DarkNet-53 backbone and processes test images of size 640x640 pixels. It demonstrates superior performance with an AP50 score of 82.68, suggesting robust object detection capabilities. SWIPENET-CMA [37], introduced in 2022, uses the SWIPENET architecture and evaluates objects in 512x512 pixel images. It achieves an AP50 score of 68, indicating moderate performance in object detection. SWIPENET-IMA [38], published in 2020, also employs the SWIPENET backbone with 512x512 pixel test images. However, it lags behind in performance with an AP50 score of 64.5 compared to more recent models.

RoIAttn [39], released in 2022, adopts the ResNet-50 backbone and tests objects in images sized 1000x600 pixels. It attains a competitive AP50 score of 82, demonstrating strong object detection capabilities. RFTM-XT101 [40], introduced in 2023, leverages the ResNetXT101 backbone and processes test images with a resolution of 640x640 pixels. It achieves an impressive AP50 score of 84.7, showcasing high accuracy in object detection. YOLOX-U [41], published in 2023, uses the Cascade Mask R-CNN with Swin-T as its backbone and evaluates objects in 640x640 pixel images. It achieves an AP50 score of 83.65, indicating excellent object detection performance. Proposed Model, employs a modified CSP Darknet backbone and operates on test images with a resolution of 512x512 pixels. It outperforms all other models in the table with an impressive AP50 score of 85.62, highlighting its exceptional object detection accuracy. The **table 2** provides a comprehensive overview of the performance metrics. The proposed model stands out as the top-performing model in this comparison, indicating its potential for real-world applications where precise object detection is essential.

5. CONCLUSION

This paper presents a robust and innovative approach to address the intricate challenges of underwater object detection, stemming from the inherent distortion and light attenuation within aquatic environments. By combining morphological edge enhancement techniques with an efficient detection model featuring bottleneck layers, a comprehensive solution is outlined. Through a critical pre-processing stage encompassing color correction and contrast enhancement, the dataset is refined to bolster the quality of input data. In an effort to fortify the model's adaptability across diverse underwater scenarios, dataset augmentation introduces variations in lighting conditions, water clarity, and object poses. Central to this approach, bottleneck layers role in reducing spatial dimensions while enhancing depth not only compresses information but also minimizes computational overhead. Experimental validation conducted on the underwater dataset yields remarkable results. The proposed model achieves impressive metrics, including a mean average precision (mAP) of 85.1%, precision of 84.4%, and recall of 79.9%. These outcomes substantially surpass the performance of existing models, underscoring the model's superior efficacy in the detection of underwater objects.

REFERENCES

1. Lee, M. F. R., & Chen, Y. C. (2023). Artificial Intelligence Based Object Detection and Tracking for a Small Underwater Robot. Processes, 11(2), 312.
2. Song, P., Li, P., Dai, L., Wang, T., & Chen, Z. (2023). Boosting R-CNN: Reweighting R-CNN samples by RPN’s error for underwater object detection. Neurocomputing, 530, 150-164.
3. Javaid, M., Maqsood, M., Aadil, F., Safdar, J., & Kim, Y. (2023). An Efficient Method for Underwater Video Summarization and Object Detection Using YoLoV3. Intelligent Automation & Soft Computing, 35(2), 1295–1310.

4. Li, M. (2023, February). Deep-learning-based research on detection algorithms for marine fish. In Third International Conference on Computer Vision and Data Mining (ICCVDM 2022) (Vol. 12511, pp. 635-639). SPIE.
5. Haider, A., Arsalan, M., Nam, S. H., Sultan, H., & Park, K. R. (2023). Computer-aided fish assessment in an underwater marine environment using parallel and progressive spatial information fusion. Journal of King Saud University-Computer and Information Sciences, 35(3), 211-226.
6. Hung, K. C., & Lin, S. F. (2022). An adaptive dynamic multi-template correlation filter for robust object tracking. Applied Sciences, 12(20), 10221.
7. Qureshi, S. A., Hussain, L., Chaudhary, Q. U. A., Abbas, S. R., Khan, R. J., Ali, A., & Al-Fuqaha, A. (2022). Kalman filtering and bipartite matching based super-chained tracker model for online multi object tracking in video sequences. Applied Sciences, 12(19), 9538.
8. Majstorović, I., Ahac, M., Madejski, J., & Lakušić, S. (2022). Influence of the analytical segment length on the tram track quality assessment. Applied Sciences, 12(19), 10036.
9. Peng, L., Zhu, C., & Bian, L. (2023). U-shape transformer for underwater image enhancement. IEEE Transactions on Image Processing.
10. Zhou, J., Sun, J., Zhang, W., & Lin, Z. (2023). Multi-view underwater image enhancement method via embedded fusion mechanism. Engineering Applications of Artificial Intelligence, 121, 105946.
11. Jiang, Q., Zhang, Y., Bao, F., Zhao, X., Zhang, C., & Liu, P. (2022). Two-step domain adaptation for underwater image enhancement. Pattern Recognition, 122, 108324.
12. Shang, Y., Guo, Y., & Tang, J. (2022). Spectroscopy and chromaticity characterization of yellow to light-blue iron-containing beryl. Scientific Reports, 12(1), 10765.
13. Ali-Bik, M. W., Sadek, M. F., & Hassan, S. M. (2022). Basement rocks around the eastern sector of Baranis-Aswan road, Egypt: Remote sensing data analysis and petrology. The Egyptian Journal of Remote Sensing and Space Science, 25(1), 113-124.
14. Wang, H., Sun, S., & Ren, P. (2023). Meta underwater camera: A smart protocol for underwater image enhancement. ISPRS Journal of Photogrammetry and Remote Sensing, 195, 462-481.
15. Zhou, J., Pang, L., Zhang, D., & Zhang, W. (2023). Underwater image enhancement method via multi-interval subhistogram perspective equalization. IEEE Journal of Oceanic Engineering, 1–15.

16. Jebadass, J. R., & Balasubramaniam, P. (2022). Low contrast enhancement technique for color images using interval-valued intuitionistic fuzzy sets with contrast limited adaptive histogram equalization. Soft Computing, 26(10), 4949-4960.
17. Zhang, W., Wang, Y., & Li, C. (2022). Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement. IEEE Journal of Oceanic Engineering, 47(3), 718-735.
18. Wang, J., Zha, F., & Bi, X. (2023). Research on Underwater Image Semantic Segmentation Method Based on SegNet. In Proceedings of the International Conference of Fluid Power and Mechatronic Control Engineering (ICFPMCE 2022). (Vol. 10, p. 131). Springer Nature.
19. Zhang, W., Zhuang, P., Sun, H. H., Li, G., Kwong, S., & Li, C. (2022). Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. IEEE Transactions on Image Processing, 31, 3997-4010.
20. 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, *527*, 204–232.
21. Mehranian, A., Wollenweber, S. D., Walker, M. D., Bradley, K. M., Fielding, P. A., Su, K. H., ... & McGowan, D. R. (2022). Image enhancement of whole-body oncology [18F]-FDG PET scans using deep neural networks to reduce noise. European journal of nuclear medicine and molecular imaging, 49(2), 539-549.
22. Azhar, A. S. B. M., Harun, N. H. B., Yusoff, N. B., Hassan, M. G. B., & Chu, K. B. (2022, May). Image Enhancement on Underwater Images for Protozoan White Spot Fish Disease Detection. In 2022 International Conference on Intelligent Systems and Computer Vision (ISCV) (pp. 1-4). IEEE.
23. Yang, G., Tian, Z., Bi, Z., Cui, Z., & Liu, Q. (2022). Adjacent Frame Difference with Dynamic Threshold Method in Underwater Flash Imaging LiDAR. Electronics, 11(16), 2547.
24. Zhou, J., Yao, J., Zhang, W., & Zhang, D. (2022). Multi-scale retinex-based adaptive gray-scale transformation method for underwater image enhancement. Multimedia Tools and Applications, 1-21.
25. Tang, C., von Lukas, U. F., Vahl, M., Wang, S., Wang, Y., & Tan, M. (2019). Efficient underwater image and video enhancement based on Retinex. Signal, Image and Video Processing, 13, 1011-1018.
26. Du, Y., Yuan, C., Li, B., Zhao, L., Li, Y., & Hu, W. (2018). Interaction-aware spatio-temporal pyramid attention networks for action classification. In Proceedings of the European conference on computer vision (ECCV) (pp. 373-389).

27. Srinivas, A., Lin, T. Y., Parmar, N., Shlens, J., Abbeel, P., & Vaswani, A. (2021). Bottleneck transformers for visual recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16519-16529).
28. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
29. Peng, W. Y., Peng, Y. T., Lien, W. C., & Chen, C. S. (2021, September). Unveiling of how image restoration contributes to underwater object detection. In 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW) (pp. 1-2). IEEE.
30. Chen, S., Wu, Y., Liu, S., Yang, Y., Wan, X., Yang, X., ... & Yan, X. (2023). Development of Electromagnetic Current Meter for Marine Environment. Journal of Marine Science and Engineering, 11(1), 206.
31. Blasiak, R., Jouffray, J. B., Amon, D. J., Claudet, J., Dunshirn, P., Søgaard Jørgensen, P., ... & Österblom, H. (2023). Making marine biotechnology work for people and nature. Nature Ecology & Evolution, 7(4), 482-485.
32. Xu, F., Wang, H., Sun, X. and Fu, X., 2022. Refined marine object detector with attention-based spatial pyramid pooling networks and bidirectional feature fusion strategy. *Neural Computing and Applications*, *34*(17), pp.14881-14894.
33. Wang, J., Qi, S., Wang, C., Luo, J., Wen, X. and Cao, R., 2022. B-YOLOX-S: A Lightweight Method for Underwater Object Detection Based on Data Augmentation and Multiscale Feature Fusion. *Journal of Marine Science and Engineering*, *10*(11), p.1764.
34. Jia, J., Fu, M., Liu, X. and Zheng, B., 2022. Underwater object detection based on improved efficientdet. *Remote Sensing*, *14*(18), p.4487.
35. Ouyang, W. and Wei, Y., 2023. An Anchor-free Detector with Channel-based Prior and Bottom-Enhancement for Underwater Object Detection. *IEEE Sensors Journal*.
36. Zhang, J., Zhang, J., Zhou, K., Zhang, Y., Chen, H. and Yan, X., 2023. An Improved YOLOv5-Based Underwater Object-Detection Framework. *Sensors*, *23*(7), p.3693.
37. Chen, L., Zhou, F., Wang, S., Dong, J., Li, N., Ma, H., Wang, X. and Zhou, H., 2020. SWIPENET: Object detection in noisy underwater images. arXiv preprint arXiv:2010.10
38. Chen, L., Liu, Z., Tong, L., Jiang, Z., Wang, S., Dong, J. and 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.006.

39. Liang, X. and Song, P., 2022, October. Excavating roi attention for underwater object detection. In 2022 IEEE International Conference on Image Processing (ICIP) (pp. 2651-2655). IEEE.
40. Fu, C., Fan, X., Xiao, J., Yuan, W., Liu, R. and Luo, Z., 2023. Learning Heavily-Degraded Prior for Underwater Object Detection. IEEE Transactions on Circuits and Systems for Video Technology.
41. Ouyang, W. and Wei, Y., 2023. An Anchor-free Detector with Channel-based Prior and Bottom-Enhancement for Underwater Object Detection. IEEE Sensors Journal.