Journal of
Engineering Sciences
Faculty of Engineering
Assiut University
JES
مجلة
العلوم الهندسية
كلية الهندسة
جامعة أسيوط
journal homepage: http://jesaun.journals.ekb.eg

Identification of Unknown Marine Debris by ROVs Using Deep Learning and Different Convolutional Neural Network Structures

Received 21 November 2023; Revised 11 December 2023; Accepted 11 December 2023

Mahmoud. Assem[1]
Ibrahim. M. Hassab-Allah[2]
Mohamed. E.H. Eltaib[3]
Mahmoud Abdelrahim[4]

Keywords
Remotely operated vehicles, Marine debris, Deep learning, Convolutional neural network, Image processing.

Abstract
We study the problem of underwater debris classification and removal by remotely operated vehicles. This task is particularly important for subsea oil and gas fields exploitation. The classification of underwater debris is a challenging and difficult problem because of the complexity of underwater environments. We investigate four different algorithms based on deep convolutional neural networks for detecting and classifying marine debris. The proposed techniques are built on Keras and TensorFlow using Python programming environment. To train the algorithm for detection, various dataset information containing different types of marine debris have been established. Four distinct classifier and activation function combinations have been compared experimentally. The dataset is consisting of fifteen categories. The suggested approach is a modified VGGNet-16 trained on the dataset. The use of a sigmoid classifier and the Relu activation function to categories marine improves classification accuracy. The overall result indicates that classification accuracy on the testing set is satisfactory.

1. Introduction

The exploration of oil and gas fields is crucial for many offshore applications to identify and remove marine debris. The task is usually carried out by a remotely operated vehicle (ROV) guided by a pilot by means of communication networks. The ROV is often equipped with a robot manipulator to perform certain subsea tasks such as debris removal. However, before the pilot can instruct the ROV to remove some debris, the object to be removed need to be clearly identified in terms of size, material, and estimated weight to ensure safe operation of the ROV. Hence, efficient detection algorithms need to be developed to assist the ROV identifying marine debris using its own vision capabilities. The traditional machine vision method to recognize unknown objects is commonly known as pattern recognition [1]. This approach consists of three phases in the traditional picture recognition approach. The proposed region (Region Proposal) is split into the original picture first,

[1] M.Sc. Graduate student, Mechatronics Engineering Dept., Faculty of Engineering, Assiut University, Assiut, Egypt. M.Assem690@gmail.com
[2] Professor, Mechanical Engineering Dept., Faculty of Engineering, Assiut University, Assiut, Egypt. Ibrahim.abdeldaiam@eng.au.edu.eg
[3] Assoc. Professor, Mechanical Engineering Dept., Faculty of Engineering, Kafrelsheikh University, Egypt. Mehaeltaib@hotmail.com
[4] Assoc. Professor, Mechatronics Engineering Dept., Faculty of Engineering, Assiut University, Assiut, Egypt.. M.abdelrahim@aun.edu.eg

https://doi.org/ 10.21608/jesaun.2023.250095.1289

and then the feature is retrieved from the region. Finally, the area is identified using the trained model. Images in the underwater environment were hazy and were heavily impacted by factors such as perspective fluctuation, distortion, lighting conditions, and background clutter. As a result, the conventional method's categorization and recognition precision was flawed. These issues can be solved using Machine Learning and deep convolutional neural networks (CNNs). The convolutional neural network (CNN) is a kind of neural network that is widely utilized in image categorization and machine vision systems nowadays. Based on shared-weights architecture and translation invariance qualities, it is an enhanced multilayer perceptron aimed to increase classification or prediction accuracy. It's been widely employed in Audio classification [2], medical diagnosis [3], pedestrian detection [4,5], natural language processing [6], and a variety of other domains in recent years. Convolution neural network (CNN)-based image identification techniques can perform more precisely than human eyes when the clarity and contrast of the vision are lacking. [7]. Convolutional, pooling, and fully linked layers are the three fundamental components of CNN architecture. Convolutional kernels are slid across all spatial points with a predetermined stride to create feature maps, which are then produced by computing the dot products between the kernel weights and tiny local patches in the input volume. The output of every convolutional kernel is a single feature map. Higher convolutional layers encode more complex, higher-level information, whereas lower convolutional layers extract general properties like lines and edges. Nonlinear activation function, commonly known as a rectified linear unit (ReLU). [8] is used to introduce nonlinearity into produced feature maps, which is useful for detecting nonlinear features [9]. Pooling layers reduce the spatial size of the input by replacing local patches in input feature maps with their maximum or mean value to give invariance to small shifts and distortions. Some CNN designs employ extra fully connected layers on top of the stacked convolutional and pooling layers before the final SoftMax layer to achieve high-level reasoning [10,11]. Before transforming the picture's pixel value to an internal representation that allows the classifier to recognize patterns in the input image, the computer is trained on massive image datasets [9]. Overfitting is a common cause of poor performance in deep learning image categorization. To optimize performance and prevent overfitting a large dataset and model are used. Because CNN models have fewer connections and hyper parameters, they are easier to train and perform somewhat worse than other models [1]. Tensorflow is one of the deep learning libraries for image classification. Tensorflow is a Google-created open-source software library for numerical computation that was released in 2015. Keras is a Python-based open-source neural network library. Deep learning, especially Convolutional Neural Networks, is particularly popular for object categorization, segmentation, and detection. Even though deep learning algorithms have made significant development in this field, methods still need to be developed for more dynamic environments, such as underwater environments. Some of the latest research on underwater object recognition based on deep learning techniques include Adaptive Foreground Extraction using CNN [12], VGG Net [13], and CNN with Transfer Learning based fine grained classification approach [14]. The previous deep learning algorithms utilized for this specific purpose of underwater objects recognition are shown in Fig .1. Despite the numerous challenges encountered underwater, the following deep learning algorithms proved to be an excellent answer for tough categorization tasks. Several studies are being conducted using deep learning algorithms to detect underwater objects. Fast R-CNN in [12], proposed a unique fish detection approach. The total study aided in the development of a massive new dataset of 24272 photos divided into 12 classes. When 2000 regions of interest (ROI) gathered from selective search are fed into the network input, it takes a long time. Even though the process is quick, it is not in real time. Using the open source Caffee CNN deep learning framework created by Berkeley AI Research (BAIR), a modified version of AlexNet, a Convolutional Neural

Network model named after its author Alex Krizhevsky [15], is used. It has five convolutional layers and three fully connected layers. Singular Value decomposition (SVD) is used to compress completely connected layers, which speeds up the identification process even further. Horizontally flipped photos are ignored during sampling, and data augmentation is performed. It converts video frames into images and therefore does not work in real time. As a result. The main contribution of this study was to: (1) develop the model for underwater debris identification and classification of different types of marine debris by the ROV single camera; (2) Create a dataset where more than 250 images of marine debris were used to categories. (3) Use a combination of various classifiers and activation functions, including softmax, sigmoid classifiers and Relu, Tanh activation functions, four distinct CNN architectures are compared on a MAC OS system. Tensorflow and the Keras framework are used for computation and processing.

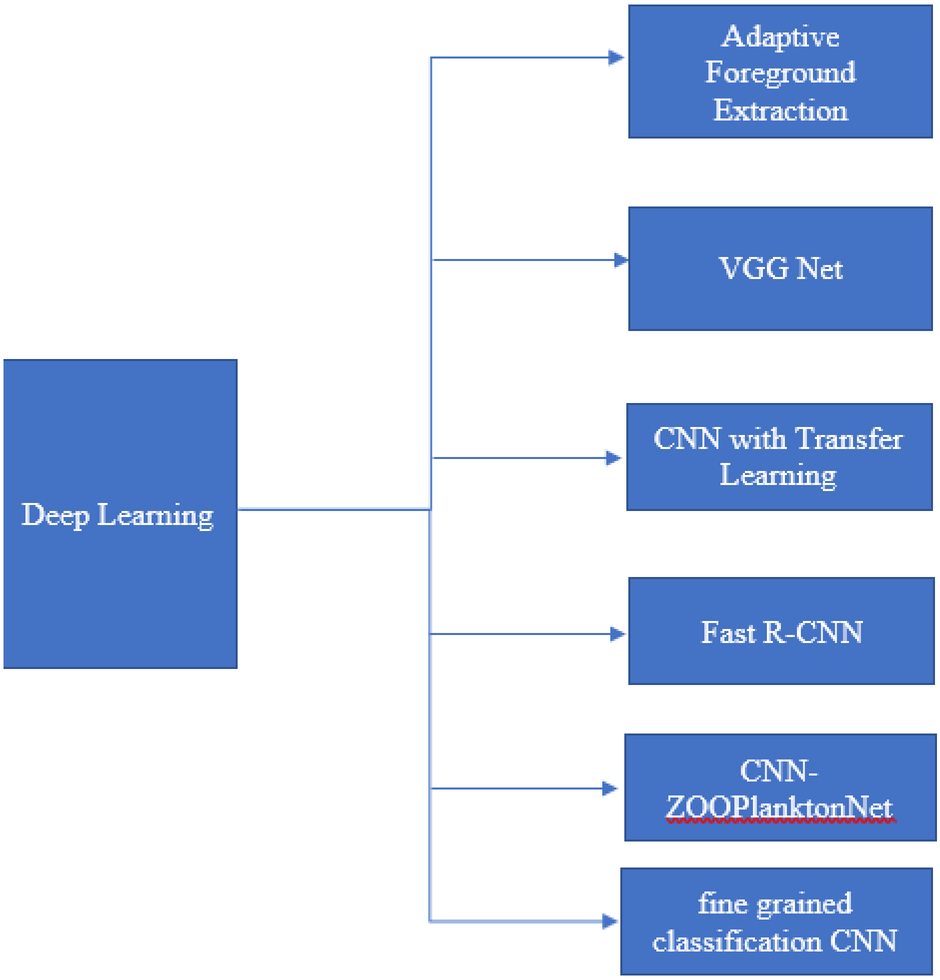


Fig.1: Deep Learning Methods

The rest of the paper is structured as follows. Section 2 describes definition of the problem. Section 3 Literature Review. Section 4 Methodology. Section 5 gives implementation details and experimental settings. Experimental results and discussion are presented in Section 6. In Section 7 concluding remarks and directions for future work are given in section 8. framework is used for computation and processing.

2. Setup Description

ROV Software and Hardware.
Vehicle contains several systems and subsystems as shown in fig.2:

1. Thrusters.
2. Control system:
3. Power supply and communication system
4. Carrier system:
   - Frame.
   - Housing for electronics.
   - Buoyancy.
5. Cameras
6. Manipulators.
7. Lights
8. Sensors

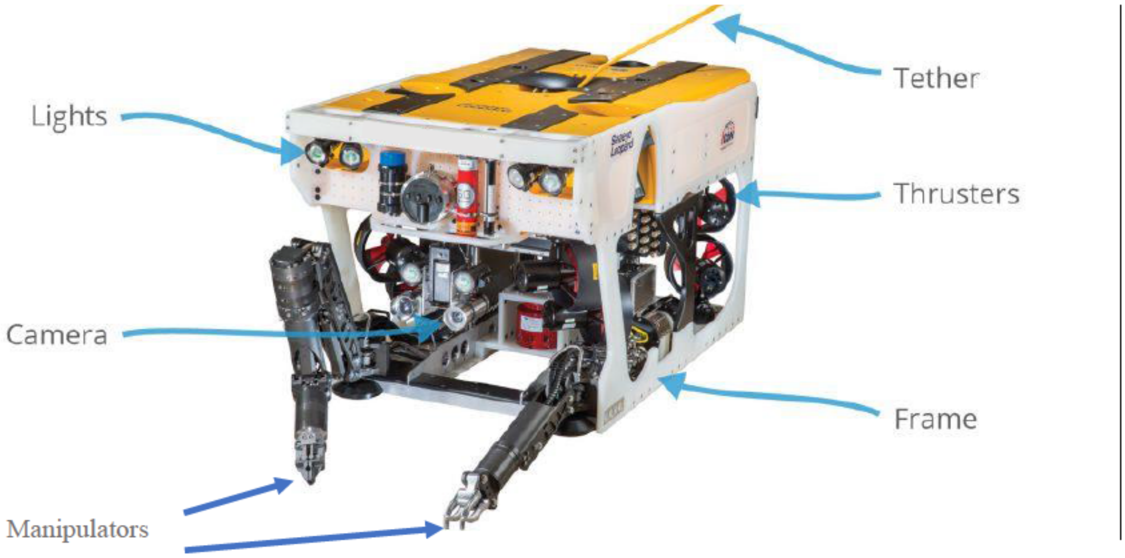


Fig.2: ROV Vehicle components

Thrusters, to move the vehicle, the thrusters use propellers that are propelled by electricity or hydraulics. To allow for movement in different directions, there are typically several thrusters present. Camera, because the vehicle goes deep underwater, the pilot's only vision is through the onboard camera, which must offer a low-latency image. Lights, the underwater camera is illuminated by the lights. Underwater, sunlight is quickly lost, and many ROV operations take place at depths where it is ordinarily completely dark. Tether, for the pilot to operate the ROV and observe the camera, nearly all of them include a tether that sends electrical power and/or communications to the surface. The ability of tethered systems using fiber optics to transfer massive amounts of data, including live video, over incredibly long distances with very little delay, is a huge benefit. Modern deep-rated ROVs, tow sleds, and sonar systems (>1000 m cable length) nearly always transmit data through fiber optics and generally use steel-armored cables with copper conductors for power transmission. Manipulators.
The subsea manipulator is a crucial component of the ROV, doing duties including collecting biological samples, removing debris, subsea intervention, mating or demitting underwater connectors, valves operation, installing hot stab, and other similar tasks. ROVs with manipulator

systems and a variety of sensors—typically HD video and sonar—are increasingly employed to coordinate underwater activities. Fig. 3 displays a Titan 4 manipulator.

(a) Master Control Unit (b) TITAN 4 Manipulator

Fig.3: Schilling T4 Manipulator seven Degree of Freedom

Pilot Controls, surface controls can be as simple as a smartphone or as complex as something that like the control panel of a spaceship. In any case, the surface controls give the pilot a real way to operate the vehicle and a way to see feedback from it, including the camera view. as depicted in Fig.4

Fig.4: ROV control room with multiple displays for monitoring vehicle subsea operations

ROV Control, This ROV's control system is separated into two sub-control systems: the underwater ROV body control system and the surface console control system. Umbilical cable connects the two control systems, allowing power and signal transmission. The control system architecture is depicted in Fig. 3, where the surface console—which has a joystick, control buttons, monitor, manipulator hand control unit, and SCC (Surface Control Computer) among other things—realizes the display of communication data and the issuance of control commands. The underwater ROV body control system receives the data from the control panel. To manage the propellers, under lights, manipulator, camera, and other sections of the ROV body as well as to gather data from the sensors and deliver it to the onshore monitor for operators to view, the ROV body employs an embedded microprocessor. The ROV control scheme is shown in Fig. 5.

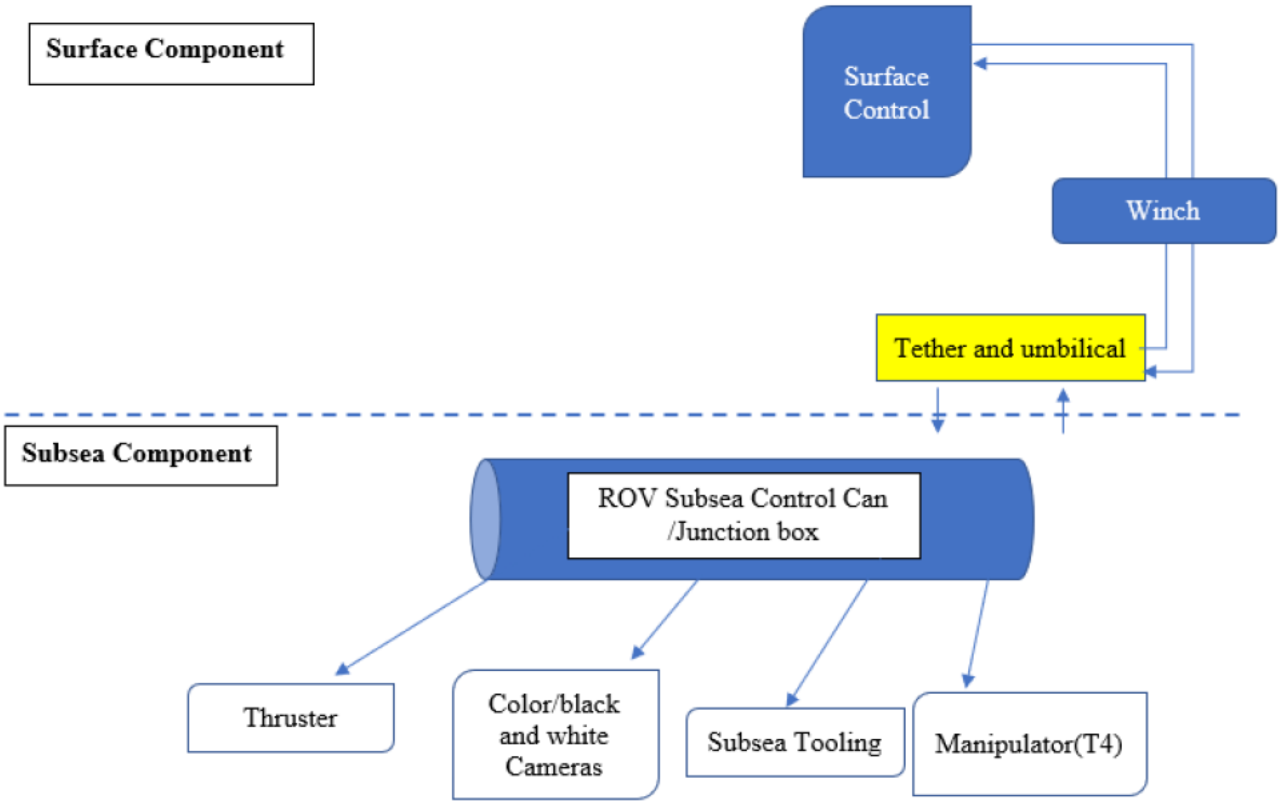


Fig.5: ROV control schematic drawing.

Despite the wide variety of ROV controllers available, the majority of industrial underwater robots employ proportional integral derivative (PID) controllers because of their straightforward design and performance in certain situations. PID-like controllers typically function well. Fig. 6 displays the block diagram of the PID control technique for fixed navigation with a double closed loop [9].

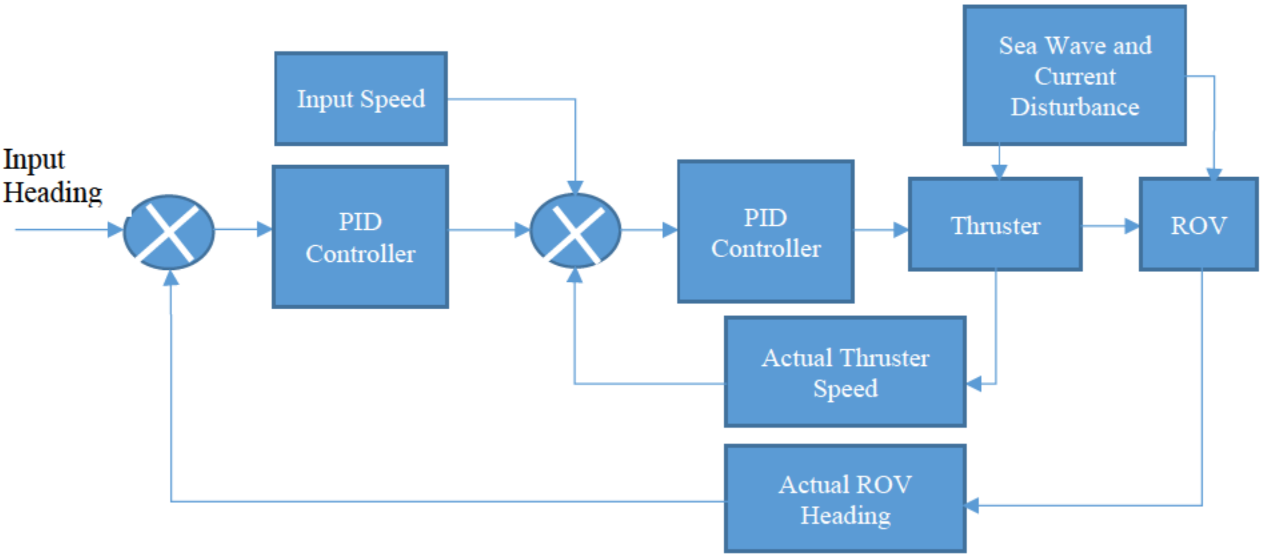
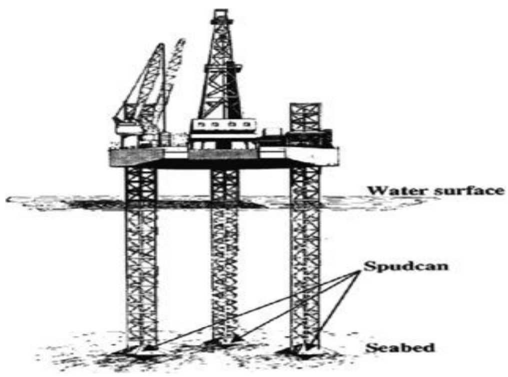


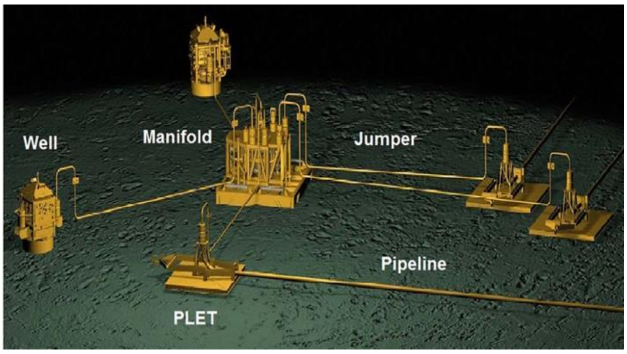
Fig.6: Speed and heading control model.

3. Definition of the problem

The existing architecture and control systems have restrictions that limit awareness and interactivity, allowing ROV actions to deviate from the operator's goal. Expert pilots have enough expertise to successfully synthesis data from various screens and direct vehicle and manipulator systems, but mistakes still happen even when operations are carried out systematically and crucial tasks are carried out slowly. As the demand for services grows, providers (businesses hired to do work overseas) are pushed to push systems to their boundaries. As a result, there is a growing tendency toward the usage of ROVs and technical advancements to improve safety and efficiency. One of the most important tasks of ROV in the offshore field is debris removal operation and seabed survey. Before standing jack-up rig on position, the seabed survey shall be conducted to ensure site is free of debris. Jack – up rig leg penetrates seabed and hard debris will break the spud-can as shown in Fig.7 (a) [13,14]. Also installing subsea structure required pre-installation survey to ensure that seabed is clear as shown in Fig.7 (b) [17]. In debris removal operation, the pilot must know the debris type to determine whether the debris can be carried by manipulator. Because the ROV manipulator can only lift a certain amount of weight, the pilot must determine whether to use it to remove the debris. For example, before installing subsea construction, area must be clear without any piece of debris. Also, jack-up rig before stand on position, area must be clear. Fig.7. shows juck-up rig, and subsea construction.



(a) Jack-up rig



(b) Subsea structure

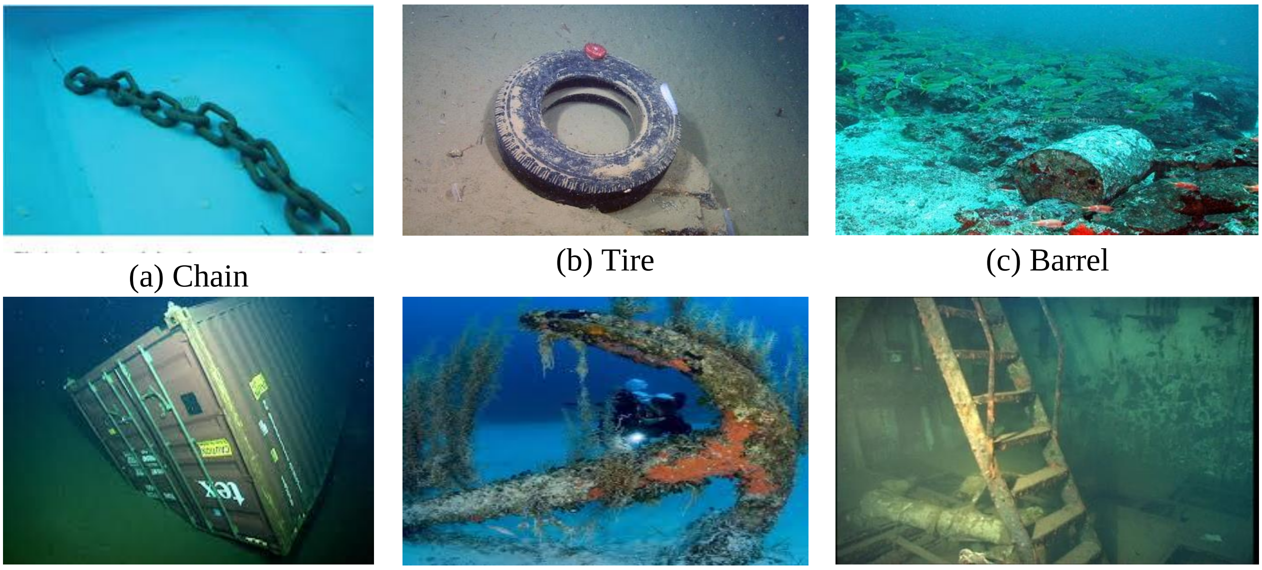
Fig 7: (a)Jack-up rig [17]. (b) Subsea structure [17].

4. Materials and Methods

4.1 Dataset

A large, annotated dataset of underwater debris is needed to utilize a deep-learning approach for marine debris detection and classification. We created a dataset with twelve different categories. Each category included more than 20 images.

The dataset contains different classes: plastic bottle, plastic _bag, metal_ship_thruster, metal_chan, metal_ anchor, metal_barrel, metal_container, metal_ladder, ceramic_toilet, tire, rope, glass bottle. To guarantee that the distribution of considered classes in the training and test sets is the same, available data is split into a training set and a test set as follows: The final model evaluation was done with 15% of the pictures from each class, while the remaining 85% were used for training. Fig.8 shows images of dataset.



(a) Chain (b) Tire (c) Barrel

(d) Container (e)Anchor (f) Ladder

Fig.8: Example of debris dataset

4.2 Deep Convolutional Neural network

The first stage is to restructure the dataset, which entails reorganizing the pictures as needed for classification. In the context of this study, the photos are extracted as basic image formats and saved in a project folder with a structure that subsequently aids in classifying them.

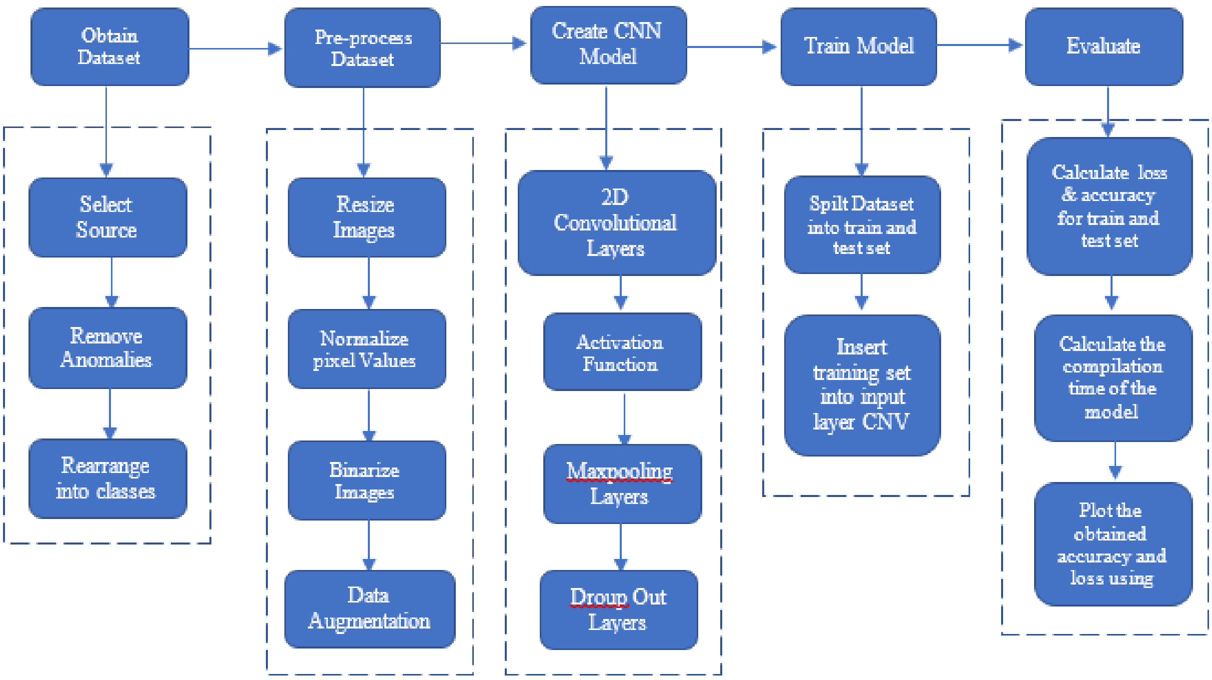


Fig.9: Proposed Methodology [18]

The training set for the dataset is then used to train this CNN. The test set for the dataset is then used to validate the model. The SoftMax classifier, a widely used classifier, is used to carry out the categorization. The effectiveness of the deep learning model presented here is then assessed using an evaluation based on computation time and accuracy, which is done to evaluate the suggested model, as depicted in Fig. 9.

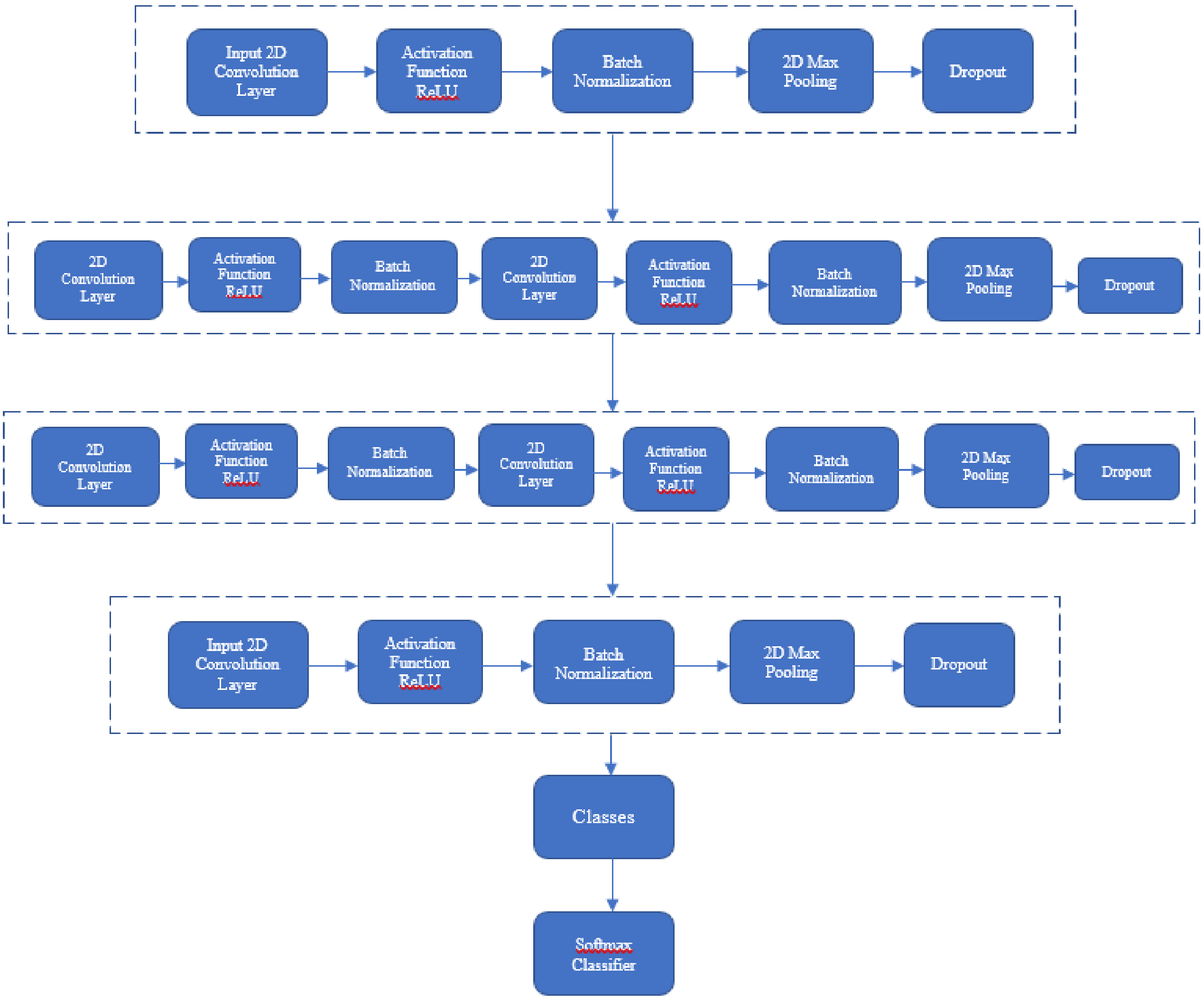


Fig.10: Proposed CNN Model [18]

The input images are obtained from the dataset. The images are obtained from the source initially. The raw pictures are preserved in the first phase, with each image placed in a folder depending on the debris name. The dataset was structured using the dataset/debris -name format. Each input picture was scaled to 96x96 in the preprocessing step, and pixel values were normalized by dividing by 255. With a 96x96x3-pixel image, the entire network will be iterated 100 times in each direction. The data is then rotated, width-shifted, height-shifted, zoomed, horizontally flipped, and filled based on the closest value. As a result of data augmentation, data shortages are addressed, and the model's overall accuracy rises. The model overcomes the challenges of underwater image detection. First, the photographs are selected based on their quality and utility. The preprocessing stage also includes measures like data augmentation, which helps overcome data shortages and improves the model's accuracy and efficiency. A variant of CNN was used in the model that is simpler, consumes less time, and produces greater accuracy. In this scenario, the Keras framework is used, which is both efficient and provides broad support for deep learning models. As a result, the offered approach provides an efficient and proper solution to the problem. The proposed VGGNet consists of four Convolutional Layers. The CNN in this case is made up of merely three 3x3 convolutional layers stacked on top of each other in increasing depth. The volume is decreased using max pooling, and the data is then fed into a SoftMax classifier for detection utilizing the fully connected layer at the network's conclusion. The approach in this case makes use of the TensorFlow backend, with 32 filters and a 3x3 kernel.

The pool layer here uses 3x3 pools to quickly decrease spatial dimensions from 96x96 to 32x32. In this case, 96x96x3 pictures are utilized as inputs to train. VGGNet was selected in this case because to its uniform architecture and ability to reliably categories dynamic objects. Only 3x3 convolutions are used in this case, resulting in excellent results with the dataset. Furthermore, the weight configuration can be easily retrieved from a variety of sources. Marine debris share similar physical features to the subject at hand. Because photos have homogeneous and identical pixel values, VGGNet is an excellent pick among similar CNNs. The input is injected in 96x96x3 dimensions in the first input convolutional layer. The addition of an activation function to the convolutional later aids in the learning of richer features. As we get deeper into the network, the image dimensions become smaller, making classification easier. A pooling layer is created using two convolutions and the RELU activation function. After that, the Dropout is applied to both the first and second layers. The classifier is described after three successful convolutional layers, and it is SoftMax Classifier, which is a widely used classifier for deep learning models.
The standard SoftMax function:

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

σ = softmax

z = input vector

e^{zi} = standard exponential function for input vector

K = number of classes in the multi-class classifier

e^{zj} = standard exponential function for output vector

The SoftMax classifier, which is similar to logistic regression in that it normalises an input pixel value into a collection of values with a probability distribution of 1, was employed in this case. As a result, the function's output ranges from 0 to 1. In the final dropout, 50 percent of the nodes of training data are processed, resulting in the removal of the majority of superfluous data. The classifier is then used to complete the model by computing the probabilities of each class label. The result of the categorization will be the class with the highest likelihood. The SoftMax classifier evaluates each class's probability and normalizes the results to a range of 0 to 1. The probabilities of all the classes add up to one in this case. As a result, the SoftMax classifier is an excellent classification option. Because of all these characteristics, the classifier is often employed with Convolutional Neural Networks. In addition to the SoftMax, a Cross Entropy Loss is calculated, which assesses the prediction's resemblance to its label. All of this results in an extremely powerful classifier that can be used in conjunction with Convolutional Neural Networks. Following the SoftMax layer, the data correctness is checked to achieve the overall accuracy. For both the training and validation sets of the dataset, the loss is initially estimated using the Mean Squared Error. Mean Squared Error (MSE) is a quadratic loss function that indicates the level of prediction loss. The following is the mathematical notation for Mean Squared Error (MSE):

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

y: actual label
ŷ: predicted value
n: number of classes
Consider a dataset with two classes (n = 2) and labels that are denoted by "0" and "1".

Now that the predicted values and actual labels are completely out of sync, we can compute the loss value and see how log-loss performs better than MSE.The MSE is calculated in this scenario by averaging the squares of the errors, which is the average squared difference between the expected and actual estimated values. The MSE value is always larger than zero because to the unpredictability of the data. The mean value is used to calculate the training and validation sets' accuracy. Accuracy is precisely defined in this context as the ratio of correct predictions to total number of input samples. The run-time of each model from start to end is then obtained to determine the speed with which the dataset is trained and tested.

5. Experimental results

Deep learning requires a large number of images and more memory space for image processing and rendering on MAC OS, which was chosen for the deep learning method. Anaconda, which is used to configure Python's entire environment, is among the software employed in the trials. The speed of training and testing of the dataset is then determined. First, the dataset was used to train the Neural Network model. The dataset contains over 250 photos, which resulted in low performance when training with the proposed model. The dataset yields the following result, which provides us with the flow of train and test loss and fall of train and test accuracy prior to any form of data augmentation. Because of the variety of data types, the overall forecast accuracy rate is greater than 94 percent after Data Augmentation. The use of data augmentation significantly improves the overall correctness of the implementation. The steady decrease in accuracy rate is then considered, and the accuracy result is gradually increased. Fig. 11 illustrates the outcomes of the pre-trained test set, which were 85% (0.85), 70% (0.70), and 86%. The technique clearly shows an improvement in accuracy. Then, in the real training and validation sets, respectively, we see a quick convergence of 97.73% and 100% (20/20) accuracy. The results of classification accuracy on the training set are 98.34% . And 82.91% accuracy on the testing set. The actual training and validation set then show a quick convergence of 97.73 percent and 100 percent (100/100) accuracy, respectively.

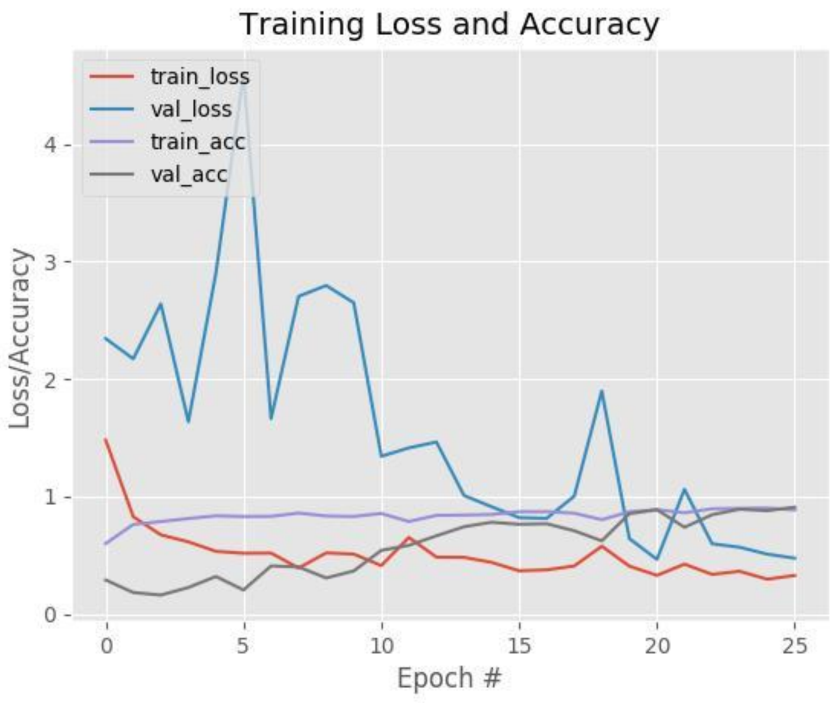


Fig.11: Training Loss and Accuracy after data Augmentation with EPOCH =25

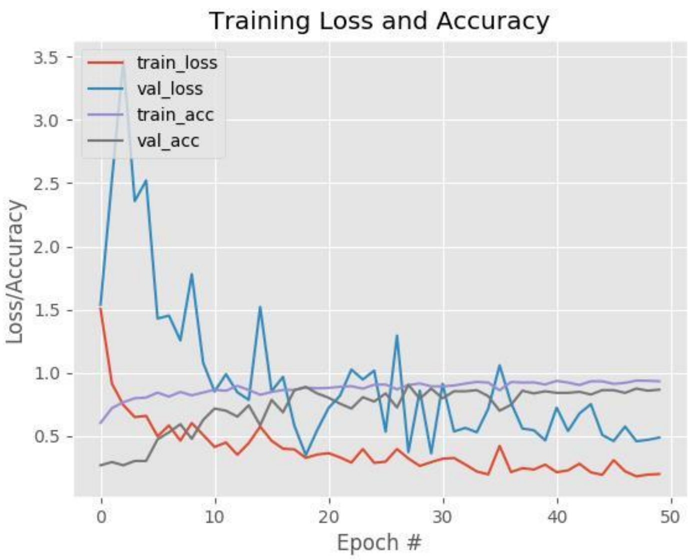
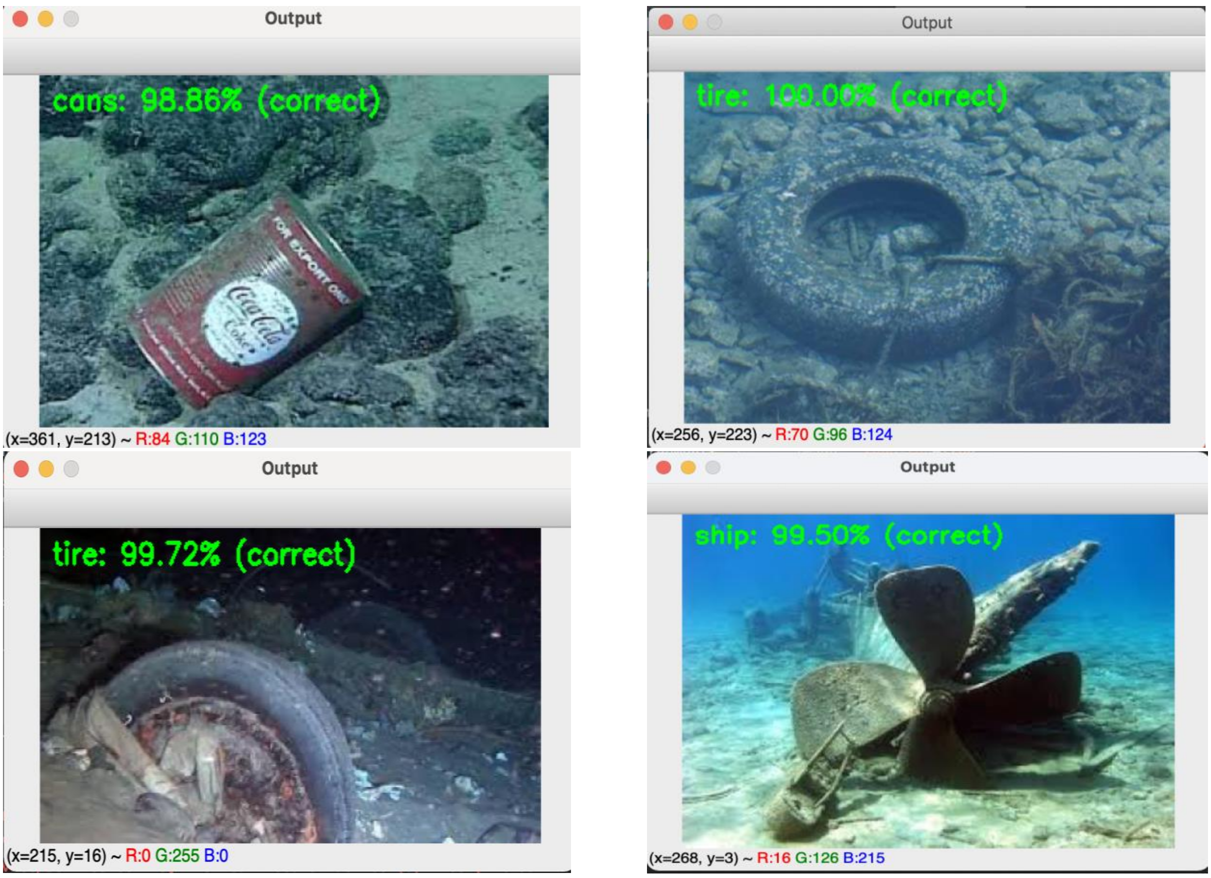


Fig.12: Training Loss and Accuracy after data Augmentation with EPOCH =50

Images were captured by an underwater robot created by Saipem Ltd. using a camera mounted on a work class ROV. The robot is around 4 metres long, 3 metres wide, and 1 tonne in weight. Figure 13 depicts the categorization result for some maritime debris. For each related marine debris, the accuracy and identified class are then output. The output is created using matplotlib.pylot, which outputs the identified picture as well as the accuracy rate. It is noticed that when the number of EPOCH rises, the overall accuracy of object identification increases to the greatest rate of 82.91%.



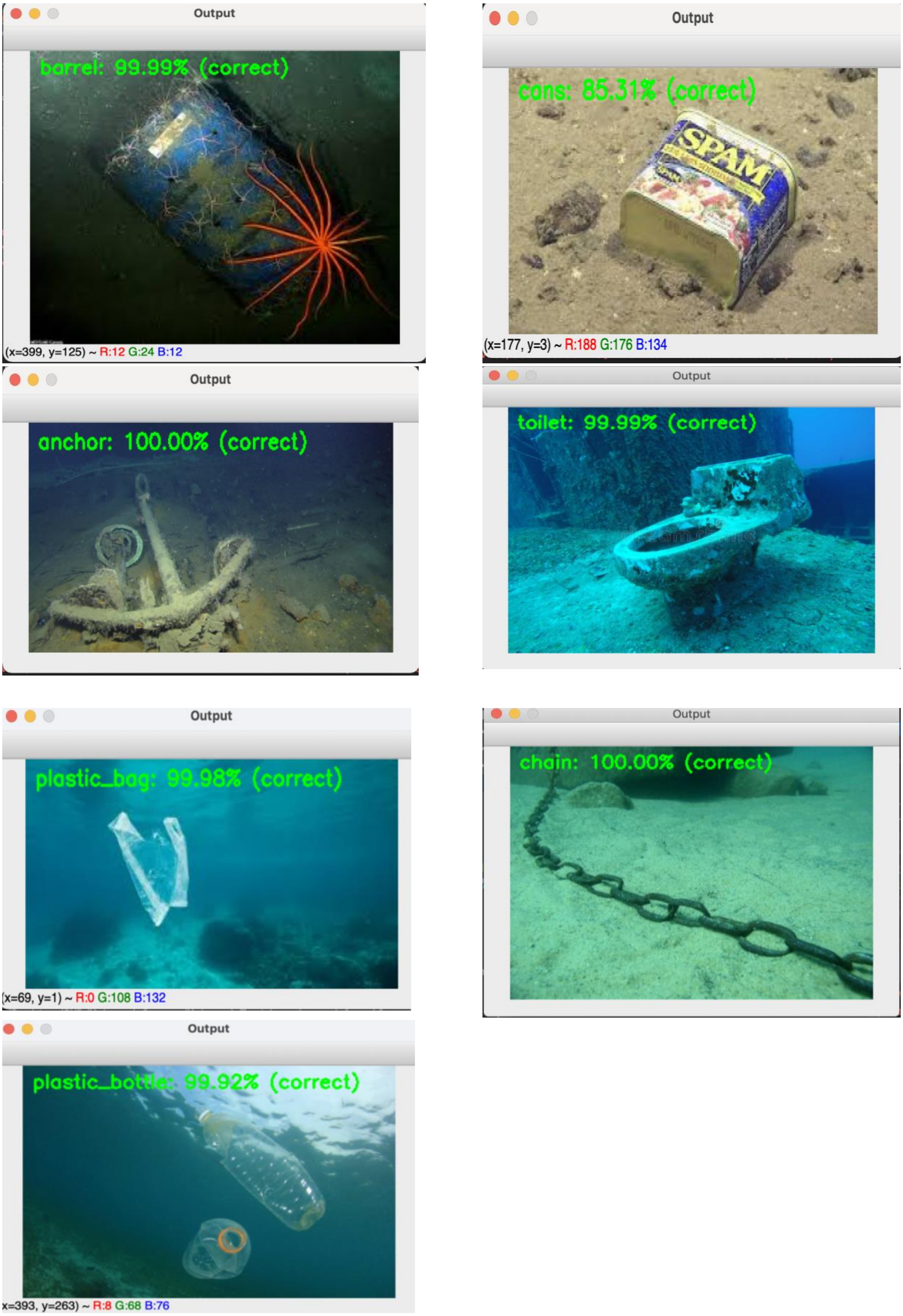


Fig13: Correct Test Results

If the system is unable to classify an object because it is not in the dataset, it will categories the object to its closest similarity in the dataset and then write the percentage of its proximity to the similar object, as well as write on the image that it is incorrect. As shown in Fig.14.

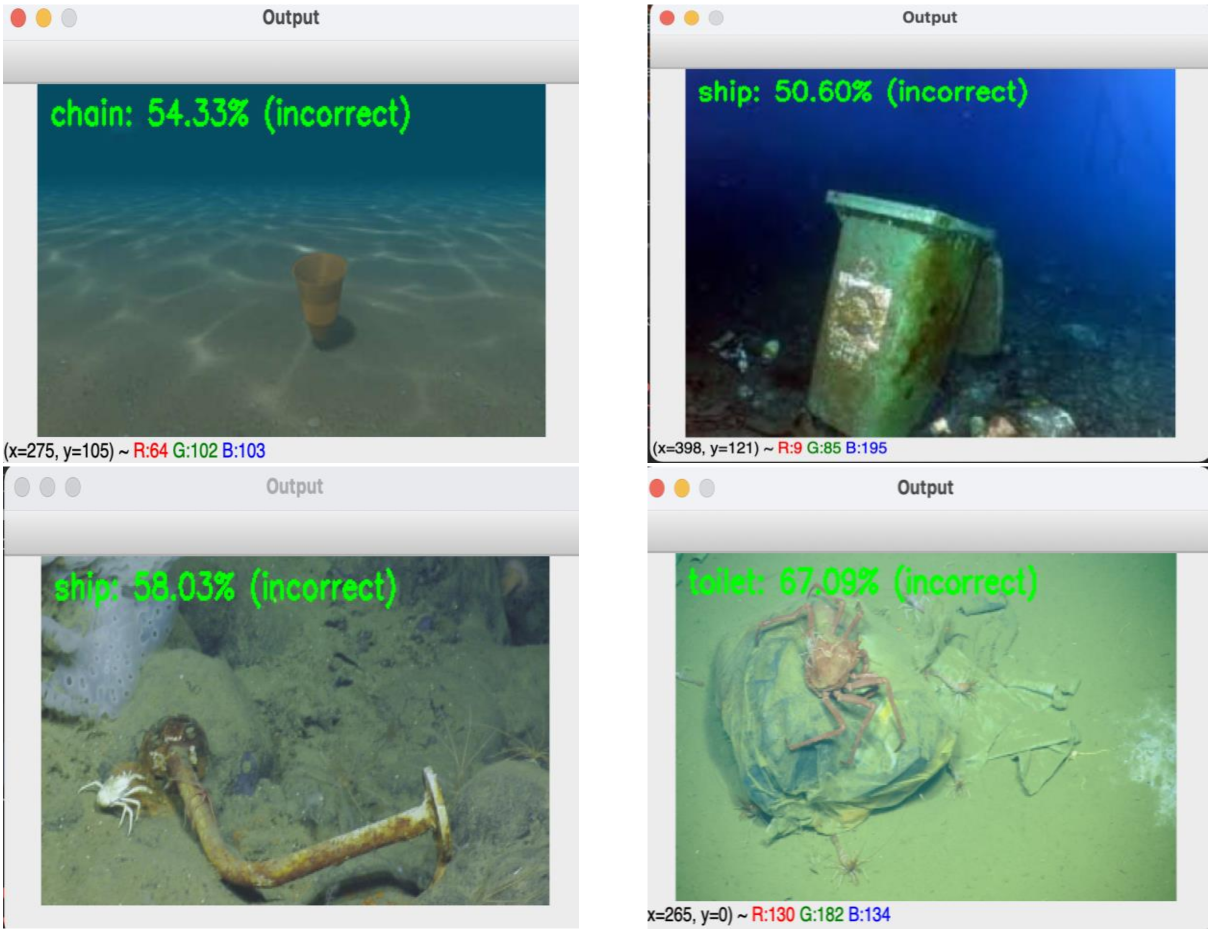


Fig.14: Wrong Recognize Output

6. Conclusions

The purpose of this study is to find out the solution for ROV to classify underwater debris which helps pilots in debris removal operation. Deep learning, a kind of Artificial Intelligence-based technology, is a field of computer Machine Learning at the computer science department. Automated technology identifies the object of interest in the image by using the principle of probability to distinguish it from its background view. In other words, we needed only a small number of images in this study as an experiment, using the Convolutional Neural Network algorithm most commonly used for image classification techniques and continuous upgraded active functions to identify images of underwater debris. In this experiment we used CNN along with Transfer Learning. Transfer learning on top of a pre-existing CNN gives a precise and accurate detection technique in this case. Furthermore, CNN constructed on top of an existing model is simpler because the process is completed by just training an existing model. Even though the approach was proven to be accurate, it failed to provide adequate detection time. The combination of CNN with transfer learning yielded a remarkable result of 98.34% accuracy and a computation time of only 7 minutes. Even though a fairly precise result was reached, this approach took a large number of resources since CNN and transfer learning both require a large amount of system resources. For large datasets, this characteristic makes the method inefficient. Second, the Fast RCNN was quicker than CNN with

Transfer Learning and had an accuracy of 81.4 %.This is due to the fact that it uses the object's region of interest rather than the entire image. Even if the time is efficient for this dataset, the accuracy rate is lower by a larger margin. As a result, this approach lacked the detection precision that is critical. The suggested model was inspired by the VGGNet, which had an accuracy of 98.34 % and a calculation time of 19 minutes. The computation time is quite long in this case, but it needs less resources and is more accurate. The dataset difficulties are handled through data augmentations, data modification through segmentation, foreground extraction, and background extraction of pictures to discover the relevant object of interest to detect. Furthermore, underwater vision in both a static and dynamic environment was investigated, and it was shown that item detection in a static environment gave superior accuracy and speed, while being more difficult to manage in real time.

7. Future Work

Here we demonstrate a model which can recognize and classify underwater objects in the image. Later it can be extended for character recognition, real-time object recognition, and measuring object sizes then objects weights which will help ROV to take a decision to remove object or it is too heavy.

References

[1] F.C.Soon,H.Y.Khaw,J.H.Chuah,J.Kanesan,Hyper parameters optimization of deep CNN architecture for vehicle logo recognition, IET Intelligent Transport Systems 12 (8) (2018) 939–946.

[2] Sinha, H., Awasthi, V., Ajmera, P. K.: Audio classification using braided convolutional neural networks. IET Signal Processing 14(7), 448–454 (2020).

[3] Umar Albalawi, S Manimurugan and R Varatharajan, "Classification of breast cancer mammogram images using convolution neural network", Concurrency and Computation: Practice and Experience, no. 1, pp. 1-12, 2020.

[4] Ferraz, P.A.P.; de Oliveira, B.A.G.; Ferreira, F.M.F.; Martins, C.A.P.d.S. Three-stage RGBD architecture for vehicle and pedestrian detection using convolutional neural networks and stereo vision. IET Intell. Transp. Syst. 2020, 14, 1319–1327.

[5] Shen, C.; Zhao, X.; Fan, X.; Lian, X.; Zhang, F.; Kreidieh, A.R.; Liu, Z. Multi-receptive field graph convolutional neural networks for pedestrian detection. IET Intell. Trans. Syst. 2019, 13, 1319–1328.

[6] Wang Q, Liao J, Lapata M, Macleod M. Risk of bias assessment in preclinical literature using natural language processing. Res Synth Methods. 2021 Oct 28

[7] Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., Harvey, E., 2016. Fish species classification in unconstrained underwater environments based on deep learning. Limnol. Oceanogr. Methods 14 (9), 570–585.

[8] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.

[9] Laila Ma‟rifatul Azizah, Sitti Fadillah Umayah, Slamet Riyadi, Cahya Damarjati, Nafi Ananda Utama "Deep Learning Implementation using Convolutional Neural Network in Mangosteen Surface Defect Detection", ICCSCE, ISBN 978-1-5386-3898-9, pp. 242-246, 2017.

[10] Guidelines for the selection and Operation of Jack-ups in the Marine Renewable Energy Industry issue 2 :2013.

[11] G. Clauss, E. Lehmann, C. ÖstergaardOffshore structures, volume I, conceptual design and hydrodynamicsSpringer-Verlag, New York (1992).

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

[13] 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.

[14] 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.

[15] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 2012, 25, 1097–1105.

[16] 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.

[17] Yong Bai, Qiang Bai, in Subsea Engineering Handbook (Second Edition), 2019, pages 320-340.

[18] Dipta Gomes , A.F.M. Saifuddin Saif . Robust Underwater Fish Detection Using an Enhanced Convolutional Neural Network. June 2021International Journal of Image, Graphics and Signal Processing 13(3):44-54