International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

YOLOv2 for Under-Water Object Detection using MATLAB

[1]V. Sri Sai, [2]P. Vithal Chandra, [3]S. Praveen Kumar, [4]V. Sai Santosh, [5]U. Suryateja

Department of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India

ABSTRACT—

A new convolution neural network architecture-based model is proposed. The model was developed using photos from the ocean. This technique uses YOLO to find objects under water. To lower the cost of underwater inspection, an autonomous underwater object detection system is required. The aim of this research is to create a model that can identify and detect underwater items using deep learning techniques. Object localization and object classification are the two components of object detection. Object Classification is the process of classifying an object into predetermined classes, whereas Object Localization is the process of identifying an object based along with its location. Our aim is to train the system with the training dataset and then test the input image by comparing the training data set items. We suggested using the YOLO model to recognize objects in images. YOLO is a method that provides real-time object detection using neural networks. You Only Look Once is known by the acronym YOLO. Because of its accuracy and quickness, this method is well-liked. MATLAB will be used to implement YOLO. MATLAB has a toolkit for deep learning. A platform for developing and deploying deep neural networks with algorithms and trained models is provided by Deep Learning Toolbox.

Keywords—YOLO, object detection, deep learning, MATLAB.

I. Introduction

The fundamental issue with the underwater imaging system is the environmental factors that have an the loss of some object boundaries or background overlapping, which makes it challenging to identify items, detect features, and correctly classify the objects in an image. Object detection is a challenging problem that requires the completion of two key tasks. In order to differentiate foreground items from background objects and give them the appropriate object class labels, the detector must first solve the recognition issue. For the detector to accurately assign bounding boxes to various objects, it must secondly solve the localization problem. These two problems are extremely challenging since the detector encounters several "close" false positives, or "close but incorrect" bounding boxes. While suppressing these nearby false positives, the detector must discover the true positives.

Currently, underwater object detection has a wide range of uses in the marine environment, including research on marine ecosystems, estimation of marine biological populations, conservation of marine species, pelagic fisheries, underwater unexploded ordnance identification, underwater archaeology, and many other potential uses. These uses make it possible to efficiently use marine resources. The detection of fishes, underwater plant species, ships submerged under water, pipelines laid inside water, garbage, etc. is included in the field of underwater object detection. We want to implement our project in MATLAB. For technical computing, MATLAB is a high-performance language. In a simple user interface, it mixes computation, visualization, and programming while expressing issues and solutions using well-known mathematical notation. There are some common uses like mathematical computation, algorithm solving, simulation, data prototyping, data exploration, analysis, and Engineering and scientific visualization, development of applications including creation of GUI or graphical user interfaces.

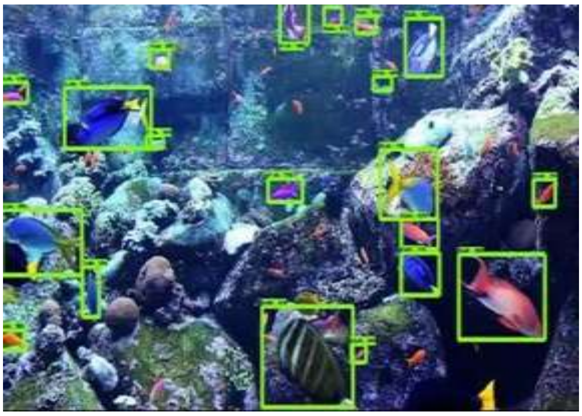

Fig.1. Objects detected under-water

The acronym MATLAB stands for matrix laboratory. A group of application-specific solutions known as toolboxes are available in MATLAB. Toolboxes, which enable you to study and use specialist technologies, are crucial for the majority of MATLAB users. There are toolboxes for many different fields. Toolboxes are extensive groups of MATLAB functions or .m extension that opens the MATLAB system to address certain problem types.

II. Related work

A. YOLOV2

You may train reliable object detectors by using deep learning techniques, a potent machine learning technology. There are numerous methods for object detection, such as You Only Look Once (YOLO) v2 and Faster R-CNN. YOLO has a number of benefits over classifier-based systems. When testing, it considers the entire image, allowing the global context of the image to influence its predictions. In contrast to R-CNN systems, which require thousands of evaluations for a single image, it also makes predictions with a single network assessment. As a result, it is incredibly quick—more than 1000 times as quick as R-CNN and 100 times as quick as Fast R-CNN.

B. MATLAB

MATLAB has deep learning toolbox which is used for object detection. A pre-trained ResNet-50 feature extraction network is automatically provided when we use the yolov2Layers (Computer Vision Toolbox) function for object identification in underwater.

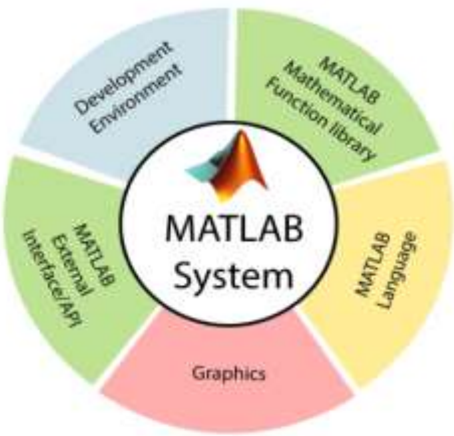

Fig.2. Matlab Applications

C. ResNet50

ResNet stands for Residual Network. A ResNet module version called ResNet50 contains 48 Convolution layers, 1 MaxPool layer, and 1 Average Pool layer. There are different operations available, the count of floating point operations available is 3.8 x 10^9. This ResNet model is frequently utilized. The foundation ResNets provided made it feasible to train ultra -deep neural networks, allowing the network to include hundreds or thousands of layers and still perform well.

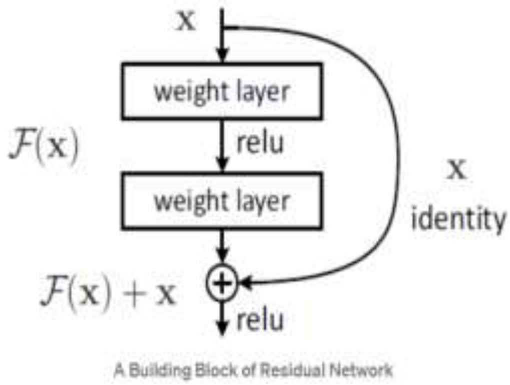

Fig.3. ReLU functionality

D. ReLU

ReLU stands for Rectified Linear Unit. It is an activation function. It is a non-linear activation function that is used in neural networks or deep neural networks which has many layers. The relu function is as:

f(x)=max(0,x)

E. CNN

Convolutional Neural network or CNN may be a sort of manufactured neural organize, which is broadly utilized for image or object acknowledgement and classification. Profound learning thus recognizes objects in a picture by employing a CNN. CNNs are playing a major part in different tasks or functions like picture handling issues, computer vision assignments like location and division, video examination, to recognize objects in self-driving cars. As CNNs are playing a critical part in these fast growing and developing zones, they are exceptionally as well known in Profound Learning.

F. Pooling Layer

Pooling layers are a method for down sampling feature maps by summing the existence of features in regions of the feature map. The average presence of a feature and the most activated presence of a feature are described by the two popular pooling techniques of average pooling and max pooling, respectively. The pooling layer is introduced after the convolution layer.

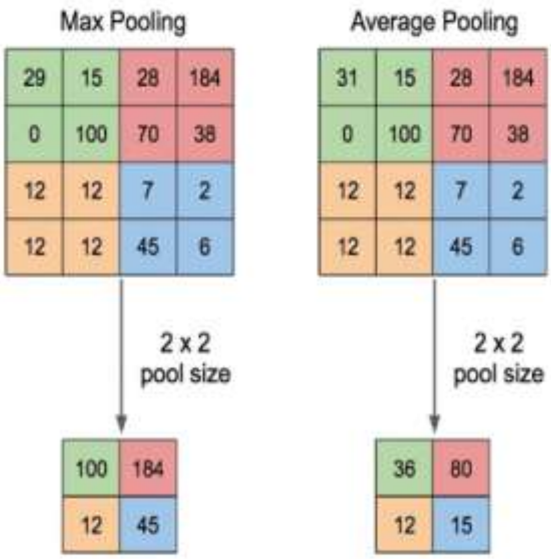

Fig.4. Max Pooling and Average Pooling Layers

G. Softmax layer

Softmax is mostly used for a multi-class problem. It assigns decimal probability to each class. The probability in decimal form must sum up to 1.0. Training converges faster than it would otherwise by adding this *additional* constraint. A neural network layer is used to implement Softmax right before the output layer of the network. The Softmax layer must have the equal number of nodes as the output layer. Softmax takes for granted that every example belongs to exactly one class.

H. Stride

Stride is a part of convolutional neural networks, which are designed by neural networks for the compression of image and video data. The neural network's filter's stride parameter determines how much movement there is across the picture or video. For instance, the filter will move one pixel or unit at a time if the stride of a neural network is set to 1. Because the filter's size influences the volume of the encoded output, stride is frequently set to a whole integer rather than a fraction or decimal.

III. Methodology

A new architecture-based convolutional neural network model is suggested. This technique uses yolo to find objects under water. The main goal of this research is to create a model that can recognize and detect underwater items using deep learning techniques.

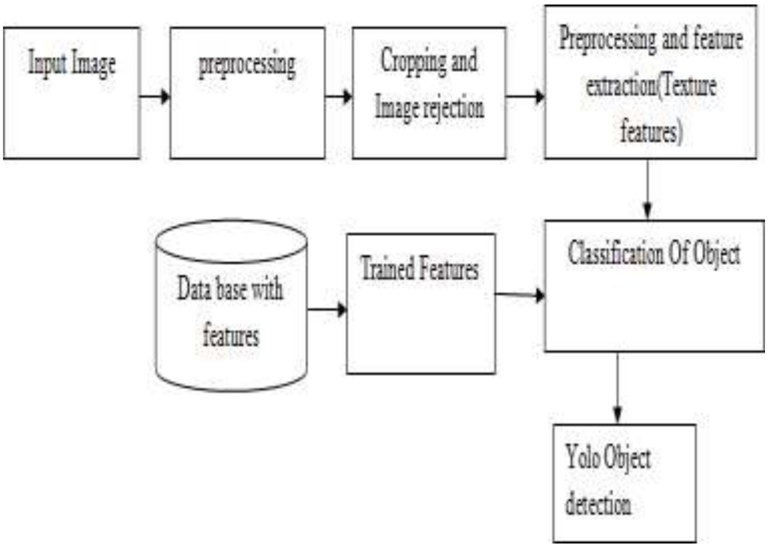

Fig.5. Flow chart of proposed model

Object localization and object classification are the two components of object detection. Object localization and object classification both involve categorizing an object into predetermined classes and separating it from other

objects based on their locations. An efficient algorithm for real-time object recognition is yolo ("you only look once").

Image Classification attempts to classify an image into one of several possible groups. Object Localization enables us to locate our object within the image. Object Detection gives users the means to draw the so-called bounding boxes around each object in a picture and locate all of its components.

A neural network predicts objects in a picture and identifies the items using bounding boxes in object detection, knowledge of image classification. There are several circumstances where we need to determine the precise bounds of our objects in the context of bounding boxes.

Instance segmentation is the name of this procedure. In YOLO, our goal is to forecast an object's class and the bounding box indicating the object's location.

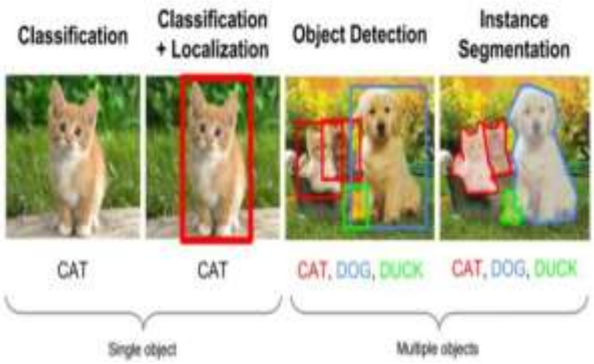

Fig.6. Object detection for multiple objects

Four descriptors can be used to explain each bounding box: the value cis corresponding to an object's class(e.g., bus, cycle, van etc.), the center of a bounding box (bxby), which represents the width and height of an object.

It is necessary to estimate the p value, which represents the chance that an object is present within the bounding box. Usually, a 19*19 grid is used to divide an image into cells. Five bounding boxes must be predicted by each cell. For one image, a sizable number of 1805 boundary boxes have come. Many of these cells and bounding boxes will not have any objects in them for most of the cases.

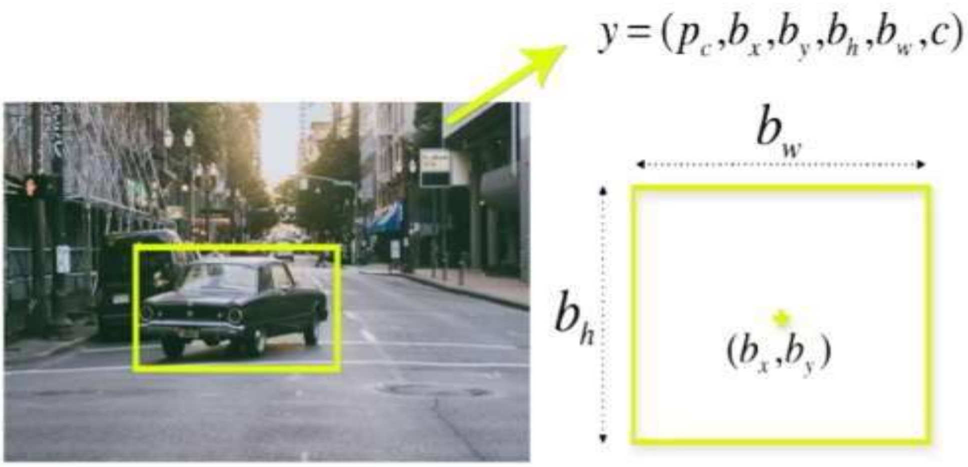

Fig.7. Descriptors of bounding box

In a way to get rid of the boxes with low chance for having object and bounding boxes with the maximum shared area, a technique known as non-max suppression is used.

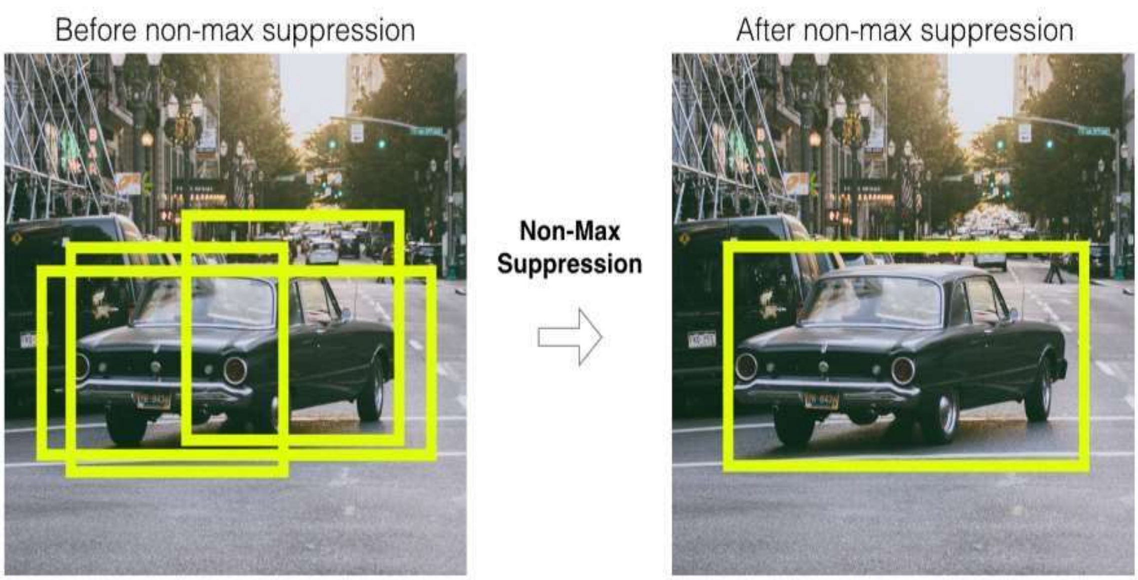

Fig.8. Non-Max Suppression

The mystery and abundance of the deep sea can be revealed in large part because to image processing.

The division of the image into segments is a crucial step in the use of image segmentation techniques.

Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W [2] in the proposed model the issue of creating potential object locations for use in object recognition is discussed. The strength of an exhaustive search and segmentation are combined in the newly introduced selective search algorithm. Selective search yields 99% recall and a Mean Average Best Overlap of 0.879 at 10,097 locations, resulting in a small group of data-driven, class-independent, high quality locations.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. [3] in the proposed model one of the largest CNNs to date was trained on the subsets of ImageNet utilized in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)-2010 and ILSVRC-2012, which is the special contribution of this research. To categorize the 1.2 million high-resolution images entered in the ImageNet LSVRC-2010 competition into the 1000 separate classes, they trained a sizable, deep convolutional neural network. The top-5 error rate for the CNN presented in this research is 18.2%. An error rate of 16.4% is obtained by averaging the forecasts of five comparable CNNs.

The objects in the photographs are found using image segmentation algorithms. It is suggested to apply morphological procedures to detect the edges in an edge-based segmentation approach, which is then followed by an object tracing algorithm. [4]

A method for face detection in an environment which is uncontrolled based on deep learning (CNN) is presented. The suggested approach entails three steps: face/non-face region of interest categorization, region of interest generation, and feature extraction using a pre-trained model. The ResNet50 pre-trained model is applied during feature extraction stage. [1]

riyadarshni, D., & Kolekar, M. H. [5] in the proposed model suggested methodology functions as a highly effective bridge between ideal and real-world circumstances. It is impossible to track both the target's matched presence and unmatched detections simultaneously. Experimental and environmental mistakes are present in the predictions at all times.

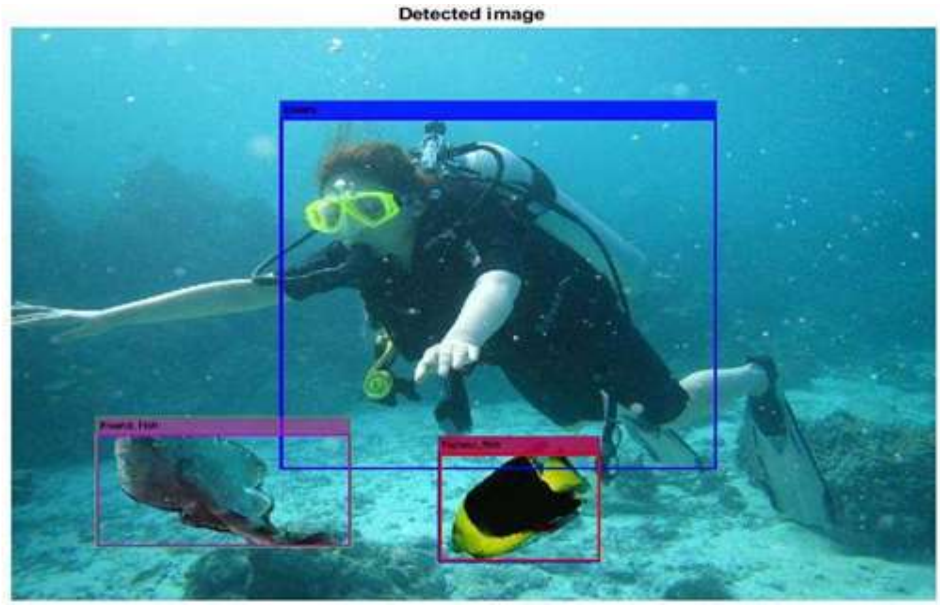

Fig.9. Result of Proposed model

Karimanzira, D., Pfützenreuter, T., & Renkewitz, H. [8] in the proposed model CNN and YOLO were used to create two networks. The first one uses the YOLO principle to locate the target object in a far-off sonar image and second one uses information from an optical camera and a region-based convolutional neural network (CNN) architecture to find the docking station in near proximity.

Sumahasan, S., Addanki, U. K., Irlapati, N., & Jonnala, A [11] in the proposed model 2500 images in the dataset utilised in this model are divided into 5 classes: Car, Cycle, Motorbike, Dog, and Flowers. Object Classification and Object Localization are the two components of object detection. The system, which wants Object Detection as a primary feature, may benefit from this approach. Accuracy obtained is 92%.

The Cascade R-CNN, a multi-stage object detection framework is used for the creation of excellent object detectors. [25]

The field of underwater picture enhancement saw the introduction of the residual learning model and the super-resolution reconstruction model VDSR.

It was suggested that Underwater Resnet (UResnet) be used to enhance the multi-term loss function's performance. [22]

CNN and YOLO were used to create two networks. The first one uses the YOLO principle to locate the target object in a far-off sonar image. Less IoU is needed for this detector. however, there are strict criteria for position update speed because the AUV is permitted to drive quickly in order to approach the docking station as quickly as feasible. [18]

Many computer vision and image processing techniques include a crucial phase called picture segmentation. It is frequently used in projects like object recognition, classification, and tracking. [19]

Because of its speed, precision, and capacity for learning, YOLOv2 is selected. The ability to classify underwater object depends on the outcomes of enhancement, segmentation, efficient features, properties of the chosen classifiers and the dataset's characteristics. [27]

The ability to classify underwater objects depends on the outcomes of enhancement, segmentation, efficiency features, properties of the chosen classifiers, and the dataset's characteristic. [26]

We are able to provide the input image at run time while using MATLAB to implement the YOLO code.

According to the pre-defined classifications, the items in the input image are identified and localized.

Paulo, D. J., Maurell, I. P., & Protas, E. V. [9] in the proposed model The goal of underwater picture segmentation in the wild is presented together with a collection of datasets for deep CNN architectures to be trained on. On a random test set of 300 genuine underwater photographs, the accuracy achieved was 91.9%.

The output image following object classification, the metrics scores, and the graph created for the metric score are used to determine the way the model is performing and are all included in the outcome. We obtained accuracy of 0.95, F1 score of 0.90, specificity of 0.97, and sensitivity of 0.90 while providing an input image at runtime.

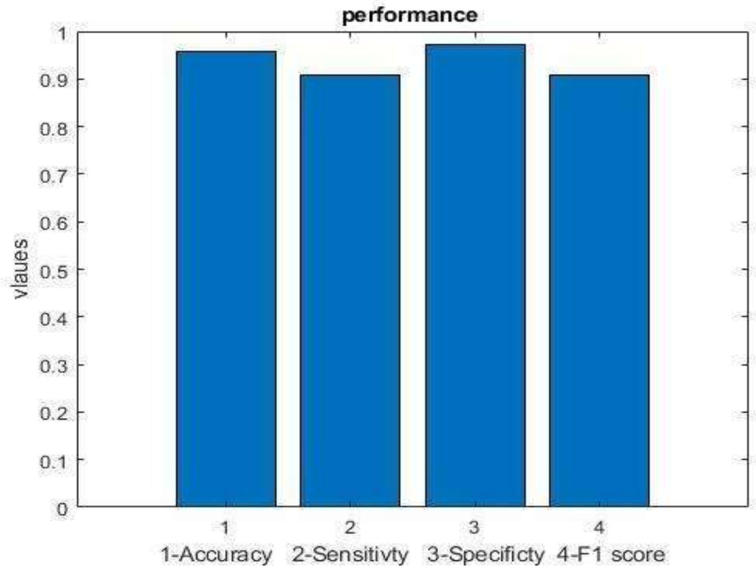

Fig.10. Visualization of metrics

A graph is generated showing the metrics such as accuracy, F1 score, specificity, sensitivity.

IV. Conclusion and Future Work

To identify the underwater objects, a model is proposed. Divers, underwater creatures, and several other items can all be identified using the model. The YOLOV2 model is utilized because to its high accuracy, quick execution, and effective performance. Object localization and object classification are the two components of object detection. Object localization and object classification both involve classifying an object into predetermined classes and separating it from other objects based on their locations. To remove boxes with minimum item probabilities and bounding boxes having the maximum common area, we employed non-max suppression. In object detection models, the intersection over union metric is employed to assess localization precision and compute localization errors.

The acronym MATLAB, which stands for Matrix Laboratory, is used to implement this approach. A deep learning toolset is available in MATLAB. Deep Learning Toolbox is a framework for creating and deploying deep neural networks using algorithms and trained models. The model's accuracy was 95.7 percent. We intend to use multi-scale ResNet models, which can accurately identify small objects, to improve the model in the future.

References

[1]. Mliki, H., Dammak, S., & Fendri, E. (2020). An improved multi-scale face detection using convolutional neural network. *Signal, Image and Video Processing*, *14*(7), 1345-1353.

[2]. Wang, X., & Yang, J. (2020). Marathon athletes number recognition model with compound deep neural network. *Signal, Image and Video Processing*, *14*(7), 1379-1386.

[3]. He, K, Zhang, X.,Ren S., & Sun J(2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778)

[4]. Priyadharsini, R., & Sharmila, T. S. (2019). Object detection in underwater acoustic images using edge based segmentation method. *Procedia Computer Science*, *165*, 759-765.

[5]. Song, Y., He, B., & Liu, P. (2019). Real-time object detection for AUVs using self-cascaded convolutional neural networks. IEEE Journal of Oceanic Engineering, 46(1), 56-67.

[6]. Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision*, *104*(2), 154-171.

[7]. Redmon, J., & Farhadi A.(2017). YOLO9000: Better, faster, wtronger. In Proceedings of thr IEEE conferences on computer vision and pattern recognition(pp 7263-7271).

[8]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[9]. Pedersen, M., Bruslund Haurum, J., Gade, R., & Moeslund, T. B. (2019). Detection of marine animals in a new underwater dataset with varying visibility. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 18-26).

[10]. Fabbri, C., Islam, M. J., & Sattar, J. (2018, May). Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 7159- 7165). IEEE.

[11]. Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprintarXiv:1511.07122.

[12]. Han, F., Yao, J., Zhu, H., & Wang, C. (2020). Underwater image processing and object detection based on deep CNN method. *Journal of Sensors*, *2020*.

[13]. Zhang, L., Li, C., & Sun, H. (2022). Object detection/tracking toward underwater photographs by remotely operated vehicles (ROVs). *Future Generation Computer Systems*, *126*, 163-168.

[14]. Priyadarshni, D., & Kolekar, M. H. (2020). Underwater object detection and tracking. In *Soft Computing: Theories and Applications* (pp. 837-846). Springer, Singapore.

[15]. Yeh, C. H., Lin, C. H., Kang, L. W., Huang, C. H., Lin, M. H., Chang, C. Y., & Wang, C. C. (2021). Lightweight deep neural network for joint learning of underwater object detection and color conversion. *IEEE Transactions on Neural Networks and Learning Systems*.

[16]. Sumahasan. S., Addanki. U. K. Irlapati. N. & Jonnala .A(2020). Object Detection using Deep Learning Algorithm CNN. International Journal for Research in Applied Science and Engineering Technology(IJRASET), 8(VIII).

[17]. Antoniou, A.,Storkey, A., & Edwards,H. (2017). Data augmentation generative adversaria lnetworks. *arXiv preprint arXiv:1711.04340*.

[18]. Karimanzira, D., Pfützenreuter, T., & Renkewitz, H. (2021). Deep learning for long and short range object detection in underwater environment.

[19]. Paulo, D. J., Maurell, I. P., & Protas, E. V. (2021). Under-water image segmentation in the wild using deep learning. *Journal of the Brazilian Computer Society*, *27*(1).

[20]. M V, J. (2020). Survey on Deep Learning Techniques Used for Classification of Underwater Sonar Images.

[21]. Mitra, V., Wang, C. J., & Banerjee, S. (2006). Lidar detection of underwater objects using a neuro SVM- based architecture. *IEEE Transactions on Neural Networks*, *17*(3), 717-731.

[22]. Liu, P., Wang, G., Qi, H., Zhang, C., Zheng, H., & Yu, Z. (2019). Underwater image enhancement with a deep residual framework. *IEEE Access*, *7*, 94614-94629.

[23]. Li, Y., Zhang, Y., Li, W., & Jiang, T. (2018). Marine wireless big data: Efficient transmission, related applications, and challenges. *IEEE Wireless Communications*, *25*(1), 19-25.

[24]. Chen, X., Yu, J., Kong, S., Wu, Z., Fang, X., & Wen, L. (2019). Towards real-time advancement of under-water visual quality with GAN. *IEEE Transactions on Industrial Electronics*, *66*(12), 9350-9359.

[25]. Fei, T., Kraus, D., & Zoubir, A. M. (2012). A hybrid relevance measure for feature selection and its application to underwater objects recognition. In *2012 19th IEEE International Conference on Image Processing* (pp. 97-100). IEEE.

[26]. 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).

[27]. Mohammed, M. S., Khater, H. A., Hassan, Y. F., & Elsayed, A. PROPOSED APPROACH FOR AUTOMATIC UNDERWATER OBJECT CLASSIFICATION.

[28]. Yu, X., Xing, X., Zheng, H., Fu, X., Huang, Y., & Ding, X. (2018, April). Man-made object recognition from underwater optical images using deep learning and transfer learning. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1852-1856). IEEE.

[29]. Katija, K., Roberts, P. L., Daniels, J., Lapides, A., Barnard, K., Risi, M., ... & Takahashi, J. (2021). Visual tracking of deepwater animals using machine learning-controlled robotic underwater vehicles. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 860-869).

[30]. Mahavarkar, A., Kadwadkar, R., Maurya, S., & Raveendran, S. (2020). Underwater Object Detection using Tensorflow. In *ITM Web of Conferences* (Vol. 32, p. 03037). EDP Sciences.