DEEP LEARNING FOR OIL SPILL DETECTION AND CLASSIFICATION IN OCEAN SURFACE

Tayná Cristina Sousa Silva[1], Monik Silva Sousa[2] and João Viana da Fonseca Neto[3]
[1,2]Electrical Engineering Graduate Program, Federal University of Maranhão, São Luís - MA, Brazil
[3]Department of Electrical Engineering, Federal University of Maranhão, São Luís - MA, Brazil
[1]silva.tayna@discente.ufma.br
[2]monik.silva@discente.ufma.br
[3]joao.fonseca@ufma.br

ABSTRACT

Maritime transport is of great economic importance due to the increasing global industrialization. The importance of environmental issues, such as oil spills on the ocean surface, has been widely studied, leading to the need for methods to solve this problem. With the availability of satellites equipped with Synthetic Aperture Radar (SAR), it is possible to work on the monitoring for detection and classification oil spills and its derivatives at ocean surface. In this work, based on deep learning approach, the detection and classification is performed via YOLOv8 family algorithms. According to the tests carried out for YOLOv8 nano, small and medium, the best performance was obtained for medium algorithm, with accuracy and mAP-50 and mAP50-90 in the validation phase equivalent to 0.891%, 0.85% and 0.716%, respectively. In the test phase, it achieved a confidence level of over 70% of the objects detected.

KEYWORDS: *Oil Spill, Deep Learning, YOLO algorithm, Detection and Classification*

I. INTRODUCTION

With the advent of global growth, the demand for the production and transportation of oil and its derivatives has become a major global competition, encouraging companies to seek commercial strategies to achieve satisfactory economic growth [1]. In this context, the growing industrialization has caused environmental pollution, especially in the aquatic environment, due to the increased flow of oil tankers on the routes [2]. In the Brazilian scenario, it is no different: there are occurrences of oil spills, crude oil and by-products, causing environmental, social and economic impacts. As mentioned in [3], the accident that occurred on August 30, 2019, is considered the largest ever documented in the world, affecting a wide area of 4,334 km in the northeast and southeast regions of the country.

As noted by the International Tanker Owners Pollution Federation Limited - ITOPF [4], these events have the potential to cause significant and often irreversible environmental consequences, posing a significant threat to marine ecology and coastal ecosystems.

Therefore, various methods based on Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) have been explored to develop tools that can help detect, classify and monitor oil spills on the ocean surface.

Some researchers have excelled in using DL in their work, as [5] used Seg-Net to produce high quality oil spill images; in [6] they presented a structure also based on DL, this method is proposed in two stages for oil spill detection and in [7] they used an object detector using SAR images. In [8], they presented a convolutional network called OSCNet for oil spill detection and [9] used the U-net algorithm based on DL, their results were also satisfactory for the detection and classification of oil slicks on the sea surface.

In paper an algorithm for oil spill detection based on a DL is presented. The proposed approach allows the authorities in the sector to have contingency plans capable of avoiding the damage caused by the problem. Therefore, the objective is to analyze the performance of the YOLOv8 algorithm for the detection of oil spills.

The paper is organized into Sections. Section 2 presents the dataset made available by the European Space Agency (ESA), the image preprocessing stage and the methodology used to apply the YOVOv8 algorithm. Section 3 describes the results and discusses the comparisons of the evaluation, precision, recall, mAP50 and mAP50-95 metrics of the experiments. Section 4 presents the conclusion.

II. PROPOSED DETECTION AND CLASSIFICATION SYSTEM

The methods, techniques and models based on the YOLOv8 algorithm are presented in this section. Figure 1 provides an overview of the methodology, which consists of the medium to be monitored, data acquisition, and the method used to detect the object of interest.

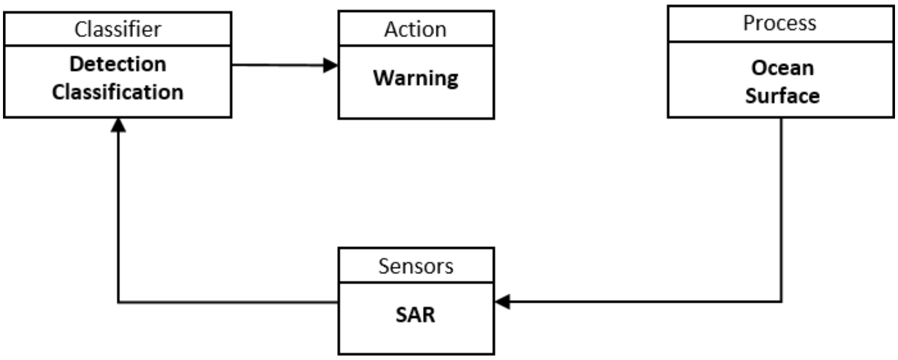


Figure 1 - Schematic of the proposed system for detecting oil slicks on the ocean surface.

According to Figure 1, it is necessary to determine the location where the monitoring will be carried out, i.e, in environments where there are oil tanker routes or in industries that use oils and petroleum as raw materials. This is how the data is acquired. In this case, the remote sensing is performed SAR images from the Sentinel-1 mission.

These SAR images were available through the European Space Agency (ESA) dataset, during a SENTINEL-1 mission from September 28, 2015 to October 31, 2017, the images contain five classes: land surface, sea surface, ships, oil slicks and oil spills. The dataset consists of 1112 grayscale images and their corresponding RGB (red, green, and blue) masks [10]. Figure 2 shows an image from the training dataset with its corresponding mask.

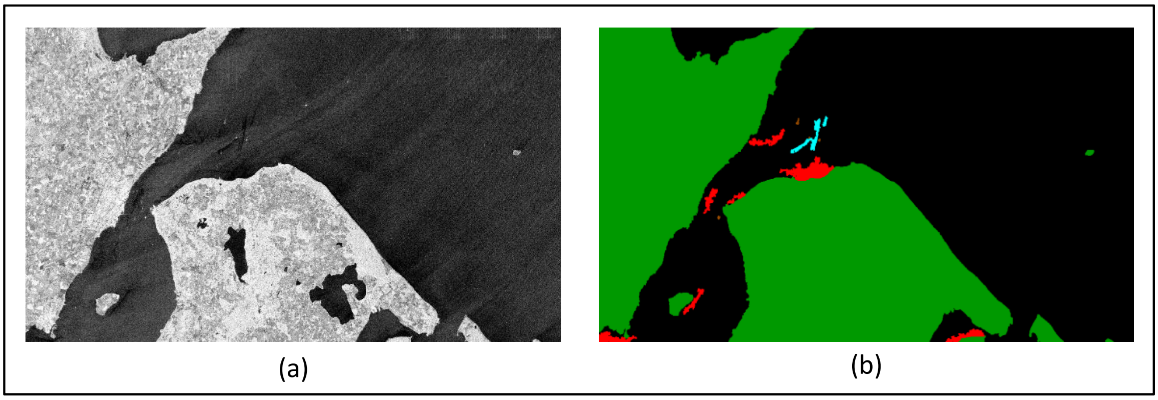


Figure 2 – Dataset image: (a) grayscale image, (b) RGB mask image.

According to Figure 2, using five different colors, the purpose of the masks is to identify the classes of images provided by the SENTINEL-1 satellite. Thus, the land surface is characterized by the green color, the sea surface is represented by the black color, ships are characterized by the brown color, oil slicks by the cyan color, and finally, sands are characterized by the red color.

However, the development of the proposed technique involves first the analysis of the data to evaluate the distinction between the two classes under study, followed by the preprocessing of the dataset.

After this stage, the dataset is prepared for training the classification model. Finally, tests are realized to evaluate the performance of the trained algorithm via evaluation metrics. These steps are presented in the following subsections.

2.1 Dataset Analysis

Analysis of the dataset is necessary before applying any preprocessing method to the images, as it provides information about the behavior of the classes. Therefore, an analysis of the pixel distribution of the images with and without oil stains was performed to determine the similarities between them and whether it is possible to separate them. Such an analysis is important to make a confident choice between the model and the classification algorithm.

2.2 Proposed Methodology

In this article, the proposed methodology aims to analyze the YOVOv8 family of algorithms. Thus, the methods of transforming images into grayscale, cropping and resizing are applied to the training and validation dataset as preprocessing. In the test dataset, only cropping and resizing were applied. Therefore, for a more detailed analysis of the pre-processing of the data set and the methodology of the YOLOv8 algorithm, Figure 3 is a representation of the stages of this process.

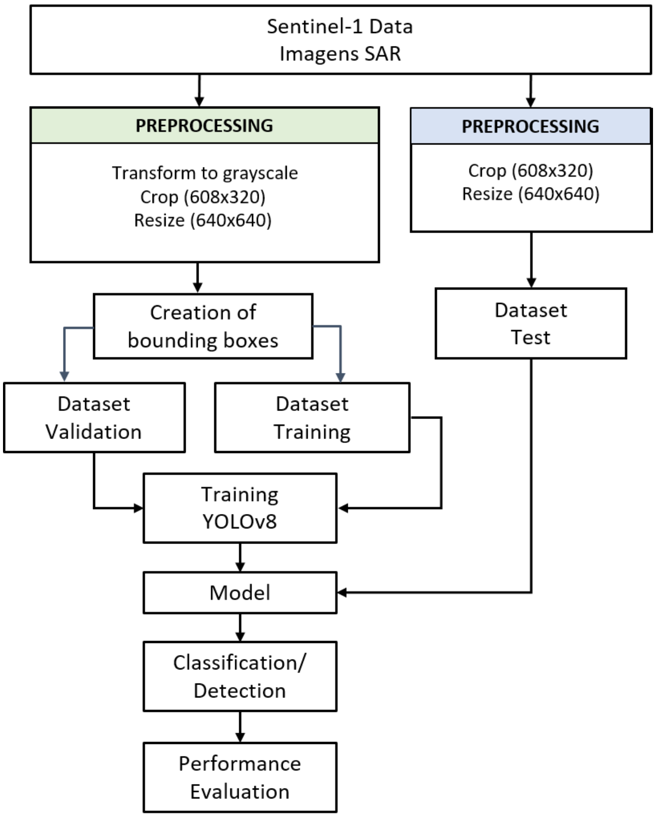


Figure 3 - Flowchart of the proposed method.

As can be seen in Figure 3, the YOLOv8 algorithm required the masks to be converted to grayscale, in which case the entire oil slick, previously cyan, was converted to gray, at pixel number 178.

To avoid overfitting and to increase the performance of the algorithm, the previously 1260x650 pixel image was cut into four parts, transformed into 625x325 pixel images, and then resized to 640x640 pixels.

Next, the step of labeling the oil spills in the objects is extremely important, because as seen in the raw dataset, there are the look-alike stains. And this labeling step serves as a way for the algorithm to identify which images have real oil stains.

Also at this stage, the Regions of Interest (ROI) were extracted based on the contour of each oil stain. In this way, each image that contains a stain according to pixel 178 has a bounding box around it, Figure 4 is an example of the result of an image with a bounding box, according to the stage of creating the bounding boxes in Figure 3.

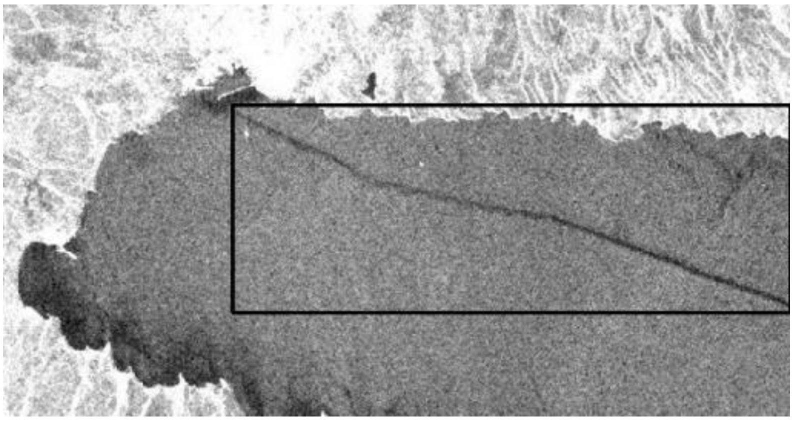


Figure 4 – Image of the training dataset with a bounding box.

In this way, the training dataset was created, consisting of 3376 files, divided into two parts: 1688 input images and 1688 files (.txt) containing the dimensions of the bounding boxes, and the validation data set with 120 files divided into two parts, with 60 input images and 60 files (.txt). For the test data set, only image cropping was used, consisting of 440 images corresponding to the input dimension of the YOLOv8 algorithm.

In the same way, after the preprocessing of the dataset, as shown in Figure 3, the training stages produce the predicted model of the classifier, which is then validated and tested in order to evaluate and compare the performance of the models used. These steps are described in detail in the following subsections.

2.3 The YOLOv8 Family

The DL-based convolutional neural network chosen to detect oil slicks on the ocean surface was You Only Look Once (YOLO), an algorithm for detecting objects in images. It divides the image into a grid and provides bounding boxes for each cell in the grid. As such, it has advantages over other neural networks, because it is able to identify and locate the object in a single pass through the network, and because of the speed of the process, it is able to provide results in real time.

The architecture of the model consists of three main layers: backbone, neck and head. The first layer is responsible for extracting features from the input images during training.

This stage can be costly, and the technique of transfer learning from pre-trained models can be used, which facilitates and teaches the neural network to see objects in general, leaving only the training stage to teach the reference to the context of the current problem [11].

The second layer is responsible for collecting maps of features from different stages in the recognition models. It is used to optimize the performance of the semantic information, thus working more effectively with the neurons that are activated with different types of perceptions [12].

The third layer, on the other hand, is responsible for the output of the model, where it is associated with the classification of labels - it uses the regression of the prediction of the bounding boxes to define the classes of objects.

In this way, the image is segmented into an $S \times S$ grid, with each individual cell within this grid performing its own recognition process. These cells are also responsible for evaluating the confidence associated with the corresponding bounding boxes. This methodology is used for all YOLOv8 models: nano, small, medium, large, and extra large.

As far as the YOLOv8 models are concerned, the image input for all of them is 640x640 pixels, while the differences are in the number of parameters. For the purposes of this article, nano, small, and medium have been used because they take into account the size of the parameters and the execution time. Table 1 compares the selected models.

Table 1 - Comparison of the YOLOv8 nano, small and medium models.

Models	Size (pixels)	Parameters (M)
YOLOv8 nano		3.2
YOLOv8 small	640	11.2
YOLOv8 medium		25.9

The calculation of the confidence of the models includes the existence or non-existence of the object. If there is no object in the cell, the confidence score is 0, otherwise it is 1. Therefore, the confidence score is the intersection of the union of the actual box and the predicted box. Figure 5 is an example of how this process works.

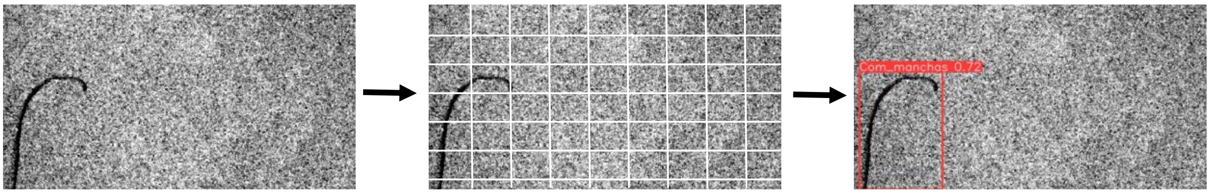


Figure 5 – Stages of the detection model flow in a test image.

According to Figure 5, the model flow for oil spill detection consists of inputting an image, creating the grids needed to compute confidence, and outputting the detection and confidence (%) of the predicted object. For the YOLOv8 training phase, the following parameters were used, as shown in Table 2.

Table 2 – Parameters of the training of YOLOv8.

Optimizer	**ADAM**
Learning rate	2e-3
Momentum	0,9
Batch size	16
Epochs	50

According to Table 2, the training set consists of 1688 images. Training is performed over 50 epochs with a batch size of 16 and a learning rate of 2e-3.

2.4 Performance Evaluation

After training, the YOLOv8 classifier may present some errors in its classification. It is therefore necessary to use evaluation metrics to assess the performance of the predicted model. The Precision (P), Recall (R) and mAP (mean accuracy) metrics is used.

Precision is defined as the ratio between the number of correct positive class predictions in relation to the total number of images classified as positive (True Positive-VP + False Positive-FP), that is given by

$$P = \frac{VP}{VP + FP}. \qquad (1)$$

Recall (R) metric indicate the proportion of samples that are correctly classified, this metric is also known as sensitivity. The Recall provides the total number of positive samples (PV + False Negative - FN) that is given by

$$R = \frac{VP}{VP + FN}. \qquad (2)$$

Mean Average Precision (mAP) is used to measure the effectiveness of object detection models. It uses the average Precision (AP) for all classes, meaning that the higher the AP, the better the precision of the detection of the object of interest, where the IoU (Intersection over Union) needs to reach or exceed 0.5. Therefore, the mAP is calculated using the ratio of precision and recall, that is given by

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k, \qquad (3)$$

where n is the number of classes to determine.

III. RESULTS AND DISCUSSIONS

The goal of this work is to analyze the performance of the YOLOv8 algorithm for the detection of oil slicks on the sea surface. This section describes the results of the classifier experiments according to the methodology described in the previous section.

First, the distribution of images with and without oil slicks was analyzed. For this, two images from each class were randomly selected and their pixel histograms is plotted.

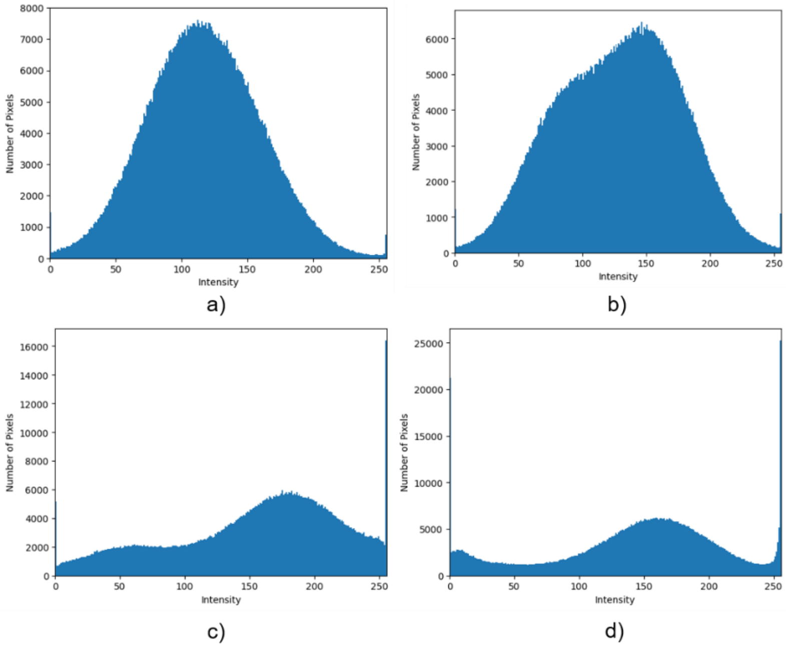


Figure 6 – Histogram of pixels in relation to image classes: (a) and (b) represent the class with oil stains; (c) and (d) represent the class without oil stains.

It can be seen that images (a) and (b) in Figure 6 correspond to those with oil stains. They have a peak of about 100 to 150 pixels, which represents an intensity of pixels in shades of black. Images (c) and (d), on the other hand, have a peak of about 150 to 200 pixels, representing color images closer to white. The network was then trained using the model and parameters seen in Subsection 2.3.

According to Table 3, three tests were performed on the YOLO v8 family: YOLO nano, YOLO small, and YOLO medium, to measure the effectiveness of each algorithm according to the performance metrics.

Table 3 – Comparison of Yolov8 family algorithm training times.

Algorithm	Training	GFLOPs
Yolo v8 n	0.655 hours	8.1
Yolo v8 s	0.857 hours	28.4
Yolo v8 m	0.947 hours	78.7

With respect to the validation test, Table 4 shows the results of these metrics calculated from equations (1), (2), and (3) that are precision, recall and mAP, respectively.

Table 4 – Performance results in the validation phase.

Models \ Metrics	Precision (P)	Recall (R)	mAP50 (%)	mAp50-90 (%)
Yolov8 n	**0.927**	0.783	0.837	0.697
Yolov8 s	0.923	0.776	0.849	0.711
Yolov8 m	0.891	**0.806**	**0.85**	**0.716**

Table 4 provides an overview of the tests performed in this article. Focusing on the mAP50-95 metric, which quantifies detection precision at all confidence levels from 50% to 95%, and the mAP50 metric, which is limited to detections with a 50% confidence level, it can be seen that the YOLOv8m model manifests the highest value, corresponding to 71.6% and 85% of these performance metrics, respectively.

Figures 6, 7 and 8 show the other evaluation metrics in the training and validation phases of the three tests of the YOLOv8 family.

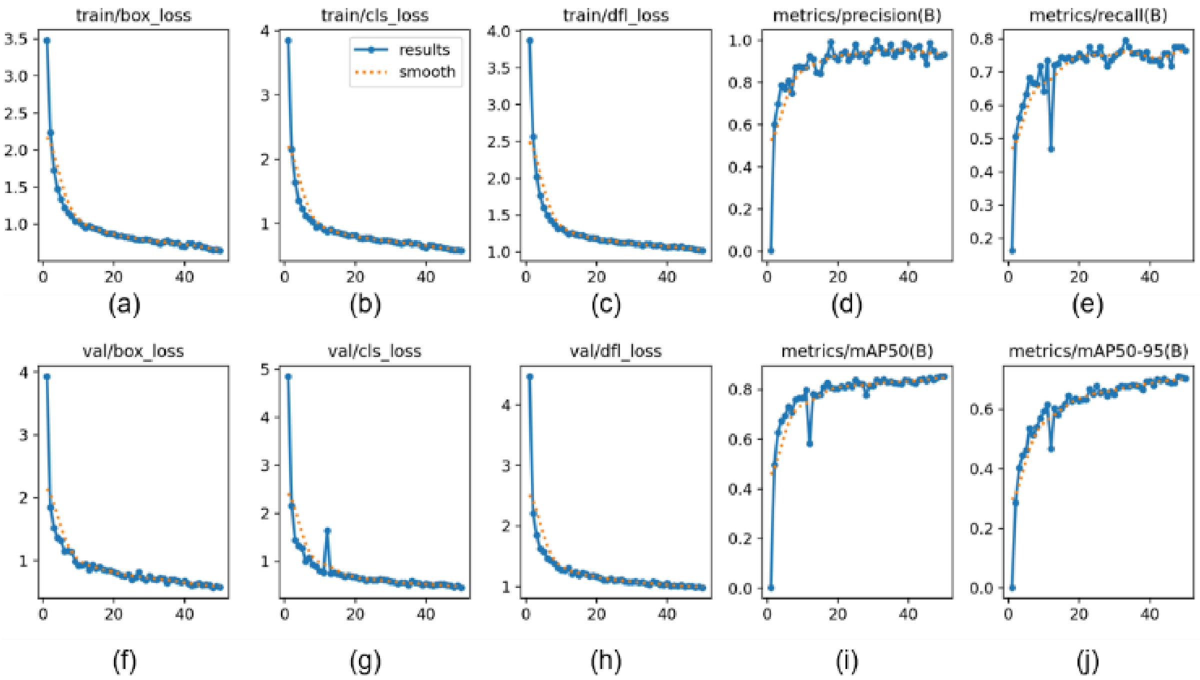


Figure 6 – Results comparing the performance metrics of precision, recall, mAP 50% and mAP 50-95% for the YOLOv8 nano algorithm.

According to Figure 6, the performance metrics for the first test are shown in graphical form, where (a) to (e) refer to the training phase and (f) to (j) refer to the validation phase.

The loss observed in Figure 6 (a) is related to the bounding boxes with respect to the objects found by the algorithm, where there is a loss associated with the coordinates with respect to the center of the object with the ends of the boxes. Figure 6 (b) is a loss associated with the classification of the boxes in relation to the objects found by the algorithm, referring to the IoU and finally, in Figure (c) is a loss associated with the Density-Free Location, where its function is to adjust the trained model and regulate the density of objects in different regions of the bounding boxes, especially when the objects are close to each other.

Thus, Figures 6 (a), (b), and (c) show an inversely proportional relationship with the number of trained epochs and the losses, indicating that the training performance of the network improves with the evolution of the 50 epochs.

Figure 6 (d) and (e) refer to the precision and recall metrics, respectively. It can be seen that the higher the number of epochs, the higher the value of the performance metrics, which are directly proportional. This means that the trained model is closer to the ideal classifier according to equations (1) and (2).

Figures 6 (e) to (j) follow the same reasoning as in the training phase, but this happens in the validation phase, which also shows a good performance with respect to the classifier, although the mAP50 and mAP50-95 metrics have fluctuated over the epochs with respect to the objects detected by the algorithm.

Similarly, Figure 7 shows the performance metrics for the second test, as mentioned in Subsection 2.2.

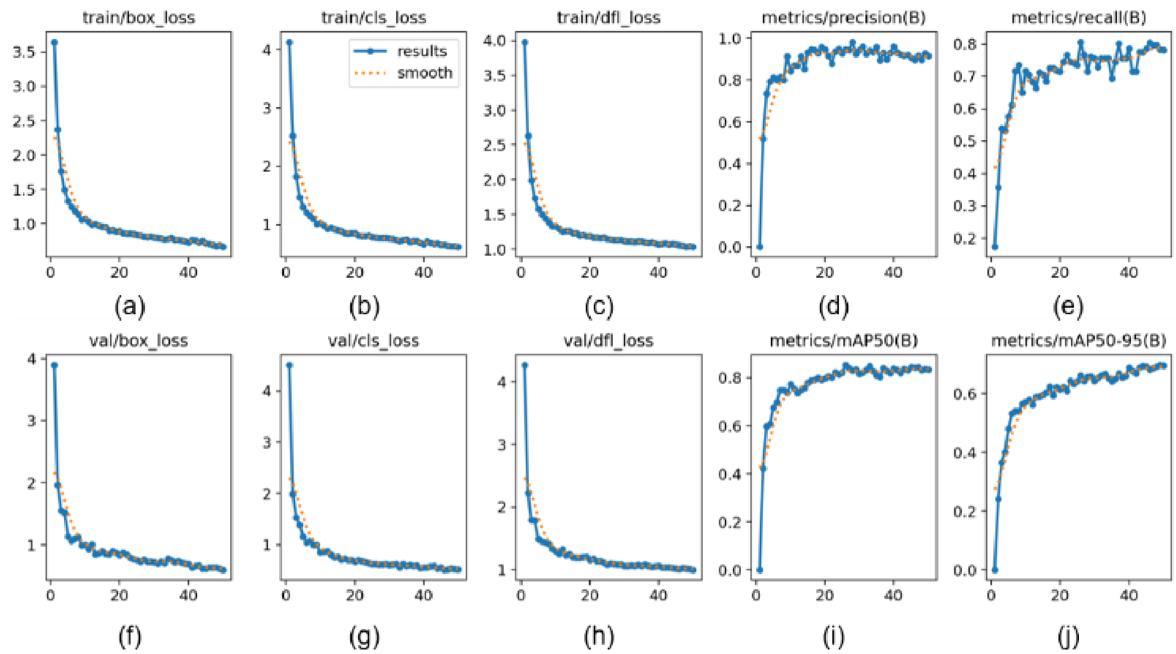


Figure 7 – Results comparing the performance metrics of precision, recall, mAP 50% and mAP 50-95% for the YOLOv8 small algorithm.

The same is true for Figure 7. Curves (a), (b) and (c) show the losses associated with the bounding boxes. However, graphs (d) and (e) show that the precision and recall metrics fluctuated strongly with the evolution of the epochs, especially in curve (e), since the proportion of true positive samples correctly classified by the classifier was not satisfactory in relation to the total number of positive samples, as seen in Equation (2).

Regarding the validation phase, Figures (f) to (h) follow the same concept regarding the losses of the bounding boxes, in curves (i) and (j) they refer to the metrics of Equation (3), i.e. they are satisfactory with respect to the first test, as seen in Figure 6 (i) and (j).

Using the YOLOv8 medium, the performance metrics are as shown in Figure 8.

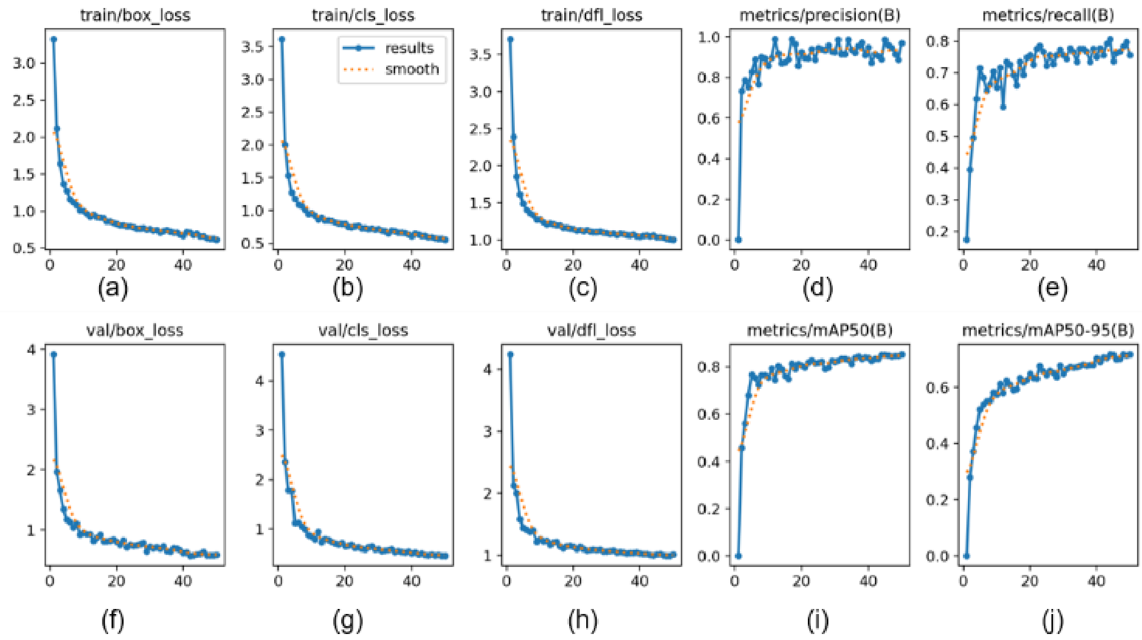


Figure 8 – Results comparing the performance metrics of precision, recall, mAP 50% and mAP 50-95% for theYOLOv8 medium algorithm.

The losses shown in curves (a), (b) and (c) follow the same concept for all trials, i.e, the trend decreases with the training periods.

The precision and recall metrics shown in curves (d) and (e) of Figure 8 fluctuated over the epochs, but reached a value corresponding to 96% for precision and 71% for recall. In the validation phase, as shown in Figure 8 (i) and (j), it reached about 85% and 71% for mAP50 and mAP50-95, respectively, according to Table 4. These values were higher than in the tests with YOLOv8 nano and small.

Comparing the tests and metrics, it can be seen that the YOLOv8 medium model is the most effective in detecting oil stains in aquatic environments. Thus, using 110 images with 98 classification instances for just one class: "0: Oil Spill", we can see the confusion matrix in Figure 7.

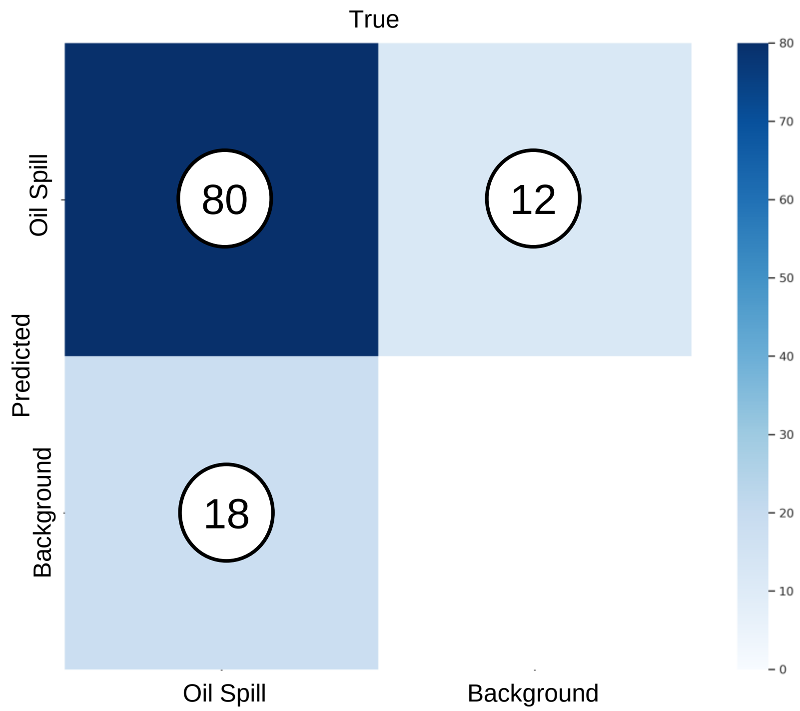
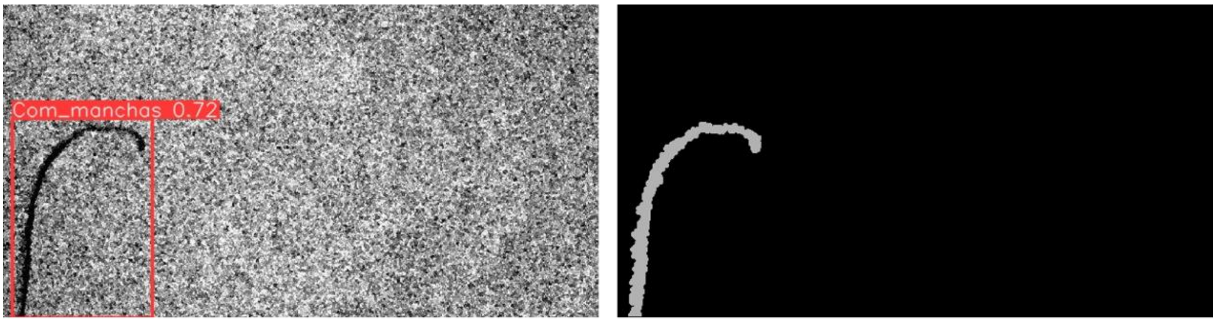


Figure 7 – Confusion matrix of the YOLOv8 medium algorithm.

Therefore, the classifier detected 80 objects called "Oil Spill", 18 images were not detected as having stains but were among those to be classified, and 12 images were considered as having stains, but not within this classification and could be associated with instances called oil stain look-alikes.

To analyze the test phase of the classifier, Figure 9(a) shows the image with the bounding boxes and its classification (%) of the predicted object, and Figure 9(b) is the representation of the oil stain detected by the mask.



a) b)

Figure 9 – Results of the YOLOv8 medium algorithm: (a) Detection of the result predicted by the classifier; (b) representation of the detection in a mask.

As shown in Figure 9, the YOLOv8 medium model recognizes and classifies the test image as expected in the training and validation phases.

3.1 Comparison With Other DL Approaches

In the literature, research using deep learning and SAR data has stood out, especially for oil spill detection in aquatic environments. In the following, we will compare some studies with the same objective and dataset as this article.

A method for detecting oil spills in the Mediterranean Sea is presented in [7], this method takes into account climatic conditions such as low light levels and the presence of shadows, the researchers used the YOLOv4 algorithm and obtained a satisfactory result with the mAP50 metric of 68.69%. In [9], their research used the U-net algorithm for detecting oil spills on the sea surface in comparison with the LDA-MLP algorithm, where his results showed an approximation in the values of the proposed algorithms, the result for U-net proved to be effective, using the precision and recall metrics of 84.7% and 69.2%, respectively.

In [10], two algorithms based on CNNs were used to semantically segment the input image for the detection and classification of each pixel in five different instances. The result for the oil slicks instance using the U-net algorithm and DeepLabv3+'s IoU metric was 53.79% and 53.38%, respectively.

The proposed algorithm in this paper, on the other hand, proved to be superior to previous studies, since the evaluation metrics obtained were 89.1% for precision and 80.6% for recall. The confidence values (IoU) related to the mAP50 and mAP50-95 metrics reached values above 70% in the detection of oil stains, showing that the proposed algorithm is satisfactory for the presented problem.

IV. CONCLUSION

With the increasing globalization of maritime transport, there is a need for techniques capable of monitoring, detecting and classifying the aquatic environment. Therefore, we presented an alternative to mitigate oil spills in these environments by applying the YOLOv8 algorithm.

In the analysis of the YOLOv8 family algorithms, the precision, recall, mAP50 and mAP50-95 metrics, compared to the other models tested, the medium model stood out with better performance, but its performance in terms of training time and computational speed was higher than the others. Therefore, its performance is still satisfactory to meet the requirements of the problem, since it achieved significant confidence values with respect to objects labeled with oil spills.

For future work, the refinement and increase of the training dataset and the improvement of the hyper parameters of the algorithm, such as – batch size, learning rate, momentum, epochs, in the training phase are valid for a more in-depth study of the proposed topic.

ACKNOWLEDGMENTS

We are grateful to the ANP - Human Resources Program (Program 54.1) - for its financial support for this research. We are also grateful to FAPEMA (Maranhão State Agency for Research and Development), the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Graduate Program in Electrical Engineering (PPGEE) of the Federal University of Maranhão (UFMA) for the infrastructure to support the activities carried out in this research at the Laboratory of Embedded Systems and Intelligent Control (LabSECI/UFMA).

REFERENCES

[1] Adhikary, S., Tiwari, S, P., Banerjee, S. (2022). Realtime Oil Spill Detection by Image Processing of Synthetic Aperture Radar Data. CONECCT - International Conference on Electronics, Computing and Comunication Technologies. DOI: 10.1109/CONECCT55679.2022.9865768.

[2] Zhang, Y., Zhu, Q., Guan, Q. (2021). Oil Spill Detection Based on CBD-Net Using Marine SAR Image. IGARSS –International Geoscience and Remote Sensing Symposium. DOI: 10.1109/IGARSS47720.2021. 9553884

[3] Pena, P. G. (2020). Derramamento de Óleo Bruto na Costa Brasileira em 2019: Emergência em Pública em Questão. *Cad Saúde Pública*. DOI: 10.1590/0102-311X00231019.

[4] ITOPF – International Tanker Owners Pollution Federation Limited: Oil Tanker Spill Statistics 2020. Disponível em: <https://www.itopf.org/knowledge-resources/data-statistics/statistics/>. Acesso em 03 de novembro de 2023.

[5] Ahmed, S; ElGharbawi, T, Salah, M; ElMewafi, M. (2023). Deep Neural Network for Oil Spill Detection Using Sentinel-1 Data: Application to Egyptian Coastal Regions, Geomatics, Natural Hazards and Risk, 14:1, 76-94, DOI: 10.1080/19475705.2022.2155998.

[6] Shaban, M.; Salim, R.; Abu Khalifeh, H.; Khelifi, A.; Shalaby, A.; El-Mashad, S.; Mahmoud, A.; Ghazal, M.; El-Baz, A. A Deep-Learning Framework for the Detection of Oil Spills from SAR Data. *Sensors* **2021**, *21*, 2351. DOI: 10.3390/s21072351.

[7] Yang, Yi-Jie; Singha, S; Mayerle, R. (2022) A Deep Learning Based Oil Spill Detector Using Sentinel-1 SAR Imagery, International Journal of Remote Sensing, 43:11, 4287 4314, DOI: 10.1080/01431161.2022. 2109445.

[8] Zeng, K.; Wang, Y. Uma rede neural convolucional profunda para detecção de derramamento de óleo a partir de imagens SAR transmitidas pelo espaço. *Sensor Remoto* **2020**, *12*, 1015. DOI: 10.3390/rs12061015.

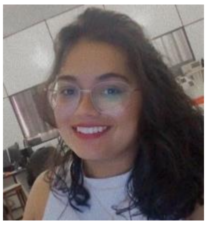
[9] Sousa, S, M. Detecção e Classificação de Perturbações no Meio Aquático via Aprendizado Profundo e Rede Neural Artificial Integrada à Métodos Estatísticos, Dissertação de mestrado: Engenharia Elétrica – Universidade Federal do Maranhão, São Luís – MA, p.94, 2023.

[10] Krestenitis, M. O. (2019). Early Identification of Oil Spills in Satellite Images Using Deep CNNs. (C. Springer, Ed.) *In International Conference on Multimedia Modeling*, 424 - 435.

[11] Kattenborn, T., Leitloff, J., and Schiefer, F. (2021). Convolutional Neural Networks (CNN) in Vegetation Remote Sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 173 (2), 24-49. DOI:10.1016/j.isorsjprs.2020.12.010.

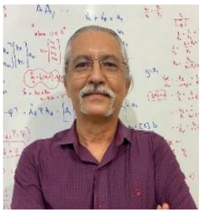
[12] Siedel, R. Detecção de Defeitos na Fabricação Têxtil Utilizando Yolov5. Dissertação de Mestrado: Computação Aplicada – Instituto Federal do Espírito Santo, Serra – ES. p.77, 2023.

AUTHORS



TAYNÁ CRISTINA SOUSA SILVA Master's degree student in Electrical Engineering from the Federal University of Maranhão (UFMA), in the field of Automation and Control, with a research line in Artificial Intelligence. He is currently working on Artificial Intelligence at the Laboratory of Embedded Systems and Intelligent Control (LabSECI/UFMA). Graduated in Mechanical Engineering from the Uniceuma University of Maranhão (UniCEUMA) with an emphasis on Internal Combustion Engines.

MONIK SILVA SOUSA Master in Electrical Engineering from the Federal University of Maranhão - UFMA (2023), with an emphasis on automation and control, degree in Electrical Engineering from UFMA (2019). PhD in progress in Electrical Engineering at UFMA (2023-2026), in the area of Automation and Control, with a line of research in Wireless Sensor Network. Currently developing work with Wireless Sensor Network in the Embedded Systems and Intelligent Control Laboratory (LabSECI). With experience in developing teaching aids for Process Control Laboratory and Automation and Control Laboratory.



JOÃO VIANA DA FONSECA NETO received the bachelor's and master's degrees in electrical engineering from the Federal University of Paraíba, in 1982 and 1986, respectively, and the Ph.D. degree in electrical engineering from the State University of Campinas, in 2000. He is currently a Titular Professor at the Universidade Federal do Maranhão (UFMA). His research interests include computational intelligence, multi-agent systems, machine learning, sensors networks, optimal control, and evolutionary computation and fuzzy.