Robust Underwater Object Detection with Autonomous Underwater Vehicle: A Comprehensive Study

Dipta Gomes
Department of Computer Science
American International
University-Bangladesh (AIUB)
Dhaka, Bangladesh
diptagomes@gmail.com

Department of Computer Science
American International
University-Bangladesh (AIUB)
Dhaka, Bangladesh
saif@aiub.edu

Dip Nandi
Department of Computer Science
American International
University-Bangladesh (AIUB)
Dhaka, Bangladesh
dip.nandi@aiub.edu

A.F.M. Saifuddin Saif

ABSTRACT

Underwater Object Detection had been one of the most challenging research fields of Computer Vision and Image Processing. Before Computer Vision techniques were used for underwater imaging, all the tasks associated with object detection had to be done manually by marine scientists making the task one of the most tedious and error prone. For this case, Underwater Autonomous Vehicles (UAV) has been developed to capture real time videos for specific object detection. Using different hardware improvements and using many varied forms of algorithms, classification of objects, mainly living objects had been carried with different AUVs and high-resolution cameras. Conventional object detection methods of Computer Vision fail to provide accurate detection results due to some challenges faced underwater. For such reasons, object detection underwater needs to be robust, real time and fast also being accurate, for which deep learning approaches are introduced. In this paper, all the works here all the trending underwater object detection techniques are discussed in details and a comprehensive comparative study is carried out.

CCS CONCEPTS

• Computing methodologies → Computer vision tasks• Computing methodologies → Image processing • Computing methodologies → Vision for robotics

KEYWORDS

Underwater Object Detection, Deep Learning, ImageEnhancement, Image Processing and Underwater Autonomous Vehicle.

ICCA 2020, January 10–12, 2020, Dhaka, Bangladesh

ACM ISBN 978-1-4503-7778-2/20/01? $15.00
https://doi.org/10.1145/3377049.3377052

ACM Reference format:
Dipta Gomes, A F M Saifuddin Saif, Dip Nandi. 2020.Robust Underwater Object Detection with Autonomous Underwater Vehicle: A Comprehensive Study. In *Proceedings of ICCA 2020, Dhaka, Bangladesh.*

1 Introduction

Underwater imaging poses as one of the most challenging domains of research due to some well-known challenges and constraints of the underwater environment. Images taken underwater have poor illumination, color degradation, dominance of blue light, haziness and unwanted obstacles making imaging and analyzing of the images difficult. Earlier researches pointed out that, mainly two forms of detection types are present for detection of underwater objects. Firstly, the detection of underwater objects that are moving on its own in real time and the second is detection of moving objects from frames extracted from videos. Both these types are of importance due to lack of good quality real time data that are required for processing. This paper will discuss some important approaches based on deep learning and conventional methods that will be used to carry out detection of underwater objects. Through this review, a comparison between conventional methods and deep learning methods is discussed that will provide a good stepping stone for further researches. For real time object detection, specific object tracking algorithm needs to be incorporated and several segmentation methods are required to find the Object of Interest (OOI) where detection and movement of AUV can be carried out at the same time. In several experiments by D. Lee et al. [7] and Aneta Nikolovska [3] and Guo-Jia Hou et al. [12] man-made objects are used to test real time object detection where images are taken as the input for classification. Here the processing time is fast, but still this method is not ideal for very dynamic environment and has low accuracy even though both works in real time, fast and robust. For real-time object detection techniques, underwater image segmentation and detection is very challenging, so several algorithms such as discriminative regional feature integration by Yafei Zhu et al. [5], background modelling using multi-feature integration framework by Srikanth Vasamsetti et al. [4] and blob analysis by Hailing Zhou et al. [11], color restoration algorithm by D. Lee et al. [7] to counterfeit color degradation, Constant false Alarm Rate and MBES by Aneta Nikolovska et al. [3] to detect OOI in dynamic underwater environment are used in previous researches. For removing haziness in real time underwater images,

approaches like enhanced fuzzy intensification operator by C. Akila et al. [8] have already been used, providing efficient object clarity for classification and detection. In case of detection of OOI from moving objects, background subtraction and frame difference methods by Hongkung Liu et al. [10] are used. Using the help of color and shape features of known man-made underwater objects [12], underwater object detection methods are also being used that provides reliable and faster results. On the other hand, the second type that deals with stationary objects are using images obtained from underwater videos that are not real time. The deep learning algorithms are used on extracted frames where the frames are fed into a neural network. Several notable deep learning algorithms based on Convolutional Neural Networks are done by Nicole Seese et al. [9], S´ebastien Villon et al. [16], A. Mahmood et al. [13], Hansang Lee et al. [15] and by Jailun Dai et al. [14]. All this research has been tested providing very promising results. Here CNN automates the steps of feature extraction and classification. Even though generic deep learning models are slower than real time detection algorithms, this method provides very promising detection accuracy for different object classification. For faster and accurate classification using CNN, the Fast R-CNN method used by Xiu Li1 Min Shang et al. [17] helps in overcoming the long time required for training the deep learning models. Several modes of CNN are used in order to detect underwater objects. Namely, ZooplanktonNet by Jialun Dai et al. [14], transfer learning alongside CNN by Hansang Lee et al. [15] have been used in previous researches. Feature-based detection is carried out by A.Mahmood et al. [13] and bounding box fusion by S´ebastien Villon [16] to increase efficiency of detection and classification of underwater objects. CNN has already been successfully used to detect coral, plankton, fish by Xiu Li1 et al. [17] along with other underwater living organisms. The corresponding researches and their findings will be evaluated for future researches and through this paper a comparison between several underwater object detection techniques will be put forward.

2 Core Research Background

Underwater object detection and underwater image enhancement methods are research domains that fascinate research enthusiasts and Computer Vision experts and thus several new paths and researches are put forward in this field of research. Object Detection is a core domain of Computer Vision as a result researches on this filed are of high value. The main research problems associated with this study involves underwater object detection and underwater image enhancement. Here, object detection is mainly of two types, one for moving objects and one for stationary objects. Underwater Object detection can then be divided into two types, one conventional methods and other deep learning methods.

Conventional methods include identification of objects using Gravity Gradient Coefficients and differential methods where the gravity gradient measurements are taken. Haar-like features are used by B. Kim et al. in [2], where the shadow part and the highlighted parts are considered Haar-like features where sonar images are used to find the intensity of light emitted from the sonar to the object. Srikanth Vasamsetti et al. [4] in 2018, proposed a feature descriptor known as MFTP (Multi frame triplet pattern) which collects data of spatiotemporal texture between two successive frames. The feature descriptor is later integrated with color and motion features to detect underwater moving objects. To detect salient objects from the foreground of underwater images Yafei Zhu et al. [5] used Discriminative Regional Feature Integration (DRFI) algorithm which integrate various data of regional contrast, regional property and regional background to create the master saliency map. Consecutive portion of the map with higher pixel values than threshold values are considered as objects. For proper illumination, background estimation is carried out by Nicole Seese et al. [9] in 2016, where he used two algorithms, one is Gaussian Mixture Model and another is Kalman filtering. Here both the algorithms are used to create an independent model, where GMM provides a background model tool and Kalman filtering as a predictive model. Hongkung Liu et al. [10] used background subtraction to detect moving objects. From the captured frame, the three-difference method is then used to find the moving objects from a set of frames captured from an image. Hailing Zhou et al. [11] in 2015 used Gaussian Mixture Model for carrying out background modelling. Using parameter of the blob analysis using bounding box, compactness and circularity, objects are recognized as each object has unique blob features. Guo-Jia et al. [12] used detection method based on color and shape features where using Color based Extraction algorithm (CEA) with an YUV model all the objects of interest from a candidate region is extracted. S´ebastien Villon et al. [16] proposed a model using both HOG+SVM and CNN where Coral fishes are detected. Here the thumbnails obtained from images are divided into 10 zones. In each zone HOG value is calculated. In case of SVR which is supported by vector regression, a Gaussian Radial basis function kernel is used to find a distinction between training sets and moreover build a classifier to find the “background” of the images.

Contrast enhancement using underwater imaging model by Yujie Li et al. [6] is carried out, where using de-scattering algorithm, images are modified and sent to a CNN model for classification. Classification with hybrid features is done by A. Mahmood et al. [13] where Convolutional Neural Networks and a pre-trained VGGNet are used for feature extraction. Here local spatial Pyramid poling is used for point annotation. Finally, color and textual features of an image with CNN features are combined for more accurate classification results. The convolutional neural network proposed by Jialun Dai et al. [14] is known as ZooplanktonNet. It is an inspired Network from AlexNet and VGGNet. Hansang Lee et al. [15] proposed a model that deals with a transfer learning-based CNN. The CIFAR10 Model [15] contains three convolutional layers following by two fully connected layers. S´ebastien Villon [16] dealt with models to

detect coral Fishes using a Neural Network based on Google Net. A motion score is calculated to detect moving fishes which differentiate moving objects based on their speed. Xiu Li1 Min Shang et al. [17] proposed a fast R-CNN architecture which is used to carry out the classification process of fishes, where the network takes RGB image as input along with its 2000 regions of interest from the image. The network ultimately produces a distribution of fish classes along with its bounding boxes that helps in the detection.

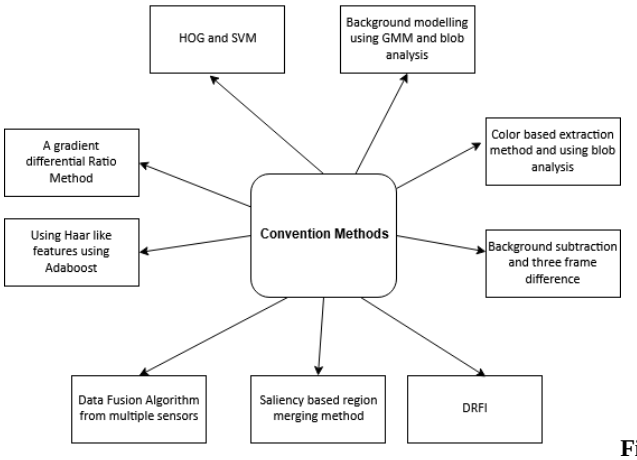

Figure 1: Conventional Methods

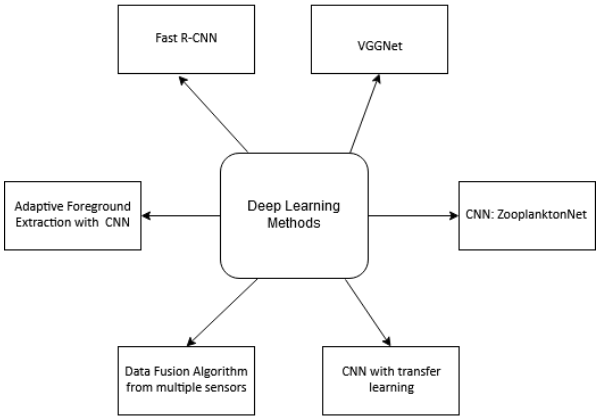

Figure 2: Deep Learning Methods

2.1 Previous Research Methods

Zu Yan et al. [1] proposed a method of detecting underwater objects by finding its gravity gradient potential. Here for the object its gravity coefficients are first calculated through the calculation of all six components of gravity gradient tensors at a specific point in a Cartesian Field of a horizontal XY axis and a Z vertical axis. The barycenter location of the object is then calculated using the Newton Raphson Method. Based on mass and barycenter location, the object of interest is identified. The methods being both efficient to be used in Autonomous Underwater Vehicles and requiring no separate gravity gradient maps. In spite of all this, the method is very sensitive to noise, as with change in object mass, and barycenter location, any slight change causes error in the detection process.

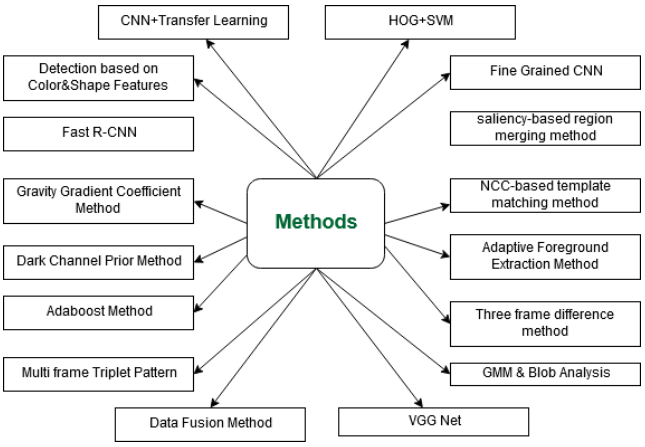

Figure 3: Previous Research Methods

B. Kim et al. [2] proposed Adaboost method of underwater object detection which is used to detect real time underwater objects from haar-like features from shadow properties taken from a forward-looking imaging sonar. Using the forward-looking imaging sonar, better quality images are taken where the haar-like features are obtained from grayscale images being both simpler to process and easier to interpret. As a result, the overall process becomes easier to be carried out.

Aneta Nikolovska et al. [3] used Data Fusion algorithm for fusing data from multiple sensors using Constant False Alarm Rate (CFAR) method and the Distance Regularized Level Set Evaluation (DRLSE) to find the object of interest. Here, immediate categorization of the object of interest is possible, as a result the method is both fast and robust even though for high resolution images the speed of the method decreases. Data from multiple sensors are complex to integrate making the overall process difficult to implement.

Srikanth Vasamsetti et al. [4] proposed a method for detecting moving underwater objects using Multi frame Triplet Pattern (MFTP). Here the feature descriptor is related to both motion and color features where from each pixel a histogram is generated which is later used in the detection process. The images used in this method only deals with grayscale images excluding any colors. Here pixel by pixel data is evaluated, thus the overall method is expensive to implement.

Yafei Zhu et al. [5] in 2016 proposed a saliency-based region merging method where haziness of the image is removed by using the dark channel prior method. At first a mean shift segmentation is used to extract the object of interest from the background. The DRFI detects salient objects from still images but ignores the overall image, thus detection of living underwater organisms are difficult to detect using this method.

Yujie Li et al. [6] used the underwater dark channel prior enhancement method, where using a depth map underwater objects are detected. Here, the depth map is refined using a guidance map which helps in overcoming the problem of de-scattering. To improve the accuracy of the depth map, the refinement of the joint is performed under a guidance image. For color correction in artificial lighting, a chromatic transfer function is used. Here, the method is ideal for turbid water. Artificial lighting is used only instead of real-life pictures as a result is not ideal for real life scenarios whereas a depth map is maintained that increases the complexity of the overall method.

D. Lee et al. [7] used Normalized Cross Correlation (NCC)-based template matching and Mean shift tracking to detect moving underwater objects, where color correction is carried out using Jaffe–McGlamery color restoration method . Here in this method cheap cameras are used to take images whereas for close range the cameras fail to provide a perfect image. Thus, the overall method has a huge downside with identifying close ranged objects.

Nicole Seese et al. [9] proposed Adaptive Foreground Extraction Method using deep Convolutional Neural Network which works very well with dynamic environment. For classification uses a deep Convolution Neural Network. It focuses on unknown illumination parameters, dynamic backgrounds and non-static imaging platform, as a result, it is very efficient in real life scenarios. For more dynamic environment a Gaussian Mixture Model is used and a separate Kalman filter is used for simple situations. As a result, the method is likely to suffer with efficiency and speed.

Moving object detection by background subtraction and three frame difference method was proposed by Hongkung Liu et al. [10]. He uses Background subtraction to handle light changes which is less intensive as well as less expensive to run, thus it is ideal for real time processing. Three frame difference handles the background noise as well as non-static background objects. It can only detect moving pixels but not a particular stationary object. Thus, works on continuous videos only but not for still images.

Object detection using background modelling method using Gaussian Mixture Models and blob analysis is put forward by Hailing Zhou et al. [11] where the intensity functions of image background is calculated. For k components, k numbers of clusters are created using Orchard-Boumann method. Then image segmentation is carried out using Otsu algorithm where the background pixels are eliminated due to segmentation. As a result, the foreground pixels are used. Here blob analysis is used for detection of object of interest.
Fast R-CNN method used by Xiu Li1 Min Shang et al. [17] is solely a method for fish detection. The method is comparatively faster than R-CNN and finds the values of Higher mean Average Precision (map). The overall research helped in constructing of a new huge dataset of 24272 images with 12 classes. When feeding in the network an input of 2000 regions of interest (ROI) are collected from selective search, which is time consuming. This process is not in real time, even though the process is fast.

Detection and Recognition of underwater man-made objects on the basis of color and shape features proposed by Guo-Jia et al. [12] is both efficient and better for complex situations. To improve the accuracy, the Average Extraction Rate is calculated for all possible pixels of the images. Improved Otsu algorithm is then used to carry out segmentation where the 2D Otsu algorithm is spilt into two 1D algorithms. Here, the method only deals with man-made objects thus not ideal while integrating into an AUV. Here regular shape with bright color man-made objects are used and not dull or objects with darker shades. Here the method deals with images taken from a short distance, thus difficult detecting objects if it is in a long distance.
The deep learning method using VGGNet proposed by A. Mahmood et al. [13] deals with hybrid features where the feature extraction proposed is based on Spatial Pyramid Pooling (SPP) approach. Here the VGGNet is pre trained and deep features from the VGGNet is combined with the texton and color-based features to improve classification. The MLC dataset is then used to train the CNN.

ZooplanktonNet, which is a Convolutional Neural Network based model is proposed by Jialun Dai et al. [14] has higher accuracy for detecting Zooplankton. It uses data augmentation to decrease data overfitting during classification. Compared to other image classification algorithms, CNN requires less pre-processing time and requires lack of dependence on prior knowledge. Here also there lacks images of Zoo Plankton for deep Neural Networks. So, this research seemed well for less data.

A Fine-grained Classification method based on Convolutional Neural Network by Hansang Lee et al. [15] uses transfer learning along with pre-trained CNN. In order to overcome class imbalance problem, multiple data augmentation techniques alongside transfer learning was used. It is convenient for large scale class imbalance dataset and so is applicable and efficient to carry out a satisfactory result.

Both Convolutional and Deep learning methods, Neural Network and HOG+SVM is used to detect underwater objects proposed by S´ebastien Villon et al. [16]. It detects underwater coral reef fishes from underwater images extracted from videos. Better detection accuracy is obtained in deep learning rather than in traditional

methods. HOG uses contours of images and is used to detect in complex situations, even in hidden in coral reef or occluded in coral reef.

Table 1. Previous Methods along with Advantages and Disadvantages

Method	Advantages	Disadvantages
Gravity gradient potential Method [1]	No separate gravity gradient maps	Very sensitive to noise data
AdaBoost Method [2]	Faster Interpretation, use of haar-like features and easy Interpretation	Not suitable for complex and dynamic environment
Data Fusion Algorithm Method [3]	Immediate Categorization of OOI, Faster & Robust	Slower for high Resolution Images, difficult integration of data from multiple sensors
Multi frame Triplet Pattern Method [4]	Feature descriptor of both motion and color	Pixel by pixel evaluation, slower and only applicable for grayscale images.
Saliency based region merging Method [5]	Removal of haziness in images, extract of OOI from background	Only salient objects are of important and difficult to detect living organisms underwater
Dark Channel Prior Enhancement Method [5]	Use of guidance map to remove de-scattering in images, use of chromatic function to improve lighting in images, used profoundly for turbid water	Use of Artificial Lighting, high complexity due to maintenance of depth maps
NCC-Based Template Matching Method [7]	Mean shift tracking to detect moving objects, color correction	Use of cheap cameras, close range fails to capture a satisfactory image, not ideal for objects in close proximity to the camera
Adaptive Foreground Extraction Method [9]	Works well for dynamic Environment, CNN for classification, focuses on unknown illumination parameters and non-static environment.	For complex and more dynamic environment use of Gaussian Mixture Model and Kalman Filter which decreases speed and efficiency
Three Frame Difference Method [10]	Moving Object detection, handle light changes and use for real time environment	Detects only moving pixels, not suitable to detect static objects, only works for continuous video footage
Blob Analysis Method [11]	Accurate, calculation of intensity function of the image background	For k-components k number of clusters for classifying thus slower and inefficient for background modelling
Fast R-CNN [17]	Faster than R-CNN, creation of fish dataset	2000 regions of interest used as input which requires huge startup time thus not applicable for real life scenarios
Color & Shape Features [12]	Average extraction rate is calculated to improve accuracy, segmentation using Otsu Algorithm, integrated with UAV	Only deals with man-made objects, not applicable for dull color objects, images that are taken from shorter distance is only taken into consideration
VGGNet [13]	Deals with Hybrid features, pre-trained with deep features	Use of MLC Dataset, which is not suitable for image classification
ZooplanktonNet [14]	High accuracy Rate, use of data augmentation for decreasing data overfitting, less preprocessing	Lack of images of Plankton as Deep Neural Network requires huge sets of data
CNN + Transfer Learning [15]	Pre-trained CNN, overcoming class imbalance problem, use of multiple data augmentation techniques	Suitable for large scale of data dependent tasks, not at all for smaller levels of tasks
HOG+SVM [16]	Used to detect hidden and occluded objects underwater	Slower in detection and less efficiency compared to deep learning techniques

2.2 Previous Research Based on Frameworks

Zu Yan et al. [1] first used the three gravity differentials which is calculated between adjacent gravity gradients, where at last a ratio of Gravity Gradient is calculated to find each unique object in the frame. The barycenter location of the object is obtained using the Newton-Raphson Method. Using this framework real time detection is carried out, but only for linear movement of AUV. It is optimum for stationary objects. Finally, due to inconsistency of class distribution, classification of small sized classes is difficult.

B. Kim et al. [2] proposes a framework where several weak classifiers are first created based on haar-like features from sonar images. Average intensity values $V(x)$ for each pixel is then calculated based on a threshold value. Using Adaboost algorithm, several weak classifiers are merged to construct a strong classifier. Due to the cascading structure, the probable region of OOI is reduced and yields a good result for any slight change of shape and size. But as sonar images are being used, where most images compose only background, so excess resources are spent unknowingly providing a less variation in the detection mechanism.

In the research of Aneta et al. [3] uses an area of interest which is first defined using a high frequency SSS sensor. The data obtained was then analyzed using the CFAR algorithm which confirms

Object of Interest from the area of Interest. Then MBES was used, where all the geological position is analyzed. Each image undergone MBES was then analyzed using DRLSE to add robustness, that helps pointing the presence of OOI features. The OOI obtained are then identified and categorized at burial level, using the BOSS method. Using data fusion, all the 3 layers data are merged to identify the specific object. Here, lower movement speed yields better SAS resolution and water must have good visibility to provide a clear view of the object. The fusion method is difficult to implement and complex in nature as each sensor uses its own template to find OOI.

Srikanth Vasamsetti et al. [4] first uses blockchain procedure to divide incoming blocks into big overlapping and non-overlapping small blocks. Then background histogram for each block is obtained using MFTP descriptor. Moving Object detection is carried out for blocks whose background patterns is greater than the threshold. The probability of small blocks is computed for every incoming frame using Otsu's method. The probabilities of foreground moving objects are extracted using color features. Then, using three-frame differencing technique, motion Information of temporal moving objects is extracted. Finally, Binary images are combined together to get the final foreground moving objects. Here block chain procedure is used to cancel noise which gives consistent object detection results requiring less execution time and memory usage. On the other hand, it gives coarse boundary detection and linkage of adjacent moving objects and the background model needs to be updated for acquiring pattern changes in each block area.

Yafei Zhu et al. [5] put forwards a framework that uses dark channel prior algorithm. Here using Discriminative Regional Integration (DRFI) algorithm a saliency map is first created. Higher saliency value than the threshold pixels are identified using Otsu method from the saliency map. Then the mean shift method is used for initial segmentation. Finally, with maximal-similarity based region merging color histogram and similarity between pixels is obtained. Here mean shift initial segmentation decreases regions to consider, thus increasing processing speed. Unfortunately, the dark channel prior used will be invalid if the scene objects are inherently similar to the atmospheric light and no shadow is cast on them. As a result, it only works for single indoor image only.

Yujie Li et al. [6] proposed a framework which uses artificial lighting where depth map is redefined using a proposed joined guidance image filter. Here at first, images are recovered using Dark channel prior de-scattering model and distortion of images are corrected based on physical spectral characteristics-based color correction method. The distortion due to de-scattering using color correction method can be improved even though serious distortion of color still remains unaffected and for using artificial lighting, this framework is not at all suitable for real life scenarios.

D. Lee et al. [7] proposed a framework where several man-made 3D objects are used as targets. Color restoration of images are carried out using Jaffe–McGlamery model of color restoration. SURF is then used to carry out feature-based detection method, where a template based matching technique is used, mainly the Normalized Cross Correlation (NCC)-based template matching. The NCC shows robust results in varying illumination environments and preprocessing such as histogram normalization adds more robustness to illumination changes. Here, Mean Shift Tracking is used to detect moving objects, which shows robustness to changes in distance. Here template-based approach used is suitable for 2D images only.

Nicole Seese et al. [9] in his framework proposed a deep learning approach to detect underwater object. Here at first, all current frames are extracted and converted to grayscale images. GMM is then used to generate a background model, along with Kalman filter in order to help predict the background model. CUDA is then used to support parallel computing, as each pixel is analyzed which is computationally expensive. Parallel computing allows faster analyzing. On the other hand, automated feature learning and classification is carried out using Google TensorFlow. Here after training a fish dataset, foreground is segmented before processed to CNN. Here, due to CNN, the feature extraction process becomes automated and GPU helps in supporting parallel computing.

Hongkung Liu et al. [10] first uses all images that are shot at an indoor space. Background modeling is carried out along with three frame difference technique which is then used to detect moving objects from a set of sequential frames. Any change to the external environment changes the background subtraction result, as a result it is best for static camera images and time interval between frames remain ambiguous which solely depends on condition of water. Blob analysis is then carried out to find the foreground objects.

Hailing Zhou et al. [11] proposed Gaussian Mixture Model (GMM) for background modelling. Here only the pixels with bin values more than the threshold are used to train. Segmentation of the images are carried out using Otsu Algorithm, and morphological erosion operations are carried out to remove noise and increase accuracy. From the foreground pixels, the object is detected using blob analysis and using Orchard Bouman algorithm, several clusters are created to complete the detection process. The clustering algorithm being fast, yields better efficiency, even though background pixels that dominate the images slows down the process and due to use of multiple GMMs, the overall process tends to be slower than other approaches.

Deep learning approach proposed by Xiu Li1 Min Shang et al. [17] used a modified version of AlexNet having five convolutional layers and three fully connected layers using the open source Caffe CNN library. The detection process is further

speeded up by using Singular Value decomposition (SVD) to compress fully connected layers. While taking sampling, horizontally flipped images are ignored and data augmentation is used. It converts video frames to images as a result not real time in nature. As a result, the whole process costs large time and space costs.

Guo-Jia Hou et al. [12] in 2015 first pre-processed images to adjust non uniform illumination. Then color-based extraction algorithm (CEA) is used to extract object of interest. The process is only applicable for grayscale images and being computationally expensive is not suitable for real time scenarios. The model requires high resolution images to classify accurately and some mixed pixels and false alarms also exists because of closeness between foreground and background.

A. Mahmood et al. [13] proposed a CNN based feature extraction where Pre-trained VGGNet is used containing 1000 classes with more than a million images. The output of the first fully connected layer is used as the feature vector. Weights are modified using the MLC dataset. Combining CNN features along with hand crafted features increases the classification performance. For classification a two-layer Multilayer Perceptron (MLP) network consisting of two fully connected layers followed by a soft-max layer with 9 output classes is trained using the MLC dataset. Use of ImageNet to train the dataset is not ideal for coral classification and MLC does not have pixel annotations to meet the input size constraint of CNN, so a lot of extra steps are required that vastly increases the complexity of the process.
Jialun Dai et al. [14] proposed a framework where first rescaling images along subtracting the mean value over the training set from each pixel be used for classification of objects. Using Artificial augmentation dataset is increased. Dataset is divided into test and train set and CNN is then trained using the training set. Data augmentation helps to overcome poor quality and small quantity of images. in spite of this the process only works with images of 256 X 256 pixels only. So, images of lower resolution or higher resolution will not work properly.

Hansang Lee et al. [15] first selected CIFAR 10 CNN model as a classifier model. A class normalized data is created based on original data with random data threshold value to decrease biasness. Then transfer learning is applied to counter with the loss of information of population caused by normalization of data. Here the classifier is trained with the normalized data along with transfer learning to detect planktons. Advantages are use of transfer learning. CIFAR 10 is more efficient than other approaches but use of multiple steps of pre-processing before classification slows down the process. For very large threshold N, classification bias will not be reduced, so detection error prevails.

S´ebastien Villon et al. [16] cropped frames of captured videos and used to create a database of 13000 fish thumbnails. The data is then widened by applying rotations and symmetries. Cropped thumbnails with labels are sent to the network, which is based on the Google Net with 27 layers, 9 inception layers and a soft-max classifier. For background a separate class is maintained which contains random thumbnails of the background and specific thumbnails around the fish as the background is highly textured helping in better detection. To improve localization accuracy, another class known as the part of fish is used which helps in processing the whole fish rather than specific part of the fish. But here, species less than 450 thumbnails are omitted from the dataset. The final detection is based on the hypothesis that most fishes are moving thus is based on the motion score, so stationary underwater objects cannot be detected.

2.3 Analysis Based on Previous Experimental Results

Zu Yan et al. [1] experimented where parameters of interests were object mass and barycenter location. The experiment gave a relative error with 6% when object is in the distance of within 360 m.

B. Kim et al. [2] in his experiment used Haar-like features of sonar images and Region of interest in the images the main parameters. The experiment gave an estimation of processing time under 25 milliseconds and processing speed about 40 fps
.
Aneta Nikolovska et al. [3] carried out experiments with parameters like geolocation points, burial depth of object of interest, distance of the DDAUV from the ocean floor and tomographic image generated from the BOSS sensor software. Mathematical estimation obtained from the experiment are data from MBES confirms presence of the OOI with confidence of above 80% and using SSS error is estimated to be in the interval of 4±2° and position identified with an estimation with an error of 5±2 m.

Srikanth Vasamsetti et al. [4] used parameters size of big block, size of small block, grey-scale intensity values, color features and motion features. Here it is found, the process performs better for blurred scenes, background having the complex environment and luminosity variations maximum F-score value obtained at $\text{Tau} = 0.05$. Max TPR=0.95 using MFI for luminosity change and Min FPR=0.0043 and F=0.0053 Using SILPT for complex backgrounds and hybrid videos and precision= 0.88 and 0.89 for camouflage foreground and hybrid background videos.

Using underwater robot yshark by D. Lee et al. [7] the parameters taken are the distances between the light sources and the objects, distances between the object surfaces and the cameras, the approaching angles of the ray vectors, the departing angles (y and y_0) for the two different mediums, Wavelength, Attenuation coefficients of the travelling mediums and distance of the object. During experimentation, target object detection had best TPR (true positive rate) of detection (for the cone image) =0.9479 and FPR=0. Target object tracking in another experiment had best TPR (sensitivity) for Sphere =0.9321 and FPR = 0. In third

experiment best TPR (sensitivity) for Sphere =0.9407 and FPR = 0.

Hailing Zhou et al. [11] used AIM's dataset to carry out the experiments. The parameters used are blobs features bounding box, compactness and circularity and foreground pixels. The experiment results are as follows, detection of Jellyfish Accuracy is 80.8% and precision is 90.3%. For detection of sea snake Accuracy is 66.7% and precision is 90.5%.

Guo-Jia Hou et al. [12] selected three representative images. The algorithm was then implemented using MATLAB. The parameters used are Illumination intensity, distance between object and the camera and shooting angle. The experiments provided the results, Detection Accuracy of 87.6%, compared to Abbadi and Saadi which was 80.2% and Sari et al. which was 86.5 %. Detection time of 15.2 milli seconds compared to Abbadi and Saadi19.6 milli seconds and Sari et al. 23.8 milli seconds.

A. Mahmood et al. [13] uses the output of the first fully connected layer of the VGGNet as the input for all 3 experiments. In Experiment 1, The classifier is trained on two-thirds of the image from the year 2008 from the dataset and tested on the remaining images of the same year. For experiment 2, training set is used the images of 2008 and the test set is the 2009 images. For experiment 3, images of year 2008 and 2009 are used as training set and 2010 as the test set. The classification accuracies along with the Average Class Precision for each experiment is obtained. The parameters used in this experiment were Color Descriptors and texture Descriptors. Classification Accuracies for Combined Features: Experiment 1= 77.9, Experiment 2= 70.1and Experiment 3= 84.5. For average Class Precision for combined features: Experiment 1: 0.69, Experiment 2: 0.63 and Experiment 3: 0.68.

Jialun Dai et al. [14] first Performed some popular architecture such as AlexNet, Caffe Net, VGGNet and Google Net on the proposed model. Depth or layers of the Network is then identified which is ideal between 8 and 16 layers to observe the final predictions. It is found that accuracy decreases with the increase of depth. The numbers of convolutions required for the ZooPlanktonNet was found to be around 384 to 512 convolutions. Parameters used in this research are Number of convolutions of the network and layers of the network. It was found for 11 layers the accuracy is 92.8 %.

Hansang Lee et al. [15] first pre-processed the WHOI-Plankton dataset by resizing to 64 × 64 with mean value padding. All data before 2014 was treated as training data and data in 2014 is treated as testing data. The Average Accuracy Rate and Unweighted average score is calculated for evaluation. The parameters used in this experiment were threshold value, ratio of five largest classes L5, average Accuracy Rate and unweighted average score. The proposed classifier yielded average Accuracy Rates (for all classes) = 92.80% and unweighted Average F_1 = 0.3339.

S´ebastien Villon et al. [16] used 4 test videos on coral reef. Biology experts select 400 frames from all over the videos and provide ground boxes of all visible fishes in the frame. The recall, precision and F-Measure of the detection is calculated for threshold=98% for both SM+HOG and deep learning. The parameters used are histogram of oriented gradient, HOG features, deep learning features obtained from the CNN and feature vectors. The estimation found in this research are, for SVM+HOG: F-measure < 49% and for CNN: F-Measure > 55%, Precision < 78% and Recall < 71 %. Obviously Deep learning is a wise choice.

Xiu Li1 Min Shang et al. [17] at first obtained the detection results and compared with two other new approaches to underwater object detection which are Deformable part models (DPM) and Regions with CNN (R-CNN). Then the Mean Average Precision (mAP) is calculated to evaluate the three methods. Here the parameters used in this experiment are bounding box and 2000 ROI per image. Fast R-CNN gives an accuracy of mAP 81.4% improving 11.2% compared to DPM baseline and slightly improving to R_CCN with bounding box regression. Test Speedup of 80.2% is observed compared to R-CNN on a single fish image.

Table 2. Summary of Review and Extimation of Experiments

Target	Methods	Dataset	Type	Accuracy
Plankton [15]	CNN +Transfer learning	WHOI-Plankton	Deep learning	92.8 %
Fish [17]	Fast R-CNN	ImageCLEF_Fish_TS	Deep learning	81.4 %
Coral [13]	VGGNet	Moorea Labelled Coral (MLC) dataset	Deep learning	84.5%
Plankton [14]	ZooPlanktonNet	ZooPlankton dataset	Deep learning	92.8%
Real life underwater images [5]	Underwater dark channel prior-based image enhancement method	JAMSTEC ImageNet	Deep learning	52.28 %
Coral Reef Fish [16]	CNN/ HOG+SVR	MARBEC database	Deep Leaning / Conventional	78% (deep learning)
Jelly Fish / Sea Snake [11]	Object detection using background modelling method using Gaussian Mixture Models	AIMS' datasets	Conventional	80.8% (Jelly Fish) 90.3 % (Sea Snake)

	and blob analysis			
Man Made Objects [12]	color-based extraction algorithm (CEA)+2D Otsu Algorithm	Own built	Conventional	87.6%,
Moving objects [4]	Multi-Frame Triplet Pattern (MFTP) feature descriptor and three-frame difference technique	Fish4Knowledge database	Conventional	89%
Man Made Objects moving objects [7]	Jaffe–McGlamery color restoration method, (NCC)-based template matching and Mean shift tracking	Own built	Conventional	94.79 %

3 Observation and Discussion

Based on the observations from the above experiments it was found, most works done recently were from the deep learning point of view. One of the very important reason for this is, deep learning gives far more accuracy in terms of detection. Most detection tasks are carried out for moving objects. Notable are Zu Yan et al. [1], B. Kim et al. [2], D. Lee et al. [7], Nicole Seese et al. [9] and Hongkung Liu et al. [10] which are resource extensive and requires more time to process and detect. Moreover, video processing in real time is far costlier in order to incorporate with a AUV. For a faster and accurate detection images from well-built datasets are more accurate and faster. Conventional methods using several important aspects are mostly built for man-made objects. Researches built on man-made objects are not perfect for real life scenarios.

Several improvements before classification are necessary for accurate detection. Transfer learning with CNN by Hansang Lee et al. [15] provided an accuracy of 92.8% where ZooPlanktonNet based on Google Net by Jialun Dai et al. [14] gives an accuracy of 92.8%. VGGNet by A. Mahmood et al. in [13] results with the highest accuracy of 84.5 % among all the results. Fast R-CNN by Xiu Li1 Min Shang [17] gave an result mAP of 81.4%. S´ebastien Villon et al. [16] for coral fish detection obtains a precision above 78% which proves deep learning gives a very accurate result if compared with other conventional methods. Detection using blob analysis on complex real-life underwater objects by background modelling using Gaussian Mixture Model by Hailing Zhou et al. [11] provided an accuracy of 90.3%. The model is very robust by fast clustering algorithm with improved Otsu algorithm, still lagging behind for being high resource extensive. Man-made object detection in Guo-Jia Hou et al. [12] provided a very promising result using Otsu algorithm with an accuracy of 87.6%. The Multi frame integration framework in Srikanth Vasamsetti et al. [4] and D. Lee et al. [7] yields a promising accuracy of 88% for moving objects. D. Lee et al. [7] carried out a manmade moving object detection with an accuracy of 94.79 %. Thus, the conventional methods are lagging behind the deep learning methods from the previous section with respect to accuracy. Man-made object detection methods by Guo-Jia Hou et al. [12] and D. Lee et al. [7] even though give better results, are still not suitable for dynamic object detection. Here, even though dealing with images, deep learning approaches are much more efficient in nature.

4 Conclusion

The research tends to provide a brief understanding of the current state of research in underwater imagery for using with Autonomous Underwater Vehicle (AUV) so that it can be used for underwater exploration and object detection. Several deep learning approaches and conventional methods are reviewed and at the end a comparison between the methods are obtained. Convolutional Neural Networks (CNN) which is used widely appreciated for computer vision models and classification in complex environments, seemed the perfect solution for underwater object detection. The problems associated with dataset can be addressed with data augmentations, data modification through segmentations, foreground extraction and background extraction of images to find the right object of interest to detect. Moreover, underwater imaging for both static and dynamic environment are observed and it was found, object detection for static environment gave more accuracy and works fast even though it is a bit difficult to manage in real life. This is a huge challenge for underwater imaging as moving objects in dynamic environment are really difficult to process and takes huge processing time. The steps associated with object detection also involves steps that are resource intensive. To modify all the above problems the paper tends to provide all possible solutions. All this leads to a solution that the above problem can be solved using deep learning. Future works that should be dealt with are development of deep learning methods for dynamic environment which is robust as well as fast for accurate object detection and can easily be integrated into an AUV. This will make the whole process of underwater object detection reliable as well as simple.

REFERENCES

[1] Zu Yan, Author, Jie Ma, Jinwen Tian, Hai Liu, Jingang Yu, and Yun Zhang 2014. A Gravity Gradient Differential Ratio Method for Underwater Object Detection. *IEEE Geoscience And Remote Sensing Letters, Vol. 11, doi: 10.1109/LGRS.2013.2279485*

[2] B. Kim and S. Yu. Imaging sonar based real-time underwater object detection utilizing AdaBoost method. 2017 IEEE Underwater Technology (UT), Busan, 2017, pp. 1-5, doi: 10.1109/UT.2017.7890300

[3] Nikolovska, AUV based flushed and buried object detection. 2015.OCEANS 2015 - Genova, Genoa, 2015, pp. 1-5, doi: 10.1109/OCEANS-Genova.2015.7271651

[4] Srikanth Vasamsetti ,Supriya Setia, Neerja Mittal1, Harish K. Sardana, Geetanjali Babbar. 2018. Automatic underwater moving object detection using

multi-feature integration framework in complex backgrounds. *IET Comput. Vis., 2018, Vol. 12 Iss. 6, pp. 770-778, The Institution of Engineering and Technology 2018*

[5] Yafei Zhu, Lin Chang, Jialun Dai, Haiyong Zheng, Bing Zheng. 2016. Automatic object detection and segmentation from underwater images via saliency-based region merging". OCEANS 2016 - Shanghai, IEEE, doi:*10.1109/OCEANSAP.2016.7485598*

[6] Yujie Li, Huimin Lu, Jianru Li, Xin Li, Yun Li, Seiichi Serikawa. 2016. Underwater image de-scattering and classification by deep neural network, *Computers & Electrical Engineering,Volume 54,2016,Pages 68-77,ISSN 0045-7906, doi: 10.1016/j.compeleceng.2016.08.008.*

[7] D. Lee, G. Kim, D. Kim, H. Myung, and H.-T. Choi. 2012. Vision-based object detection and tracking for autonomous navigation of underwater robots. *Ocean Engineering, vol. 48, pp. 59–68, Jul. 2012, ACM, doi:10.1016/j.oceaneng.2012.04.006*

[8] C. Akila and R. Varatharajan. 2018. Color fidelity and visibility enhancement of underwater image de-hazing by enhanced fuzzy intensification operator. *Multimed Tools Appl, vol. 77, no. 4, pp. 4309–4322, Feb. 2018, February 2018, Volume 77, Issue 4, pp 4309–4322, Springer, doi:*

[9] Nicole Seese, Andrew Meyers, Kaleb Smith, Antony O. Smith. 2016. Adaptive Foreground Extraction for Deep Fish Classification. *2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI), Cancun, Mexico, IEEE, doi: 10.1109/CVAUI.2016.016*

[10] Hongkung Liu, JailunDia, Ruchen Wang, Haiyong Zheng, bing Zheng. 2016. Combining background subtraction and three-frame difference to detect moving object from underwater video. *OCEANS 2016-Shanghai, IEEE, 2016, doi:10.1109/OCEANSAP.2016.7485613*

[11] Hailing Zhou, Lyndon Llewellyn, Lei Wei, Doug CreightonandSaeidNahavandi. 2015. Marine Object Detection using background Modelling and blob analysis, *2015 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 2015, doi: 10.1109/SMC.2015.86*

[12] Guo-Jia Hou, Xin Luan, Da-Lei Song§, and Xue-Yan Man (2015). Underwater Man-made Object Recognition on the basis of color and shape features. *Journal of Coastal Research, Coastal and Research Foundation, doi: 10.2112/JCOASTRES-D-14-00249.1*

[13] A. Mahmood, M. Bennamoun, S.An, F.Sohel, F.Boussaid, R.Hovey, G.Kendrick, R.B.Fisher. 2016. Coral Classification with hybrid feature representations. *2016 IEEE International Conference on Image Processing (ICIP), doi: 10.1109/ICIP.2016.7532411*

[14] Jialun Dai, Ruchen Wang, Haiyong Zheng, Guangrong Ji, XiaoyanQiao. 2016. ZooplanktonNet: Deep Convolutional Network for Zooplankton Classification, *OCEANS 2016 - Shanghai, IEEE, doi: 10.1109/OCEANSAP.2016.7485680*

[15] Hansang Lee, Minseok Park, Junmo Kim. 2016. Plankton Classification On Imbalanced Large Scale Database Via Convolutional Neural Networks With Transfer Learning. *2016 IEEE International Conference on Image Processing (ICIP), doi: 10.1109/ICIP.2016.7533053*

[16] S´ebastien Villon, Marc Chaumont, G´erardSubsol, S´ebastienVill´eger, Thomas Claverie, and David Mouillot. 2016. Coral Reef Fish Detection and Recognition in Underwater Videos by Supervised Machine Learning: Comparison Between Deep Learning and HOG+SVM Methods. *International Conference on Advanced Concepts for Intelligent Vision Systems, ACIVS 2016: Advanced Concepts for Intelligent Vision Systems pp 160-171, doi: 10.1007/978-3-319-48680-2_15*

[17] Xiu Li1 Min Shang, Hongwei Qin, Liansheng Chen. 2015. Fast Accurate Fish Detection and Recognition of underwater images with Fast R-CNN. 19-22 Oct. 2015, IEEE Xplore, doi: *10.23919/OCEANS.2015.7404464*

[18] Md. Moniruzzaman, Syed Mohammed Shamsul Islam, Mohammed Bennamoun, and Paul Lavery. 2017. Deep Learning on Underwater Marine Object Detection: A Survey. *International Conference on Advanced Concepts for Intelligent Vision Systems ACIVS 2017: Advanced Concepts for Intelligent Vision Systems pp 150-160, 2017*

[19] Lin Wu, Xin Tian, Jie Ma, and Jinwen Tian. 2010. Underwater Object Detection Based on Gravity Gradient. IEEE Geoscience And Remote Sensing Letters, Vol. 7, No. 2, April 2010

[20] Byeongjin Kim, Hyeonwoo Cho and Son-Cheol Yu. 2016. Development of imaging sonar based autonomous trajectory backtracking using AUV. 2016 IEEE/OES Autonomous Underwater Vehicles (AUV)

[21] Juhwan Kim and Son-Cheol Yu. 2016. Convolutional Neural Network-based Real-Time ROV Detection using Forward-Looking Sonar Image. 2016 IEEE/OES Autonomous Underwater Vehicles (AUV)

[22]Nina S. T. Hirata, Mariela A. Fernandez&Rubens M. Lopes. 2016. Plankton Image Classification based on Multiple Segmentations. 2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)

[23] Ouyang py, Hu Hong, Shi zhongzhi. 2016. Plankton Classification with Deep Convolutional Neural Networks.2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference