



A Multiscale GAN- SOD Yolov7 Framework for Robust Small Object Detection in Complex Underwater Environment

Atul Patil^{1,3*} and Vinay Kumar Singh²

¹ Research Scholar, Dept. Computer Science and Engineering, Amity School of engineering and Technology, Amity University Raipur, Chhattisgarh, India

² Professor, Department Computer Science and Engineering Amity University Raipur Chhattisgarh, India

³ Associate Professor, Dept. of Computer science SVKM's NMIMS, MPSTME Shirpur, India
*atulrpatil1@gmail.com
vksingh@rpr.amity.edu
atul.patil@nmims.edu.com

Abstract. The detection of underwater objects is very important in such areas as the marine search, underwater surveillance and environmental monitoring. But it is a difficult task because underwater domain is complex. Where traditional object detection methods can fall short in these cases, due to problems such as differing illumination conditions, low contrast, turbidity and light scattering, obfuscation/occlusion or biological camouflage of marine life with respect to their background. These difficulties contribute to the degradation of performance of traditional algorithms. Against this background, in this paper we propose an advanced underwater object detection framework that combines a Multiscale GAN with the SOD-YOLOv7 detector. The proposed model is comprised of several main components: Adaption Denoise Block, CNN-based Dual Attention Selective (DAS) Network, synthetic data augmentation and Multi-scale Unpaired GAN. The Adaption Denoise Block efficiently diminishes noise and makes the critical aspects in the underwater photos stand out. The DAS network enhances feature extraction via spatial and channel attention mechanism. In this case, we use the multi scale GAN to generate super-resolved images which are more closer to those generated under real-world settings. We evaluate the proposed method on the publicly available UTDAC2020 dataset which includes a total 6164 images that have different resolutions, including (3840×2160), (1920×1080), (720×405), (704×576) and (586×480). The results show significant improvement over the classical methods, with a mAP = 97.40%, precision = 97.18% and recall = 94.85 %. The model works at fast 115 FPS detection speed and about 105.4 GFLOPs computational complexity. In summary, the SOD-YOLOv7 model put forward in this paper offers a competitive way to detect small object in complex underwater.

Keywords: DAS, Data augmentation, YOLOv7, Multiscale GAN, Underwater Object Detection

© The Author(s) 2026

S. Bhalerao et al. (eds.), *Proceedings of the 2nd International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2026)*, Advances in Intelligent Systems Research 207,

https://doi.org/10.2991/978-94-6239-678-4_7

1 Introduction

In marine research and conservation underwater to underwater monitoring, the underwater object detection is a key field of research with a wide range of applications [1]. Locating and recognizing objects under the ocean floor play important roles in numerous areas [2][5], including national security, offshore engineering and marine biology. Nevertheless, object identification in water is a challenging issue that introduces a specific range of hardships and challenges [6][7]. One of the largest challenges in detection of items in underwater is the intrinsic nature of the aquatic environment. The difference between the underwater scenes and those which are above water is that they are defined by dynamic lighting, varying conditions of water, varied degrees of visibility and turbidity.

The detection of underwater objects is one of the challenging tasks because of the nature of aquatic environment. Dynamic lighting as well changing water conditions induce different turbidity and visibility are some of the reasons underwater photography differs from terrestrial or airborne one, in particular due to significant amount of non-waterborne particulates [8][10]. All of that together reduces image quality to such a degree that it's difficult to see what's in the water. Furthermore, objects in underwater scenes usually present varied textures as well as irregular shapes and sizes [11][12], which makes the detection process even more challenging [13]. Occlusion is another significant barrier. Detection in underwater is even more difficult because objects may be concealed by columns of water, vegetation (or animals) or submerged objects including other targets [14]. If so, this may contribute to the poor matching in some cases observed as water filtering light and color sometimes lends towards lower contrast / differentiation [15] making objects less easily recognized in photographs. The method developed for detecting underwater objects on the UTDAC 2020 is very promising. The main objectives are to increase sensitivity - particularly for small objects, better denoising in the data, and to enable domain adaptation for better detection. One of the strong points of this approach is its capability to generalize across a variety of underwater scenarios, on which it can be attacked. Moreover, it is important to construct a holistic approach that exploits its advantage in the feature extraction.

2 Literature Review

Task of detection of underwater object is difficult due to such problems as unbalanced light, low contrast and occlusion. A number of detectors often yield unreliable detection results when applied to ambiguous underwater targets. In response [16] proposed a Boosting R-CNN, which integrates hard example mining and uncertainty tuning. Namely, we are including such voting strategies as a probabilistic inference pipeline for predictions, boosting reweighting for robustness, or Retina RPN for high quality proposals. However, once a truly complex underwater scenery is encountered the method can still be limited.

The research presented in 2022 [17], presented a novel deep neural network for anomaly detection in underwater environments that includes an autoencoder, an ARN, and a patch learning paradigm. By efficiently reducing the impact of global information

and improving anomaly identification with local information, the network performs exceptionally well. The need for thorough labelled data, difficulties adjusting to quickly shifting underwater conditions, and the requirement for additional study to guarantee successful generalization to many and unknown underwater scenarios are some of its possible drawbacks, though. This effort is a major step toward enabling curiosity-driven autonomous underwater exploration, despite these obstacles.

In [18], presented a new training paradigm termed CMA together with a SWIPENET. Image resolution, feature maps, and enhanced object detection are provided by SWIPENET. Inspired by human education, the CMA training facilitates generalization and convergence by starting with easy to tough levels. SWIPENET+CMA has a greater time complexity because of its ensemble nature and does not respond to scenarios with significant noise, even if it successfully manages noise issues in submerged object detection.

The [19], introduced a class-weighted YOLO network-based underwater target detection technique with an emphasis on improving outcomes while balancing sample difficulty. The suggested method entails creating a loss function that is weighted by class. An adaptive dimensional clustering approach is also presented for target box refining in order to enhance detection. First, side-scan sonar data are preprocessed to reduce noise and use speckle detection to extract putative target areas. The target scale is then estimated, spurious targets are removed, and object pixel detection is carried out. The approach reduces non-static background noise by combining background and three-frame differences and applying morphological procedures. The technique does, however, recognize its limitations when it comes to managing small, fuzzy, and complicated underwater images, highlighting the difficulties in attaining high recall rates and addressing morphological variations among underwater targets.

Yeh et al. [20] designed a deep underwater detecting network, which is relatively light. The objectives of this paper are to pursue the reduction of color absorption and improve the UWD performance and at the same time keep computation economic by jointly training color conversion and object detection. Tests on a Raspberry Pi model confirm that the proposed model is useful. The effects of image degradation are mitigated by the approach. This approach can cause generalization, speed and scalability limits.

In the research [21], proposed an underwater target detection model that integrated adversarial network based on the Faster RCNN detector. Through the artificial occlusions, the adversarial network challenges the detection network and enhances its robustness for underwater scenes. The adversarial network introduces extra training examples with modified features instead of pixel space. This approach solves the problems with fewer undersea data and further enhancing detection capabilities. Nonetheless, highly visualized underwater images may exhibit the limitations of this approach and suggest some potential directions for future research and improvement. The [22], introduced a deep learning-based sea cucumber detection method tailored for underwater environments. It tackles issues unique to underwater environments, such as the requirement to manage differences in sea cucumber appearances and the scarcity of statistics. The suggested S-FPN and PFL methods greatly enhance detection performance, and the strategy is tailored for practical mariculture applications. To solve

data imbalance, improve model performance, and account for dynamic changes, more study is still possible.

The researcher [23], addressed the challenge of multi-object tracking when provided with unreliable object finding. It introduces RCT, a search algorithm that use unfiltered data detection and is guided by detection confidence levels. RCT is superior than other tracking techniques, according to the study, especially when dealing with subpar detectors. This study is one of the few MOTs that compares a lot of baselines across several domains. Particularly when it comes to fish monitoring in underwater areas, the FISHTRAC dataset enhances the research. The usage of detection confidence is a fundamental part of the suggested methodology. Nevertheless, it is not feasible to scale or generalize to new situations.

In 2021, the author [24] gave a modification of faster mode of R-CNN, the study aims to improve detection accuracy. It incorporates selectable kernel units, modifies the feature pyramid network, and streamlines anchor selection in the Region Proposal Network. Although the research emphasizes the benefits of the suggested method, it is important to take into account any possible drawbacks, such as the precision of object location and generalization to various underwater environments. Accurate detection of marine benthos requires the removal of these limitations.

The [25], presented a learning-based approach that prioritizes edge computing, quick data transport, and high-resolution photos for visual surveillance in the deep sea. It is important to note that deep-sea mining situations present particular difficulties and environmental issues, even with the considerable developments in the suggested technology. The adaptability of AI models and data transfer may be constrained by such intricate underwater environments.

3 The Proposed Methodology

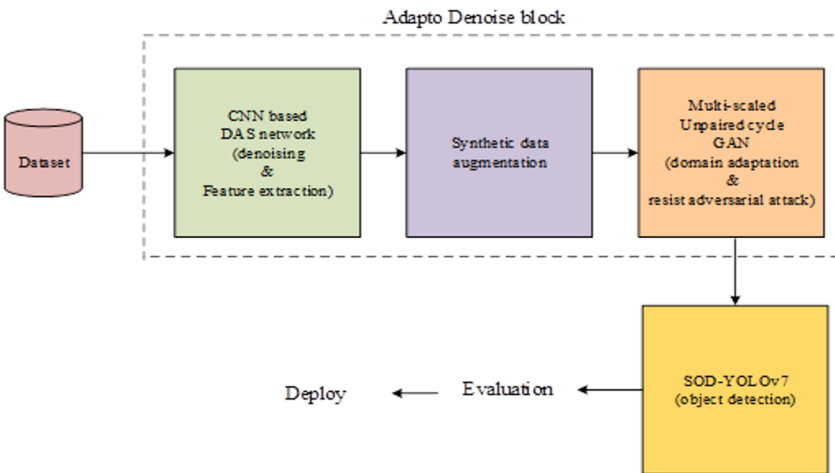


Fig. 1. Proposed methodology

The proposed method shown in the fig.1 for detecting small objects in underwater environment using data from underwater cameras. This approach works on the various stages to make detection more accurate, reliable and also to improve precision and flexibility. Once the raw data is extracted from the UTDAC 2020 dataset the images

are pre-process using Adapto Denoise Block (ADB), it consist of CNN-based Dual Attention Skip Connection (DAS) network which is use to capture the low level and high level features crucial for object detection. The process images then applied to synthetic data augmentation it generates diverse dataset through different transformation. The Multi-scaled Unpaired Cycle GAN with domain adaption, multiscale network helps to capture features at various level while the Cycle GAN facilitates domain adaptation, allowing for translation between source and target domains without requiring one-to-one correspondences. It also improve resilience to adversarial attacks and control domain flexibility. After that, SOD-YOLOv7, a modified YOLOv7 architectural version, receives the updated data. The YOLOv7 with Small Object Detection (SOD) is innovative approach to detect small underwater objects This innovative approach incorporates a combination of dilated convolutions, common convolutions and the PAN (Path Aggregation Network) module, situated within the YOLOv7 model's "neck" before the feature pyramid pooling (FPN) module. This technique significantly increases accuracy and flexibility for small objects in underwater environment.

3.1 The Dataset

The UTDAC 2020 dataset is a significant and varied collection of images used for underwater object detection including more than 20,000 among diverse underwater objects such as fish, coral reef, submarines etc. These were shot at various depths and under water conditions from clear to dark. The training set contains more than 16,000 images and the test set more than 4000, and had made sure none of the training examples appears in the test. The objects within each image have been meticulously annotated with bounding boxes. Despite the inherent challenges caused by the oftentimes noisy, low-quality photos and the fact that objects are often buried or partially veiled, the UTDAC 2020 dataset collection is particularly valuable for researchers and developers trying to improve underwater photography. However, even with the challenges that the sometimes noisy and low-quality photos bring to an imprecise localization of the objects (as well as in many cases they are buried or partially hidden), UTDAC 2020 dataset is very interesting for researchers and developers who aim at performing a better set up of underwater object detection systems by collecting the images at different scales, resolutions. The greatest and most comprehensive selection of online learning products anywhere in the industry, this resource provides an essential groundwork for building artists' skills.

3.2 Adapto Denoise Block

The building block of the proposed underwater object detection technology is called "Adapto Denoise". This block includes the CNN-based DAS network, multi-scaled unpaired Cycle GAN and synthetic data augmentation that plays several pivotal roles. Meanwhile, a DAS network is designed for gathering the related details to support the model's perception of object features and noise reduction. Synthesis data enhancement, such as DNN adding synthesized data to the training set, enhances the model's capability to work under different underwater situations. This block also covers domain adaptation, where it is desirable that model can identify objects trustworthily

and consistently while adapting to different underwater scenarios. By hardening itself against adversarial attacks, the ‘‘Adapto Denoise’’ block further improves the model’s real-time performance of object detection.

3.3 CCN Based DAS Network

CNN based DAS network in fig. 2 contains Convolutional Neural Network (CNN) with skip connections to process UTDAC 2020 data as the primary input. With these skip-connections, both low-level and high-level features necessary for object detection can be included in the network. The CNN produces in a stack of feature maps indicating different levels of abstraction of the data. Following that, a channel attention layer receives the feature maps, suppresses less important features, and assigns attention weights to different feature channels. A spatial attention layer simultaneously assesses and pinpoints significant spatial regions within the feature maps. Concatenating the results of both attention algorithms yields an attention-guided feature map. This combined feature map, which has been improved by both channel-wise and spatial attention, is processed by another CNN to further refine and improve the data. Updated feature maps representing feature-extracted and denoised data comprise the result. Finally, up sampling is utilized to generate data that is prepared for further object detection or visualization activities by improving the spatial resolution. For accurate underwater object detection, this painstaking procedure enhances the image data by removing noise and focusing on key features.

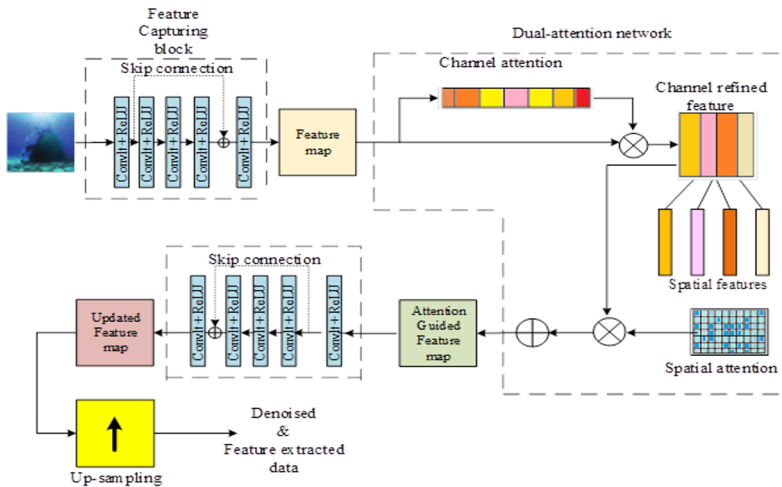


Fig. 2. CNN based DAS network

3.4 Data Augmentation

Synthetic data augmentation is crucial for the underwater object detection, as due to practical and environmental constraints it is often challenging to obtain enough high-quality labelled data. Due to phenomena such as light absorption, bisexual scattering, turbidity, occlusion and redundancy among species appearance of underwater contents

vary greatly. Such diversities make it hard for detection models trained on small datasets to be reliable under different conditions.

To address these issues, synthetic data augmentation techniques are exploited artificial samples are generated by warping endogenous images or employing sophisticated generative models. Some of these manipulations include rotating, scaling and flipping the images, injecting noises into the data during training process augmenting random noise at run time, standardizing image colors to a common mean value. More sophisticated methods such as GANs can be used to augment the data, and add new photo-realistic underwater images even. Such augmentations target to expand the diversity of training data without involving expensive and time-consuming collecting or hand annotation of real underwater photos. Synthetic augmentation contributes to the robustness and generalization of the detection model in real-world underwater conditions by letting it see more diverse visual transformations during training.

In addition, this approach also solves the problem of data imbalance that there may be fewer examples available for some certain types of objects, and at the same time significantly reduces the cost and effort in collecting a large amount of data. Thus, synthetic data augmentation ensures that the trained model can be flexible scalable and reliable in diverse unpredictable underwater environments, besides improving detection performance.

3.5 Multi-Scaled Unpaired GAN

In order to address the challenges emerging from underwater imaging, such as noise and distortions caused by the aquatic environment, denoising methods are further applied for the image restoration. And the denoised images are adapted to a related yet different undersea environment and prevent potential adversarial attack by using a multi-scaled unpaired cycle GAN. The Cycle GAN supported domain adaptability, allowing the translator to translate across source and target domains even when one-to-one correspondence is not required, a NSCA multi-scale network helped attain features at different scales. Adversarial training and other defense techniques can be added to make the model more robust against adversarial samples. As a result, we obtain a series of domain-adapted feature extracted and denoised underwater images that are robust against potential adversarial attacks therefore can be applied to various underwater imaging tasks.

The distortions caused by the water medium, such as light scatter, color loss, turbidity and particle drops in suspension become a difficult to handle problem for underwater imaging. Those features limit the computer vision models for underwater operations, including navigation and detection, made difficult by low visibility and noisy background images. To address it, denoising techniques are employed that reduces unwanted noise while retaining the structural information. Convolutional neural network (CNN)-based denoisers and filters are two such representatives which enhance contrast and sharpness while creating cleaner features, serving as a good basis for further processing.

Domain shift is due to the difference between blue water and green water in a deep environment compared with shallow water. The adaptability is guaranteed by employing a multi-scaled unpaired Cycle Generative Adversarial Network (CycleGAN) that is the unpaired CycleGAN can be used to translate cross several

undersea domains without building one-to-one paired images and multi-scale architecture which considers not only large-scale features but fine-grained textures. The model learns strong mappings across the source-target domains by enforcing cycle-consistency, which minimizes domain-agnostic variances and maintains characteristics coherent between different environments.

Models for underwater environments should be robust to adversarial attacks, such as small, carefully designed impostors resulting in misclassification, on top of regular variance. Adversarial robustness methods are employed to alleviate this. Adversarial training, where the model is trained on both clean and perturbed examples, ameliorates models' susceptibility to attacks. The resistance is outperformed through some defense strategies, such as input modification and feature denoising. As a result, the final outputs here are sets of underwater images that have been denoised, domain-adapted and adversarially-nonthreatened. This leads to better generalization and reliability, and the pipeline is well suited for critical underwater applications such as object detection, marine ecosystem monitoring, infrastructure inspection and AUV navigation.

3.6 SOD-YOLOv7

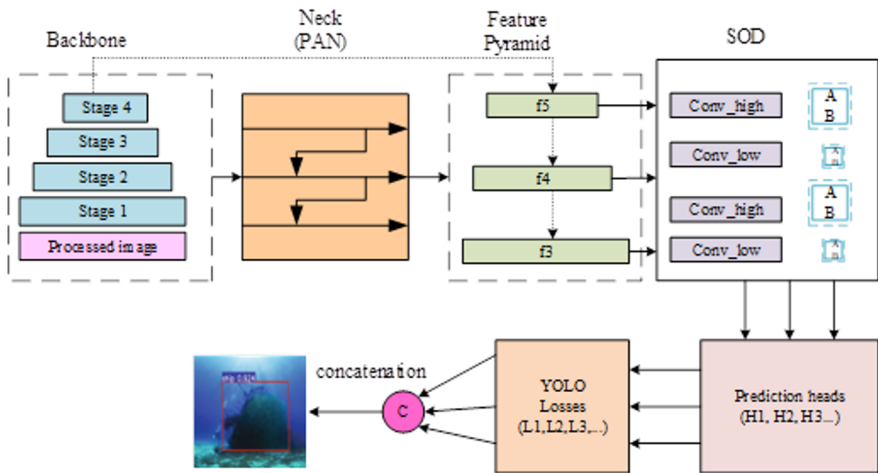


Fig. 3. Proposed SOD-YOLOv7 Architecture

As depicted in Fig.3, SOD-YOLOv7 is a state-of-the-art Small Object Detection (SOD) method and is implemented based on YOLOv7 object detection model specially designed for the challenging task of small object detection. Introducing a mixture of dilated, ordinary convolutions and PAN (Path Aggregation Network) module in the YOLOv7 model "neck", which follows the FPN module. The PAN module strengthens the ability of the model to gather information from different levels of backbone layers, which can also help with capturing small objects in different scales. Rate dilated convolution involves long range dependencies, ordinary convolutions bias short term dependencies and the PAN module helps powerful feature collection. In training, the

same loss function adopted in YOLOv7 is utilized by SOD-YOLOv7 to decrease the prediction mistake. After training, the model is good at quickly finding small objects in new images of new classes with accurate predictions about both where each of those objects are and what they are. SOD-YOLOv7 offers substantial improvements in small object detection performance, heightened robustness in noisy and occluded scenarios and faster inference times, making it a powerful enhancement for YOLOv7's capabilities in small object detection tasks.

4 Experimental Results

In this section experimental environment, hyper parameters, ablation study, comparison with traditional object detection models and some YOLO series models is performed. The dataset and optimization of anchor boxes on different underwater images are described. The experimental results shows that the proposed model of SOD-YOLOv7 with multiscale GAN improves the accuracy, precision, and detection speed in underwater environment.

4.1 Experimental Environment

The above method is implemented using Python on the dataset collected from the UTDAC 2020 dataset it consist of 5168 training and 1293 validation images with an NVIDIA RTX 3090 GPU, 64GB RAM and combination of SSD and HDD storage for efficient data handling and processing.

4.2 Simulating Parameters – are shown in table 1

Table 1. Hyper parameters

Parameter	Proposed Value
Image Size	640×640
Learning rate	0.001
Batch size	16
Optimizer	SGD
Momentum	0.937
Number of epochs	200
Weight decay	0.0005

4.3 Dataset Description

To check the performance of the proposed method UTDAC2020 dataset, derived from the Underwater Target Detection Algorithm Competition 2020 is used which contains total images of 6164 with varying resolutions such as 3840×2160, 1920×1080, 720×405, 704×576, and 586×480. The total images are divided as 5168 training and 1293 testing or validation images which have four different object categories like Echinus, Holothurian, Starfish, Scallop with image size of 640×640.

4.4 Performance Metrics

To check the performance of object detection the commonly used metrics are Precision, Recall, and mean Average Precision (mAP). Above performance metrics are calculated using equations 1, 2, 3 and 4 respectively:

$$\text{Precision} = \frac{\text{TruePositives(TP)}}{\text{TruePositives(TP)} + \text{FalsePositives(FN)}} \times 100 \tag{1}$$

Precision is used to measures the true positive predictions made by the model.

$$\text{Recall} = \frac{\text{TruePositives(TP)}}{\text{TruePositives(TP)} + \text{FalseNegatives(FN)}} \times 100 \tag{2}$$

Recall measures the proportion of true positive predictions among all actual positive cases.

$$\text{mAP} = \left(\frac{1}{\text{number of classes}} \right) * \sum (\text{avg. precision for each class}) \tag{3}$$

Mean Average Precision measures average precisions across all classes.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The F1-Score is a balanced measure that combines the accuracy of Precision and the comprehensiveness of Recall.

4.5 Experimental Results and Analysis on UTDAC2020 dataset

The performance of the proposed model on the UTDAC2020 dataset captured in complex underwater conditions. These images has different scenarios like clear water, low visibility water condition etc. In initial processing the input image need to preprocess to get clear image for object detection. The adapto denoise block is used to refine the image so that the small object can clearly visible. The clear preprocessed image in applied to SODYOLOv7 for detecting objects using bounding boxes as shown in the figure 4.

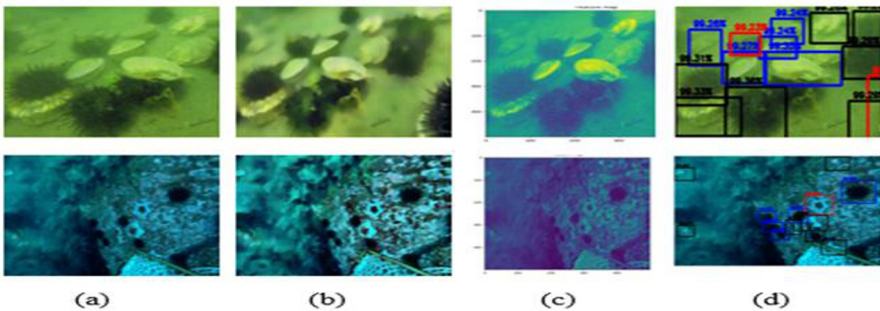


Fig. 4. Proposed method output (a) Input image (b) Pre-processed image(c) Adapto denoise image (d) Final detection output

4.6 Ablation Experiment

Table 2. Ablation experiment on UTDAC2020 dataset

Model	Precision (%)	Recall (%)	F1 Score (%)	mAP(%)
Baseline YOLOv7	81.8	76.3	78.95	79.2
YOLOv7+ADB	82.1	76.9	79.41	80.6
YOLOv7+Multiscaled Unpaired GAN	86.7	79.5	82.94	83.8
YOLOv7+Multiscaled Unpaired GAN+SOD	90.3	88.9	89.59	90.5
YOLOv7+ADB+Multiscaled Unpaired GAN+SOD (Proposed)	97.18	94.85	96.0	97.4

Table 2 shows the ablation experiment conducted on baseline model of YOLOv7 which provides poor performance parameter this model achieves mAPs of 79.2% with precision and recall values of 81.8% and 76.3% respectively. By adding Adaption Denoise Block (ADB) the performance parameter gets improved. The Multiscaled Unpaired GAN and SOD modules contributed to improvement of mAP to 97.4%. Resulting in improvement of 18.2% over the baseline YOLOv7 model.

4.7 Comparison with YOLO series model on UTDAC 2020 dataset

The proposed model is compared with some traditional objects detection models like Faster-RCNN, SSD, Cascade R-CNN, RetinaNet, and Deformable-DETR. The comparison shown in the Table 3 that the proposed model SOD-YOLOv7 achieves the AP50 of score 87.8% on UTDAC2020 dataset. The proposed model shows significant advantages in computational performance, detection capabilities. This model not only provides excellent accuracy but also provides remarkable detection speed of 115 frames per second. This finds the proposed model has great potential in the field of object detection in the underwater environment. This model achieved moderate parameter count and lightweight so it can be operate in any environment.

Table 3. Comparisons with other object detection models on UTDAC 2020 dataset

Model	Param(M)	GFLOPs	FPS	UTDAC2020 (%)		
				AP 50	AP 75	AP
Faster-RCNN [26]	41.36	210.32	8	82.4	46.8	46.0
SSD [27]	26.28	62.74	21	77.5	36.5	40.0
Cascade-RCNN [28]	69.16	236	16	82.3	50.4	47.6
RetinaNet[29]	36.39	207	17	80.4	42.9	43.9
Deformable-DETR 30]	40.0	200.71	19	84.1	45.6	46.1
SOD-YOLOv7 (Ours)	36.5	105.4	115	87.8	55.3	53.3

4.8 Comparison with YOLO series model on UTDAC 2020 dataset

We also perform experiments to compare our model with some YOLO series models on UTDAC2020 dataset. The result in the Table 4 demonstrates that the proposed model SOD-YOLOv7 achieves an impressive mAP@0.5 of 97.40%, also it improves mAP metrics across all the classes. The detection accuracy of the model for starfish and remarkably good as compare to YOLOv5, YOLOv7 and YOLOv7-tiny. The smallest object in the dataset Holothurianis efficiently detected by the different stag-es of backbone layers and SOD layers in the YOLO model.

Table 4. Comparisons with other YOLO series models on UTDAC 2020 dataset

Model	mAP@0.5	mAP@0.95	Holothurian	Scallop	Starfish	Echinus
YOLOv3 [31]	84.4	50.5	74.8	84.3	89.9	89.2
YOLOv5	78.8	43.2	66.7	74.5	97.7	86.7
YOLOv6 [32]	75.5	41.2	61.9	70.6	84.5	84.9
YOLOv7 [33]	85.5	49.5	76.2	84.4	89.6	91.4
YOLOv7-tiny	81.6	45.0	68.8	78.8	88.8	90.0
SOD-YOLOv7 (Ours)	97.4	59.25	97.3	96.2	98.5	97.4

4.9 Visualization Results

The confusion matrices provide classifications performance on our model on UTDAC2020 dataset at mAP@0.5, mAP@0.95 the figure 5 shows how model predict the actual class across the different categories. Each cell represents proportion of samples from particular class. The diagonal cells show correctly classified data and other elements shows misclassified data. Higher the value of diagonal element misclassification rate is less which shows good model performance. The model shows remarkable classification accuracy for Holothurian, Starfish and Echinus but faces difficulty in classifying Scallop samples. The matrix analysis shows that correct prediction rate of each class ‘‘Holothurian’’, ‘‘Scallop’’, ‘‘Starfish’’, ‘‘Echinus’’ is 84.%,

89.3%, 90.1% and 88.90% respectively. The misclassification is more in some of the classes due to smaller size, fluctuating light conditions and presence of natural elements like rocks, corals and seaweed.

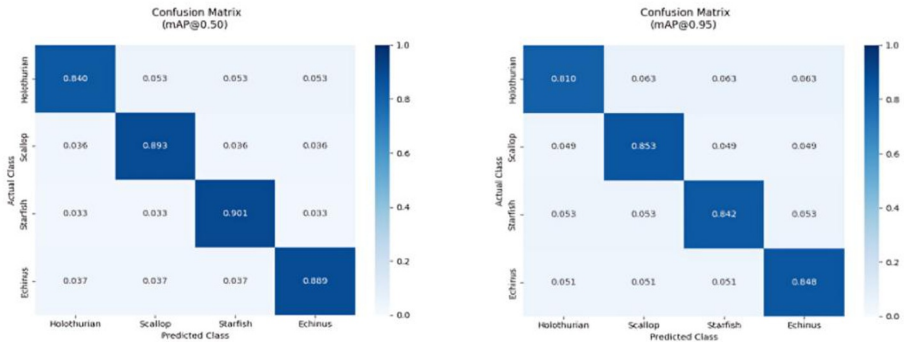


Fig. 5. Confusion matrix of the proposed method for (a) mAP@0.5 (b) mAP@0.95

The figure 6 shows the precision vs recall curve at mAP@0.5, mAP@0.95 on UTDAC2020 dataset on various object categories, the Echinus category which own average precision of 98.9% and the mean average precision (mAP) was 97.10%. Figure 7 shows the comparison of different YOLO models with the proposed models which shows the improvement in precision, recall and mAP.

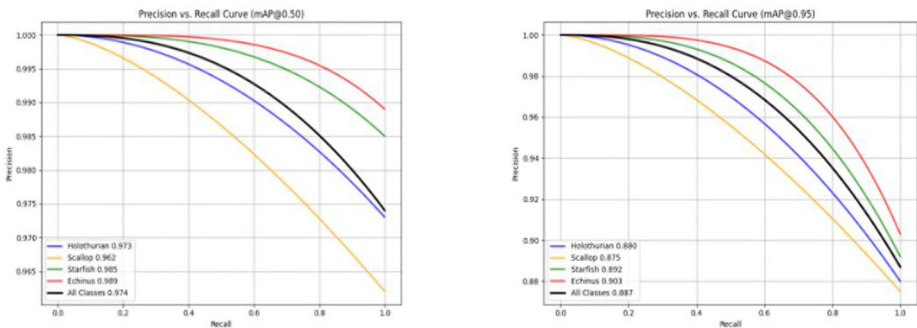


Fig. 6. Precision vs Recall Curve of the model on UTDAC dataset for (a) mAP@0.5 (b) mAP@0.95

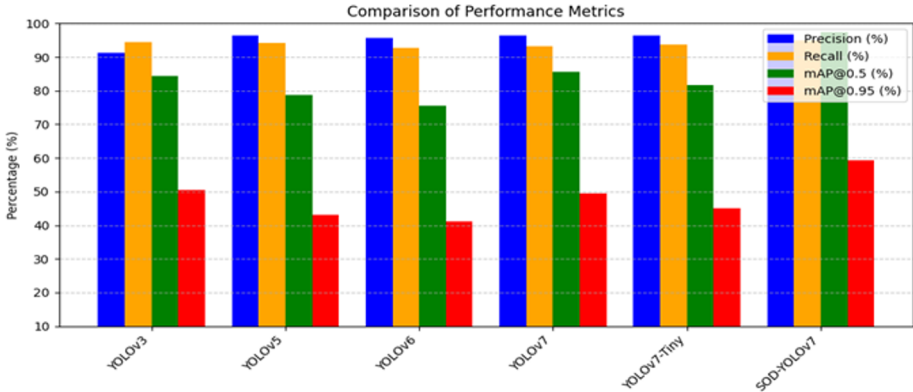


Fig. 7. Performance comparison of existing YOLO models with proposed method

5 Conclusion

This paper presents a proposal of a YOLOv7-based network, which is called SOD-YOLOv7, to identify small objects in turbid underwater scenarios. The study performed based on the UTDAC2020 dataset points to an original and complex solution to the problem of detecting objects under water. The combination of denoising, domain adaptation, and small-object detection enhancements and robust feature extraction into YOLOv7. PAN in SOD-YOLOv7 allows the model to have the ability to aggregate information on different levels that will assist it in detecting small objects under different underwater scenarios. SOD YOLOv7 model recorded the state of art performance and out did both conventional object detection method. The experimental results indicate that SOD YOLOv7 algorithm performs better with a precision of 97.18% and a recall of 94.85% and a mAP of 97.40% and improves the mAP by 18.2 as compared to the baseline model This model has a high speed of detection of 115FPS. The findings were compared to the other popular objects detection methods and showed large enhancements in precision, recall and mAP. These features render SOD YOLOv7 a perfect model to the underwater object detection. The strategy can vastly expand the capacities of SOD YOLOv7 and offer useful usage in marine and aquatic settings, which will eventually lead to a more efficient process of detecting and exploring underwater objects. The findings that were obtained suggest that the given model is very capable even of functioning in complicated underwater conditions. Nonetheless, the scarcity of high-quality underwater datasets and images remains an important problem in developing the target detection in multifaceted underwater settings. Consequently, upcoming studies will be aimed at creating underwater large and diversified datasets in dissimilar settings and implementing image enhancement methodologies to enhance image quality that is very crucial in underwater target detection.

References

1. Fayaz S, Parah SA, Qureshi GJ. Underwater object detection: architectures and algorithms—a comprehensive review. *Multimedia Tools and Applications*. 2022 Jun;81(15):20871-916.
2. Han F, Yao J, Zhu H, Wang C. Underwater image processing and object detection based on deep CNN method. *Journal of Sensors*. 2020 May 22;2020.
3. Fan B, Chen W, Cong Y, Tian J. Dual refinement underwater object detection network. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16 2020* (pp. 275-291). Springer International Publishing.
4. Lin WH, Zhong JX, Liu S, Li T, Li G. Roimix: Proposal-fusion among multiple images for underwater object detection. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020 May 4* (pp. 2588-2592). IEEE.
5. Chen L, Liu Z, Tong L, Jiang Z, Wang S, Dong J, Zhou H. Underwater object detection using Invert Multi-Class Adaboost with deep learning. In *2020 International Joint Conference on Neural Networks (IJCNN) 2020 Jul 19* (pp. 1-8). IEEE.
6. Yu Y, Zhao J, Gong Q, Huang C, Zheng G, Ma J. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. *Remote Sensing*. 2021 Sep 7;13(18):3555.
7. Li M, Mathai A, Lau SL, Yam JW, Xu X, Wang X. Underwater object detection and reconstruction based on active single-pixel imaging and super-resolution convolutional neural network. *Sensors*. 2021 Jan 5;21(1):313.
8. Liu H, Song P, Ding R. WQT and DG-YOLO: Towards domain generalization in underwater object detection. *arXiv preprint arXiv:2004.06333*. 2020 Apr 14.
9. Mathias A, Dhanalakshmi S, Kumar R, Narayanamoorthi R. Deep Neural Network Driven Automated Underwater Object Detection. *Computers, Materials & Continua*. 2022 Mar 1;70(3).
10. Gomes D, Saif AS, Nandi D. Robust underwater object detection with autonomous underwater vehicle: A comprehensive study. In *Proceedings of the International Conference on Computing Advancements 2020 Jan 10* (pp. 1-10).
11. Karimanzira D, Renkewitz H, Shea D, Albiez J. Object detection in sonar images. *Electronics*. 2020 Jul 21;9(7):1180.
12. Zhang X, Fang X, Pan M, Yuan L, Zhang Y, Yuan M, Lv S, Yu H. A marine organism detection framework based on the joint optimization of image enhancement and object detection. *Sensors*. 2021 Oct 29;21(21):7205.
13. Wang CC, Samani H. Object detection using transfer learning for underwater robot. In *2020 International Conference on Advanced Robotics and Intelligent Systems (ARIS) 2020 Aug 19* (pp. 1-4). IEEE.
14. Chen X, Lu Y, Wu Z, Yu J, Wen L. Reveal of domain effect: How visual restoration contributes to object detection in aquatic scenes. *arXiv preprint arXiv:2003.01913*. 2020 Mar 4.
15. Nguyen HT, Lee EH, Bae CH, Lee S. Multiple object detection based on clustering and deep learning methods. *Sensors*. 2020 Aug 7;20(16):4424.
16. Song P, Li P, Dai L, Wang T, Chen Z. Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection. *Neurocomputing*. 2023 Apr 14;530:150-64.
17. Zhou Y, Li B, Wang J, Rocco E, Meng Q. Discovering unknowns: Context-enhanced anomaly detection for curiosity-driven autonomous underwater exploration. *Pattern Recognition*. 2022 Nov 1;131:108860.
18. Chen L, Zhou F, Wang S, Dong J, Li N, Ma H, Wang X, Zhou H. SWIPENET: Object detection in noisy underwater scenes. *Pattern Recognition*. 2022 Dec 1;132:108926.

19. Zhang L, Li C, Sun H. Object detection/tracking toward underwater photographs by remotely operated vehicles (ROVs). *Future Generation Computer Systems*. 2022 Jan 1;126:163-8.
20. Yeh CH, Lin CH, Kang LW, Huang CH, Lin MH, Chang CY, Wang CC. Lightweight deep neural network for joint learning of underwater object detection and color conversion. *IEEE Transactions on Neural Networks and Learning Systems*. 2021 Apr 26;33(11):6129-43
21. Zeng L, Sun B, Zhu D. Underwater target detection based on Faster R-CNN and adversarial occlusion network. *Engineering Applications of Artificial Intelligence*. 2021 Apr 1;100:104190.
22. Peng F, Miao Z, Li F, Li Z. S-FPN: A shortcut feature pyramid network for sea cucumber detection in underwater images. *Expert Systems with Applications*. 2021 Nov 15;182:115306.
23. Mandel T, Jimenez M, Risley E, Nammoto T, Williams R, Panoff M, Ballesteros M, Suarez B. Detection confidence driven multi-object tracking to recover reliable tracks from unreliable detections. *Pattern Recognition*. 2023 Mar 1;135:109107. Liu Y, Wang S. A quantitative detection algorithm based on improved faster R-CNN for marine benthos. *Ecological Informatics*. 2021 Mar 1;61:101228.
24. Liu Y, Wang S. A quantitative detection algorithm based on improved faster R-CNN for marine benthos. *Ecological Informatics*. 2021 Mar 1;61:101228.
25. Ma C, Li X, Li Y, Tian X, Wang Y, Kim H, Serikawa S. Visual information processing for deep-sea visual monitoring system. *Cognitive Robotics*. 2021 Jan 1;1:3-11.
26. Ren, He, K. Girshick, R., Sun, J., 2015. Faster R-CNN : Towards real-time object detection with region proposal network. *Adv. Neural Inf. Proces. syst.* 28.
27. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II* 4. Springer, pp. 21–37.
28. Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: delving into high quality object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154–6162.
29. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125.
30. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J., 2020. Deformable detr: deformable transformers for end-to-end object detection. *arXiv preprint. arXiv:2010.04159*.
31. Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, realtime object detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 779–788.
32. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., et al., 2022. Yolov6: a single-stage object detection framework for industrial applications. *arXiv preprint. arXiv:2209.02976*.
33. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464–7475.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

