Critical object recognition in underwater environment

Alexandra Nunes
Electrical and Computer Engineer
FEUP
Porto, Portugal
up201402644@fe.up.pt

Ana Rita Gaspar
Electrical and Computer Engineer
FEUP
Porto, Portugal
up201402645@fe.up.pt

Anibal Matos
Electrical and Computer Engineer
FEUP
Porto, Portugal
anibal@fe.up.pt

***Abstract*—Nowadays, ocean exploration is far from complete and the development of suitable recognition systems are crucial, to allow that the robots perform inspection and monitoring tasks in diverse conditions. The online available datasets are incomplete for these kinds of scenarios and, so it is important to build datasets that covered real condition in a simulated environment. Thus, it was developed a dataset with some man-made objects presents in the underwater environment. Moreover, it is also presented the developed method (Convolutional Neural Network) and its evaluation in diverse conditions is performed. It is also presented a comparative analysis and a discussion between the proposed algorithm and the ResNet architecture. The obtained results showed that the developed method is appropriate to classify 7 critical different objects with good performance.**

***Index Terms*—Classification, objects, visual, underwater**

I. Introduction

Currently, the object recognition is an attractive area due to the close relationship with video analysis and image comprehension. Otherwise, ocean exploration is far from complete and the development of recognition systems able to acquire in diverse conditions is crucial. This environment is still unknown, but into the ocean, there are some critical and dangerous objects, as the mines or chains, that automatic identification is very important in different types of tasks. In the last years, many of these tasks have been performed by the robotic vehicles and, therefore, they need to adapt its behaviour according to the objects that found. In addition, it is important that the robot autonomously recognizes diverse objects, for example in shipwreck situations, that involve many risks for Human. There are several sensors able to assists in object recognition, but the visual sensors present obvious advantages, such as their cost, the good performance when the acquisition is close to structures and the amount of information provided. However, they are limited when applied in underwater environment, due to some phenomena related to the propagation of light in water. The object identification is realized with the resort to their limits, which in this environment it is a challenge because they are noisier and more unfocused. The turbidity rate of the water, the different conditions of the luminosity on the images, and the background confusion due to reef features and underwater plant life are a priori unknown, making the process of image classification a challenging task.

In the last few years, deep learning has emerged as a powerful machine learning tool with the ability to overcome the shortcomings of the conventional image classification approaches. The challenges previously described difficult conventional methods to model and to adapt to the features of the interest objects in such images. Thus, multilayer deep neural networks provide such an opportunity to extract unique, invariant and robust features in the presence of the distortions and variability in images. Thus, in this work is presented a developed approach for solving this challenge. The proposed method resorts to Neural Networks approaches in order to learn the main features of some critical objects in diverse conditions. In addition, it is also presented a comparative analysis between the proposed approach (In-house) and a state-of-art method (ResNet). Nowadays, the online available datasets are incomplete for this type of scenarios, because the more commons in this environment only include fishes, other maritime species and reefs, that means without critical objects. Thus, the experimental results are obtained with a dataset developed by the authors with different classes, acquired in an environment that simulates the real conditions.

Therefore, the main contributions of this paper include a novel dataset called UWObjects@CRAS, provided as a public repository from CRAS (Center for Robotics and Autonomous Systems) - https://rdm.inesctec.pt/dataset/nis-2019-001 . Moreover, a preliminar progress in the development of algorithms to classify critical objects present in underwater environment, ensuring a balance between computational resources and accuracy.

This paper is organized as follows: Section II presents the evolution of state-of-art machine learning methods and presents some works that demonstrated good results and new perspectives for the future. Section III presents the developed dataset as well as the data acquisition conditions. Moreover, the methodology used in the development of the algorithm to classify the objects was described. Afterwards, in section IV the obtained results in terms of the performance and robustness were demonstrated. Section V presents a discussion about results, the major conclusions about this work and some issues that can be improved in future work.

978-1-7281-1450-7/19/$31.00 ©2019 IEEE

II. Related works

The underwater environment is unknown and unstructured. So, it needs to be explored and monitored according to diverse applications. Thus, there is a need for automatic systems that reliably detect, track and classify marine species or objects presented in underwater images or videos without human intervention. It is also crucial improves the efficiency of image analysis, the cost-effectiveness and the availability of numerical data. There are some conventional supervised learning techniques to image classification, such as Minimum Distance, Maximum Likelihood (ML), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) [1]. This kind of techniques requires a training dataset in order to teach the classifier to define the decision boundary. Although these techniques require a large time for the training phase, the errors can be easily identified and solved. The ML is an efficient technique but requires a large computational time. The SVM classifier optimizes the use of training data since it assumes that there is no prior information on how to classify the data. In addition, it is less prone to overfitting and requires less computational memory compared with the previous one. However, it does not suitable for large and noisy datasets. The challenging factors in underwater scenes (such as turbidity of the water, changes in light intensity, changes in the orientation of species or objects and the background confusion) impose a non-linearity in the image, which difficult conventional methods to adapt to the features of the interest objects on the images. In this context, data-driven classification models, like Neural Networks, are more suitable. They are inspired by the human nervous system and they are used to detect trends or patterns. Although require prior training (time-consuming) and present the overfitting problem, these methods present a high computation rate and deal with the noisy inputs efficiently. In [2] and [3] is proposed a classification technique in the natural environment based on capturing the texture pattern and shape of the fish using image processing. Later, a method to fish species classification based on Convolution Neural Network (CNN) together with KNN and SVM was proposed and demonstrated results with a correct classification rate more than 90% [4]. Recent trends are moving toward the use of machine learning algorithms in the video. In [5] is proposed a cross-layer pooling using a pre-trained CNN with underwater video imagery. In this work, classification accuracies of 89% and 96,7% with the proposed dataset and LifeClef dataset, respectively, were obtained. The fish species identification is important for monitoring the status and trends in the relative abundance, composition, size, and biomass fish assemblages. In addition, it is also relevant to detect and recognize the critical objects that there are in underwater environment (like pipelines, mines, anchors, etc) for that being possible to realize inspection tasks even in places that are difficult to reach. However, this area is still under study and, therefore, no state-of-art dataset is available.

III. Object classification

Inherent to object classification process there are three main phases: data acquistion, training and evaluation. Thus, due to a shortage of underwater dataset with critical objects, it was necessary to create a dataset, described in section III-A. Moreover, in section III-B are presented the main steps of developed Convolutional Neural Network and in section III-C are detailed the ResNet and its variations that allowed to evaluate the performance and robustness of the proposed method.

A. Underwater Dataset

One of the most important steps in solving any real-world problem of classification is to get the data. Many times, the quality of input data is the main factor to accuracy of the results. The developed dataset comprises a set of images that includes different man-made objects considered critical in underwater operations. The inspection and monitoring tasks of the seafloor have been more and more conducted by autonomous robotic vehicles and the behaviour of these platforms has to be different in relation to each object. For example, if the robot to find a mine is dangerous and the robot must take a pre-programmed action (to escape or to destroy). Otherwise, if the robot sees a pipeline or a chain, can getting around the object and continue the mission. Thus, it is important that the robots recognize different objects with high accuracy.

In this context, to try represents some common objects that are present in underwater scenarios and to determine in the future an action plan that the robot should apply, the following critical objects to create the dataset were chosen:

- Anchors and chains - Usually, these objects are in the ocean floor and are confused with the background due to its colour and because can be in part hidden (buried in the sand);
- Pipelines and flutuation modules - These objects may be together or separately. The first one presents a grey colour (that hamper the identification) and presents some fluctuation, which means they may not always be static on the bottom of the sea. In contrast, the flutuation modules can have diverse colours and shapes;
- Mines - These objects present a characteristic shape and they are one of the most dangerous objects present in underwater environment. They can be in the ocean floor or afloat;
- Buoys - In general, they are coloured to ensure its visibility conditions by the maritime vehicles, for example, when they are applied to limit an area in the sea. Otherwise, it can be help in the location of the underwater vehicles;
- 3D Marker - This object helps the vehicle navigation, namely in the tracking, and for docking tasks during missions. In addition, it allows to obtain the 3D position of the robot.

The mine and marker were developed with an appropriate scale according to acquisition conditions. The data acquisition phase was divided into two steps: pure water and turbid water. The last one was used to better simulate the real environment since that in the underwater environment the water is not

crystal clear (which difficult the image processing). In both cases, the objects were captured in colour (reef) and white backgrounds. The images were acquired by a Mako G-125 with an appropriate enclosure, in small videos sequences with diverse perspectives of each object. The main features of the developed dataset are summarized in table III-A.

Resolution	Frames/sec	Classes	Rectified
1280x960	12	7	Yes

TABLE I
MAIN FEATURES OF DEVELOPED DATASET.

Figure 1 presents some illustrated perspectives of the critical objects included in the dataset in both conditions (pure and turbid water).

Fig. 1. Some examples of the critical objects in diverse acquisition conditions.

B. Convolutional Neural Network Approach

For many cases of real-world is rare to develop novel architectures, because there are different approaches that can be used to solve the problems. However, sometimes, it is important to build Convolutional Neural Networks to evaluate its performance and computational resources in other contexts and conditions. In networks that only use fully connected layers all nodes in a layer are fully connected to all nodes in the previous layer, producing a complex model. This complexity is bad to training phase and, many times, does not provide additional benefits (many features are localized). Otherwise, CNN applies small size filters and, so the number of training parameters is tiny, allowing CNN to use many filters simultaneously. However, it is important to notice that the depth of the network increases the complexity of the model and the ideal is to maintain a compromise between complexity and performance. The developed network architecture contains a combination of different layers, see figure 2.

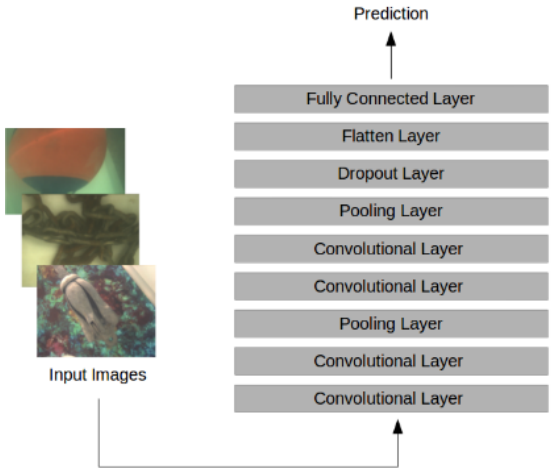

Fig. 2. Process overview of proposed CNN method.

The convolutional layers have as main purpose to extract features from the input image, preserving the spatial relationship between pixels. To reduce the size of features and to ensure that the algorithm not focus on small changes of orientation and position, the pooling layers were added. To improve the model generalization capacity and to prevent overfitting in final results a dropout layer was applied. The process overview is presented in figure 2.

C. Deep Residual Learning Architecture: ResNet

Many times, the CNN has to deal with the degradation problem, that means when the training error increases as depth increases. On the other hand, deeper neural networks are more difficult to train. For these reasons, in [6] was presented this residual learning framework to ease the training of the network. In this work was proved that the extreme deep residual net is easy to optimize, achieving better results up to now than the other deep methods in the literature. Thus, the ResNet-18 was selected to compare with results obtained by developed CNN. This framework allows the use of weights trained on the ImageNet dataset to start the model, which usually improves the accuracy of final results.

IV. RESULTS

A set of experiments using diverse splits from the developed dataset was conducted. In this section, the general performance of the in-house algorithm with a validation set is demonstrated and a comparative analysis with the state-of-art method is also presented (section IV-A). Moreover, the robustness evaluation of the proposed method also was verified (section IV-B). Otherwise, after the initial performance evaluation, the proposed algorithm was tested with the best model. Thus, in each study case, four evaluation parameters were obtained: training accuracy (TA), validation accuracy (VA), test accuracy (TestA) and processing Time (PT).

A. Performance evaluation

In this section, the evaluation of the proposed Convolutional Neural Network with an initial set of images is performed. This set was selected randomly by an appropriated algorithm and the existence of different perspectives between the training and validation set, to avoid overfitting, was the only aspect proved by the user.

The main aspects of this dataset, named DatasetA (on the tables), are:

- Training and validation with mixture of images (pure and turbid water);
- Each class is divided by an 80/20 proportion (train/validation);
- Total training images: 3525;
- Total validation images: 967.

Figure 3 presents the obtained performance in both phases: training and validation, as well as the losses values obtained in the training phase for 50 epochs. The obtained accuracy, in the best epoch, was 76% which means that the algorithm found correctly more or less 735 images.

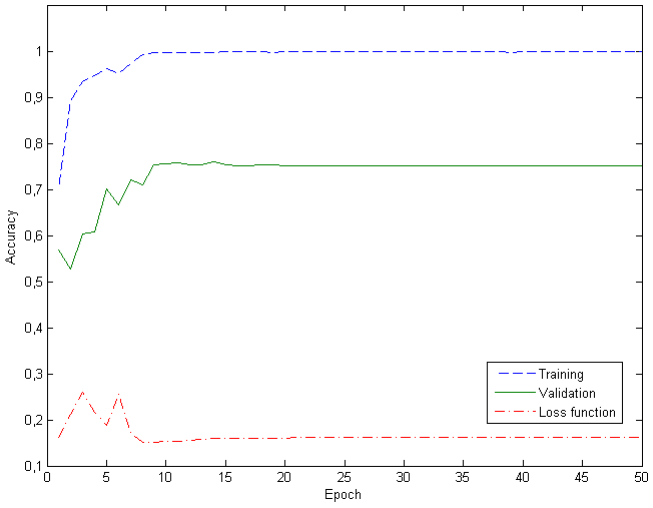

Fig. 3. Process overview of proposed CNN method.

Otherwise, Figure 4 illustrates the behaviour of the object classification along the different epochs. Thus, it is possible to see that the anchors and the pipelines are the classes harder to classify. This aspect can be explained by their grey colour that can be confused with the background in some situations. Conversely, the mine is the object with higher accuracy, which helps to improve the obtained final accuracy. Although also has grey colour, it presents a characteristic shape.

To prove the obtained accuracy results, the dataset was validated by the ResNet algorithm without pre-trained weights (ResNetv1), see table IV-A. In terms of learning conditions, this ResNet variation is the only can be directly compared with the proposed approach since that does not include any a prior knowledge about the objects.

It is possible to conclude that the achieved results by the proposed approach do not distant much in relation to

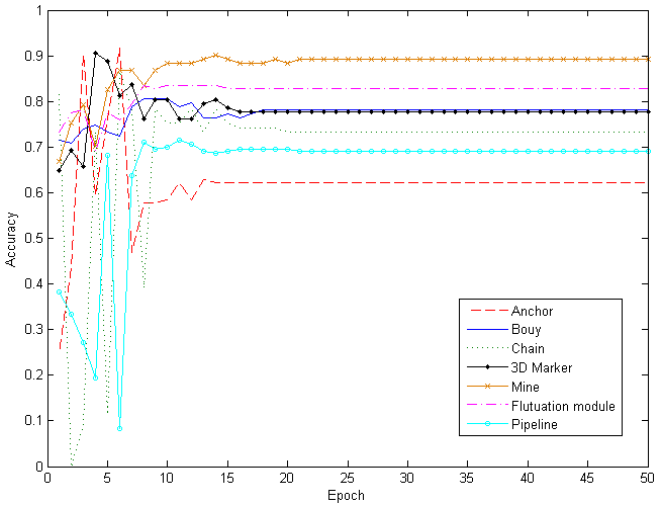

Fig. 4. Process overview of proposed CNN method.

DatasetA	In-House	ResNetv1
TA	99%	99%
VA	76%	86%
PT	37m22s	89m25s

TABLE II
TA, VA AND PT OBTAINED BY PROPOSED METHOD AND RESNETV1 FOR THE DATASETA.

ResNet results. Otherwise, in terms of the processing time, the ResNet method requires more than double time. Thus, this fact has a significant impact on robotic applications that involves real-time operations. Once again, the anchors and pipelines were the classes that obtained the worst accuracy results, as expected. Thus, to evaluate the proposed algorithm performance without these classes the training and validation phases are again performed but, now, with only 5 object classes. In this case, the performance of the developed method increases to 84% and the ResNet performance increases to 96% (about 10% in both cases).

B. Robustness evaluation

To evaluate the robustness of the proposed approach to a new condition, a novel dataset (named Dataset B) was created to difficult the classification process. Therefore, was possible to train the model with only images acquired in ideal conditions (pure water, without suspensions) and test in turbid water to simulate the real conditions. The proportion of training and validation images was maintained. This experiment is the more complex (illustrating the worst conditions for this evaluation), so is expected that the performance decreases significantly. Table IV-B presents the comparative results between the proposed method and three ResNet variations: ResNet without weights pre-trained (ResNetv1), ResNet with weights pre-trained but only training the last layer - Fully Connected (ResNetv2) and ResNet with weights pre-trained (ResNetv3).

As expected, the validation accuracy decreases sharply in all cases except in ResNetv3 that presents 94% of accuracy.

DatasetB	InHouse	ResNetv1	ResNetv2	ResNetv3
TA	81%	98%	99%	99%
VA	25%	47%	30%	94%

TABLE III
TA AND VA OBTAINED BY PROPOSED METHOD AND RESNET VARIATIONS FOR THE DATASETB.

In this case, the fact of the training is performed in all layers allowed the learning of new features about the different objects, during the epochs. Moreover, the use of weights pre-trained helps in the initial step since some interest features are included, such as limits and shapes (that can be similar to objects present in the developed dataset). This fact can also be observed in the ResNetv2 results (30%) that presents the worst accuracy, as expected because only the last layer is trained. So, this approach does not learn new features along with the training phase. However, the initial values demonstrated some knowledge about the classes, which allowed to achieve the obtained accuracy.

C. Object classification

Until now, the developed approach only was validated. But a crucial step is the prediction using the best model obtained from the best epoch (in a total of 30 epochs) in the validation phase. Thus, a new dataset (named Dataset C) with the train images selected randomly was created and, so all conditions were included: pure and turbid water, white and colored (reef) background. In the validation set, only images in pure water were used and in test set only images in turbid water were used (to simulate the real test conditions). For this case, the data was split as follows: 70%, 20% and 10% for training, validation and test phase, respectively. Table IV-C summarises the obtained results for this experiment.

DatasetC	InHouse	ResNetv1
TA	97%	98%
VA	75%	91%
TestA	74%	76%

TABLE IV
TA, VA AND TESTA OBTAINED BY PROPOSED METHOD AND RESNETV1 VARIATIONS FOR THE DATASETC.

In the test phase, the developed method maintained the accuracy obtained in the validation phase (74%). In contrast, the ResNetv1 (without weights pre-trained) presented a good accuracy in the validation phase, but in the test phase, its accuracy decreased 15%. This situation can be explained by difficulty present in the test images (turbid water and some in reef background). Moreover, probably these type of images in the training set were not enough to the model learning the intended features. Lastly, was created the DatasetD with images selected by the user. It was attempted that the training set was heterogeneous as possible, to learn correctly the main features of the objects in different conditions. Moreover, the validation and test set were selected according to the training, but with other perspectives (to avoid the overfitting). In addition, it was intended that the majority of images included in the test set contemplates the simulated real conditions. TaIV-C presents the behaviour of the proposed method and ResNETv1.

DatasetD	InHouse	ResNetv1
TA	98%	98%
VA	70%	87%
TestA	67%	92%

TABLE V
TA, VA AND TESTA OBTAINED BY PROPOSED METHOD AND RESNETV1 VARIATIONS FOR THE DATASETD.

It is possible to observe that the performance of In-house implementation decreased. This can be explained by the difficult imposes in the test set. Otherwise, the ResNetv1 presents a higher validation and test accuracy. Thus, it is noticeable its learning capacity during the training phase, which is reflected in the final performance (92%).

Finally, two study cases were conducted:

1) In the training phase was used images with white background (pure and turbid water) and the validation was performed with reef background (pure and turbid water). The results showed the impact of the training phase in the classification process since the validation accuracy decrease significantly in both methods: the In-house approach presented 20% of accuracy and ResNetv1 presented 37% of accuracy (a decrease of 50% in both cases).
2) The training was performed with only 6 objects and 7 classes as input were used . In this case, the In-house approach achieved a 51% of accuracy and the ResNetv1 achieved 64% of accuracy. Thus, it is crucial that the approaches include mechanisms that are able to conclude that there is a new object in the test phase (but not included in training).

V. CONCLUSIONS AND FUTURE WORK

From the obtained results, it is possible to conclude:

- The performance of the developed method is good enough to recognize correctly the objects in most of the conditions with an accuracy of 75%. Although all classes can be classified by the developed method, there are some classes where this task is more difficult. This fact can be explained by the colour that can be confused with the background (reef). The object classification without these classes (anchors and pipeline) improves to 84%;
- The obtained results by the ResNet without weights pre-trained (ResNetv1) can be compared with the developed method since that in both situations the weighs are initialized randomly. The performance of this ResNet variation is better but presents a higher computational cost and requires more processing time, which sometimes is very worrisome, namely in real time operations;
- ResNet with weights pre-trained (ResNetv3) presents an increase in performance. Otherwise, the obtained results in ResNetv2 proves that the initial weights are very

important during the classification process, since that allows to obtain an accuracy of the 30%, even without training of the convolutional layers;
- In the robustness evaluation, the results demonstrated that the training in ideal conditions and the validation and test in the simulated real conditions does not work correctly, as expected. In general, the underwater classification difficult this task, namely in the turbid water (a situation that most resembles the real environments - causes unfocused limits on the objects). In that extreme situation the more appropriate way to solve the problem is training the model with the ResNet with weights pretrained (ResNetv3) when the computational cost and processing time are not a problem;
- In the final experiment, it was concluded that the developed method can be used in many situations since the training was performed with heterogeneous images. So, the model learning the features that better differentiate the objects;

Many times, a good accuracy of classification methods comes at the cost of high computational complexity, which implies the use of graphics processing units (GPU). So, nowadays the works should be focused on inproving the energy-efficiency without sacrificing the accuracy. Thus, to evaluate the properties of a neural model should be considered some metrics: network architecture (number of layers, filter sizes, etc), number of weights and the accuracy of the model. In robotic applications, it is crucial the existence of a balance between computational resources and precision in the results.

To future work, the authors will conduct novel datasets including new objects and conditions to simulate more and more the challenges present in the underwater environment. These datasets are crucial because they allow supporting and help the scientific community on the development of approaches for diverse applications. In addition, the authors intend to improve the proposed method, to apply in real environments, namely the recognizing of new objects and introducing mechanisms that are able to conclude that there is a new object in the test phase but not included in the training. Together with new assumptions intended, it is objective to maintain a reduced computational cost to be possible to apply the method in robotic platforms.

REFERENCES

[1] N. Thakur and D. Maheshwari. “A Review on Image Classification Approaches and Techniques”, International Research Journal of Engineering and Technology (IRJET), vol. 04(11), pp. 1588–1591, 2017.

[2] A. Rova, G. Mori and L. M. Dill, “Automatic fish classification for underwater species behavior understanding,”ACM Workshop on Analysis And Retrieval of Tracked Events and Motion in Imagery Streams, pp. 45–50, 2010.

[3] C. Spampinato, D. Giordano, R.D. Salvo, Y.H. Chen-Burguer, R.B. Fisher and G. Nadarajan, “Automatic fish classification for underwater species behavior understanding,” in ACM Workshop on Analysis And Retrieval of Tracked Events and Motion in Imagery Streams, pp. 45–50, 2010.

[4] A. Salman, M. Shortis, J. Seager and E. Harvey, “Fish species classification in unconstrained underwater environments based on deep learning,” in Limnology and Oceanography: Methods, vol. 14, pp. 570–585, 2016.

[5] S. Siddiqui, A. Salman, I. Malik, F. Shafait, A. Mian, M. Shortis and E. Harvey, “Automatic fish species classification in underwater videos: Exploiting pretrained deep neural network models to compensate for limited labelled data,” in ICES Journal of Marine Science, in press, vol. 77, pp. 374–389, 2018.

[6] K. He, X. Zhang, S. Ren and J. Sun I., “Deep Residual Learning for Image Recognition,” in Microsoft Research, pp. 1–12, 2015.