Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 5741–5746

Detection and classification of man-made objects for the autonomy of underwater robots

A. Gentili *,** M. Bresciani *,** F. Ruscio *,** S. Tani *,** A. Caiti *,**,*** R. Costanzi *,**,***

* *Dipartimento di Ingegneria dell'Informazione - Università di Pisa, Pisa, Italy*
** *Interuniversity Center of Integrated Systems for the Marine Environment (ISME), Italy*
*** *Centro di Ricerca "E.Piaggio" - Università di Pisa, Pisa, Italy*

Abstract: Recent developments in marine technologies allow underwater vehicles to perform survey missions for data collection in an automatic way. The scientific community is now focusing on endowing these vehicles with strong perception capabilities, aiming at full autonomy and decision-making skills. Such abilities would bring benefits to a wide range of field applications, e.g. Inspection and Maintenance (I&M) of man-made structures, port security, and marine rescue. Indeed, most of these tasks are currently carried out employing remotely operated vehicles, making the presence of humans in water necessary. Projects like Metrological Evaluation and Testing of Robots in International CompetitionS (METRICS), funded by the European Commission, are promoting research on this field by organising events such as the Robotics for Asset Maintenance and Inspection (RAMI) competition. In particular, this competition requires participants to develop perception techniques capable of identifying a set of specific targets. Within such context, this paper presents an algorithm able to detect and classify Objects of Potential Interest (OPIs) in underwater camera images. First, the proposed solution compensates for the quality degradation of underwater images by applying color enhancement and restoration procedures. Then, it exploits deep-learning techniques, as well as color and shape based methods, to recognize and correctly label the predefined OPIs. Preliminary results of the implemented neural network using restored images are provided, and a mean Average Precision (mAP) of about 92% was achieved on the dataset provided to the RAMI competition participating teams by the NATO Science and Technology Organization Centre for Maritime Research and Experimentation (STO CMRE).

Keywords: Neural Networks, Perception and Sensing, Intelligent Autonomous Vehicles, Autonomous Underwater Vehicles, Underwater Robotics.

1. INTRODUCTION

In the past decades, researchers focused their studies on the development and optimisation of Autonomous Underwater Vehicles (AUVs) able to autonomously navigate and perform data collection missions in underwater environment. Despite satisfying results were achieved in this field, tasks performed by AUVs are usually based on predefined missions characterised by poor decision-making capabilities. Therefore, underwater robots are often joined by divers when performing Inspection & Maintenance (I&M) tasks in complex scenarios, risking jeopardising the safety of the latter due to the hazards of the environment in which the operations are conducted. For this reason, the ability to perform fully autonomous decisions to successfully complete a mission is becoming a key requirement for AUVs in many industrial scopes, such as I&M of man-made infrastructures, environmental monitoring, disaster response intervention, and port protection. Within this scope, the Metrological Evaluation and Testing of Robots in International CompetitionS (METRICS (2022)) project, funded by European Union's Horizon 2020 research and innovation program, organises robotics competition to promote and motivate research in this field. In particular, the first virtual edition of the Robotics for Asset Maintenance and Inspection (RAMI) competition (Ferri et al. (2021)) took place this year (July 2022) and was led by NATO Science and Technology Organization Centre for Maritime Research and Experimentation (STO CMRE). The competition required participating teams to develop detection and classification algorithms capable of identifying a predefined set of Objects of Potential Interest (OPIs), consisting of colored buoys, digits and red markers used to simulate hazardous and hostile environments where human intervention is impracticable.

Due to the characteristics of the underwater environment, developing an accurate object detection and classification

⋆ Work partially supported by the Italian Ministry of Education and Research (MUR) in the framework of the FoReLab and CrossLab projects (Departments of Excellence).

10.1016/j.ifacol.2023.10.532

algorithm able to deal with images obtained by underwater cameras is a challenging task. Indeed, underwater images are affected by light attenuation and scattering phenomena, resulting in limited distance visibility, contrast degradation, hazy, and greenish or bluish scenes. Existing image processing techniques for improving the quality of underwater images can be grouped into two classes: image restoration and image enhancement. The former attempts to remove haze from images, relying on the physics of light propagation and exploiting the similarities between light propagation in fog and underwater. In particular, the Dark Channel Prior (DCP) (He et al. (2011)), based on the studies of Jaffe (1990), was introduced and several algorithm DCP-based were developed. DCP is based on the observation that, in most of the non-sky patches of outdoor haze-free images, some pixels often have low intensity in at least one color channel. Thus, in hazy images, the main intensity contribution is done by the atmospheric light. Based on the observation that in underwater environment the attenuation of the red channel is very strong and that the predominant source of information lies in the blue and green color channels, Drews Jr et al. (2013) proposed to exploit only the latter channels to estimate and remove the image haze. Galdran et al. (2015) developed the Red Channel Prior (RCP) relying on the observation that red intensity rapidly decays and its weight in the red Channel image decreases. Instead, Luczynski and Birk (2018) introduced the underwater-ready DCP, that shifts the color space (so that blue becomes white) before estimating the underwater atmospheric light.
On the other hand, image enhancement improves contrast and colors of images exploiting pixel intensity redistribution and white-balancing techniques; it uses qualitative subjective criteria to produce a more visually pleasing image. Among these methods, Contrast Limited Adaptive Histogram Equalisation (CLAHE) (Zuiderveld (1994)) and γ-correction, are the most used algorithm to improve contrast and visibility in low-light images. Gray World Assumption (GWA), developed by Buchsbaum (1980), is one of the most used white-balancing process, together with Gray-Edge Assumption (GEA) introduced by van de Weijer et al. (2007). However, due to the low intensity of RGB components occurring in underwater images, these methods could introduce artifacts, halos, and color distortion. In the attempt to minimize the occurrence of these issues, Codruta et al. (2017) showed that better results are obtained if a red and blue channel pre-compensation is applied before the white-balancing procedure.

Several efforts were made trying to overcome the challenges that underwater object detection and classification algorithms have to face. Vasamsetti et al. (2018) proposed a detector based on the collection of space-time data textures from successive frames. Zhou et al. (2015) and Seese et al. (2016) employed Gaussian Mixture Models (GMMs) for background modelling and blob analysis to recognise fish, jellyfishes, and sea snakes. Bazeille et al. (2012) developed a color matching algorithm to identify man-made objects, based on the modeling of the color modification by the water, while Hou et al. (2016) developed a detection method based on color and shape features. Nevertheless, due to the extreme variability and inconstancy of the underwater environment conditions, it is hard to obtain good performance of detection and classification solutions relying only on geometric and physical characteristics, such as shape and color. Artificial Intelligent (AI)-based techniques were therefore employed in an effort to address such problems. For example, Mahmood et al. (2016) proposed the first application of deep learning to the coral reef classification problem using a pre-trained VGGNet. The ZooplanktonNet, inspired by VGGNet, were developed by Dai et al. (2016) to detect and classify zooplankton. Recently, many attempts were carried out on fish species classification problem, exploiting Faster Region-based Convolutional Neural Network (R-CNN) (Ren et al. (2015)) and You Only Look Once (YOLO v3) (Redmon and Farhadi (2018)). In particular, Sung et al. (2017), proposed a real-time fish detection algorithm based on YOLO, achieving a classification accuracy of 93%. Zeng et al. (2021) utilised Faster R-CNN together with adversarial occlusion network to improve the performance of a classic Faster R-CNN on the marine organism detection problem. Recently, employing the Mask R-CNN, proposed by He et al. (2017), Conrady et al. (2022) developed an automated detection and classification algorithm for African Roman sea bream.

Many results were achieved exploiting deep neural network and computer vision-based techniques in the field of marine biology. Furthermore, the application of such solutions to the problem of detection and classification of artificial structures is getting more and more attention from the scientific community. Some relevant results were presented in Zacchini et al. (2020). Two neural networks inspired by the Single Shot Multibox Detector (SSD) and the Faster R-CNN were trained and validated to detect artificial structures from acoustic and optical images, respectively. Other results were achieved by employing deep learning solutions on acoustic images obtained from a Forward-Looking Sonar (FLS). In Palomeras et al. (2022) the authors implemented a detection and classification algorithm based on CNNs and probabilistic maps capable of identify mine-like objects, while in Zacchini et al. (2022) the authors compared the performances of an SSD and a Mask R-CNN aiming at identifying and localizing OPIs in FLS imagery.

This work proposes an object detection and classification solution implementing deep-learning, as well as color-based, and shape-based algorithms, aimed to recognize OPIs in underwater environment. The dataset used during the development of the algorithm is the one provided by NATO STO CMRE to the RAMI competition participating teams as support for software development. The paper is organised as follows: in Section 2 the dataset is illustrated and the proposed algorithm pipeline is described, as well as details on the Neural Network (NN) fine-tuning procedure. In Section 3, preliminary results about the performance of the implemented NN are presented and commented. At last, in Section 4 conclusions are drawn and future developments are proposed.

2. DETECTION AND CLASSIFICATION

2.1 Dataset description

The dataset provided as support for the software development is composed of 957 images in `.png` format, with dimension of 1280×720. All the images were acquired in underwater environment, in different conditions, at NATO STO CMRE in La Spezia, Italy. The whole dataset is divided into five OPIs classes:

- `class 1`: colored buoys;
- `class 2`: black numbers on yellow pipes;
- `class 3`: black numbers over red markers;
- `class 4`: red markers over yellow pipes;
- `class 5`: no OPI.

Specifically, `class 1` contains 253 images of underwater buoys characterised by three different colors: 121 red buoys, 86 white buoys, and 46 yellow buoys. `class 2` includes 288 images of four different black digits over yellow pipes: 52 images with digit "1", 56 with digit "2", 85 with digit "3", and 95 with "4". `class 3` consist of 84 images with four different black digits located over red markers. In this case, the dataset involves only sample images with digit "6". Regarding digits "3", "4", and "5", only single digital drawings are provided. Lastly, 80 images of red markers comprises `class 4`, while 252 images without any OPI constitutes `class 5`. In Figure 1, samples extracted from the dataset are shown.

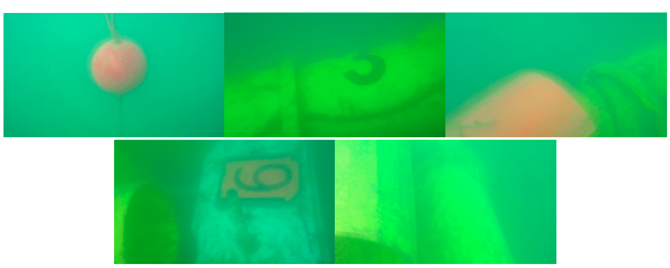


Fig. 1. Sample images extracted from the dataset.

2.2 Algorithm overview

In this section, an overview of the algorithm pipeline, summarised in Figure 2, is presented. At first, the input image is pre-processed aiming at compensating the quality degradation mainly caused by particle suspension and scattering phenomena. This was done to improve the performance of the subsequently applied techniques.

As depicted in Figure 3, the color reconstruction stage is composed by four different steps. As done by Codruta et al. (2017), a red and blue channels compensation is executed before the white-balancing procedure, so that the occurrence of red artifacts is minimised. Then, GWA (Buchsbaum (1980)) is applied with the aim to: computing the white point, shifting the mean value of each RGB channel and compensating the color cast induced by the underwater environment. To solve contrast degradation and haziness, two further algorithms are applied in series: through the CLAHE algorithm (Zuiderveld (1994)) the image contrast is increased, while, applying the DCP algorithm (He et al. (2011)) the image haze is reduced. After the color processing stage, an R-CNN inspired by the Faster R-CNN (Ren et al. (2015)), is fine-tuned and used to predict the probability that one or more OPIs are present in the image. This network outputs the pixel coordinates in the image plane of the bounding box containing the OPI and the associated discrete probability that the detected object belongs to a target class. Due to the limited number of images contained in the dataset, transfer learning and

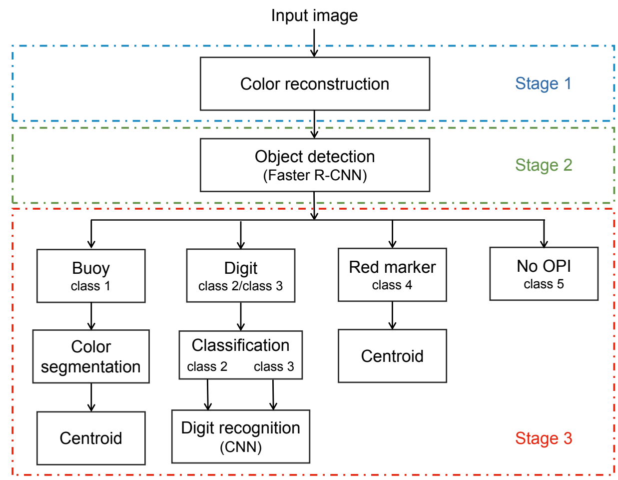


Fig. 2. Scheme of the proposed algorithm for OPIs detection and classification.

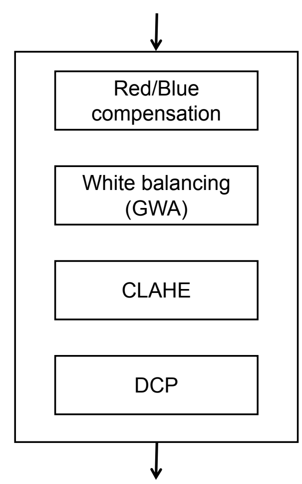


Fig. 3. Steps of the color reconstruction algorithm.

fine-tuning procedures were applied. With this aim, the OPIs were distinguished into the following classes: i) colored buoys, ii) a unique class composed of OPIs with black numbers either on yellow pipes or on red markers, iii) red markers, and iv) no OPIs. In the event that at least one of the OPIs is detected, by the Faster R-CNN, the patches containing the object are extracted. Hence, color and shape based algorithms are applied aiming at identifying the exact instance of the recognised object.
In particular:

- If a buoy (`class 1`) is detected, a Hue-Saturation-Value (HSV) filter, together with checks on the roundness of the located object, are employed to estimate the buoy color. Then, the position in pixels of the OPI centroid is estimated applying the Hough circle transform (Hough (1962));

- If a digit is found (`class 2` or `class 3`) an HSV filter is employed to detect the presence of red color in the extracted patch and to figure out whether the digit is located on a yellow pipe or on a red marker (`class 2` and `class 3`, respectively). At last, to recognise the digit, the image patch is binarised through Otsu's method (Otsu (1979)) and a CNN, trained on MNIST (Lecun et al. (1998)) and fine-tuned through a dataset created from-scratch exploiting the provided data, is used;

- If a red marker is detected (`class 4`), an HSV filter is applied to the extracted patch to detect the presence of red color and to refine the selection rectangle from which the OPI centroid is estimated;
- Lastly, if none of the OPIs is found, the image is classified as belonging to `class 5`.

2.3 *Faster R-CNN fine-tuning*

In this work, a `keras` (Chollet et al. (2015)) implementation of the Faster R-CNN, based on the code released by Ren et al. (2015), was used and the fine-tuning procedure was carried out exploiting Google Colaboratory (Bisong (2019)) and the GPU made available in this framework (NVIDIA Tesla T4). Moreover, a ResNet50 (He et al. (2015)) pre-trained on ImageNet (Deng et al. (2009)) was employed as backbone of the Faster R-CNN. As first step, the provided dataset was enriched by doubling it with a processed version of the available images, using the color restoration and color enhancement algorithms described in section 2.2. Then it was entirely labeled and ground-truth boxes were defined through an open-source labelling tool, named LabelImg (Tzutalin (2015)). The latter automatically creates an annotation file in `.xml` format for each image, in which labels and ground-truth bounding boxes coordinates are saved. The available dataset was split into training, validation and test set with a ratio of 70:20:10, respectively. Given the limited number of available image samples, data augmentation techniques, i.e. random rotation, vertical flip and horizontal flip, together with early-stopping procedures were employed to increase the training set and to prevent over-fitting. At last, the test set was used to evaluate the performance of the NN. Table 1 shows the hyperparameters selected for the training procedure.

Table 1. Training hyperparameters

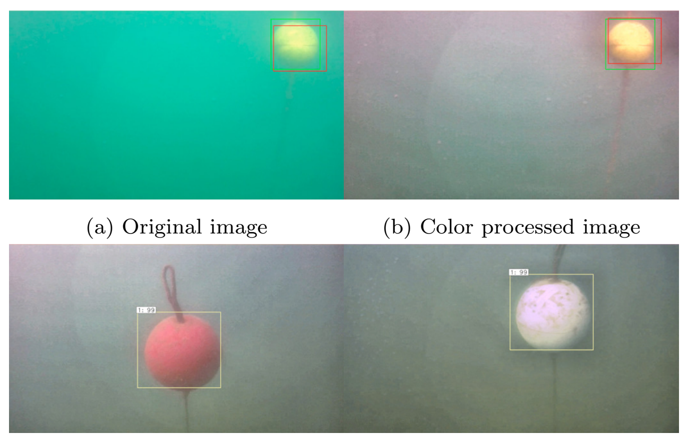
Hyperparameters	Values
Image size	[416,416]
Anchor box scales	128, 256
Anchor box ratios	[1,1], [1,2], [2,1]
Learning rate	$1e-5$
Optimizer	Adam
Epochs	37

3. RESULTS

In this section preliminary results regarding the detection and classification performance of the Faster R-CNN network on the compensated images are presented and commented. To rank the network results, Average Precision (AP) and mean Average Precision (mAP) were used according to PASCAL VOC evaluation metrics (Everingham et al. (2015)). These metrics, here computed exploiting the tool provided by Padilla et al. (2021), are generally used by the research community to evaluate and quantify the performance of detection algorithms in different fields.

3.1 *Class 1: buoys*

Figure 4 shows some prediction results for `class 1`, where the ground-truth and prediction bounding boxes are depicted in green and red, respectively, in Figure 4a and 4b. As it can be seen by Figure 4c and 4d, the Faster R-CNN is able to detect and classify the buoys despite the different environment conditions and their colors. The network achieved an average prediction time of about 3 frame per second for each input image and an AP of about 98%. It can be noticed how the color processing algorithm helps obtaining a slightly more accurate predicted bounding boxes with respect to the ground-truth one, increasing the intersection over union value from 71.5% to 73.8%. Furthermore, Figure 4c shows how the deep-learning approach is able to accurately detect the buoys despite of the cable-mooring presence. The parameters for the adjustment of the red and blue channels were tuned on the entire dataset. This could result in a red channel overcompensation, as visible in the brightest areas of Figure 4b and 4c.



(a) Original image (b) Color processed image

(c) Prediction on red buoy (d) Prediction on white buoy

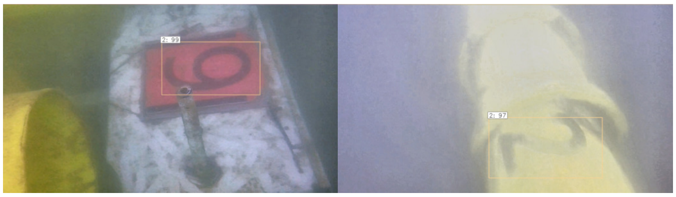
Fig. 4. Faster R-CNN results on the `class 1` OPIs.

3.2 *Class 2 and class 3: digits*

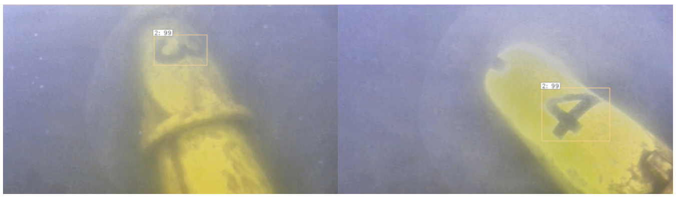
For what concerns the OPIs belonging to `class 2` and `class 3`, considered as a unique class labelled as "2", the Faster R-CNN achieved an AP of about 87%, resulting in less accurate performance with respect to the buoy detection. This could be caused by the presence of marine organisms and dirt on the pipes that make the task more challenging. However, results show the ability of the network to correctly classify the OPIs despite different shape and orientation of the digits, as well as high variability of the surrounding environment. In Figure 5 some examples of prediction for the digit class are shown. Furthermore, as shown in Figure 6, the color correction process yields to an improvement of the object detection performance, raising the intersection over union value from 62.9% to 74.2%.

3.3 *Class 4: red markers*

The network obtained a class AP of about 92% on the detection and classification of the red markers of `class 4`.



(a) Class 3 digit ”6” (b) Class 2 digit ”2”



(c) Class 2 digit ”3” (d) Class 2 digit ”4”

Fig. 5. Examples of network prediction on digits.

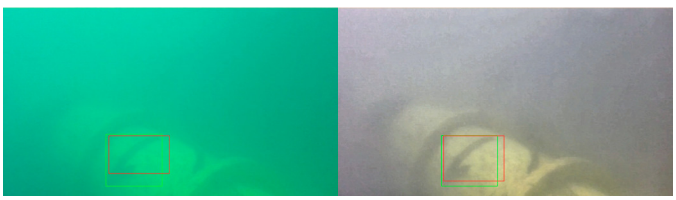


Fig. 6. Comparison between predicted bounding boxes in original (left) and corrected (right) images containing `class_2` OPIs.

Furthermore, Figure 7 shows that the solution is able to detect the OPIs even if their area varies considerably and they are partially out of the field of view of the camera in most of the images used for the fine-tuning procedure. Results on the detection of red markers show that the network achieved smaller AP values with respect to buoys detection. Probably due to the limited number of images (80) present in the dataset for this class.

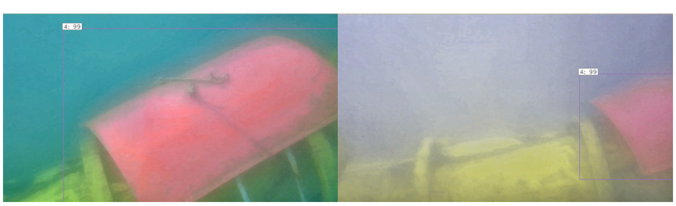


Fig. 7. Examples of prediction on `class_4` samples.

3.4 Class 5: no OPIs

At last, some tests were carried out exploiting images without any OPIs, aiming at quantifying the occurrence of False Positives (FPs). As aforementioned, the strategy adopted was to distinguish this class of images by exclusion among the others, i.e. an image is classified as belonging to `class_5` if the Faster R-CNN does not detect any OPI. Thus, all the images belonging to `class_5` provided by the dataset were duplicated exploiting the color reconstruction algorithm, described in Section 2.2, and were then applied as input to the network. As a result, we have obtained an overall accuracy of about 86% with 70 FPs predictions over 504 samples, of which 46 produced by the 252 original images, and 24 produced by the 252 color-corrected ones. This means that by applying the color correction algorithm to the input images before performing the prediction, the performance of the network improves. Therefore, computing the accuracy on just the corrected images, it increases from 82% to 90%. Figure 8 demonstrates how the network output changes if the input image were previously corrected, thus avoiding a false positive detection.

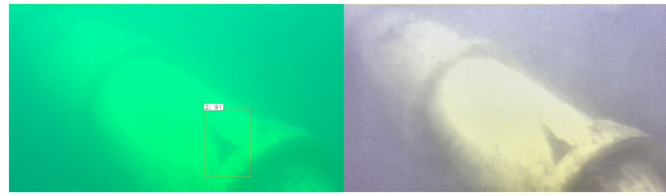


Fig. 8. Example of how the color correction algorithm allows to avoid a false positive OPI in underwater images.

4. CONCLUSION & FUTURE WORKS

This paper proposes an algorithm that exploits AI-based computer vision techniques together with physical and geometrical characteristics of target objects to detect and classify OPIs in underwater environment. Due to the challenges faced by underwater cameras, the acquired images are firstly processed using color restoration and color enhancement methods with the goal of compensate the quality degradation. The color-reconstructed images are then applied as input to a detection and classification algorithm that employs deep-learning methods, as well as color and shape based techniques to recognise and correctly label the detected OPI. Preliminary results about the implemented Faster R-CNN show that the network achieves a class AP of about 98%, 87%, and 92% in detecting buoys, digits and red markers, respectively. Hence, yielding to a mAP over all the classes of about 92%. Tests performed on both color-processed and original images shown that color restoration and color enhancement are effective methods capable of increasing the detection accuracy and decrease the amount of occurring false positive.

Future works will focus on assessing the performance of the entire presented algorithm and on the development of real-time detection and classification capability. Moreover, other networks (e.g. belonging to YOLO-family) will be tested and the results compared with the one obtained for the Faster R-CNN. Since results showed that color processing could be used to improve the classification performance, the development of a refined and computationally efficient color reconstruction method will be considered.

It is worth mentioning that the described algorithm was submitted by Università di Pisa to the 1st Marine Cascade Campaign: the virtual edition of the RAMI competition (METRICS (2022)), achieving the second place. [1]

REFERENCES

Bazeille, S., Quidu, I., and Jaulin, L. (2012). Color-based underwater object recognition using water light attenuation. *Intelligent Service Robotics*, 5(2), 109–118.

[1] The algorithm scored 107 out of 125 points, and details about each assignment can be found on the official website of the competition: https://metricsproject.eu/inspection-maintenance/rami-cascade-campaign-results-marine/.

Bisong, E. (2019). *Google Colaboratory*, 59–64. Apress.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1), 1–26.

Chollet, F. et al. (2015). Keras. `https://keras.io`.

Codruta, A., Ancuti, C., Vleeschouwer, C., and Bekaert, P. (2017). Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, PP, 1–1.

Conrady, C.R., Şebnem Er, Attwood, C.G., Roberson, L.A., and de Vos, L. (2022). Automated detection and classification of southern african roman seabream using mask r-cnn. *Ecological Informatics*, 69, 101593.

Dai, J., Wang, R., Zheng, H., Ji, G., and Qiao, X. (2016). Zooplanktonet: Deep convolutional network for zooplankton classification. In *OCEANS 2016 - Shanghai*, 1–6.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Drews Jr, P., do Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). Transmission estimation in underwater single images. In *2013 IEEE International Conference on Computer Vision Workshops*, 825–830.

Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1), 98–136.

Ferri, G., Ferreira, F., Faggiani, A., and Fabbri, T. (2021). From erl to rami: Expanding marine robotics competitions through virtual events.

Galdran, A., Pardo, D., Picón, A., and Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, 26, 132–145.

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 2341–2353.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

Hou, G.J., Luan, X., Song, D.L., and Ma, X.Y. (2016). Underwater Man-Made Object Recognition on the Basis of Color and Shape Features. *Journal of Coastal Research*, 32(5), 1135 – 1141.

Hough, P.V. (1962). Method and means for recognizing complex patterns.

Jaffe, J. (1990). Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2), 101–111.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

Luczynski, T. and Birk, A. (2018). Underwater image haze removal and color correction with an underwater-ready dark channel prior. *CoRR*, abs/1807.04169.

Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., and Fisher, R.B. (2016). Coral classification with hybrid feature representations. In *2016 IEEE International Conference on Image Processing (ICIP)*, 519–523.

METRICS (2022). Metrological evaluation and testing of robots in international competitions. URL `https://metricsproject.eu/contact/`.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.

Padilla, R., Passos, W.L., Dias, T.L.B., Netto, S.L., and da Silva, E.A.B. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, 10(3).

Palomeras, N., Furfaro, T., Williams, D.P., Carreras, M., and Dugelay, S. (2022). Automatic target recognition for mine countermeasure missions using forward-looking sonar data. *IEEE Journal of Oceanic Engineering*, 47(1), 141–161.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks.

Seese, N., Myers, A., Smith, K., and Smith, A.O. (2016). Adaptive foreground extraction for deep fish classification. In *2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*, 19–24.

Sung, M., Yu, S.C., and Girdhar, Y. (2017). Vision based real-time fish detection using convolutional neural network. In *OCEANS 2017 - Aberdeen*, 1–6.

Tzutalin (2015). Labelimg git code. URL `https://github.com/tzutalin/labelImg`.

van de Weijer, J., Gevers, T., and Gijsenij, A. (2007). Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9), 2207–2214.

Vasamsetti, S., Setia, S., Mittal, N., Sardana, H.K., and Babbar, G. (2018). Automatic underwater moving object detection using multi-feature integration framework in complex backgrounds. *IET Computer Vision*, 12, 770–778(8).

Zacchini, L., Ridolfi, A., Topini, A., Secciani, N., Bucci, A., Topini, E., and Allotta, B. (2020). Deep learning for on-board auv automatic target recognition for optical and acoustic imagery. *IFAC-PapersOnLine*, 53(2), 14589–14594.

Zacchini, L., Topini, A., Franchi, M., Secciani, N., Manzari, V., Bazzarello, L., Stifani, M., and Ridolfi, A. (2022). Autonomous underwater environment perceiving and modeling: An experimental campaign with feel-hippo auv for forward looking sonar-based automatic target recognition and data association. *IEEE Journal of Oceanic Engineering*, 1–20.

Zeng, L., Sun, B., and Zhu, D. (2021). Underwater target detection based on faster r-cnn and adversarial occlusion network. *Engineering Applications of Artificial Intelligence*, 100, 104190.

Zhou, H., Llewellyn, L., Wei, L., Creighton, D., and Nahavandi, S. (2015). Marine object detection using background modelling and blob analysis. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 430–435.

Zuiderveld, K. (1994). Viii.5. - contrast limited adaptive histogram equalization. In P.S. Heckbert (ed.), *Graphics Gems*, 474–485. Academic Press.