U-YOLOv3: A Model Focused on Underwater Object Detection

Pratima Sarkar[1], Sourav De[2], Sandeep Gurung[1]
[1]Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Majhitar, Sikkim Manipal University, Gangtok, India
[2]Department of Computer Science and Engineering, Cooch Behar Government Engineering College, Vill-Harinchowrah, Cooch Behar, 736170, West Bengal, India
E-mail: psmoon2@gmail.com, dr.sourav.de79@gmail.com, sandeep.gu@smit.smu.edu.in

Keywords: Image enhancement, underwater object detection, MIRNet, YOLOv3

Received: July 11, 2024

Underwater image enhancement and object detection has great potential for studying underwater environments. It has been utilized in various domains, including image-based underwater monitoring and Autonomous Underwater Vehicle (AUV)-driven applications such as underwater terrain surveying. It has been observed that underwater images are not clear due to several factors such as low light, the presence of small particles, different levels of refraction of light, etc. Extracting high-quality features from these images to detect objects is a significant challenging task. To mitigate this challenge, MIRNet and the modified version of YOLOv3 namely Underwater-YOLOv3 (U-YOLOv3) is proposed. The MIRNet is a deep learning-based technology for enhancing underwater images. while using YOLOv3 for underwater object detection it lacks in detection of very small objects and huge-size objects. To address this problem proper anchor box size, quality feature aggregation technique, and during object classification image resizing is required. The proposed U-YOLOv3 has three unique features that help to work with the above specified issue like accurate anchor box determination using the K-means++ clustering algorithm, introduced Spatial Pyramid Pooling (SPP) layer during feature extraction which helps in feature aggregation, and added downsampling and upsampling to improve the detection rate of very large and very small size objects. The size of the anchor box is crucial in detecting objects of different sizes, SPP helps in aggregation of features, while down and upsampling changes sizes of objects during object detection. Precision, recall, F1-score and mAP are used as assessment metrics to assess proposed work. The proposed work compared with SSD, Tiny-YOLO, YOLOv2, YOLOv3, YOLOv4, YOLOv5, KPE-YOLOv5, YOLOv7, YOLOv8 and YOLOv9 single stage object detectors. The experiment on the Brackish and Trash ICRA19 datasets shows that our proposed method enhances the mean average precision for both datasets by 10% and 9%, respectively, compared to the original YOLOv3 and other existing work. This enhancement demonstrates that our proposed model is more appropriate for identifying submerged items and is also capable of recognizing closely clustered, small, very large objects on the ocean floor.

Povzetek: Razvit je model U-YOLOv3, ki izboljšuje zaznavo podvodnih objektov z uporabo MIRNet za izboljšanje slike, K-means++ za optimalno izbiro sidrnih okvirjev in SPP za združevanje značilnosti.

1 Introduction

The importance of underwater biology is increasing due to the expansion of marine ecology and aquaculture. Current research is increasingly focused on creating reliable algorithms for underwater scenes and subsequently identifying underwater species [1]. Underwater exploration is carried out mainly via Remotely Operated Vehicle (ROV) [2] and Autonomous Underwater Vehicle (AUV) [3] as underwater environment is dangerous for humans. During underwater object detection, image enhancement is a crucial step due to the challenge of unclear images [4].

Traditional image processing techniques for enhancing underwater images include color correction algorithms and contrast enhancement algorithms. The white balance method [5], the gray world hypothesis [6], and the gray edge hypothesis [7] are typical color correction methods, and contrast enhancement algorithms include histogram equalization [8] and restricted contrast histogram equalization [9]. The results produced by these technologies are inadequate for underwater vision when compared with conventional image processing techniques. Contrast enhancement is one of the most popular techniques for underwater image enhancement. The contrast enhancement technique is classified into local enhancement and global enhancement techniques. Histogram Specification (HS) [10] is a global approach that works on the whole image. The fuzzy and image partition-based approach is also used for image enhancement [11]. Ahmad Shahrizan et al. used local and global contrast to improve underwater images. In this paper, dual-intensity images are produced by global contrast correction and subsequently merged to yield contrast-

enhanced final images. These images are then processed locally to improve details [12]. Generative Adversarial Networks (GANs) and Convolution Neural Networks (CNNs) have recently shown impressive performance in a range of image-to-image translation tasks, such as super-resolution, dehazing, and image denoising. Li et al. [13] presented the design of the WaterNet gated fusion network, which uses images produced by three enhancement techniques to help the network identify the most important aspects of the input image. FSpiral-GAN, [14] which may significantly speed up the processing of large-size images while preserving the excellent quality of the improved images. The model comprises N discriminators and one generator and is built on a generative adversarial architecture. The author used an encoder and decoder structure with equal upsampling and downsampling blocks to create a lightweight generator structure that would increase model efficiency and retain high-quality generated images. UICE-MIRNet [15] is an approach that improves the detection of underwater objects by improving the quality of images. In this work, improve the colorfulness of images which enables quality feature extraction.

The Conventional object detectors are used to extract features using artificial feature extractors, and then combine those features with classifiers to produce the desired detection outcomes. Many moving object detection techniques are successfully proposed by various researchers in last few decades [16]. Recent advancements have been made progress in the deep learning-based object detection system like, deep convolution neural network (DCNN) which is used for feature extraction. It is capable of extracting a large amount of image details independently by learning features at various layers. Object detectors are mainly categorised into two-stage detectors and single-stage detectors. Region Convolutional Neural Network (RCNN) [17], Fast R-CNN [18], Faster R-CNN [19], etc are commonly known two-stage detectors. Using the sliding window technique or the Regional Proposal Network (RPN) [20], the two-stage detectors algorithm first creates a number of region proposals. Based on these proposals, it then classifies and locates objects. Without generating further region suggestions, the single-stage detectors perform object categorization and localization and work as an end-to-end network for object detection. When compared to a two-stage detector, the second scanning of images streamlines the detection procedure and boosts detection effectiveness. Recently, some anchor-free detection techniques have been proposed by researchers in addition to the aforementioned anchor-based detectors.

Different single-stage detectors are Single Shot multibox Detectors (SSD) [21] and YOLO [22] series. The different YOLO versions are YOLOv2, YOLOv3, YOLOv4, YOLOv5 and YOLOv7 etc. YOLOv2 is popularly known as YOLO9000 [22] which is a real-time object detection technique capable of detecting 9000 categories of object but not achieved good mean Average Precision(mAP). Joseph Redmon et. al. [23] modified SSD to improve detection speed and named the model as YOLOv3. The YOLOv3 is three times faster than SSD and the accuracy is same as SSD. In the year 2020 Alexey Bochkovskiy et. al. proposed a model for real-time object detection i.e. YOLOv4 [24] which is popular due to proper balancing of speed and accuracy. YOLOv4 introduced some new approches like Weighted-Residual-Connections (WRC), Cross mini-Batch Normalization (CmBN), Cross-Stage-Partial-connections (CSP), Mish-activation and Self-adversarial-training (SAT) to achieve better accuracy. YOLOv5 [25] and YOLOv7 [26] is proposed to enable a real-time object identification even on devices with limited resources. It keeps good precision while achieving high frame rates. All of these methods perform incredibly well on land images, but the indistinct images provide a very difficult problem when it applied on underwater images. Feature extraction from underwater images is challenging task due blur images so that, Yong Liu et. al. [27] used ResNet101 [28] to enhance image quality. The authors also added a compact and lightweight Selective Kernel unit, which helps the CNN extract features more effectively without adding more layers. The algorithm also incorporates feature fusion for better sharing of information. This article [29] suggests a low computation deep underwater object detection network. Also presenting a deep model for concurrently learning object detection and color conversion for underwater images are crucial. The purpose of the image color conversion module is to convert color images into their equivalent gray scale images in order to improve object detection performance while reducing computational complexity. This solves the issue of underwater color absorption. YOLO-Fish, [30] a fish detection model powered by deep learning. Two models, namely YOLO-Fish-1 and YOLO-Fish-2, have been proposed in this work by addressing the problem of upsampling step sizes to lessen the misdetection of small fish, YOLO-Fish-1 improves YOLOv3 by incorporating Spatial Pyramid Pooling into the initial model, YOLO-Fish-2 enhances the model even more and adds the capacity to identify fish appearance in those dynamic environments. Another model performs detection of underwater organism from underwater images called U-YOLOv7 [31]. Initially in U-YOLOv7, a network is built that combines an effective squeeze-excitation module with CrossConv. This network improves the network's feature fusion and increases the extraction of channel information while decreasing parameters. Secondly, prior to feature fusion, additional semantic information about underwater images are obtained using a lightweight Content-Aware ReAssembly of FEatures (CARAFE) operator. A similar kind of work is used for coral reef detection by Jingyao Wang et. al. [25]. This work modified YOLOv5 model incorporates multiple stages of CSP and channel attention mechanism. KPE-YOLOv5 [32] is an approach which enables small object detection. Shenming Qu et. al. [33] proposed a model which modifies the original upsampling method used in YOLOv8 and proposed a new upsampling technique, CARAFE which improves in small object detection. To lessen the effect of underwater image quality problems on the detection job, the Qiming Li

et. al. [34] proposed an image improvement module. They also suggested GEBlock with YOLOv8, an attention-based fusion module that suppresses noise from lower-level feature layers and gathers long-range contextual information.

But when comes to underwater object detection, it becomes very challenging as the underwater image background contains different organisms, including plankton, starfish, and schools of fish etc. As background and objects are sometimes having similar colors, the detection of various sizes of objects is a challenging task. Table 1 presents some works on underwater object detection using YOLOv3. It is also found that YOLOv3 is a robust architecture to detect underwater objects. Still, the work mentioned above in Table 1 lacks in small objects, dense objects from underwater images. The contribution of the work includes:

a. The underwater images are not clear because of the presence of multiple particles and insufficient light. To improve the visibility of images, MIRNet image enhancement is used to improve the quality of images.

b. The selection of anchor box sizes is a challenging issue during object detection. Usually K-means clustering is applied for selection of anchor box sizes but it selects centroid randomly, which may lead to sub-optimal solution. By addressing the problem of random initialization, K-means++ improves upon the original K-means method. The initial centroids in K-means++ are chosen to guarantee that they are evenly dispersed throughout the dataset.

c. YOLOv3 may have difficulty accurately identifying small items, particularly when they are collected. To address this problem, extracted features need to be aggregated. So, the feature aggregation layer, namely the Spatial Pyramid Pooling (SPP) layer, added to the existing work after feature extraction.

d. To improve detection accuracy, the proposed work modified the classification layer of YOLOv3. Detection of very large objects improves by incorporating downsampling in large object prediction layer where as upsampling added in small object detection layer to improve small object detection accuracy.

The organization of the work is as follows: Section 2 explains related work, while Section 3 provides an explanation of the proposed methodology. Sections 4 and 5 contain details of the description of the database and the assessment method, respectively. Sections 5 and 7 provide an analysis of the results and its discussions, respectively. Ultimately, section 8 provide conclusion and explores potential future paths.

2 Related works

In this section, an overview of the underwater image enhancement and object detection process is described. Here, emphasis is placed on a brief architecture of YOLOv3 and MIRNet. The sub-section 2.1 is dedicated for YOLOv3 and a small description of MIRNet has been established in subsection 2.2, respectively.

2.1 YOLOv3 architecture

Joshep Redmon and Ali Farhadi created YOLOv3 [23] in 2018 proposed YOLOv3 is the third version of the YOLO series. The YOLOv3 provides an end-to-end object detection framework. This algorithm predicts bounding boxes with some probability value, depending upon these probability value objects classified. The following are the main three steps to predict objects:

1. Feature extraction
2. Bounding-box prediction
3. Class prediction

2.1.1 Feature extraction

Feature extraction is most important phase of object detection. After feature extraction based on these feature performed bounding box prediction and classification of objects. YOLOv2 used Darknet-19 for feature extraction whereas modified version of Darknet-19 is Darknet-53 which is used by YOLOv3. It has been established that this network is more effective and potent than ResNet-101 or ResNet-152. But Darknet-53 has several flaws like lacks in detection of small objects and feature aggregation technique. So proposed work modified selection of anchor box size which covers all size objects. Introduced new feature aggregation layer into original Darknet53.

2.1.2 Bounding box prediction

The bounding box sizes in the YOLOv3 network are obtained using the K-means clustering method. There are nine clusters in all, and three fixed scales that are based on ground truths as bounding box priors are chosen at random. The network selects 4 co-ordinates for each anchor box these are a_x, a_y, a_w, a_h. Top most corner of anchor box is (t_x, t_y) , width and height is p_w, p_h respectively then predicted bounding box co-ordinates are:

$$b_x = f(a_x) + t_x \tag{1}$$

$$b_y = f(a_y) + t_y \tag{2}$$

$$b_w = p_w \times e^{a_w} \tag{3}$$

$$b_h = p_h \times e^{a_h} \tag{4}$$

Total 9 bounding box are predicted, then for each of them calculated sum square loss. YOLOv3 used logistic regression for prediction of bounding boxes. Depending on overlapping of ground truth and prediction, its gives score 1.

Table 1: Some of the previous work used for underwater object detection based on YOLO

Authors	Description of methodology	Dataset	mAP	Limitations
K. Cai et al. [35]	This study integrates YOLOv3 with MobileNetv1 for the purpose of detecting fish in a live breeding farm environment.	self created Real fish farm Dataset	78.3%	The work is not tested on real marine images where detection of objects are more difficult due to image quality.
H. Yang et al. [36]	YOLOv3 and Faster R-CNN is compared for detecting and recognizing targets in underwater environment and it is found YOLOv3 is suitable for real time object detection.	URPC dataset	76.1%	The work is comparative study between two networks.
X. Li et al. [37]	In this work the original anchor boxes are adjusted by K-means clustering algorithm to improve accuracy of the detection model.	Own data set of garbage	84.1%	Authors are considered the surface images and suface images does not face the issue related to image clarity.
P. Athira et al. [38]	This paper introduces a methodology that employs the YOLOv3 architecture, coupled with the Darknet framework and deep learning.	Fish4 Knowledge	96.0%	Authors have used YOLOv3 directly no modifications are made on existing network.
Shenming Qu et al. [33]	In order to maximize spatial feature extraction and greatly reduce redundant computation and storage needs, this work devised a Lightweight Efficient Partial Convolution (LEPC) module that processes input channels selectively.	URPC2021	84.2%	The heightened feature extraction method causes YOLOv8-LA to confront constraints, because it slows down the system.
P. Sarkar et al. [39]	This paper performs object detection using YOLOv3 and YOLOv4 on underwater dataset and provide details of challenges present during underwater object detection.	Roboflow object detection dataset	40.5%	Lacks in detection of dense objects.
Kun Liu et al. [40]	The TC-YOLO network is modification of YOLOv5 and compared with YOLOv3. To improve feature extraction, the new network's neck and backbone, respectively, adopted transformer self-attention and coordinate attention.	RUIE2020	45.6%	The work does not incorporates detection of various size of objects.

2.1.3 Classification of objects

Each anchor box predicts the class for an object. YOLOv3 does not use the Softmax function for the classification of objects. The softmax function classifies a particular object in one category but it's not always true. YOLOv3 uses a logistic regression classifier for the prediction of classes. Each predicted anchor box must contain the probability value of belongs to a particular class. YOLOv3 uses 106 layers for the detection of objects from an image. Out of all the layers 82, 94, and 106 are object detection layers. YOLO v3 predicts 9 anchor boxes for each object. Three anchor boxes are predicted for each scale (13x13, 26x26, 52x52 strides). Finally calculates Intersection over Union (IoU) to detect the objects.

2.2 MIRNet architecture

Learning enriched image enhancement technique MIRNet [41] is used for image enhancement. MIRNet is convolution network-based architecture and works very well with low resolution images. Underwater images are usually low-light images so that the proposed work used MIRNet for image enhancement. The special feature of the network is it preserves contextual of images. In MIRNet used three Recursive Residual Groups (RRG). Each RRG consists of

multiple Multi-scale Residual Blocks (MRB). The MRB has two major components Selective kernel Feature Fusion (SKFF), Dual Attention Unit (DAU). Input image I is given as input to Convolution layer which generates feature map FM. The RRG blocks consists of three MRB for our proposed work. The FM passed as parameter to the MRB block. MRB extracts high contextual information from low resolution areas. On FM applied down sampling and generates two more feature matrix FM1, FM2. This feature matrix passes through SKFF and DAU multiple times to consolidate high resolution features from low resolution areas as shown in Figure 1. Each SKFF consists of two operations fuse and select. Fuse aggregates the features from FM, FM1, FM2 i.e. $F = FM + FM1 + FM2$. Select uses softmax function to recalibrate multi-scale feature map F. The DAU share information along the spatial and channel domain. DAU has two components that work in parallel manner. These two are used to suppress less needed features and pass more informative features. Finally, MRB returns the image to RRG block. This process will be repeated three times to generate enhanced image.

3 Proposed methodology

This section presents an approach for detecting underwater objects. The AUV collected images on marine environment are unclear due to presence of small particles and lack of lights in underwater images. So, one of the main components of the proposed work is image enhancement. In this work, first, the images were enhanced by using MIRNet proposed in 2020 by Zamir et al. [41] It is also found that small and dense object detection is one of the major issues during object detection. To address these problems, the proposed work modified YOLOv3 and named as Underwater-YOLOv3(U-YOLOv3). Modifications are made based on the anchor box selection technique and feature extraction model. In enhanced images, U-YOLOv3 was applied to perform real-time object detection. The work consists of two main stages: image enhancement and item detection utilizing the enhanced images.

Enhanced images are applied to U-YOLOv3 for object detection. It is a single-stage object detector and a modified version of YOLO9000. The U-YOLOv3 used the following steps:

1. Bounding-box size prediction by K-means++
2. Modified Darknet53
3. Modified Class Prediction

3.1 Bounding box size prediction by K-means++

In this proposed work, K-means++ clustering is applied to predict the sizes of anchor boxes. K-means++ clustering is a modified version of K-means clustering. One of the drawbacks of K-means algorithm is sensitive initialization of centroid. If centroid is very far it may not include any point into the cluster. K-means clustering has another issue that always starts with arbitrary selection of centroid, which may not be the best selection of centroid. The main reason behind not selecting Gaussian Mixture Models(GMM) is that, it is difficult to work with categorical features. So we have selected K-means++ for bounding box size prediction. K-means++ changed the technique of centroid selection and improves the quality of centroid selection. K-means++ algorithm as follows:

Algorithm 1 K-means++ Algorithm

Step1: Select an initial centre C1 from U.
Step2: Identify next centre Ci where $Ci = X_{new} \in U$ with probability $(D(X_{new}))/(\sum_{x \in U} D(x))$ where D(x) is shortest distance from a data x from centre Ci .
Step3: Next choose K centre using step2.
Step4: After selecting initial centroid, sum square distance to be calculated from each centroid.
Step5: Those points are close to a centroid are placed in same cluster.
Step6: Then recalculate centroid for each cluster by averaging all of the data points associated with each cluster. If change occurs in centroid then in continue with step4 otherwise stop.

The main objective of this clustering technique is to minimize the sum-squared distance between the centroid of the cluster and all points. The objective function is shown in equation 5.

$$dist = \sum_{i=1}^{k} \sum_{j=1}^{n} \mid x_i^{(j)} - c_j \mid^2 \tag{5}$$

Where, j is the cluster number, n is the number of cluster $x_i^{(j)}$ is point belongs to j^{th} cluster. c_j is the centroid of the j^{th} cluster.

3.2 Modified Darknet53

Darknet53 used as backbone of YOLOv3 for feature extraction. The input image size for all the images is 416×416 pixels and is used for feature extraction. Darknet53 is a modified version of Darknet19 which is used in YOLO9000. The modified Darknet-53 is shown in Figure 2 in which the main components are Convolution, batch normalization, and Leaky ReLU together known as CBL with it modified Darknet-53 has residual blocks and the Spatial Pyramid Pooling (SPP) layer. Convolution layers are using either Leaky activation functions as it can overcome vanishing gradient problem. The network applies different sizes of filter, i.e. 32, 64, 128, 256, 1024 during object detection. The network also uses different sizes of strides, which usually are 1×1, 3×3. A new layer SPP is added at the end of Darknet53 for feature aggregation purpose. It helps to get over the network's fixed-size restriction of images and allow processing of images of different sizes. By using this pooling technique, the size of the image can be

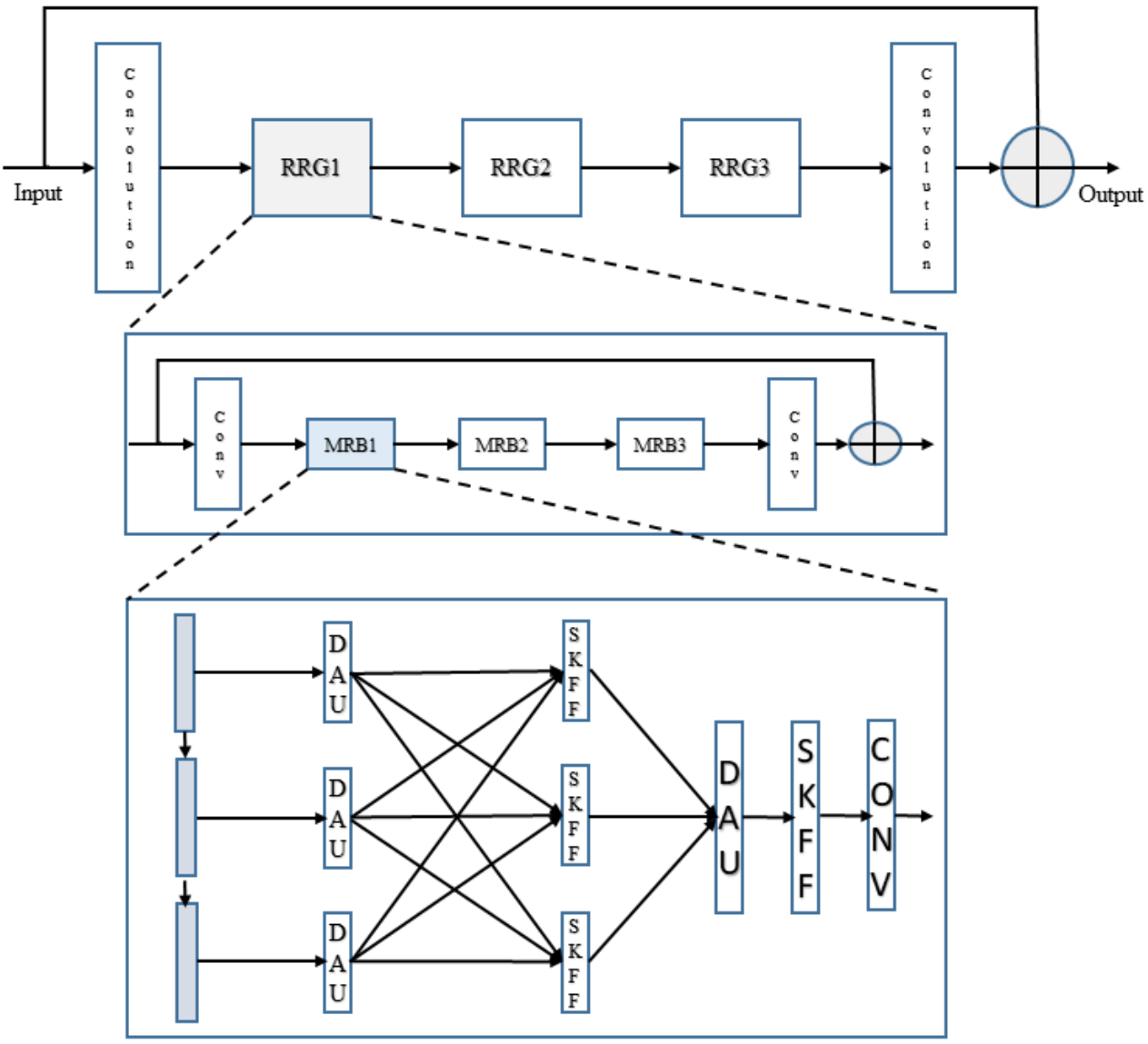


Figure 1: MIRNet architecture

reduced without losing important features. In the proposed work, SPP has three maxpool layers, two CBL (convolution, batch normalization, LeakyRelu) as shown in Figure 3. This layer receives features from the convolution layer. SPP layer generates a fixed length representation from different sizes of features. The proposed work added SPP at the end of Darknet53 to aggregate the generated features generated from Draknet53 as depicted in Figure 2.

3.3 Modified class prediction

The classes of a bounding box must be predicted after the bounding box prediction process to provide the classification information of the object. There are three prediction layers available for object detection. 13×13 is used for large object detection, 26×26 is used for medium-sized object prediction, and 52x52 is used for small-sized object detection. If a very large object exists in an image, it is very difficult to accommodate in predicted bounding box size similarly for very small objects until unless the size of the of object increases. To address this issue, we modified the existing YOLOv3 classification of objects technique. In layer 13×13, we have used downsampling so that very large objects can also be accommodated again in the 52x52 layer, adding upsampling four times to increase the size of very small objects, as shown in Figure 2. In this way, the proposed work helps to improve small and large objects.

During object prediction, a multilevel approach is introduced. A logistic regression classifier is used to detect the probability value or the objectness score for each of the bounding boxes. An approach to binary classification, called logistic regression, determines the likelihood

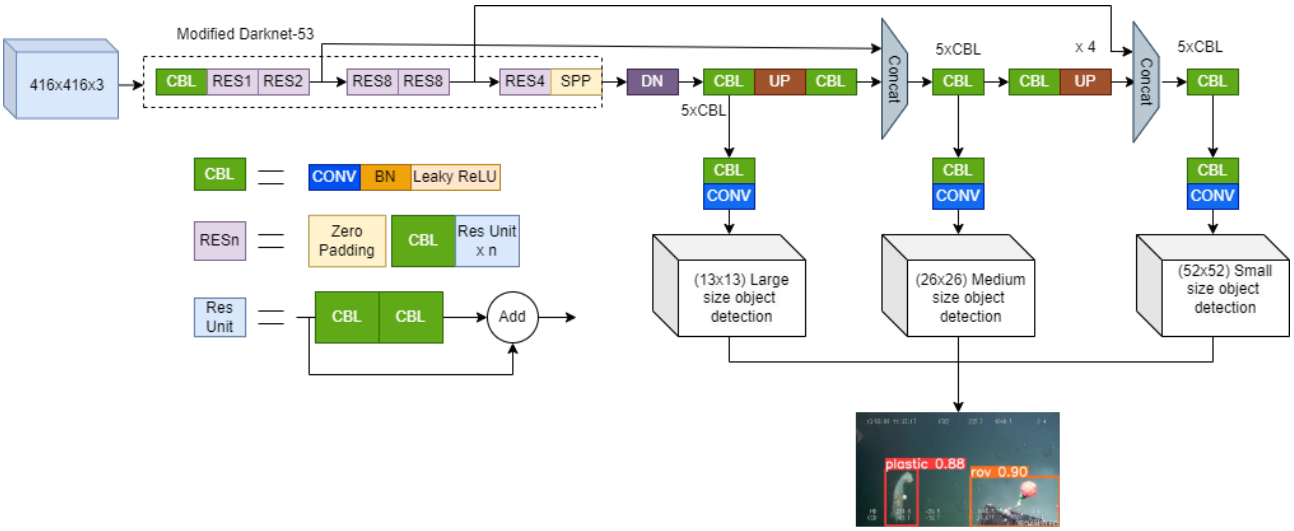


Figure 2: Underwater-YOLOv3 architecture

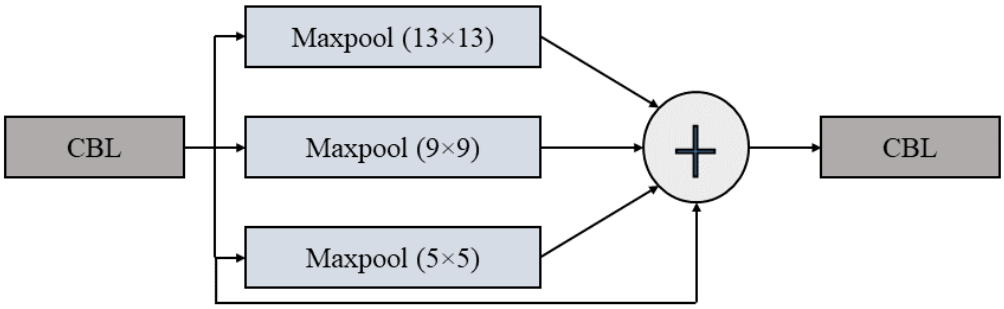


Figure 3: Spatial pyramid pooling layer

that an input will fall into a particular class. For class prediction in YOLOv3, independent logistic regression classifiers are employed rather than softmax classifiers, which enable multi-class classification. YOLOv3 is capable of handling complicated datasets with overlapping labels due to the use of logistic classifiers. Formula for logistic regression:

$$\log[p(X)/(1-p(X))] = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_p X_p \quad (6)$$

where, X_j is j^{th} predictor variable and α_j coefficient for j^{th} variable.

For class predictions, YOLOv3 employed binary cross-entropy (BCE) loss during training. It is a binary classifier that is used to measure the difference between ground truth and predicted probabilities. BEC is as follows:

$$BEC = -[x * log(p) + (1-x) * log(1-p)] \quad (7)$$

where, x is the ground truth value and p is the predicted probability value so that the ground truth is 1. For each of the boxes, some probability value is used for the selection of the best anchor box. However, selecting the maximum probability value for an object might result in incorrect detection in the case of multiple objects of the same type. Intersection over Union (IoU) is used to choose the perfect anchor box.

$$IOU = Intersectarea/UnionArea \quad (8)$$

4 Dataset description

Two open-source data sets, the Brackish Underwater Dataset [42] and the Trash ICRA19 Dataset [43] are used to perform experiments. The motivation behind selecting these two dataset is covering all different size and densely packed objects. Also, the dataset is consists of all real images not artificial images.

The Brackish dataset contains 14,674 images. This is the first publicly available European underwater image dataset with bounding-box annotations of fish, crabs, small fish, etc. The second dataset consists of trash images collected by Marine Earth Science and Technology (JAMSTEC). The underwater videos are collected between the year 2000 to

2017. This dataset contains 5,720 images. Majorly this dataset is classified into plastic (like a bottle, plastic bags, etc.), Remotely Operated Vehicle (ROV), bio (plants, fish other leaving organism).

5 Assessment method

The proposed work has two major components: image enhancement and underwater object detection. The quality of the enhanced images is evaluated using two parameters Underwater Image Quality Measure (UIQM) [44] and Underwater Color Image Quality Evaluation (UCIQE) [45]. Precision, recall, mean Average Precision(mAP) are used for measuring the quality of object detection.

5.1 Underwater image quality measure (UIQM)

Underwater Image Quality Measure (UIQM) [44] is one of the most basic performance metrics that does not need any reference images. The UIQM has three underwater image component measures: the underwater image colorfulness measure (UICM), the underwater image sharpness measure (UISM), and the underwater image contrast measure (UIConM).

$$UIQM = c1*UICM+c2*UISM+c3*UIConM \quad (9)$$

where, $c1 = 0.0282, c2 = 0.2953, c3 = 3.5753$. A bigger value of UIQM represents better good quality of the enhanced image and vice-versa.

5.2 Underwater color image quality evaluation (UCIQE)

Another most popular underwater image enhancement metric is UCIQE [45]. It is based on contrast, standard deviation, and saturation of the image. The formula for calculating UCIQE is as follows:

$$UCIQE = c1 * Sd + c2 * Col + c3 * Sat \quad (10)$$

where Sd is the standard deviation, Col is the contrast of luminance and Sat is the average of saturation, and $c1, c2, c3$ are weighted coefficients.

5.3 Object detection assessment methods

Object detection assessment is done using precision, recall, F1-score, Intersection over Union(IoU) and mAP. Precision, recall, IOU, F1-score are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$IOU = \frac{AreaofIntersection}{AreaofUnion} \quad (13)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

Category-wise Average Precision is computed as equation (15)

$$AP(C_i) = (1/n) \times (\sum_{i=1}^{n} P_i) \quad (15)$$

where P_i is i^{th} the image of the C_i category and n is number of iterations.

Mean Average Precision is calculated as equation (16)

$$mAP = (1/N) \times \sum_{i=1}^{n} AP(C_i) \quad (16)$$

where N is number of classes.

6 Experimental results

This section presents the results of the image enhancement and object detection experiments in various scenarios: (1) quantitative analysis of the MIRNet image enhancement technique; (2) qualitative analysis of the MIRNet image enhancement technique; (3) results of object detection using the proposed work and detailed analysis of it.

The results obtained by MIRNet image enhancement and object detection are reported in this section. The experiment is carried out on the Brackish and Trash ICRA19 datasets. To evaluate the quality of enhanced images used average UIQM and UCIQE performance metrics and results are shown in Table 2.

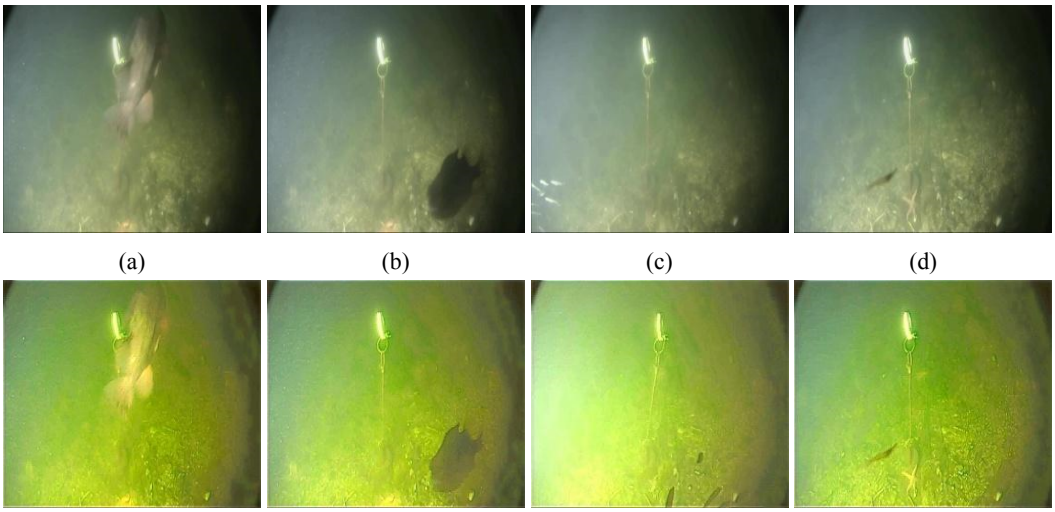
Table 2 tabulated data regarding quantitative analysis image enhancement by MIRNet for both datasets. From Table 2 we can conclude that after image enhancement, UIQM and UCIQE improved by $3 - 5\%$. MIRNet generates high-resolution images and preserves contextual information. This unique property of MIRNet helps to achieve better UIQM and UCIQE values.

Figures 4 and 5 show the qualitative analysis of MIRNet in two different diagrams. It is evident from the enhanced images that resolution and visibility have improved in both data sets. Since MIRNet is built on deep learning methods, it is more stable to obtain image enhancement methods based on it.

The experiments for object detection models are tested with CUDA version 11010, cuDNN 7.6.5, and OpenCV 3.2.0 on a GPU runtime in the Google Colab environment. To make the experiment unbiased, all the models are evaluated in the same environment and for the same number of iterations. Details of the hyper parameters are shown in Table 3. Experiments are conducted through training of dataset and pre-trained weights are not used. The models in the existing work used to compare with U-YOLOv3 trained

Table 2: Average UIQM and UCIQE of the original and enhanced dataset

Parameter	Brackish Dataset		Trash ICRA19	
	Original Image	Enhanced Image	Original Image	Enhanced Image
Average UIQM	0.19	0.23	0.21	0.34
Average UCIQE	4.86	4.91	3.86	4.12



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: (a-d) original images, (e-h) enhanced images with MIRNet for Brackish underwater dataset

Table 3: Hyper-parameter for performing experiments

Parameters	Specification
Training-Test split ratio	7:3
Batches	64
Steps	9600, 10800
Final layer filters	32
Classes	6
Iterations	8000
Learning Rate	0.001
Momentum	0.9
Channels	3
Classes	6

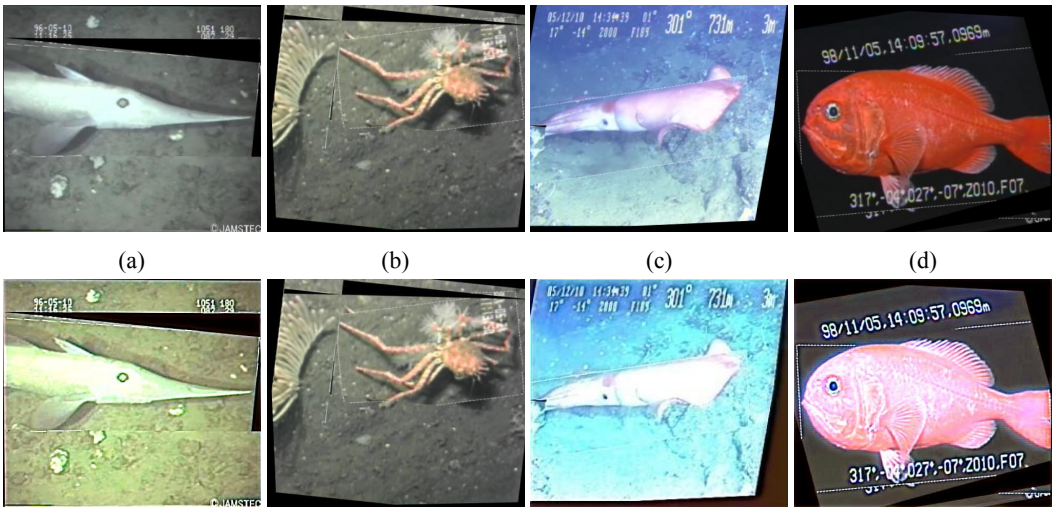
using the same dataset and parameters only the number of epochs are different as we wait until the mAP are steady.

Underwater images suffer from blurring effects, low contrast, and grayed-out colors because of the absorption and scattering effects under water. So we have enhanced the images before object detection to increase the visibility of the image. We can observe from Table 4 that the proposed U-YOLOv3 gives 20% precision, 8% recall, 14% F1-score and 7% Average IoU better performance than YOLOv3 for the Brackish data set. Reported data from Table 5 show that 2% precision, 6% recall, 4% F1-score, and 14% Average IoU improved than YOLOv3 for the proposed work while using Trash ICRA19 dataset.

Tables 6 and 7 demonstrate the statistical analysis of the proposed work. We have performed a paired t-test between YOLOv3 and U-YOLOv3 to validate the performance of the proposed works. The parameters used to perform statistical analysis are mean, Standard Deviation (SD), Standard Error of Measurement (SEM) and p-value. We have assumed YOLOv3 as the null hypothesis, U-YOLOv3 as an alternative hypothesis, and the significance level 5%. From Table 6 and 7, it is clear that all p-values are less than 5% hence alternate hypothesis is accepted and null hypothesis is rejected.

Tables 8 and 9 contain class-wise average precision for Brackish dataset and Trash ICRA19 dataset respectively. In most of the cases, the average precision increased, but only in one case the average precision value decreased after using the enhanced dataset. This is because during enhancement, some images lose their contextual information.

Training for this study is conducted on a GPU platform using an 8000 iteration setting. The indicators of the experimental ablation network are shown in Table 10. Precision increased by 10%, recall increased by 6%, F1-score increased by 9%, and mAP increased by 4% after employing K-means ++ and image enhancement for the Brackish dataset. After inclusion of SPP for feature aggregation, precision improved by 4%, recall improved by 1%, F-1 score improved by 2%, mAP improved by 3%. Then finally, after modification of the classification layer, we got an improvement of 4% in precision, 1% in recall, 3% in F1-score and



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5: (a-d) original images, (e-h) enhanced images with MIRNet for Trash ICRA19 dataset

Table 4: Performance comparison of YOLOv3 and proposed U-YOLOv3 for Brackish underwater dataset

Iterations	YOLOv3				U-YOLOv3			
	Precision	Recall	F1-Score	Avg IoU	Precision	Recall	F1-score	Avg IoU
2000	0.32	0.52	0.42	0.32	0.44	0.56	0.49	0.49
4000	0.62	0.74	0.68	0.51	0.83	0.77	0.80	0.71
6000	0.67	0.78	0.72	0.60	0.82	0.89	0.85	0.71
8000	0.68	0.79	0.73	0.67	**0.88**	**0.87**	**0.87**	**0.74**

2% in mAP.

We have performed similar experiments on the trash ICRA19 dataset shown in 11. After using K-means++ for anchor box size detection and using enhanced images in the base model, the precision improved by 6%, recall improved by 1%, the F-1 score improved by 4% , mAP improved by 8%. After inclusion of SPP for the aggregation of features, precision improved by 1%, recall improved by 2%, F-1 score improved by 1% , mAP improved by 0.4%. Lastly, after incorporation of modifications into the classification layer 1% improvement in precision, recall, F1 score, and 1. 6% improvement in mAP.

The article highlights the difficulties in identifying small objects in highly opaque and blurry underwater landscapes, such as fish, jellyfish, shrimps, crabs, and jellyfish. It is very clear from Figure 6 that the proposed work can also detect small and dense objects from underwater images with higher recognition precision. U-YOLOv3 can identify the two fish that are significantly dense as shown in Figure 6 (d) and (h) with good precision. It is also challenging for the human eye to discriminate between different objects in such images with complicated backgrounds. The U-YOLOv3 suggested in this study performs exceptionally well, correctly recognizes the items in the images, particularly the small object and large objects, and is more equipped to handle the difficulties associated with underwater detection. It is also shown in Figure 6 that the sizes of the anchor box are perfect and incorporate entire objects.

7 Discussions

In this section the performance of proposed work is assessed, with respect respect to popular single-stage object detectors. U-YOLOv3 compared with other current state-of-the-art and the results are shown in Tables 12 and 13. Ten different object detection networks SSD, Tiny-YOLO, YOLOv2, YOLOv3, YOLOv4, YOLOv5, KPE-YOLOv5, YOLOv7, YOLOv8, and YOLOv9, were compared with the proposed U-YOLOv3. The comparison is made for the trash ICRA19 and the Brackish dataset. The assessment carried out based on the mAP achieved by different models. Each network was derived using an image size of 416 × 416 and it is found that the proposed work achieved 2-10% better mAP than the existing work. The underwater images contains small particles because this light gets reflected multiple times as a result images captured in the underwater environment are not clear. The extraction of quality features from underwater images is very difficult, so we have used MIRNet to improve the image resolution. Another important issue during object detection is the selection of the bounding box instead of K-means the proposed work used K-means++ during object detection and gives a better result than original YOLOv3 as shown in Table 10 and 11.

Table 5: Performance comparison of YOLOv3 and proposed U-YOLOv3 for trash ICRA19 dataset

Iterations	YOLOv3				U-YOLOv3			
	Precision	Recall	F1-Score	Avg IoU	Precision	Recall	F1-score	Avg IoU
2000	0.32	0.06	0.10	0.31	0.37	0.19	0.25	0.43
4000	0.46	0.56	0.43	0.44	0.52	0.60	0.53	0.50
6000	0.64	0.61	0.62	0.47	0.71	0.73	0.71	0.59
8000	0.66	0.75	0.70	0.57	**0.74**	**0.79**	**0.76**	**0.69**

Table 6: Paired t-test between YOLOv3 and U-YOLOv3 for Brackish dataset

	YOLOv3			U-YOLOv3			
Parameter	Mean	SD	SEM	Mean	SD	SEM	p-value
Precision	0.5575	0.1665	0.0802	0.7425	0.2034	0.1017	**0.0074**
Recall	0.7075	0.1269	0.0634	0.7725	0.1511	0.0755	**0.0390**
F1-Score	0.6375	0.1466	0.0733	0.7525	0.1775	0.0887	**0.0051**
AverageIoU	0.5250	0.1525	0.0758	0.6625	0.1159	0.0579	**0.0182**

Feature reduction without losing significant information of object detection is necessary, so we have introduced SPP in Darknet-53. In addition, we have incorporated down sampling during large objects detection and up sampling during small objects detection, which improves mAP for object detection.

The proposed work selected two different datasets Brackish data set consists mostly dense and small objects and Trash ICRA19 mostly consists of large, medium size objects. It is found from Table 8 that for all the different classes U-YOLOv3 performed better than YOLOv3 with small objects. From Table 12 it is also clear that U-YOLOv3 performed better than other existing techniques for the Brackish dataset with respect to precision recall, IOU, F1-score and mAP. The ICRA19 trash data set consists of 5,700 images, and it is clear from Table 13 with a smaller number of training images the proposed work achieved a mAP of more than 4% better than the existing work. In addition, it successfully detects large and medium-size objects. In addition, it is evident from Figure 6 (d) and (h) that the proposed work detects small dense objects with good precision. Similarly, for large and medium-sized objects, they are also perfectly detected as shown in Figure 6 (e), (f) and (g). We have tried to make a generalized model which helps in detection in both the datasets. To make the proposed model generalized, added image enhancement which is a requirement of underwater images due to clarity of images with it tried to achieve detection of all size including dense objects.

The main limitation of the proposed work is that training time is higher than other existing techniques and taining time is almost similar as YOLOv9. U-YOLOv3 works with almost 63.25 million parameters, so it takes a long time to train the images, but the object detection time from an image is 22 ms. In the future, a model can be designed in such a way that it can balance training time with precision.

8 Conclusion

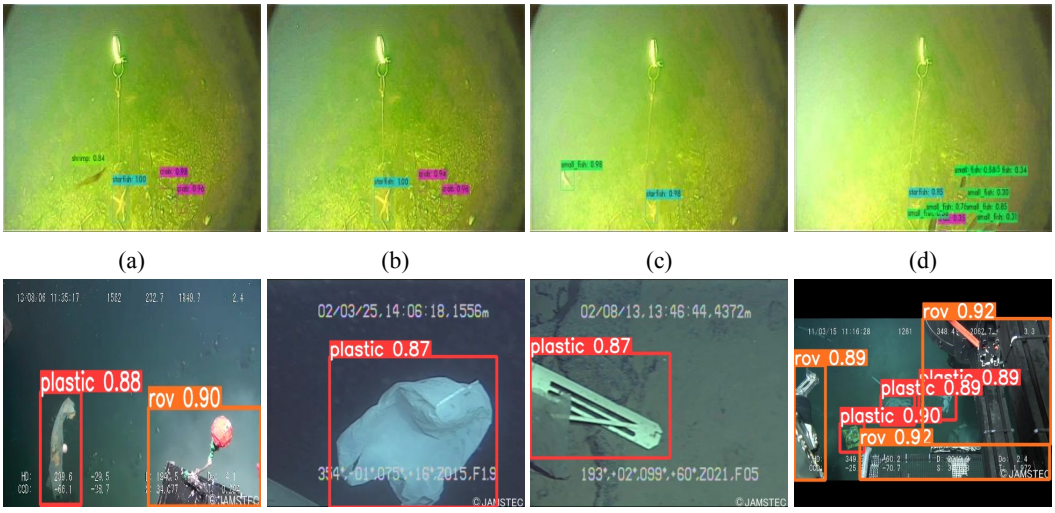
This research proposes an architecture which is a modified version of YOLOv3 and aims to detect underwater images. As underwater images are not clear the work included a pre-processing step i.e. image enhancement before object detection. Image enhancement is based on the deep learning technique MIRNet that helps to improve the resolution of images without losing contextual information, and for object detection the proposed work proposed a modified version of YOLOv3 i.e. U-YOLOv3 which is suitable for underwater object detection. U-YOLOv3 incorporates three modifications on YOLOv3 first one is the proposal of bounding box size by K-means++ clustering, the addition of SPP for feature aggregation, resizing of features during classification of objects. Together, these changes help us to improve mAP by increasing the detection rate of very large, small, and dense objects. The entire experiment is conducted on two datasets, that is, the Brackish and Trash ICRA19 dataset. In both the dataset average UIQM and average UCIQE values increased by 3-10%. These enhanced images are used for training purposes. The work proposed U-YOLOv3 which concentrates on better selection of bounding box size which is capable of detection of different size objects. Also proposed work includes an efficient feature aggregation technique, which enables reducing feature size. U-YOLOv3 achieved a better 1- 10% mAP compared to existing work.

Acknowledgement

Authors are thankful to Marine Analytics using Computer Vision project for the Brackish dataset. The authors also thank Data Repository for the University of Minnesota for the Trash-ICRA19 dataset.

Table 7: Paired t-test between YOLOv3 and U-YOLOv3 for Trash-ICRA19 dataset

	YOLOv3			U-YOLOv3			
Parameter	Mean	SD	SEM	Mean	SD	SEM	p-value
Precision	0.5200	0.1608	0.0804	0.5850	0.1733	0.0867	**0.0021**
Recall	0.4950	0.3009	0.1505	0.5775	0.2702	0.1431	**0.0441**
F1-Score	0.4652	0.2669	0.1334	0.5625	0.2706	0.1153	**0.0128**
AverageIoU	0.4475	0.1072	0.0536	0.5525	0.1127	0.0563	**0.0060**



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 6: (a-d) are object detection results for Brackish dataset and (e-h) are object detection results for Trash ICRA19 dataset by U-YOLOv3

Table 8: Class wise average precision for YOLOv3 and proposed U-YOLOv3 for Brackish dataset

	YOLOv3	U-YOLOv3
Class	Avg Precision	Avg Precision
Crab	0.43	**0.87**
Fish	0.62	**0.64**
Jellyfish	0.77	**0.81**
Shrimp	0.68	**0.72**
Small Fish	**0.76**	0.72
Starfish	0.99	**1.00**

Table 9: Class wise average precision for YOLOv3 and proposed U-YOLOv3 for Trash ICRA19

	YOLOv3	U-YOLOv3
Class	Avg Precision	Avg Precision
Unidentified Object	0.66	**0.80**
Bio	0.71	**0.81**
Plastic	0.73	**0.78**
Rov	0.42	**0.58**
Timestamp	0.82	**0.94**
Wood	0.50	**0.58**

Funding and conflicts of interest

Conflict of Interest: The authors of this paper have no conflicting interests. The authors have not received funding for their research.

Data availability statements

This article uses the following datasets that are accessible to the public:
1. Brackish dataset: `https://public.roboflow.com/object-detection/brackish-underwater`
2. Trash ICRA19 dataset: `https://conservancy.umn.edu/handle/11299/214366`

Table 10: Indicators for the ablation experiments for Brackish Dataset

Module	Precision	Recall	F1-Score	mAP@50:95
A:YOLOv3	0.68	0.79	0.73	71%
B:A+ Enhanced Images +K-means++	0.78	0.85	0.82	75%
C:B+SPP	0.82	0.86	0.84	78%
D:C + Modified classification	0.88	0.87	0.87	80%

Table 11: Indicators for the ablation experiments for Trash ICRA19 Dataset

Module	Precision	Recall	F1-Score	mAP@50:95
A:YOLOv3	0.66	0.75	0.70	64%
B:A+ Enhanced Images +K-means++	0.72	0.76	0.74	72%
C:B+SPP	0.73	0.78	0.75	72%
D:C+ Modified Classification	0.74	0.79	0.76	74%

Table 12: Comparison with other techniques for Brackish dataset

Network	Precision(%)	Recall(%)	IOU(%)	F1-score(%)	mAP@50:95(%)
SSD [21]	52.1	73.2	65.6	60.0	68.39
Tiny-YOLO [46]	42.3	34.4	41.9	42.3	33.64
YOLOv2 [22]	58.6	57.9	65.0	57.6	52.54
YOLOv3 [23]	68.9	79.9	67.8	74.8	72.49
YOLOv4 [24]	82.3	71.1	68.7	76.8	78.23
YOLOv5 [47]	83.3	81.3	78.6	82.1	81.18
KPE-YOLOv5 [32]	48.6	47.4	47.0	46.2	47.47
YOLOv7 [26]	76.9	75.9	75.7	65.2	69.40
YOLOv8 [48]	87.3	86.7	86.6	79.3	73.30
YOLOv9 [49]	91.4	89.6	89.3	80.6	78.50
U-YOLOv3	**88.2**	**87.5**	**78.1**	**87.6**	**80.39**

Table 13: Comparison with other techniques for Trash-ICRA19 dataset

Network	Precision(%)	Recall(%)	IOU(%)	F1-score(%)	mAP@50:95(%)
SSD [21]	66.1	71.8	68.6	68.1	66.47
Tiny YOLO [46]	48.2	47.2	45.7	47.2	46.95
YOLOv2 [22]	53.3	50.1	62.4	51.3	48.84
YOLOv3 [23]	66.1	75.9	67.6	70.0	69.33
YOLOv4 [24]	86.1	0.64	58.6	73.3	71.82
YOLOv5 [47]	56.0	45.2	45.5	44.4	46.96
KPE-YOLOv5 [32]	66.1	46.3	46.3	44.5	44.85
YOLOv7 [26]	48.2	46.2	46.7	55.6	44.65
YOLOv8 [48]	42.1	36.1	38.8	70.3	36.68
YOLOv9 [49]	46.5	45.6	45.5	57.9	45.36
U-YOLOv3	**74.8**	**79.8**	**69.5**	**76.8**	**74.98**

References

[1] D. L. Rizzini, F. Kallasi, F. Oleari, and S. Caselli, "Investigation of vision-based underwater object detection with multiple datasets," *International Journal of Advanced Robotic Systems*, vol. 12, no. 6, p. 77, 2015 https://doi.org/10.5772/60526.

[2] Y. Zhai, "River ship monitoring based on improved deep-sort algorithm," *Informatica*, vol. 48, no. 9, 2024 https://doi.org/10.31449/inf.v48i9.5886.

[3] X. Qi, "Event-triggered predictive control algorithm for multi-auv formation modeling," *Informatica*, vol. 48, no. 9, 2024 https://doi.org/10.31449/inf.v48i9.5890.

[4] P. Sarkar, S. De, and S. Gurung, "A survey on underwater object detection," in *Intelligence Enabled Research: DoSIER 2021*, pp. 91–104, Springer, 2022 $https://doi.org/10.1007/978-981-19-0489-9-8$.

[5] E. Y. Lam, "Combining gray world and retinex theory for automatic white balance in digital photography," in *Proceedings of the Ninth International Symposium on Consumer Electronics, 2005.(ISCE 2005).*, pp. 134–139, IEEE, 2005 doi: 10.1109/ISCE.2005.1502356.

[6] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin institute*, vol. 310, no. 1, pp. 1–26, 1980 https://doi.org/10.1016/0016-0032(80)90058-7.

[7] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 687–698, 2010 DOI: 10.1109/TPAMI.2010.93.

[8] R. Hummel, "Image enhancement by histogram transformation. comp. graph," 1977 https://doi.org/10.1016/S0146-664X(77)80011-7.

[9] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, pp. 474–485, 1994 https://doi.org/10.1016/B978-0-12-336156-1.50061-6.

[10] D. Garg, N. K. Garg, and M. Kumar, "Underwater image enhancement using blending of clahe and percentile methodologies," *Multimedia Tools and Applications*, vol. 77, pp. 26545–26561, 2018 https://doi.org/10.1007/s11042-018-5878-8.

[11] P. Sarkar, S. Gurung, and S. De, "Underwater image segmentation using fuzzy-based contrast improvement and partition-based thresholding technique," in *Evolution in Computational Intelligence: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021)*, pp. 473–482, Springer, 2022 https://doi.org/10.1007/978-981-16-6616-2-46.

[12] A. S. A. Ghani and N. A. M. Isa, "Enhancement of low quality underwater image through integrated global and local contrast correction," *Applied Soft Computing*, vol. 37, pp. 332–344, 2015 https://doi.org/10.1016/j.asoc.2015.08.033.

[13] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019 DOI: 10.1109/TIP.2019.2955241.

[14] Y. Guan, X. Liu, Z. Yu, Y. Wang, X. Zheng, S. Zhang, and B. Zheng, "Fast underwater image enhancement based on a generative adversarial framework," *Frontiers in Marine Science*, vol. 9, p. 964600, 2023 DOI: 10.1109/ACCESS.2020.3041280.

[15] P. Sarkar, S. De, S. Gurung, and P. Dey, "Uice-mirnet guided image enhancement for underwater object detection," *Scientific Reports*, vol. 14, no. 1, p. 22448, 2024 https://doi.org/10.1038/s41598-024-73243-9.

[16] A. Mondal, "Supervised machine learning approaches for moving object tracking: a survey," *SN Computer Science*, vol. 3, no. 2, p. 146, 2022 https://doi.org/10.1007/s42979-022-01040-0.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014 doi: 10.1109/CVPR.2014.81.

[18] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015 https://doi.org/10.48550/arXiv.1504.08083.

[19] H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, and A.-Y. Zang, "Faster r-cnn for marine organisms detection and recognition using data augmentation," *Neurocomputing*, vol. 337, pp. 372–384, 2019 https://doi.org/10.1016/j.neucom.2019.01.084.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015 doi: 10.1109/TPAMI.2016.2577031.

[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–*

ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21–37, Springer, 2016 https://doi.org/10.48550/arXiv.1512.02325.

[22] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, 2017 https://doi.org/10.48550/arXiv.1612.08242.

[23] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018 https://doi.org/10.48550/arXiv.1804.02767.

[24] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020 https://doi.org/10.48550/arXiv.2004.10934.

[25] J. Wang and N. Yu, "Utd-yolov5: A real-time underwater targets detection method based on attention improved yolov5," *arXiv preprint arXiv:2207.00837*, 2022 https://doi.org/10.48550/arXiv.2207.00837.

[26] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "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, 2023 https://doi.org/10.1109/CVPR52729.2023.00721.

[27] Y. Liu and S. Wang, "A quantitative detection algorithm based on improved faster r-cnn for marine benthos," *Ecological Informatics*, vol. 61, p. 101228, 2021 https://doi.org/10.1016/j.ecoinf.2021.101228.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016 doi: 10.1109/CVPR.2016.90.

[29] C.-H. Yeh, C.-H. Lin, L.-W. Kang, C.-H. Huang, M.-H. Lin, C.-Y. Chang, and C.-C. Wang, "Lightweight deep neural network for joint learning of underwater object detection and color conversion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6129–6143, 2021 doi: 10.1109/TNNLS.2021.3072414.

[30] A. Al Muksit, F. Hasan, M. F. H. B. Emon, M. R. Haque, A. R. Anwary, and S. Shatabda, "Yolo-fish: A robust fish detection model to detect fish in realistic underwater environment," *Ecological Informatics*, vol. 72, p. 101847, 2022 https://doi.org/10.1016/j.ecoinf.2022.101847.

[31] G. Yu, R. Cai, J. Su, M. Hou, and R. Deng, "U-yolov7: A network for underwater organism detection," *Ecological Informatics*, vol. 75, p. 102108, 2023 https://doi.org/10.1016/j.ecoinf.2023.102108.

[32] R. Yang, W. Li, X. Shang, D. Zhu, and X. Man, "Kpe-yolov5: An improved small target detection algorithm based on yolov5," *Electronics*, vol. 12, no. 4, p. 817, 2023 https://doi.org/10.3390/electronics12040817.

[33] S. Qu, C. Cui, J. Duan, Y. Lu, and Z. Pang, "Underwater small target detection under yolov8-la model," *Scientific Reports*, vol. 14, no. 1, p. 16108, 2024 https://doi.org/10.1038/s41598-024-66950-w.

[34] Q. Li and H. Shi, "Yolo-ge: An attention fusion enhanced underwater object detection algorithm," *Journal of Marine Science and Engineering*, vol. 12, no. 10, p. 1885, 2024 https://doi.org/10.3390/jmse12101885.

[35] K. Cai, X. Miao, W. Wang, H. Pang, Y. Liu, and J. Song, "A modified yolov3 model for fish detection based on mobilenetv1 as backbone," *Aquacultural Engineering*, vol. 91, p. 102117, 2020 https://doi.org/10.1016/j.aquaeng.2020.102117.

[36] H. Yang, P. Liu, Y. Hu, and J. Fu, "Research on underwater object recognition based on yolov3," *Microsystem Technologies*, vol. 27, no. 4, pp. 1837–1844, 2021 https://doi.org/10.1007/s00542-019-04694-8.

[37] X. Li, M. Tian, S. Kong, L. Wu, and J. Yu, "A modified yolov3 detection method for vision-based water surface garbage capture robot," *International Journal of Advanced Robotic Systems*, vol. 17, no. 3, p. 1729881420932715, 2020 https://doi.org/10.1177/1729881420932715.

[38] P. Athira, T. M. Haridas, and M. Supriya, "Underwater object detection model based on yolov3 architecture using deep neural networks," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, pp. 40–45, IEEE, 2021 doi: 10.1109/ICACCS51430.2021.9441905.

[39] P. Sarkar, S. De, and S. Gurung, "Fish detection from underwater images using yolo and its challenges," in *Doctoral Symposium on intelligence enabled research*, pp. 149–159, Springer, 2022 https://doi.org/10.1007/978-981-99-1472-2-13.

[40] K. Liu, L. Peng, and S. Tang, "Underwater object detection using tc-yolo with attention mechanisms," *Sensors*, vol. 23, no. 5, p. 2567, 2023 https://doi.org/10.3390/s23052567.

[41] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pp. 492–511, Springer, 2020 https://doi.org/10.48550/arXiv.2205.01649.

[42] M. Pedersen, J. B. Haurum, R. Gade, T. B. Moeslund, and N. Madsen, "Detection of marine animals in a new underwater dataset with varying visibility," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019 DOI: 10.1109/CVPRW.2019.00089.

[43] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," in *2019 international conference on robotics and automation (ICRA)*, pp. 5752–5758, IEEE, 2019 doi: 10.1109/ICRA.2019.8793975.

[44] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015 doi: 10.1109/JOE.2015.2469915.

[45] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015 doi: 10.1109/TIP.2015.2491020.

[46] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved yolov4-tiny," *arXiv preprint arXiv:2011.04244*, 2020 https://doi.org/10.48550/arXiv.2011.04244.

[47] J. Zhang, J. Zhang, K. Zhou, Y. Zhang, H. Chen, and X. Yan, "An improved yolov5-based underwater object-detection framework," *Sensors*, vol. 23, no. 7, p. 3693, 2023 https://doi.org/10.3390/s23073693.

[48] R. Varghese and M. Sambath, "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pp. 1–6, IEEE, 2024 doi: 10.1109/ADICS58448.2024.10533619.

[49] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024 https://doi.org/10.48550/arXiv.2402.13616.