Robust Underwater Fish Detection Using an Enhanced Convolutional Neural Network

Dipta Gomes
American International University-Bangladesh (AIUB), Dhaka, Bangladesh
Email: diptagomes@aiub.edu

A.F.M. Saifuddin Saif
American International University-Bangladesh (AIUB), Dhaka, Bangladesh
Email: saif@aiub.edu

Abstract: Underwater Object Detection is one of the most challenging and unexplored domains in this area of Computer Vision. The proposed research refines the image enhancement of under-water imagery by proposing an improvement of already existing tools for underwater Object detection. The comparative study clearly depicts the enhancement of the proposed method with respect to the existing methods for underwater object detection. Moreover, a framework for detection of underwater organisms such as fishes are proposed, which will act as the steppingstone for re-serving the ecosystem of the whole fish community. Mostly the object detection using deep learning has been the prime goal to this research and the comparison between other preexisting methods are compared at the end. As a result, techniques that are already well established will be used for overall enhancement of those images. Through this enhancement and through finding a healthy environment for their breeding ground, the extinction of selected species of fishes is can be diminished and decreased. All this is carried out by overcoming difficulties underwater through a novel technique that can be integrated into an Underwater Autonomous Vehicle and can be classified as robust in nature. Robustness will depend on three important factors in this research, first is accuracy, then fast and lastly being upgradeable. The proposed method is a modified VGGNet-16, which is trained using the ImageCLEF FISH_TS dataset. The overall result provides an accuracy of 96.4% which surpasses all its predecessors.

Index Terms: Underwater Object Detection, VGGNET, Convo-lutional Neural Networks, Data Augmentation

1. Introduction

Underwater imaging faces some terrible consequences during object detection. Degrading of images and presence of certain anomalies poses a very difficult task to detect. Almost 70% of the Earth is water and the ecosystem underwater is still a mystery. As a result, observance of the underwater environment is required. The underwater image processing and enhancing algorithm for detecting underwater organisms that can be integrated into a mapping payload of an UAV needs to be implanted to detect and observe underwater ecosystem. As a result, through observance several aspects of underwater object detection as well as enhancement of existing methods can be achieved. This provides a wider range of possibilities to explore underwater to discover more objects. To preserve the ecosystem a far more efficient method needs to be used. Based on the findings, it is possible to determine the right track of research and in the meanwhile better results of the problem may be achieved, to contribute towards more accurate and faster method of object detection underwater. Beside detection, it is also required to find the most suitable track of further research in this very specific domain of research.

This Research is a steppingstone of detecting underwater organisms using basic yet robust deep learning mechanism. A proper enhancer and detector imaging technology for marine organism needs improvement. For Further improvement of underwater imaging, this research has some vital importance. The proposed enhancing tool will be an innovation for marine life as this enhancement is solely for seabed waterbed organisms.

Object detection had been one of the most vital functionalities of Computer Vision and for a very long time, various types and forms of objects are studied and detected, taking this field into a completely new level. Even then, object detection had not been so easy for each type of environment. Most importantly, underwater object detection is still at developing stage, and each day new methods are proposed to make the overall process accurate and faster. Previously researches regarding underwater environment required experts to manually annotate objects for detection. Due to huge data explosions all over the world, image data are of huge abundance, mainly for marine objects.

Automating the detection process now needs to be addressed, in order to enhance our knowledge about the underwater environment. The overall study of the paper includes past approaches based on conventional methods and as well as deep learning methods. Various data sets such as ImageCLEF [1] which are Wikipedia Image Retrieval Data sets, JAMSTEC Japan Agency for Marine- Earth Science and Technology data set [2], Moorea Labelled Coral (MLC) data set [3], AIMS's data set [4] and various others are previously for various and underwater object detection. Annotations for each of the datasets are carried out through various computer vision techniques that will be further explored in this paper.

For real time classification and segmentation of the under- water images, robots such as Autonomous Underwater vehicles are used which helps to monitor and carry out classification process rapidly and accurately. But a mechanism to integrate along with the AUV is a major challenge for researchers as the process should be robust, accurate as well as simple. To meet all the requirements, conventional methods even though provided a very promising result, failed to provide an accurate result for specific dynamic environments, as underwater environments are extremely diverse and full of constraints. Various constraints such as poor light illumination, excessive blue light, haziness, light attenuation, color degradation and presence of various obstacles makes the detection even more difficult. To counterfeit this huge barrier, researchers put forward deep learning approaches that give more accuracy regarding underwater object detection. Deep learning algorithms based approaches are taken by Nicole Seese et al. [1], Sebastien Villon et al. [5], A.Mahmood et al. [3], [6] and [7]. Here overall accuracy of over 90% is found for each of the researches proving deep learning as a wise choice for underwater object detection.

CNN has already successfully used to detect, coral, Plankton, Fish and other living organisms living underwater. The paper will address a Convolutional Neural Network that will classify fish living underwater. The data set used was IM- AGECLF_fish_ts. Where more than 250 images of fishes were used to classify. The data augmentation method was used to make the images more diversified in order to make the classification process more accurate. Here, the deep learning library Keras was used on top of TensorFlow to build the network. Here a 5 2D Convolutional Layered Network is used which is inspired from the VGGNET-2014 [8] with a SoftMax layer for classification and Rectified Linear Activation Function (ReLU) as the activation for each Convolutional layer. The implemented network is made as simple and concise as possible so that the overall process can easily be included into the payload of Autonomous Under- water Vehicle. Main contributions to the research as follows, own built CNN solely for fish detection, Supplication of VG- GNET for smaller datasets and finally increasing the overall accuracy of a CNN for underwater object detection to above90%.

2. Literature Review

The research is based on some of the milestones of underwater object detection methods that put forward some very promising results. Notably some of the research that had very high accuracy rate as well as efficiency is studied.

At first some conventional object detection methods, most notably by Zu Yan et al. [9] where the gravity differential are calculated between adjacent gravity gradients to detect unique objects. B. Kim proposed a framework composed of multiple classifiers [10] to detect objects using haar like features from sonar images. Aneta et al. [11] searches for object of interest from a defined area using a high frequency SSS sensor. The data is then analyzed using the CFAR algorithm and then categorized using BOSS method. Srikanth Vasamsetti et al. [12] uses block chain procedure and analyzed each block using MFTP descriptor. Plankton image classification is carried out by Nina S. T. Hirata et al [13] based on multiple segmentations. Yafei Zhu et al. [14] put forward 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 pix- els is obtained. Yujie Li et al. [2] proposed a framework which uses artificial lighting where depth maps is redefined using a proposed joined guidance image filter. D. Lee et al. [15] proposed a framework where several man-made 3D objects are used as targets. Color restoration of images is carried out using Jaffe–McGlamery model of color restoration. SURF is then used to carry out featurebased detection method. Hongkung Liu et al. [16] first use 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. Hailing Zhou et al. [4] proposed Gaussian Mix- ture 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. Guo-Jia Hou et al. [17] in 2015 first pre-processed images to adjust non uniform illumination. Then color-based extraction algorithm (CEA) is used to extract object of interest. A survey for multiple other approaches are discussed in [18] that has very relevance to this research.

Deep Learning methods are nowadays have transformed into an ideal solution to very complex problems that deals with classification. Most importantly deep learning, mainly Convolutional Neural Networks are widely popular for object classification, segmentation, and detection purposes. Even though deep learning methods have showed huge progress in this fields, still deep learning methods needs to be developed for more dynamic environment like underwater environment. Some of the recent pioneer researches on underwater object detection based on deep learning approaches are Adaptive Foreground Extraction using CNN [19], VGG Net [3], CNN with Transfer Learning based fine grained

classification method [7]. The following figure 1 provides the previous deep learning methods used for this specific purpose of underwater object detection. In spite of all the obstacles faced underwater the following deep learning methods provided an ideal solution for difficult classification purposes.
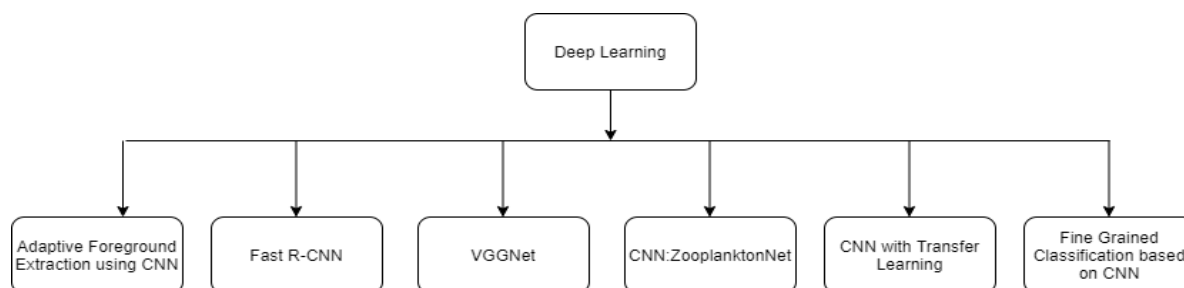

Fig. 1. Deep Learning Approaches

Several other research are carried out to detect underwater objects using deep learning methods. Most notably, Fast R-CNN in [19] where, a specific fish detecting method is proposed. The overall research helped in constructing of a huge new dataset of 24272 images with 12 classes. When feeding in the network input 2000 regions of interest (ROI) collected from selective search are fed, which is time consuming. This process is not in real time, even though the process is fast. Here a modified version of AlexNet which is a Convolutional Neural Network model named after its founder Alex Krizhevsky is used having five convolutional layers and three fully connected layers using the open source Caffee CNN which is a deep learning framework developed by Berkeley AI Research (BAIR). 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. Here Mean Accuracy Precision (mAP) is 81.4% is obtained thus improving 11.2% compared to DPM baseline and slightly improving to Regions with Convolutional Neural Networks (RCNN) with bounding box regression. Test Speedup of 80.2% is observed compared to R-CNN on a single fish image. VGGNet in [3] was used pretrained, as a result the accuracy rate was higher than that of Fast Region with CNN (Fast R-CNN) in [19]. In case of very complex environment with Hybrid Features, VGGNet provided much superior results than that of Fast R-CNN. This is because, in case of Fast R-CNN, 2000 regions of interests are used as input to detect Region of Interest (ROI) from the image. This process speeds up the process but decreases the efficiency by requiring more time thanneeded by VGG Net.

The deep learning method using a Convolutional Neural Network by Oxford's renowned Visual Geometry Group (VGG) known as Visual Geometry Group Network (VG- GNet) in [3] deals with hybrid features where the feature ex- traction 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 texture and color-based features to improve classification. The Moorea Labelled (MLC) dataset is then used to train the CNN. The CNN based feature extraction model is composed of a Pretrained VGGNet containing 1000 classes with more than a million images. The output of the first fully connected layer is used as the feature vector and 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) net- work consisting of two fully connected layers followed by a soft-max layer with 9 output classes is trained using the MLC dataset. Use of Image Net 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. The method provides a high accuracy of 84.5% during classification of underwater objects. The Adaptive Foreground Extraction Method extract features au- tomatically unlike VGG Net [1]. It deals with images with Luminosity better than that of VGG Net, along with detec- tion of Camouflaged Foreground objects. Overall it works better for dynamic environment than VGG Net.

Nicole Seese (2016) et al. [1] proposed Adaptive Fore- ground Extraction Method using deep Convolutional Neural Network which works very well with dynamic environment and uses a deep Convolution Neural Networks for classifying. It focuses on unknown illumination parameters, dynamic backgrounds and nonstatic 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. Zooplankton Net[6] based on Convolutional Neural Networks, which was unlike Adaptive Foreground Extraction where the CNN was solely used for dynamic environment. In adaptive Foreground Extraction method, Gaussian Mixture Model (GMM) is used for background extraction. As a result background estimation requires longer time. On the other hand, in Zooplankton Net, methods such as data augmentation to decrease overfitting was used. As a result, the method gave accurate data. Even in case of classification, deep CNN is used which speeds up the process, giving morespeed than Adaptive Foreground Extraction Method.

Zooplankton Net a convolutional Neural Network based on Zooplankton architecture [6] has higher accuracy for detecting ZooPlanktonNet. It uses data augmentation to de- crease data over fitting during classification. Compare to another image classification algorithms, CNN requires less pre- processing and requires lack of dependence on prior knowledge. Here also there are lacks of images of Zoo Plankton for deep Neural Networks. So this research seemed well for less data. The numbers of convolutions required for the Zoo- PlanktonNet 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%. For hugely dynamic or changing environment, the accuracy rate of ZooplantonNet CNN falls rapidly. But, the Fine-Grained Classification overcomes this problem by using Transfer Learning on top of the Convolutional Layers. As a result, the overall accuracy is increased, and many more objects were able to be detected.

A Finegrained Classification method based on Convolutional Neural Network by Hansang Lee (2016) et al. [7] uses transfer learning by pre-training CNN in order to overcome class imbalance problem and also uses multiple data augmentation techniques alongside transfer learning. It is convenient for large scale class imbalance dataset and so is applicable and efficient to carry out a result. Hansang Lee (2016) et al. [7] 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 using transfer learning with CIFAR 10 is that it 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. Fine Grained Classification Method on the other hand, uses transfer learning even though gives an accurate result but fails to carry out the classification task faster. As a result, the Google Net method used is far faster as well as accurate. As a result, gives a more efficient solution to the problem.

Coral Fishes are detected using a Neural Network (Neural Network) based on Google Net [5]. Here a motion score is calculated, as it to detect moving fishes. Both Convolutional Neural Network and Histogram of Oriented Gradients (HOG) along with Support Vector Machines (SVM) is used to detect underwater objects by Sebastien Villon (2016) et al.

[5] which detects underwater coral reef fishes from underwater images extracted from videos. It provides better detection 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. 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. Deep Convolutional Neural Network [20] which is a Sparse Classification Method, uses color enhancement and also acts to overcome the degradation of underwater images. This acts as a good modification to the Neural Network based on Google Net. On the other hand [20] the method can be used to per- form well in every aspect and condition of image rather than the NN based on the Google Net.

Mohammad Elawady et al. [20] proposed a Sparse Classification Method based on Convolutional Neural Networks. Here the CNN model was trained using the Moorea Labelled Dataset created from Heriot-Watt University's Atlantic Deep Sea dataset. In the preprocessing state, the Color Enhancement by stretching histograms and Hybrid patching is carried out. On the other hand, Feature Map based on Zero Component Analysis Whitening is proposed along with a Weber Local descriptor. Using normalized data, the input is given to the convolutional layer which yields an accuracy of 88%. Deep Convolutional Networks requires huge resources and is not a very good solution to real time scenarios that works directly with an ROV. This can be overcome by using a ROV detection from sonar images and using a CNN to classify the images. It helps to be accurate as well as real time in nature. Even the proposed object detection is faster with respect to other methods. Here, the Cyclopes UAV provides all the images necessary where the process is further speeded up, more than the Sparse Classification Method based on CNN.

A real time ROV detection from sonar images using Convolutional Neural Network [13] is proposed where mecha- nism for agent vehicle localization is proposed. A Conventional Neural Network is used for classification, where the input images are used as sonar images. The fast object detection makes the system more real time. While using the Autonomous Underwater Vehicle "Cyclops", the overall processing speed of the method using the YOLO Algorithm increased by 107.7 fps.

Other similar researches included by Dalal AL-Alimi et al. [21] who proposed a method using Mask R-CNN for Geospatial Object Detection where objects outlines are detected at the pixel level. The dataset Seg-VHR-10 has been used to train the model, which is easier to segment and further detect object from Geospatial Images. Rafflesia Khan Rameswar Debnath in [22] where an efficient approach to detect fruits using a proposed Fruit Detection and Recognition Method (FDR) based on an improved Convolutional Neural Networks. Fei Cai et al. in [23] to detect man-made objects based on Texture Clustering and Geometric Structure Feature Extraction. The method helps to detect any man-made object in dynamic environment and any man-made objects like vehicles.

3. Proposed Research Methodology

As proposed in the research framework, at first the dataset is obtained from the source, where the data was raw images of various fish species. The first step involves rearranging the dataset, which is restructuring the images as per needed for classification. In case of this research the images are extracted as basic image formats and stored within a project folder maintaining a hierarchy that later helps to classify into classes. Data is then pre-processed and some major preprocessing steps such as data augmentation, image resizing and image denoising are carried out. A Convolutional Neural Net- work, CNN is then developed using the Keras framework based on TensorFlow. This CNN is then trained with the training set of the dataset. The model is then further evaluated using the test set of the dataset. The classification is carried out using a widely used classifier which is known as the SoftMax classifier. In order to evaluate the proposed model, an evaluation based on computation time and accuracy, which later helps to identify the efficiency of the deep learning model here proposed.
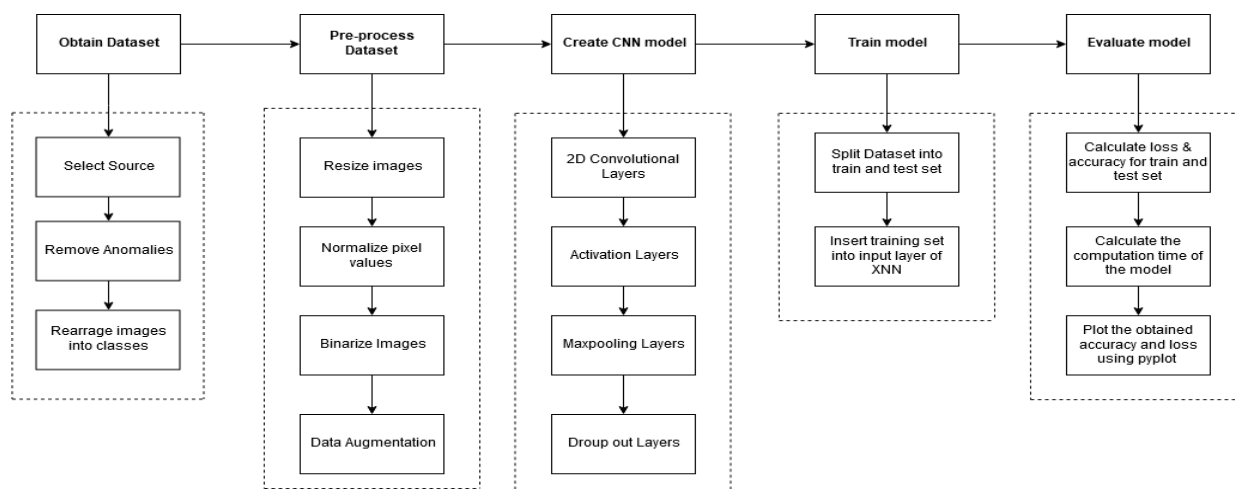

Fig. 2. Proposed Methodology

The input images are obtained from the IMGACLEF web- site, where the images of interest were only fishes. Image Cross language Evaluation Forum (ImageCLEF) is a plat- form that provides dataset for Cross-Language Evaluation. Here the dataset Fish_ts of the ImageCLEF was used. The images are first retrieved from the source. In the first step the raw images are archived where each image was positioned into folder based on the species name. The dataset was organized using the format dataset/species-name/fish-name. In the preprocessing state each input image was resized into 96x96 and pixel values are normalized by dividing with 255. The overall network will be iterated for 100 iterations from both directions with image dimension of 96x96x3 was main- tained. After this state, the data augmentation is carried out where the data is rotated, width_shifted, height_shifted, zoomed, horizontally flipped and filled based on the nearest value. Due to data augmentation, data shortages are overcome; as a result the overall accuracy of the model increases.

The model solves all the challenges faced for underwater image detection. Firstly, the images are selected based on their quality and usability. The preprocessing state also involved step such as data augmentation which helps in over- coming data shortages and increases accuracy and efficiency of the model. A variant of CNN was used in the model, which is far simpler requiring far less time and achieving far more accuracy. Here the framework Keras is used which is both efficient and has huge support for deep learning models. The proposed method as a result provides a both efficient and accurate solution to the problem.

The proposed VGGNet is composed of 4 Convolutional Layers. Here, the CNN is consist of only 3x3 convolutional layers stacked on top of each other in increasing depth. Through max_pooling the volume is reduced and finally using the fully connected layer at the end of the network, the data is then inserted to a SoftMax classifier for detection. Here, the implementation is done using the TensorFlow backend, with 32 filters and a 3x3 kernel. Here, the pool layer uses 3x3 pools to reduce spatial dimensions quickly from 96x96 to 32x32. To train, here 96x96x3 dimension images are used as inputs.

Here, VGGNet was used because of its uniform architecture and capability to accurately classify dynamic objects. Here, only 3×3 convolutions are used providing great results with the ImageCLEF dataset. Moreover, the weight configuration can be obtained easily from various sources. A smaller version of the VGGNet is proposed in this paper, which is even though small but provides a very robust nature with a high accuracy rate. In case of the problem at hand, fishes and other small living organisms have similar physical characteristics. Since images have uniform and similar pixel values, VGGNet poses a very good choice amidst of other similar CNN.

In the first input convolutional layer, input are inserted in 96x96x3 dimensions. Using an activation function alongside the convolutional later helps to learn more enriched features. As we go deeper into the network, the dimension of the images gets smaller thus become easier to classify. Using 2 convolutions and RELU activation function, a pooling layer is used. After the Dropout is then performed in both the first and the second layer. After 3 successful convolutional layers, the classifier is specified, which is SoftMax Classifier, which is a widely used classifier for deep learning models.
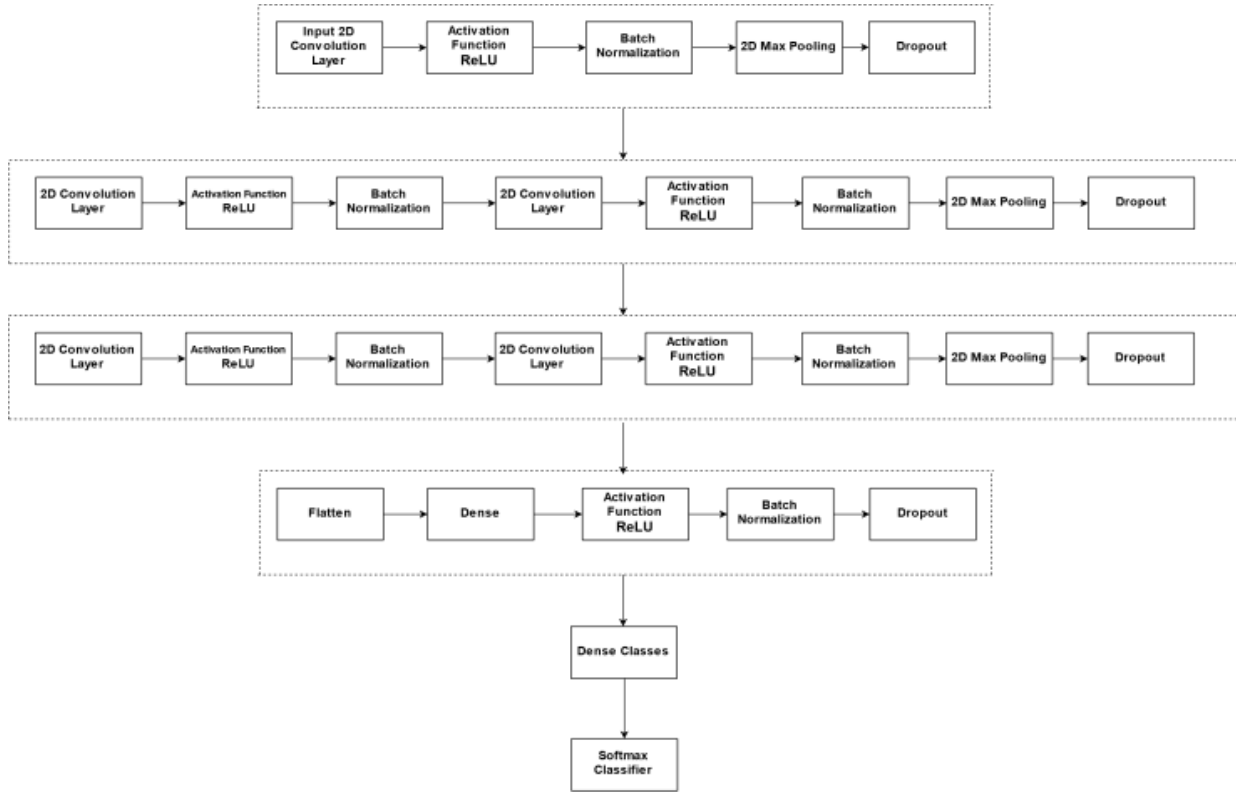

Fig. 3. Proposed CNN Model

The standard SoftMax function:

$$\sigma(s)_i = \frac{e^{2i}}{\sum_{j=1}^{k} e^{2j}} \qquad \text{for } j = 1, \ldots, K \qquad (1)$$

Here, the used SoftMax classifier which is similar to logistic Regression where it normalizes an input pixel value into a set of values that has a probability distribution of 1. So the output of the function is between 0 and 1.

In the final dropout, 50% of the nodes of training data are carried out, as a result results in most unnecessary bits of data to be removed. The model is then completed using the classifier that computes the probabilities of each class labels. The class with the highest probability will be considered the result of the classification. SoftMax classifier calculates probabilities of each class and normalizes the values in a range between 0 and 1. Here the probabilities of all the classes sum up to 1. As a result, SoftMax classifier is a very good choice for classification. The classifier is widely used with Convolutional Neural Networks because of all this features. Along with the SoftMax, a Cross Entropy Loss is also calculated which measures whether the prediction similarity with its respective label. All this provides a very strong classifier to be used alongside Convolutional Neural Networks.

After the SoftMax layer, the accuracy of the data is then carried out so that the overall accuracy can be obtained. At first, the loss is calculated using the Mean Squared Error for both the training and the validation set of the dataset. Mean Squared Error which is a quadratic loss function which denotes the extent of loss faced for prediction. The mathematical notation for Mean Squared Error (MSE) is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_n - \bar{y}_n)^2 \qquad (2)$$

Here, the MSE is calculated by finding the average of the squares of the errors, which is the average squared difference between the predicted and actual estimated values. The value of MSE always is above 0 due to the randomness of the data. The accuracy of the training and the validation set is calculated by finding the mean value. Here, accuracy is

a measurewhich is strictly the ratio of number of correct predictions tothe number of total input samples.

Then the run-time of each model from the start and of execution is calculated in order to identify speed of training and testing of the dataset. At the end, the model with the least Run time and Computation Time poses a good evaluation metrics value to be used for finding the most suitable model for the problem at hand.

4. Experimental Results and Discussions

A   Experimental Setup

The setup for deep learning requires huge bulk of hard- ware and requires greater memory space for image processing and rendering. For the deep learning methods, a computer with Intel i5 3470 with 16 GB DDR3 RAM is selected. The software associated with the experiments include Anaconda that is used to set the complete environment of Python

B   Dataset

The IMAGECLEF dataset [19] was first used to train the dataset. The IMAGECLEF dataset composed of only 250 images as a result yielded very poor performance while training with the proposed model. For the IMAGECLEF dataset, the following result is obtained that provides us the flow of train and test loss and fall of train and test accuracy. The dataset used to train the model was IMAGECLEF Fish dataset [19]. Through data augmentation and several pre processes more accuracy is obtained. Unlike some complex algorithms used for object detection, the proposed method is both simple and fast, whereas a efficient dataset organization in the research is also proposed. Several other approaches are then compared with the prosed method at the ending part of the paper and the corresponding result is put forward for evaluation. WHOI-Plankton [7], Moorea Labelled Coral (MLC) dataset [13] and MARBEC dataset [16] is then used to evaluate the proposed framework and each of the dataset implemented along with this model is observed. The overall result is then tabulated. The long run time is due to the fact of using CPU instead of GPU that would have boosted the runtime. Some images that are considered to be used for evaluation are stored in a separate folder named examples, which is then classified using the classify.py. The classification result for some fishes is given in Table 1. The accuracy along with the detected class is then given as output for each corresponding fish. The output is generated using the matplotlib.pylot that provides the detected image along with theaccuracy rate as output.

Table 1. Accuracy And Computation Using Proposed Method For Various Underwater Target Objects FromVarious Datasets.

Dataset	Target	Accuracy	Computation Time
WHOI-Plankton [7]	Plankton	88	18
ImageCLEF_Fish_TS [19]	Fish	88	18
Moorea Labelled Coral (MLC)[3]	Coral	90	38
MARBEC database [5]	Coral Reef Fish	89	30

C   Experimental results

The IMAGECLEF dataset was first used to train the Neural Network model. The IMAGECLEF dataset is composed of only 250 images as a result yielded very poor performance while training with the proposed model. For the IMAGE- CLEF dataset, the following result is obtained that provides us the flow of train and test loss and fall of train and test accuracy before any sort of data augmentation. After Data Augmentation the accuracy rate of the overall prediction is above 94% which is due to the variety of data forms. Data Augmentation greatly increases the overall accuracy of the implementation. The gradual fall of the accuracy rate is then taken into account, and a gradual increase of the accuracy result is obtained. The dataset yielded a run time of 25 minutes with an accuracy of 85%. The long run time is due to the factof using CPU instead of GPU that would have boosted the runtime.

Here it is observed, with the increase of EPOCH, the over- all accuracy of detection increases to the highest rate of 94%. Several datasets used in the proposed network are put forward. Here at first, WHOI Plankton dataset is used to evaluate the model, where an accuracy of 88% is obtained, due to huge redundancy of data. The IMAGECLEF dataset yielded the best result, which is almost 95%. The Moorea Labelled dataset to detect Coral required a run time of 38 minutes with an accuracy rate of 90% which was great. The MARBEC dataset, to detect coral reef fishes were then proposed that provided an accuracy of 89% requiring 31 minutes. All the above datasets provided very decent results on the proposedmethod of Simplified Visual Geometry Group Network (VG- GNet).

The classification result for some fishes is given in Figure3. The accuracy along with the detected class is then given as output for each corresponding fish. The output is generated using the matplotlib.pylot that provides the detected image along with the accuracy rate as output. Here it is observed, with the increase of number of EPOCH, the overall accuracy of object detection increases to the highest rate of 94%.
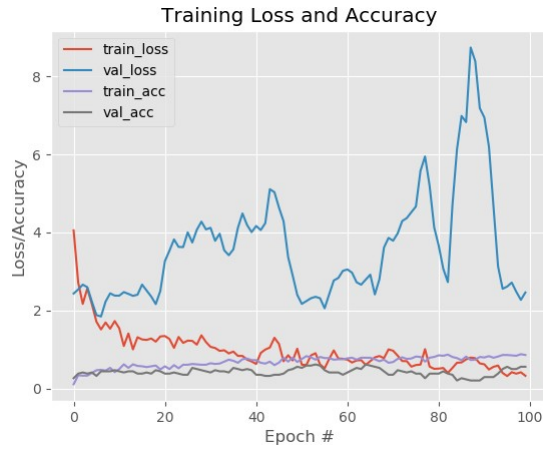
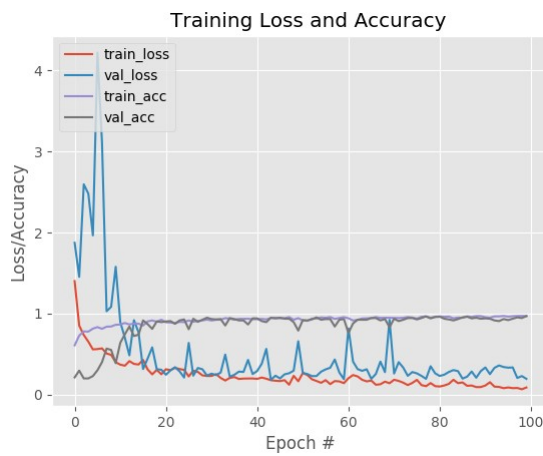Fig. 4. Experimental Result prior Data Augmentation

Fig. 5. Experimental Result after Data Augmentation

Fig. 6. Detection results of some input fishes

The methods used in this experiment include CNN along with Transfer Learning. Here Transfer learning on top of a preexistent CNN provides a accurate and precise detection mechanism. On top of that CNN developed on top of a preexistent is easier as only training an existing model carries out the task. The method even though proved accurate failed to give acceptable detection time. CNN along with transfer learning provided an astonishing result of 92.8% accuracy and a Computation Time of 7 minutes. Even though a very accurate result is obtained but this process required huge resource as CNN along with transfer learning requires huge system resource. Previously it was found, images when used with transfer learning based on a previous dataset yielded great results but suffer greatly with computation time, where it is also the case. This property makes the process not at all efficient for huge datasets. Secondly, the Fast RCNN provided an accuracy of 81.4% and being faster than CNN with Transfer Learning. This is due to its property of

using region of interest of object rather than the whole image. Here, for this dataset even though the time is efficient, but the accuracy rate falls in a wider margin. As a result, this method lacked the detection accuracy which is of prime importance. The VGGNet, which was the inspiration of the proposed model, provided an accuracy of 84.5% and computation time of 14 minutes. Here, the computation time is very huge, but requires less resources with more accuracy. Our proposed model overcame the computation time with a much higher accuracy rate. In CPU, the method obtained a computation time of 25 minutes which was very large yet very efficient with respect to CPU only. The most promisingimprovement to the research is the accuracy which is 94.6% above all the other methods.

Table 2. Comparison With Existing Methods From Various Data Sets.

Method Name	Reference	Accuracy	Computation Time
Fast r-CNN [19]	Plankton	88	18
VGGNet [3]	Fish	88	18
Proposed Method	Fish	89	30

D   Analysis and Discussion

Previous Methods such as CNN with Transfer learning provides an accuracy of 98.2% which was by far the highest. The Fast R-CNN provided an accuracy of 81.4% and VGGNet of 84.5%. The simplified proposed CNN based on VGGNet yielded an accuracy of 94.6%, thus correctly detects objects in the test dataset. The only downside to the proposed method is the run time, which even though great in a CPU still lags in that of other approaches that are done in GPUs.
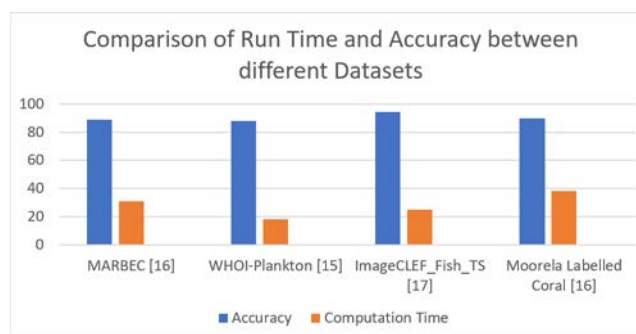

Fig. 7. Comparison between Computation Time and Accuracybetween datasets

Three other datasets are used to train our model. Where for the ImageCLEF_Fish_TS the highest accuracy and computation time is observed. As a result, the dataset was initially selected to train the model. The other datasets did not yield such promising results even though WHOI-Plankton provided the second-best accuracy and computation time which is almost even to ImageCLEF_Fish_TS dataset. Several other datasets are then used in the proposed network. Here at first, WHOI Plankton dataset is used to evaluate the model, where an accuracy of 88% is obtained, due to huge redundancy of data. The IMAGECLEF dataset yielded the best result, which is almost 95%. The Moorea Labelled dataset to detect Coral required a run time of 38 minutes with an accuracy rate of 90% which was really great. The MARBEC dataset, to detect coral reef fishes were then proposed that provided an accuracy of 89% requiring 31 minutes. All the above datasets provided very decent results on the proposed method of Simplified VGGNet.

Some of the most important contributions to the research includes using of a lightweight deep learning model to train and test images properly and accurately. Then obtaining an accuracy of 96.4% in detection of underwater objects mainly fishes. All the achievement provides a detection mechanism that depends on VGGNET, yet faster and more accurate. Here, the data augmentation is put forward as one of the prime enhancing method for deep learning methods. The proposed method that will work accurately for underwater organisms, which is the cornerstone to this research. All the above factors provide a firm standing that data augmentation step poses a very important enhancement step for deep learning methods. Finally carrying out tests using latest methods for underwater object detection gives a very broad comparison field and a complete picture for the overall condition of the research. The result obtained from the research helps to confirm the accuracy of the model, which was mainly the prime research objective, to obtain a accurate and robust deep learning model.

5. Conclusion

Through the research, an optimum solution to the current state of research of underwater imagery for using in Autonomous Underwater Vehicle (AUV) is addressed, and a method is proposed that resulted an accuracy of 94.6%. Thus this solution provides a good method to 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 is obtained showing that the proposed method overcomes the other approaches in a large margin. The problems associated with dataset can be addressed with data augmentations, data modification through segmentation, 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 are fast even though a bit difficult to manage in real life. This is a huge challenge for underwater imaging as moving objects in dynamic environment are difficult to process and takes huge time. The steps associated with object detection also involve steps that are resource intensive. To modify all the above states the paper tends to provide all possible solution. All this leads to a solution that the above problem can be solved using deep learning. Future works involved are, deep learning method development for dynamic environment which is robust as well as fast for accurate object detection, can easily be integrated in an AUV. Currently the VGGNet is being used to detect several types of underwater organisms. The proposed VGGNet will be able to enhance the existing VGGNet and make it more sophisticated for fish detection. In future the deep learning model may be used to detect other similar objects and their habitat to preserve from extinction. Making the whole process of underwater object detection reliable as well as simple, so that the overall detection tasks become faster and powerful as well. More over until now, the research is based on a IMAGECLEF_FISH_TS dataset, which is only composed of images of fishes. The method used for training the model is VGGNET which is based on ALEXNET. VGGNET being very powerful but being too bulky posed a threat to the speed of the identification process. To build a faster and accurate method the small VG- GNET is proposed that will go through an ongoing process of improvement. Even though the research until now posed a very positive impact on the underwater object detection, but still some very important limitations exists. The most important limitations being used upon CPU, which gave poor computation time and using a smaller dataset for training. Even though an accuracy of 96.4% is reached, but it is only possible for fishes, but for more complex organisms with poor or unclear dataset, this method will not provide an optimum result. Thus, the research will also focus on a generic method that will help in detecting underwater organisms and objects of all shapes, textures and sizes accurately. In the midst of all the limitations, still it can be stated that Convolutional Neural Networks (CNN) which are used widely appreciated for computer vision models and classification in complex environments, seemed the perfect solution for underwater object detection.

References

[1] Nicole Seese, Andrew Myers, Kaleb Smith, and An- thony O Smith. Adaptive foreground extraction for deep fish classification. In *2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*, pages 19–24. IEEE, 2016.

[2] Yujie Li, Huimin Lu, Jianru Li, Xin Li, Yun Li, and Seiichi Serikawa. Underwater image de-scattering and classification by deep neural network. *Computers & Electrical Engineering*, 54:68–77, 2016.

[3] Ammar Mahmood, Mohammed Bennamoun, Senjian An, Ferdous Sohel, Farid Boussaid, Renae Hovey, Gary Kendrick, and Robert B Fisher. Coral classification with hybrid feature representations. In *2016 IEEE In- ternational Conference on Image Processing (ICIP)*, pages 519–523. IEEE, 2016.

[4] Hailing Zhou, Lyndon Llewellyn, Lei Wei, Doug Creighton, and Saeid Nahavandi. Marine object detec- tion using background modelling and blob analysis. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 430–435. IEEE, 2015.

[5] Sébastien Villon, Marc Chaumont, Gérard Subsol, Sébastien Villéger, Thomas Claverie, and David Mouil- lot. Coral reef fish detection and recognition in un- derwater videos by supervised machine learning: Com- parison between deep learning and hog+ svm methods. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 160–171. Springer, 2016.

[6] Jialun Dai, Ruchen Wang, Haiyong Zheng, Guangrong Ji, and Xiaoyan Qiao. Zooplanktonet: Deep con- volutional network for zooplankton classification. In *OCEANS 2016-Shanghai*, pages 1–6. IEEE, 2016.

[7] Hansang Lee, Minseok Park, and Junmo Kim. Plank- ton classification on imbalanced large scale database via convolutional neural networks with transfer learn- ing. In *2016 IEEE international conference on image processing (ICIP)*, pages 3713–3717. IEEE, 2016.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recogni- tion. *arXiv preprint arXiv:1409.1556*, 2014.

[9] Zu Yan, Jie Ma, Jinwen Tian, Hai Liu, Jingang Yu, and Yun Zhang. A gravity gradient differential ratio method for underwater object detection. *IEEE Geoscience and Remote Sensing Letters*, 11(4):833–837, 2013.

[10] Byeongjin Kim and Son-Cheol Yu. Imaging sonar based real-time underwater object detection utilizing adaboost method. In *2017 IEEE Underwater Technol- ogy (UT)*, pages 1–5. IEEE, 2017.

[11] Aneta Nikolovska. Auv based flushed and buried object detection. In *OCEANS 2015-Genova*, pages 1–5. IEEE, 2015.

[12] Srikanth Vasamsetti, Supriya Setia, Neerja Mittal, Har- ish K Sardana, and Geetanjali Babbar. Automatic un- derwater moving object detection using multi-feature integration framework in complex backgrounds. *IET Computer Vision*, 12(6):770–778, 2018.

[13] Oscar Beijbom, Peter J Edmunds, David I Kline, B Greg Mitchell, and David Kriegman. Automated an- notation of coral reef survey images. In *2012 IEEE Conference on Computer Vision and Pattern Recogni-tion*, pages 1170–1177. IEEE, 2012.

[14] Yafei Zhu, Lin Chang, Jialun Dai, Haiyong Zheng, and Bing Zheng. Automatic object detection and segmen- tation from underwater images via saliency-based re- gion merging. In *OCEANS 2016-Shanghai*, pages 1–4.IEEE, 2016.

[15] Donghwa Lee, Gonyop Kim, Donghoon Kim, Hyun Myung, and Hyun-Taek Choi. Vision-based object de-tection and tracking for autonomous navigation of underwater robots. *Ocean Engineering*, 48:59–68, 2012.

[16] Hongkun Liu, Jialun Dai, Ruchen Wang, Haiyong Zheng, and Bing Zheng. Combining background sub- traction and three-frame difference to detect moving object from underwater video. In *OCEANS 2016-Shanghai*, pages 1–5. IEEE, 2016.

[17] Guo-Jia Hou, Xin Luan, Da-Lei Song, and Xue-Yan Ma. Underwater man-made object recognition on the basis of color and shape features. *Journal of Coastal Research*, 32(5):1135–1141, 2015.

[18] Md Moniruzzaman, Syed Mohammed Shamsul Islam, Mohammed Bennamoun, and Paul Lavery. Deep learn- ing on underwater marine object detection: A survey. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 150–160. Springer, 2017.

[19] Xiu Li, Min Shang, Hongwei Qin, and Liansheng Chen. Fast accurate fish detection and recognition of underwater images with fast r-cnn. In *OCEANS 2015- MTS/IEEE Washington*, pages 1–5. IEEE, 2015.

[20] Mohamed Elawady. Sparse coral classification using deep convolutional neural networks. *arXiv preprint arXiv:1511.09067*, 2015.

[21] Dalal AL-Alimi, Yuxiang Shao, Ahamed Alalimi, Ahmed Abdu, "Mask R-CNN for Geospatial Object Detection", *International Journal of Information Technology and Computer Science,* Vol.12, No.5, pp.63-72, 2020.

[22] Rafflesia Khan, Rameswar Debnath, " Multi Class Fruit Classification Using Efficient Object Detection and Recognition Techniques", *International Journal of Image, Graphics and Signal Processing*, Vol.11, No.8, pp. 1-18, 2019.

[23] Fei Cai, Honghui Chen, Jianwei Ma, "Man-made Object Detection Based on Texture Clustering and Geometric Structure Feature Extracting", *International Journal of Information Technology and Computer Science*, vol.3, no.2, pp.9-16, 2011.

Authors' Profiles

Dipta Gomes received his undergraduate in Computer Science at American Inter- national University-Bangladesh (AIUB). Then he completed his Masters in Intelligent Systems at AIUB in 2019. Most of his current and ongoing contributions are in the fields of Machine Learning, Computer Vision and Algorithms. Currently working as a Lecturer at the department of Computer Science, Amer- ican International University-Bangladesh (AIUB). His research research interests include Artificial Intelligence, Computer Vision, Image Processing, Pattern Recognition and Machine Learning.

A.F.M. Saifuddin Saif received PhD from Center For Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, University Kebangsaan Malaysia (UKM) - The National University of Malaysia, M.Sc. in Computer System Engineering (Software System), School of Architecture,Computing and Engineering (ACE) from University of East London, UK and B.Sc. (Eng.) in Computer Science and Engineering , Applied Sciences Technology, Shahjalal University of Science and Technology, Bangladesh. Most of his current and previous contributions are in Artificial Intel- ligence, Computer Vision, Pattern Recognition, Image Processing research domains. Currently serving as Senior Assistant Professor, department of Computer Science, American International University – Bangladesh. Before joining the university, completed his Post Doctorate at MyXLab (Mixed Reality and Pervasive Lab), Faculty of Information Science and Technology, University Kebangsaan Malaysia. His research research interests include Artificial Intelligence, Computer Vision, Image Processing, Pattern Recognition.