A Multi-Layer Perceptron Model for Underwater Object Recognition

Igor Vilas-Bôas Silveira[1*], Roberto Monteiro[2], Oberdan Pinheiro Rocha[2], Alex Álisson Bandeira Santos[2]
[1]Marinha do Brasil; Rio de Janeiro, RJ; [2]SENAI CIMATEC University Center; Salvador, Bahia Brazil

The more human gets interested in sea exploration, the more research to detect objects under the water is done. Therefore, the ability to detect, classify, recognize and track all kinds of objects is evolving daily. This paper aims to introduce a computer model for underwater object classification and recognition based on a Multilayer Perceptron network. The model was constructed with a mixed dataset for the training phase, combining artificial and natural objects, and it reached approximately 99.97% classifying accuracy.
Keywords: Computer Model. Underwater Objects. Classification. Recognition. Accuracy.

Introduction

Underwater target (object) recognition has a great variety of research purposes: monitoring underwater life sustainability, underwater gas pipeline leak detection as an industrial or environmental prevention application, identifying the presence of manufactured archeological objects for archeological research, underwater detection for mineral exploration, among others [1-4]. In addition, military activities such as Mine Counter Measures (MCM) [5] and Search and Rescue (SAR) operations [6] may also take advantage of this capability.

Also known as target detection, target recognition is a kind of computer vision task that aims to identify specific types of visual objects [7]. A system able to perform this task may have at least three essential elements: a sensor, a vehicle, and a data processing unit. The sensor is responsible for collecting data from the environment, codifying and sending it as an electrical signal to the data processing unit. The vehicle carries one or more sensors that can either be attached or tugged. Finally, the processing unit is responsible for interpreting the data collected by the sensor. As the element that decodes the electrical signal sent by the sensor, the processing unit is the 'brain' that makes all the necessary calculations to identify, classify, and, therefore, recognize one or more targets.

Received on 12 September 2022; revised 21 November 2022.
Address for correspondence: Igor Vilas-Bôas Silveira. ComemCh. Mocanguê Island - S/N° - Centro. Niterói/RJ. Zipcode: 24040-300. E-mail: igor.silveira@marinha.mil.br.
DOI 10.34178/jbth.v5i4.250.

J Bioeng. Tech. Health 2022;5(4):266-271

A generic diagram (Figure 1) describes the process that occurs since the very beginning: when the sensor captures information from the environment (data carried either by acoustic or visible light waves), passing through the transmission of this information to the processing unit, and reaching, at last, the signal interpretation and object recognition stage.

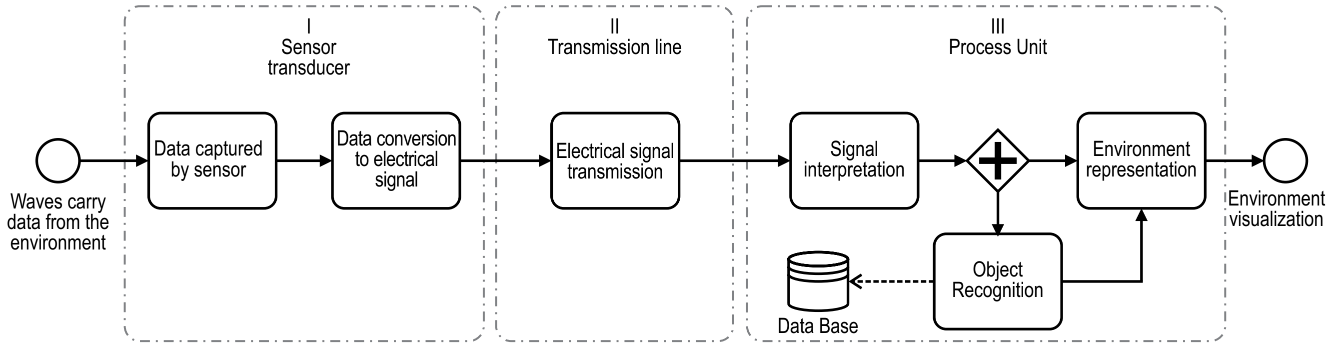
The recognition task finds its way in several algorithms and techniques developed and improved through the last years, and it is the core of this paper. This paper is divided as follows: Section 2 will give an insight into some of these techniques in use nowadays; section 3 will explain the methodology applied for the computer modeling; section 4 will discuss the results acquired by the model.

Related Works

From conventional signal processing methods using Fourier Transform to modern Machine Learning (ML) algorithms [8], many techniques and methods were developed and improved to perform accurate target recognition, as well as Automatic Target Recognition (ATR).

Image classification based on Deep Learning (DL) algorithms has been actively studied in recent years [9]. Indeed, as Teng and Zhao [10] has highlighted, significant improvements in the digital

Figure 1. Environment representation process diagram.



Source: Authors

image for target recognition and classification were reached by using DL.

While traditional object detection methods' performance quickly stagnated, DL developed powerful tools that can learn semantic, high-level, and deeper features [11].

According to Neupane and Seok [8], conventional signal processing and ML algorithms come up with some limitations that deep learning algorithms can overcome. Considered a subset of ML, DL works based on the Artificial Neural Networks (ANNs) concept, especially with multiple layers. For example, the Multi-Layer Perceptron (MLP) is a kind of Artificial Neural Network (ANN) with at least three layers: the input, the output, and a hidden layer. Another important DL method to be mentioned is the Convolutional Neural Network (CNN), which has become a modern approach for object detection and classification. For example, Makantasis and colleagues [12] developed an accurate and automated manufactured object detection approach using a CNN to encode spectral information and an MLP for the classification task. Jin and colleagues [13] proposed a method for ATR using a forward-looking sonar based on Deep Convolutional Neural Networks (DCNNs). Girshick and colleagues [14] developed a new approach that combines region proposals with CNNs. Thus this method was called Regions with CNN, or simply, R-CNN. Later, the same author developed the Fast R-CNN by improving the previous approach. R-CNN and Faster CNN explore the idea that an image may contain multiple objects, and they use regions to find a particular object. The first one to present a different approach was Redmon and colleagues [15], called You Only Look Once (YOLO). The idea is to divide the image into an $S \times S$ grid, and for each grid cell, predict B bounding boxes and C class probabilities. A further version of YOLO, called YOLOv4, was tested for underwater target recognition by Chen and colleagues [16], achieving improvements in recognition accuracy and speed and reducing hardware requirements.

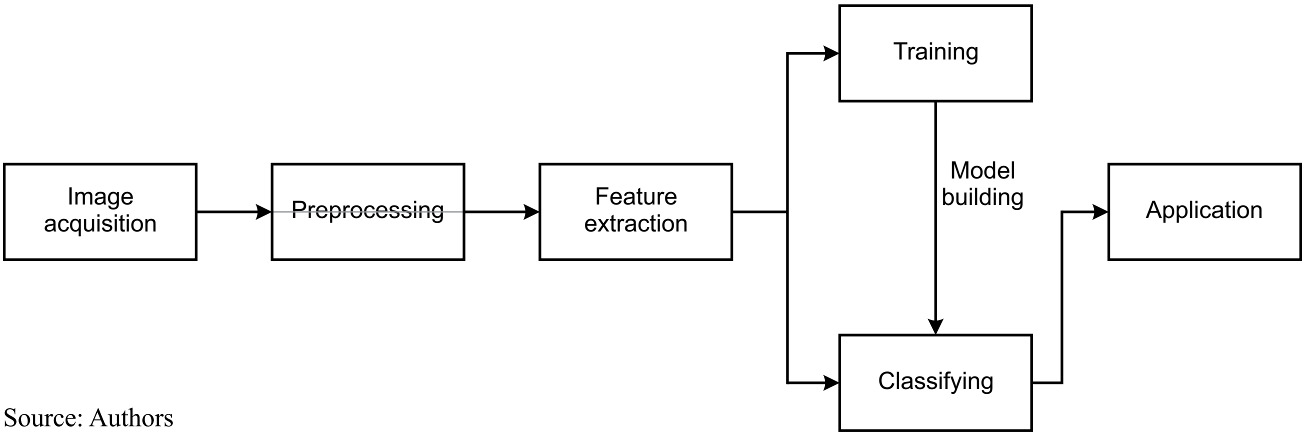
Materials and Methods

Computer Modeling

In order to build the proposed computer model, six components were taken into account as the essential ones: image acquisition, preprocessing, feature extraction, training, classifying, and application. Figure 2 depicts these six components as six blocks and the modeling procedure.

The first step of the modeling procedure, the image acquisition, consists of reading and converting the original dataset images into a bidimensional matrix so that each pixel on the original image corresponds to an RGB value. After that, at the preprocessing stage, a Sobel filter is applied to the data matrix, generating a binary

Figure 2. computer modelling procedure.



Source: Authors

matrix and making easier the identification of image contours.

Once the image is filtered, it is possible to calculate(statistically) the attributes that contain the input patterns for the features vector. This process is performed during the feature extraction stage, which transforms the binary matrix into a features vector matrix containing input patterns for training and classifying. Four attributes were defined to compound the features vector: areaFraction (percentage of non-zero pixels), mean, stdDev (standard deviation), and area (in pixels). These attributes are represented in the figure 3 as the inputs of an MLP.

The classifying process is performed in two phases: training (learning phase) and data classification (test). In the training phase, the classifying model is built by describing a set of the following classes: pipe, turtle, and shark. Hence the model extracts the attributes from the features vector to perform the training by using the training dataset. The training elements are associated with the class labels to which each belongs. Since the class label of each element is provided to the classifier, this step is known as supervised learning. The output of the classifier will indicate a command for an application.

Network Architecture

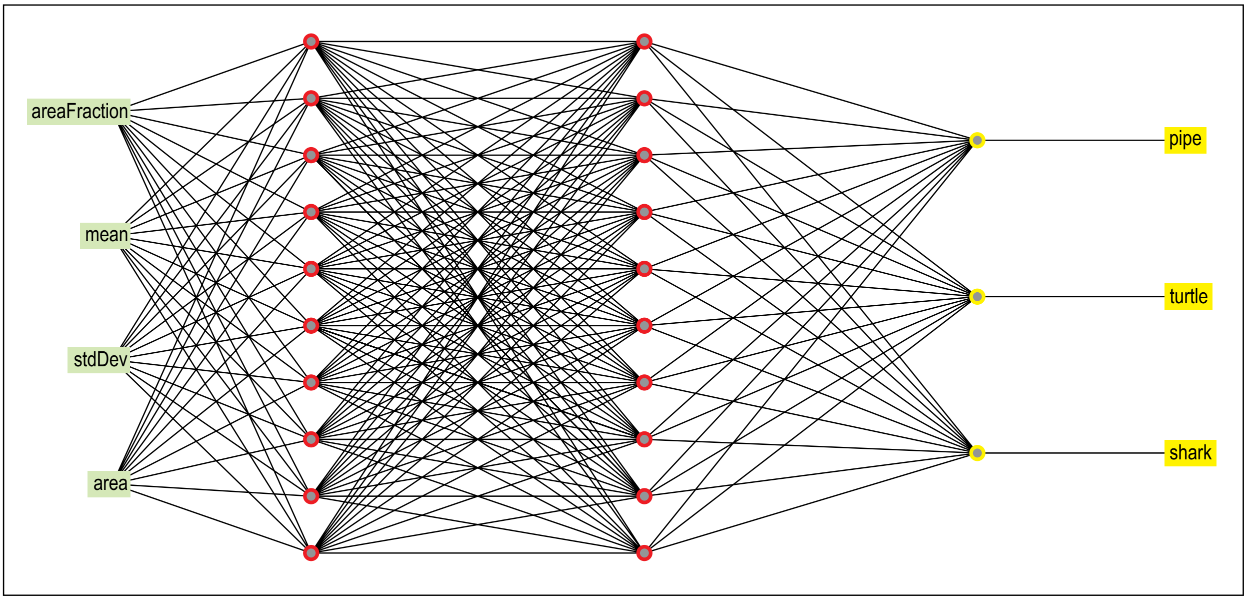
An MLP was the ANN chosen to perform the computer modeling of the object recognition algorithm. Figure 3 shows the architecture of the network used. In the first layer, green-colored nodes represent the described input layers. Right after, two hidden layers are represented by red-colored nodes, each containing ten nodes. The output layer, at last, is composed of three yellow nodes so that each one equals a class.

Table 1 describes the network in terms of its hyperparameters.

Dataset

A dataset was built by combining two other existing datasets: MARIS and UOT-100. The MARIS dataset originates in the Italian national Project MARIS (Marine Autonomous Robotics for InterventionS), whose goal is the development of beneficial technologies for underwater intervention in the offshore industry [17]. It contains 9600 stereo images of both single and multiple pipes with different colors ranging from 5 to 6 cm. Oleari and colleagues [18] and Kallasi and colleagues [19] are some of the many works unfolded by Project MARIS. The UOT-100 (Underwater Object Tracking) dataset was created to facilitate the developing of tracking algorithms well-suited for underwater environments. It contains 104 underwater video sequences and more than 74,000 annotated frames from natural and artificial underwater videos. Panetta and colleagues [20] and Kezebou and colleagues [21] are some essential papers

Figure 3. MLP architecture.



Source: Authors

Table 1. Network hyperparameters.

Hyperparameter	Value
Number of layers	4
Number of hidden layers	2
Number of hidden units (for each hidden layer)	10/10
Batch size	100
Epochs	1000
Activation function	Sigmoid
Learning rate	0.03
Momentum	0.2

about object tracking and the UOT dataset. The resultant dataset used for this paper contains 150 images and three classes: a class ‘pipe’ from MARIS and the classes ‘shark’ and ‘turtle’ from UOT-100.

Results and Discussions

In order to analyze the network performance, the following metrics were used:

- Mean absolute error (MAE): computes the mean absolute error between the labels and predictions;
- Root mean squared error (RMSE): computes root mean squared error metric between y_true and y_pred;
- Relative absolute error (RAE): computes the ratio of the absolute error of the measurement to the actual measurement;

- Root relative squared error (RRSE): computes the square root of the sum of squared errors.

For the 150 images selected to compound the training dataset, the model could classify 145 instances correctly, achieving an accuracy of approximately 96,667%. Therefore, 5 of the 150 images, i.e., only 3,33%, were incorrectly classified. The confusion matrix shown in FIgure 4 might shed some light on these incorrect instances. As revealed, three turtles were classified as sharks; two were classified as turtles.

Table 2 shows the returned values for each mentioned metric.

Figure 4. Confusion matrix.

```
  a   b   c   <-- classified as
 50   0   0 |   a = pipe
  0  47   3 |   b = turtle
  0   2  48 |   c = shark
```

Source: Authors.

Conclusion

Many kinds of research and developments in underwater object detection have recently appeared. The advances in deep learning methods brought a new perspective on the recognition modeling process. This work created a computer model for underwater object detection based on an MLP through a supervised learning DL process and using a mixed dataset. The results returned by the model, as presented in section 4, were above the expectations, with the classifying accuracy reaching about 96,97%. Further studies may be done by exploring other techniques, such as CNN, R-CNN, or even YOLO, as well as other combinations of datasets.

Table 2. Performance metrics.

Metric	Value
MAE	0.0308
RMSE	0.1388
RAE	6,9191%
RRSE	29.4353%

References

1. Rosli MSAB et al. Underwater animal detection using yolov4. In: IEEE. 2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE) 2021;[S.l]:158-163.
2. Nadimi N, Javidan R, Layeghi K. Efficient detection of underwater natural gas pipeline leak based on synthetic aperture sonar (sas) systems. Journal of Marine Science and Engineering 2021;9(11):1273, 2021.
3. Moroni D et al. Underwater manmade and archaeological object detection in optical and acoustic data. Pattern recognition and image analysis, Springer 2014;24(2):310-317.
4. Siddhartha JB, Jaya T, Rajendran V. Rdnn for classification and prediction of rock/mine in underwater acoustics. Materials Today: Proceedings, 2021.
5. Sulzberger G, Bono J, Manley RJ, Clem T, Vaizer L, Holtzapple R. Hunting sea mines with UUV-based magnetic and electro-optic sensors. Oceans 2009:1-5. Doi: 10.23919/OCEANS.2009.5422086.
6. Casalino G et al. MARIS: A national project on marine robotics for interventions. 22nd Mediterranean Conference on Control and Automation 2014:864-869. Doi: 10.1109/MED.2014.6961482.
7. Chen L et al. Underwater target recognition based on improved YOLOv4 neural network. Electronics 2021;10(14):1634.
8. Neupane D, Seok J. A review on deep learning-based approaches for automatic sonar target recognition. Electronics MDPI 2020;9(11):1972.
9. Xiongwei W, Doyen S, Steven CHH. Recent advances in deep learning for object detection. Neurocomputing 2020;396.
10. Teng B, Zhao H. Underwater target recognition methods based on the framework of deep learning: A survey. International Journal of Advanced Robotic Systems, SAGE Publications Sage UK: London, England 2020;17(6):172.
11. Zhao Z-K, Zheng P, Xu S-T, Wu X. Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems 2019;30(11):3212-3232.
12. Makantasis K, Karantzalos K, Doulamis A, Loupos K. Deep learning-based man-made object detection from hyperspectral data. In:, et al. Advances in Visual Computing. ISVC 2015. Lecture Notes in Computer Science(), vol 9474. Springer, Cham. https://doi.org/10.1007/978-3-319-27857-5_64.

13. Leilei J, Hong L; Changsheng Y. Accurate underwater atr in forward-looking sonar imagery using deep convolutional neural networks. IEEE Access, IEEE 2019;7:125522-125531.
14. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. UC Berkeley, 2014.
15. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. University of Washington, Allen Institute for AI, 2016.
16. Chen L, Zheng M, Duan S, Luo W, Yao L. Underwater target recognition based on improved YOLOv4 neural network. College of Mechanical Engineering and Automation, Fuzhou University, China, 2021.
17. Rizzini DL et al. Investigation of vision-based underwater object detection with multiple datasets. InTech, -, n. -, p. –, may 2015.
18. Oleari F, Kallasi F, Rizzini DL, Aleotti J, Caselli S. An underwater stereo vision system: From design to deployment and dataset acquisition. OCEANS 2015 - Genova, 2015:1-6. DOI: 10.1109/OCEANS-Genova.2015.7271529.
19. Kallasi F, Rizzini DL, Oleari F, Aleotti J. Computer vision in underwater environments: A multiscale graph segmentation approach. OCEANS 2015 - Genova, 2015:1-6, doi: 10.1109/OCEANS-Genova.2015.7271531.
20. Panetta K, Kezebou L, Oludare V, Agaian SS. Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN. IEEE Journal of Oceanic Engineering, June 2021.
21. Kezebou L, Oludare V, Panetta K, Agaian SS. Underwater object tracking benchmark and dataset. IEEE International Symposium on Technologies for Homeland Security (HST), Woburn, MA, USA, 2019:1-6.