RESEARCH

Deep learning based deep-sea automatic image enhancement and animal species classification

Vanesa Lopez-Vazquez[1*], Jose Manuel Lopez-Guede[2], Damianos Chatzievangelou[3] and Jacopo Aguzzi[4,5]

*Correspondence:
vlopez017@ikasle.ehu.eus

[1] University of the Basque
Country (UPV/EHU), Nieves Cano,
12, 01006 Vitoria-Gasteiz, Spain
[2] Department of System
Engineering and Automation
Control, University of the Basque
Country (UPV/EHU), Nieves Cano,
12, 01006 Vitoria-Gasteiz, Spain
[3] Department of Renewable
Marine Resources, Instituto de
Ciencias del Mar (ICM-CSIC),
08003 Barcelona, Spain
[4] Instituto de Ciencias del Mar
(ICM-CSIC), 08003 Barcelona,
Spain
[5] Stazione Zoologica
of Naples (SZN) Anton Dohrn,
80121 Naples, Italy

Abstract

The automatic classification of marine species based on images is a challenging task for which multiple solutions have been increasingly provided in the past two decades. Oceans are complex ecosystems, difficult to access, and often the images obtained are of low quality. In such cases, animal classification becomes tedious. Therefore, it is often necessary to apply enhancement or pre-processing techniques to the images, before applying classification algorithms. In this work, we propose an image enhancement and classification pipeline that allows automated processing of images from benthic moving platforms. Deep-sea (870 m depth) fauna was targeted in footage taken by the crawler "Wally" (an Internet Operated Vehicle), within the Ocean Network Canada (ONC) area of Barkley Canyon (Vancouver, BC; Canada). The image enhancement process consists mainly of a convolutional residual network, capable of generating enhanced images from a set of raw images. The images generated by the trained convolutional residual network obtained high values in metrics for underwater imagery assessment such as UIQM (~ 2.585) and UCIQE (2.406). The highest SSIM and PSNR values were also obtained when compared to the original dataset. The entire process has shown good classification results on an independent test data set, with an accuracy value of 66.44% and an Area Under the ROC Curve (AUROC) value of 82.91%, which were subsequently improved to 79.44% and 88.64% for accuracy and AUROC respectively. These results obtained with the enhanced images are quite promising and superior to those obtained with the non-enhanced datasets, paving the strategy for the on-board real-time processing of crawler imaging, and outperforming those published in previous papers.

Keywords: Underwater image enhancement, Deep learning, Neural networks, Image classification

Introduction

The use of Machine Learning (ML) techniques by Artificial Intelligence (AI) is growing at a constant pace in several scientific fields, with applications in medicine [15], [55], agriculture [23], industry [53], and marine ecology [6, 7, 59, 81]. As a matter of fact, the implementation of AI-based algorithms for tracking and classification of animals in seafloor (i.e., benthic) realms has grown spectacularly in the past decade both in data from

cabled observatories [60, 61, 71, 105], Remotely Operated Vehicles (ROVs) [65], [100] and Autonomous Underwater Vehicles (AUVs) [16], [101]. Thus, AI processing innovation is rather conditioned by the different operational scenarios, as for example the monitoring of deep-sea biodiversity and stock assessment [1, 2, 24] or ecosystem recovery from different disturbances [13, 75, 82].

Marine imaging acquisition

Although most of the surface of this planet is covered by seawater, a large part of its volume and projected seafloor remains unexplored [25, 40, 92]. This is principally a consequence of the particular characteristics of the ocean environment such as high pressure, low temperature and absence of light, which make it hostile to humans and pose practical challenges to its exploration [79]. However, advances in robotic platforms and sensor technologies have made possible to dive in virtually any realm of the deep marine biosphere [12, 33, 67], obtaining relevant information about many marine environments and their inhabiting species [3, 22, 41, 68]. Oceans provide us with goods and services that, if exploited without control, can be depleted [29, 99]. Therefore, much more effort is required for the acquisition of baseline knowledge on marine ecosystems in terms of species and their habitats characteristics, in order to promote sound and scientifically sustained management policies [27, 28]. This data acquisition is up to date with the framework of the UNESCO Ocean Decade Initiative (https://en.unesco.org/ocean-decade) (C. N. [64]. The study of imaging data provided by new robotic platforms is progressively playing a central role for fisheries [1, 2, 14], with evident industrial applications in terms of impact quantifications at off-shore decommissioning and projected mining activities [8, 43, 46], [65, 87].

During the last two decades, underwater imaging assets have increased with the improvement of High-Definition (HD) optics and the introduction of low-light equipment that, in association with acoustic multi-beam cameras, are presently allowing vision in darkness [3]. Complex sensor packages are being installed on cabled observatories and their docked mobile platforms such as crawlers (i.e., Internet Operated Vehicles, IOVs), combining oceanographic and geo-chemical assets with cameras [20, 20, 21, 21], [32, 50, 93]. This combined image and multiparametric environmental data acquisition is allowing to link visual counts of animals for the different species within a marine community with concomitant habitat quality changes, as an experimental field measure of their ecological niche [1, 2]. The technological developments are also paralleled by vessel-assisted robotic technologies such as AUVs and ROVs [17, 26, 81], [101].

Underwater image quality challenges

The automation of image acquisition and processing for animal tracking prior the classification is of relevance for the prolonged and continuous monitoring of marine biodiversity in any different operational scenario [6, 7, 10, 49, 81]. Although the use of better imaging technologies has made it possible to obtain HD outputs of increased quality, in almost all cases a pre-processing is necessary, due to strong environmental variability (i.e., "real world" scenarios, [61, 86]). While in controlled laboratory environments the bottom is static and with hardly any details [36, 37, 56, 88, 89, 103], in uncontrolled field settings such as the seas, images are acquired under variable environmental light or artificial lighting (for costal to

Lopez-Vazquez *et al. Journal of Big Data*   (2023) 10:37

Page 3 of 26

deep-sea applications), floating particles and variable substrates as background [18, 31, 57, 62]. Such a variability represents a challenge in the identification and tracking of animals within the Field of Views (FOVs) and the extraction of their morphological features for classification [5, 57, 60, 61, 71, 85].

In this framework, image pre-processing is essential to improve animals' detection and their posterior classification. Image enhancement encompasses both Computer Vision (CV) methods (i.e. noise removal, contrast change or colour adjustment: e.g., [30] and Deep Learning (DL) methods such as neural networks, capable of generating a new enhanced image from the original one [54], [96, 105]. However, given the amount of images collected, manual processing becomes infeasible. Therefore, an automatic process is needed to improve these huge datasets, either to analyse them manually or to use them later in a detection and classification process to obtain better results. In order to manage the processing difficulties of image treatment in real world case scenarios some authors [57, 88, 89, 105] established a pipeline based on different automated treatment steps toward quality enhancement at cabled observatories. Such an effort has not yet been done with their docked mobile platforms such as IOVs, although some work has already been developed for ROVs (e.g., [100] and AUVs (e.g., [16]. The images and videos obtained by crawlers in particular are composed of different scenarios, as IOVs are often in constant motion. Typically, multiple sediment particles can be seen in the images, as they are lifted by the movement of the crawler and can impede visibility [20, 21, 32].

Objectives

The objective of this paper is to design an automatic image enhancement process pipeline using Deep Learning techniques, in order to enhance the images and subsequently obtain better animals' classification results. Based on previous experience at cabled observatories [57], this process can be applied to ROV and AUV data. This enhancement was carried out on images acquired by the deep-sea crawler "Wally" with innovative deep-learning techniques, as baseline condition for improved animal classification which is required by monitoring and conservation strategies [28]. This processing provides a solution to the problem of enhancing very dark underwater images, since existing enhancement and classification solutions are still too dependent upon high illumination levels.

The article is organised as follows: Sect. "Materials and methods" describes the image set used in this work, some existing methods for image enhancement, and the chosen evaluation metrics (both for image quality assessment and species classification). Sect. "Proposed methods" presents the proposed methods, including the image enhancement pipeline and the classification process. Sect. "Experimental design" details the experimental design used to carry out the different experiments. Sect. "Results" shows all the obtained results, first regarding image generation and then animal classification, while Sect. "Discussion" discusses the results. Finally, Sect. "Conclusions and Future Work" gives our conclusions and future work.

Materials and methods

The IOV and the study area

The crawler operates since 2009 at the Hydrates site (~ 870 m depth; 48° 18′ 46″ N, 126° 03′ 57″ W) of the Barkley Canyon Node of the North-East Pacific Undersea Networked

Experiments (NEPTUNE) seafloor cabled observatory, operated and maintained by the Ocean Networks Canada (ONC; www.oceannetworks.ca). The site is a soft sediment plateau, characterised by the presence of outcrops of methane hydrates forming small mounds, with chemosynthetic bacteria forming thin mats suspect to erosion by currents [94]. A prevalent down-canyon current flow of intermediate velocity, i.e., seldom higher than 0.3 m/s [20, 21] and mainly SW direction can generate turbidity and phytodetritus fluxes [20, 21] with varying quantities of particles, potentially interfering with the FOV. The tidal regime is mixed semi-diurnal, with two unequal pairs of highs and lows [47], determining the image capturing protocol (see Sect. "Data collection").

Data collection

The image capturing process is described in [20, 21]. In brief, a total of 18 imaging transects (i.e., 9 forth and 9 back) of ~ 30 m length were carried out between 2 November and 2 December 2016. The currents in the area are generally stronger than the crawler's established speed (i.e., ~ 0.04 m/s), meaning that moving towards the same direction as the current would constantly place the entire sediment cloud in front of the camera. Moving against the current was an efficient strategy to avoid that, but did not impede the generation of sediment cloud, parts of which would occasionally interfere with the camera's field of view.

Imaging was performed at 1 Hz (i.e., 1 image/s) with a Basler dart USB 3.0 camera (daA1600-60 μm/μc; 1600 × 1200 pixel), which was mounted on a structured light system (i.e., pan/tilt unit; PTU) developed by DFKI Bremen, with illumination provided by two 33 W, 2000 m rated LED lamps. For standardization purposes, angles were set to − 76° left for pan and − 2° up for tilt in all but one transect (for further details on camera calibration see Supplementary Table S2 in [20, 21]. The camera was facing towards the right of the crawler at all times, so that the background was different between the back and forth transects.

The targeted group of species

The main megafauna (i.e., animals with body size above 2–3 cm) from a total of 14 morphospecies (i.e., identified down to species level or to higher taxonomic ranking, based on general morphology) present in the dataset, were identified by visual inspection with the help of the NEPTUNE Canada Marine Life Field Guide in [38]. However, species for which few records were available were combined into one class which was later added. The remaining considered species were:

- Demersal fish of the family Sebastidae. This group included rockfishes of the genus *Sebastes* and thornyheads of the genus *Sebastolobus*, which are mainly observed inactive on the seabed and are characterised by their orange-red colour pattern.
- The blackfin poacher (*Bathyagonus nigripinnis*) a small, thin, dark coloured fish also often observed as inactive on the seabed.
- The Pacific hagfish (*Eptatretus stoutii*), with a characteristic grey-violet colour and long, slender body, observed either laying on the seabed or swimming in sudden bursts.

Lopez-Vazquez *et al. Journal of Big Data*   (2023) 10:37

Page 5 of 26

- The grooved tanner crab (*Chionoecetes tanneri*) of varying size between adults and juveniles, orange body and 4 pairs of walking legs, observed both as immobile or walking.
- Sea stars of the class Asteroidea, appearing as white and stationary (compared to the temporal scale of each transect).
- The ctenophore *Bolinopsis infundibulum*, transparent and drifting with the current flow at varying velocities or actively swimming across the FOV.
- The jellyfish *Poralia rufescens*, orange-red and round-bell shaped, also moving with a combination of active swimming and drifting with the current.

Moreover, three classes were added a posteriori due to their recurrent occurrence: firstly, the class encompassing floating and resuspended sediment particles in water, because the movement of the IOV leads to their occurrence on countless occasions and the currents drag visible phytodetritus from shallower waters; on the other hand, the class containing transect plastic metric reference marks (for remote navigation) has been added, as they can be observed in many of the images; finally, the class which includes unclassified species, as too dark or too far in the FOV to be recognised, or because they belong to the group of species for which few records were available in the current dataset. All species and classes used can be seen in Figs. 1 and 2.

Table 1 shows the number of samples per class (before data augmentation) considered for the generation of the datasets.

Image enhancement methods

To enhance the images, a pipeline has been generated using different techniques to deliver an input for a neural network that would then generate properly enhanced images.

Although many techniques were finally discarded, it was necessary to carry out a preliminary survey on those techniques to seek for best enhancement results. The following list contains a brief description of each of the techniques that were used:
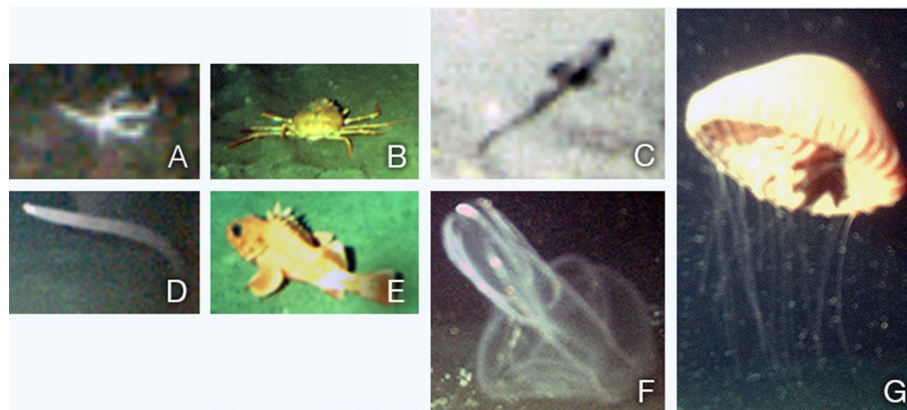
Fig. 1 An example of the species in the dataset: **A** Asteroidea, **B** *Chionoecetes tanneri*, **C** *Bathyagonus nigripinnis*, **D** *Eptatretus stoutii*, **E** Sebastidae, **F** *Bolinopsis infundibulum*, **G** *Poralia rufescens*
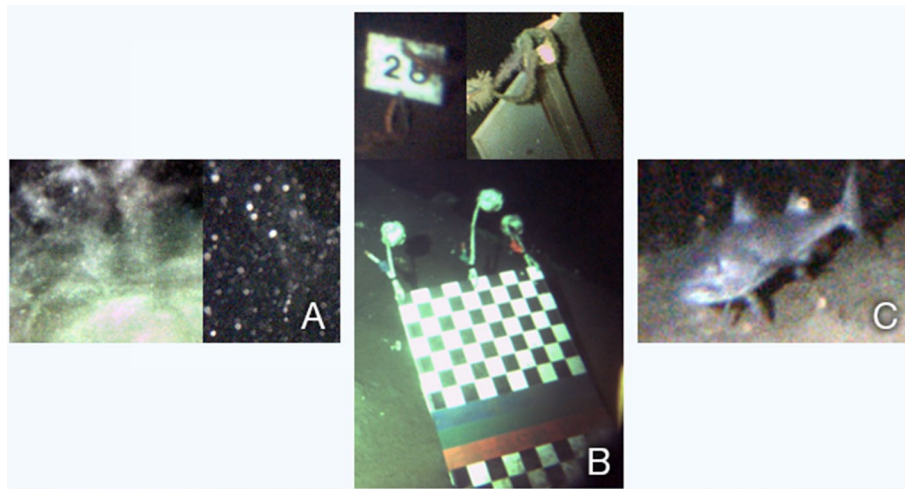
Fig. 2 An example of other classes added to the dataset: **A** floating particles and **B** transect plastic metric reference marks and **C** other species (unclassified or few records available)

Table 1 Number of samples of species used for building the dataset for reference at automated classification

Species	Samples in dataset
Asteroidea	731
Chionoecetes tanneri	793
Bathyagonus nigripinnis	604
Eptatretus stoutii	651
Sebastidae	1102
Bolinopsis infundibulum	849
Poralia rufescens	748
Sand particles	520
Objects	656
Other species	318

- The Contrast Limited Adaptive Histogram Equalization (CLAHE) [106] reduces the problem of its predecessor, the traditional Adaptive Histogram Equalization [76], which tends to amplify the noise in constant and homogeneous regions.
- Gamma Correction (or Gamma Encoding) is the non-linear operation to encode and decode the luminance values in images or videos, used to compensate human vision in order to maximize bit bandwidth in relation light/colour perception and details hidden in dark images can be appreciated [77].
- Colour Balance (i.e., as white balance) is the global adjustment of colour intensities to correct the representation of neutral colours [9, 84].
- Convolutional Neural Networks (CNNs) are a type of neural network models commonly used for image recognition and classification [51, 52].

- Autoencoders are used as unsupervised learning neural networks and have three main components: the encoder, the code (also known as latent space representation) and the decoder [11, 39].

Metrics

To evaluate the improvements in image quality after applying the processes described above and also to evaluate the further classification, we selected a set of evaluation metrics.

Image quality metrics

For the evaluation of the images, the following image quality assessment metrics have been used:

- Structural Similarity Index (SSIM) (Z. [97]
- Peak Signal-to-Noise Ratio (PSNR) [44]
- Underwater Image Quality Measure (UIQM) [72], [73]
- Underwater Colour Image Quality Evaluation (UCIQE) [102]

Classification metrics

To evaluate the performance of the classifiers, the following metrics were used:

- Accuracy [34]
- The Area Under the Receiver Operating characteristic Curve (AUROC) [34]
- Loss
- The confusion matrix [34]

Proposed methods

Image enhancement pipeline

In this subsection we present the image enhancement process in terms of composing steps and the description of the residual network that mainly constitutes this process. The complete pipeline can be seen in Fig. 3. The original images were in raw format and appeared in grayscale, so a chromatic interpolation algorithm by demosaicking or debayering, a digital image process used to reconstruct an image in colour, was applied.

Since the images still retained the greenish hue characteristic, we model and train a residual CNN network to generate the enhanced images and thus eliminate the greenish hue characteristic. Those neural networks are known for skip connections, or shortcuts to jump over some layers. The omitted connections aim to avoid the problem of vanishing gradients or mitigate the problem of Degradation (accuracy saturation), where adding more layers to a deep model leads to a larger training and test error [42]. This network has the structure of an autoencoder, which usually presents a structure made by three parts: the encoder, which extracts features from the input image, a central part that performs feature processing, and the decoder, the final part, which decodes
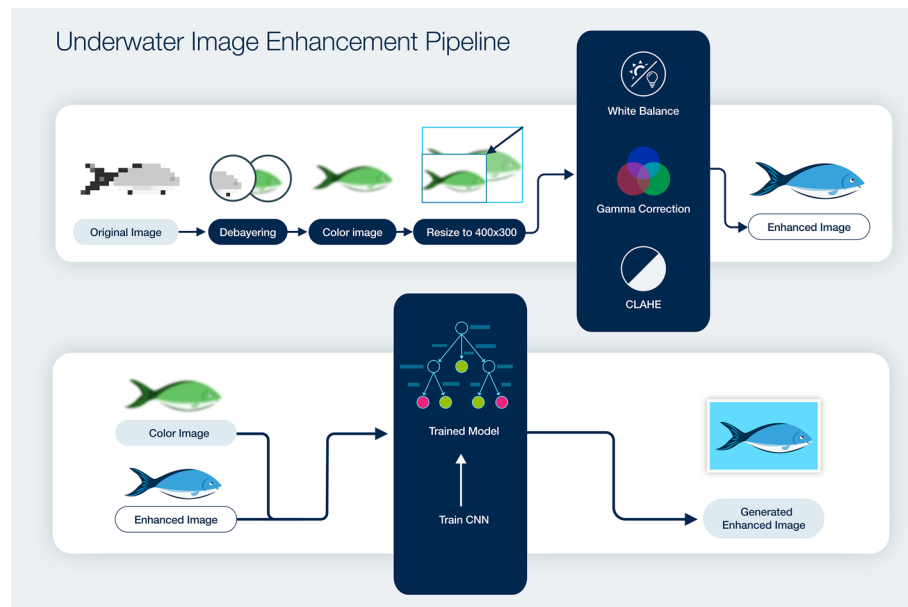
Fig. 3 Underwater image enhancement Pipeline

the processed features into an output image. In the elaborated residual CNN network, the optimizer, batch size and layers of were modified until the results were improved. Techniques such as White Balance, Gamma Correction and the CLAHE algorithm were applied to generate the images with which the network would be trained.

Each convolutional layer was followed by a ReLU activation layer [70], a linear function whose output, if positive, will be the same as the input value, while if negative, the output will be zero, as indicated in Eq. (1):

$$f(x) = \max(0, x) \tag{1}$$

After the input layer, there were two pairs of convolutional and ReLU layers followed by a max pooling layer. Next, there was a larger block consisting of three convolutional and ReLU layers and a max pooling layer. This was followed by a group of four convolutional layers. The last group was composed of three convolutional layers. The optimiser chosen for this network was Adam [48], while the loss function chosen was MSE loss. The layered structure can be seen in Fig. 4. This residual network has two residual blocks which skip connections. In this way, these shortcuts perform identity mapping, where their outputs are added to the outputs of the stacked layers.

Classification pipeline

For the detection and classification of animals within the different species as categories, a modified version of the pipeline previously proposed in [57] was used, omitting the application of CLAHE at the image processing step. However, here, background subtraction procedure did not work properly since FOV characteristics slightly changed over
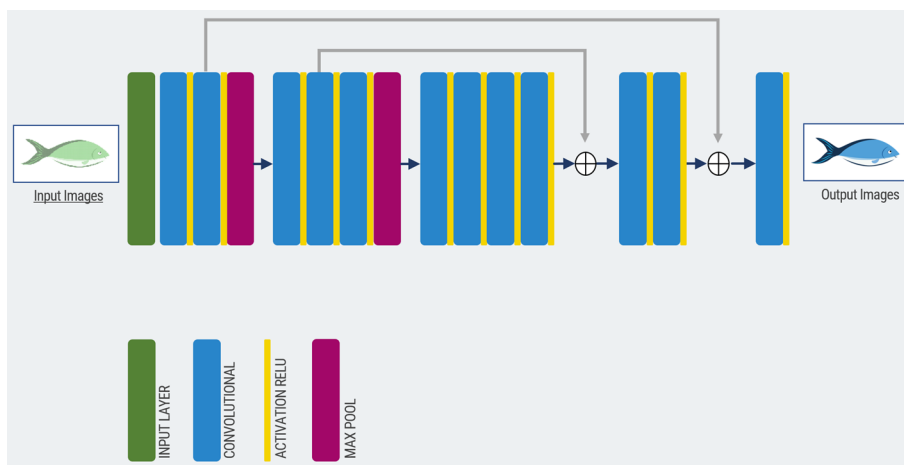
Fig. 4 Structure of the residual convolutional neural network

consecutive frames/image due to the crawler's motion. Therefore, the frame difference technique was applied, where each frame was subtracted from the previous one [66].

The used classification algorithms are detailed in subSect. "Metrics" of [57]. On one hand, 8 classical algorithms: two versions of Support Vector Machine (LSVM and SVM_SGD), two K-Nearest Neighbours (K-NN1 and K-NN2), two Decision Trees (DT1 and DT2) and two Random Forests (RF1 and RF2). On the other hand, 8 neural networks: four Convolutional Networks (CNN1, CNN2, CNN3 and CNN4) and four Deep Neural Networks (DNN1, DNN2, DNN3 and DNN4) with different configuration and structure parameters. The parameters chosen for training were the same as in [57].

Experimental design

This section explains the experimental setup of the experiments carried out using both proposed pipelines, i.e., the image enhancement pipeline and the classification pipeline.

As the original images were too large and slowed down the training too much, they were resized to 400×300 pixels.

The implemented residual CNN had a 14-layer structure, separated in different blocks. The network for image enhancement was configured to train a maximum of 50 epochs of 100 iterations each. An epoch is one complete presentation of the data set to be learned, while an iteration is the number of batches needed to complete one epoch. The batch size is the total number of training examples present in a single batch. The higher is this parameter, the more memory space is needed. In addition, it was also designed to save the model every time the loss value decreased, and also to stop training if the loss value did not improve within 5 epochs (to avoid the over-fitting problem).

To train the residual CNN that was part of the image enhancement process, a dataset of 13,548 images was used to create the training and test data sets for the new datasets. 80% of the images (10,838 images) constituted the training dataset, while the remainder was used to test the network. Different datasets were incrementally generated, to test the techniques that best affected the final classification results. For the purpose of simplify the names of the datasets used and generated and used in this paper, the following table has been created in which all the names and their description are listed in Table 2.

Table 2 Simplified names of datasets used and their description

Techniques	Short name	Description
RAW	Dataset0	Dataset with the original images
DBY	Dataset1	Dataset with the original images, but in colour (after the application of the debayering technique)
DBY + WB	Dataset2	Dataset containing images of the DBY dataset to which the White Balance (WB) method has been applied
DBY + WB + GC	Dataset3	Dataset containing images of the DBY + WB dataset to which the Gamma Correction (GC) method has been applied
DBY + WB + GC + CLAHE	Dataset4	Dataset containing images of the DBY + WB + GC dataset to which the CLAHE algorithm has been applied
DBY + WB + ResidualCNN	Dataset5	Dataset generated by the trained Residual CNN, using as input the colour images of the DBY dataset to which the WB method has been applied
DBY + WB + GC + ResidualCNN	Dataset6	Dataset generated by the trained Residual CNN, using as input the images of the DBY + WB + GC dataset
DBY + WB + GC + CLAHE + ResidualCNN	Dataset7	Dataset generated by the trained Residual CNN, using as input the images of the DBY + WB + GC + CLAHE dataset

To evaluate the CNN-generated underwater images, we chose SSIM, PSNR, UIQM and UCIQE values.

As for the classification, the size of the collected set was only of 6972 elements. As there were not too many images, we decided to apply data augmentation techniques to 80% of the images (a total of 5573 images), which are the ones that made up the training set. After applying data augmentation techniques, the training set increased from 5573 to 35,020 images, obtaining 3502 samples per class.

All the selected classifiers were tested by tenfold cross-validation by considering that the elements of each class were distributed evenly in each fold [45, 58, 90]. The performance of the models was evaluated by the accuracy, the loss (both train and testing), the AUROC average scores [34], as well as by the confusion matrix. The accuracy and AUROC values were calculated by the multiclass implementation from Scikit-learn, which estimates the metrics for each label, without considering the label imbalance.

All experiments were conducted in Python. The implementation of all the classical algorithms used is within the Scikit-learn library [74] (https://scikit-learn.org), while the neural networks were implemented with the Keras and Tensorflow libraries. The environment used for training the selected algorithms and the defined models was Google Colaboratory (also known as Colab). It operates currently under Ubuntu 18.04 (64 bits) and it is provided by an Intel Xeon processor and 12 GB RAM. It is also provided with Nvidia K80, T4, P4 and P100 GPUs.

On the one hand, a classification of 4 datasets (the original-coloured dataset, Dataset1, and the three generated by the network, Dataset5, Dataset6 and Dataset7), whose number of classes amounts to 7, was carried out. In addition, the dataset for which the lowest loss value and the highest accuracy and AUROC test values were obtained was selected for another classification, in which three more classes will be added to make a total of 10.

Results

In this section, the results obtained following the experimental design are presented: the results of the image enhancement pipeline are presented in SubSect. "Image enhancement pipeline results", while the results of the classification pipeline in SubSect. "Classification results".

Image enhancement pipeline results

The network was finally trained for 15 epochs (due to the configured stop condition) and approximately each epoch took 1100 s to run (approx. 20 min). The visual comparison between the different image datasets is reported in Fig. 5, where the original, colourised and CNN-generated images are shown.

Regarding the visual aspect, the greenish colour that characterised the images to which techniques such as WB, GC and CLAHE were applied (Dataset2, Dataset3 and Dataset4), was almost eliminated. The images generated by the residual network did anyway maintain a bluish tone, but the colours of the visible animals were more pronounced. In any case, at a simple visual inspection, the images generated by CNN were somewhat blurrier compared to those processed by CV methods.

Figures 6 and 7 show in detail the comparison between two images as an example. Figures 6A and 7A are part of the input dataset to which CV methods were applied, while Figs. 6B and 7B are images of the dataset generated by the enhancement network. Figure 6 shows an example of image processing where the generated result is quite similar to the input image. As a result of that processing, the two animals belonging to two species (i.e., a floating ctenophore and a rockfish lying on the seabed) can be better visualised by the naked eye.

Similarly, a different comparison between two images can be seen in Fig. 7 also for two species (i.e., a hagfish and another rockfish). This time, the residual CNN generated a superior colour and illumination quality in the output image, except for warm colours, which maintained some of the bluish tone, as common characteristic of untreated underwater images.

Figure 8 shows other two images from input Dataset4 and the output Dataset7. As can be seen in both images, a floating jellyfish can be observed. In this case, the network has transformed the orange tones while maintaining some bluish tones.

The values of the UIQM and UCIQE metrics for the evaluation of images of the two scenes are summarised in Table 3. The UIQM and UCIQE mean values were slightly higher for the input images of the network. The difference is even greater for coloured images (by debayering).

The values of the SSIM and PSNR metrics for the evaluation of the quality and similarity of the images corresponding to the different datasets are shown in Tables 4 and 5 respectively. As for the SSIM values (Table 4), we can observe that the highest values have been obtained by the datasets generated by the network, while the datasets to which CV techniques were applied (used to train the residual network) obtained lower values. In the same form, the PSNR values have been slightly higher for the datasets generated by the residual network, as can be seen in Table 5
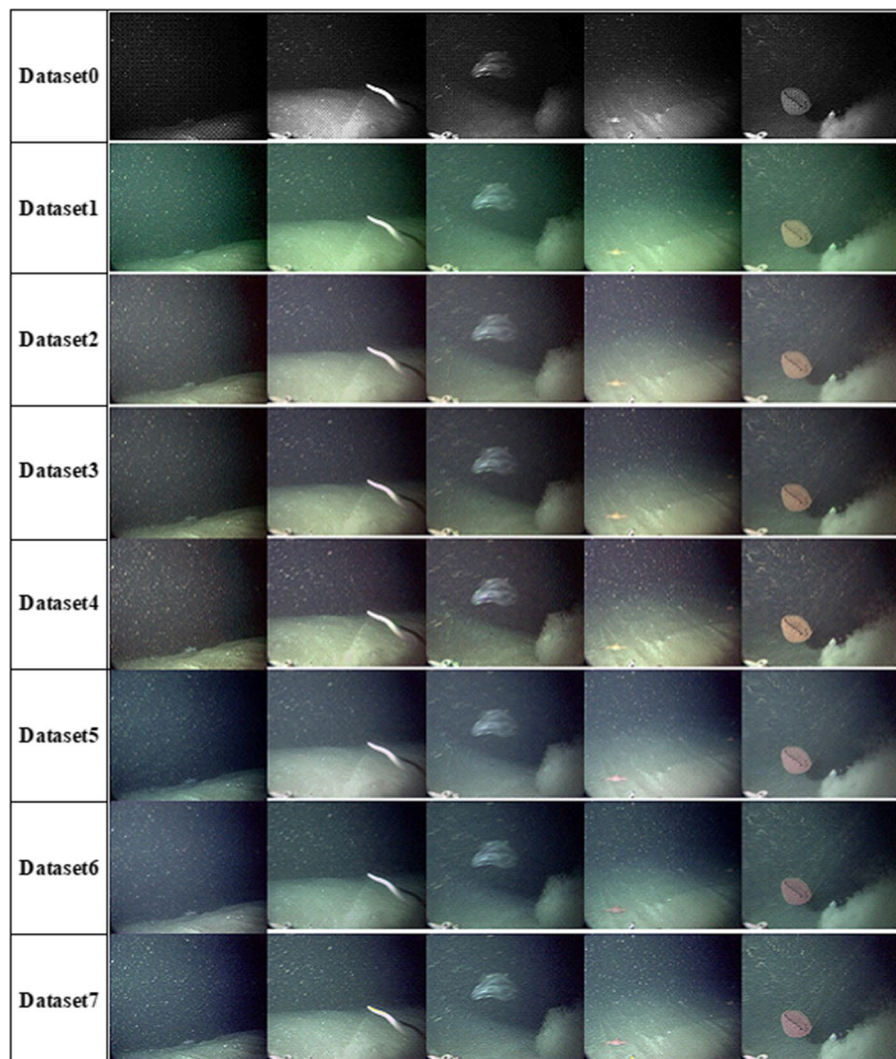
Fig. 5 Comparison among the different image datasets: original, processed with CV techniques and generated by the CNN. Dataset0: dataset of original images in RAW format. Dataset1: dataset of colourised images. Dataset2: dataset of images to which the WB technique has been applied. Dataset3: dataset of images with WB and GC applied. Dataset4: dataset of images to which the WB, GC and CLAHE techniques have been applied. Dataset5: dataset of images generated by the CNN having as input the images of the Dataset2 dataset. Dataset6: dataset of images generated by the CNN having as input the images of the Dataset3 dataset. Dataset7: dataset of images generated by the CNN having as input the images of the Dataset4 dataset

Classification results

The classification results have been divided in two parts: first the results are shown for the seven-class dataset and then for the modified dataset containing ten classes.

The 7-classes classification results

The results obtained in the classification of the first 7 selected classes (Sebastidae, *Bathyagonus nigripinnis*, *Eptatretus stoutii*, *Chionoecetes tanneri*, Asteroidea, *Bolinopsis infundibulum*, *Poralia rufescens*) are shown in Tables 6 and 7.
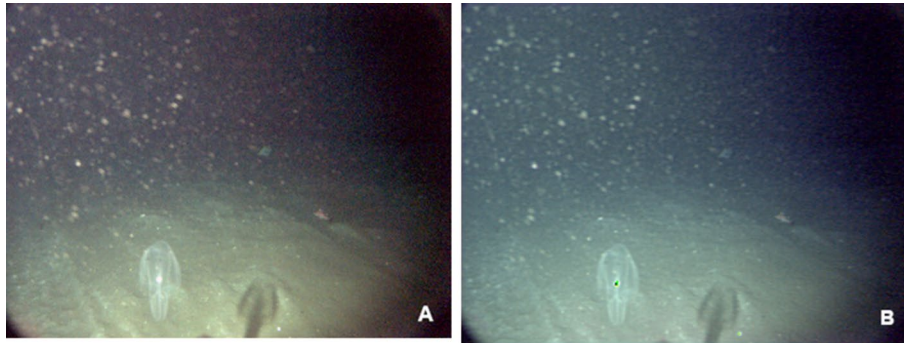
Fig. 6 Comparison of CNN input and output images. **A** Shows an input image as after WB, GC and CLAHE processes (from Dataset4). **B** Shows an image generated by the CNN (from Dataset7)
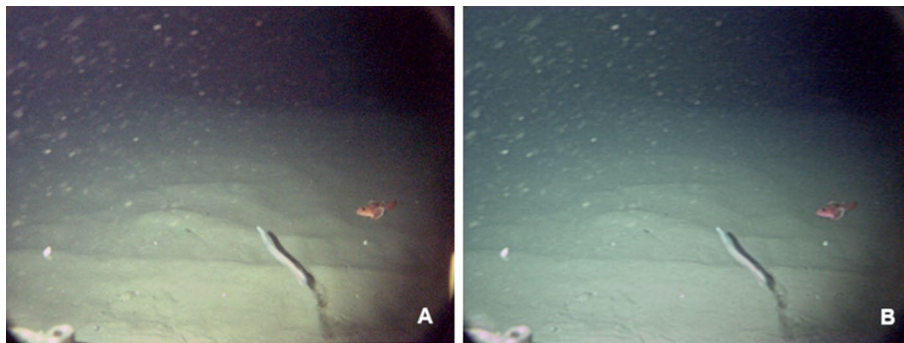
Fig. 7 Comparison of CNN input and output images. **A** Shows an input image after WB, GC and CLAHE processes (from Dataset4). **B** Shows an image generated by the CNN (from Dataset7)
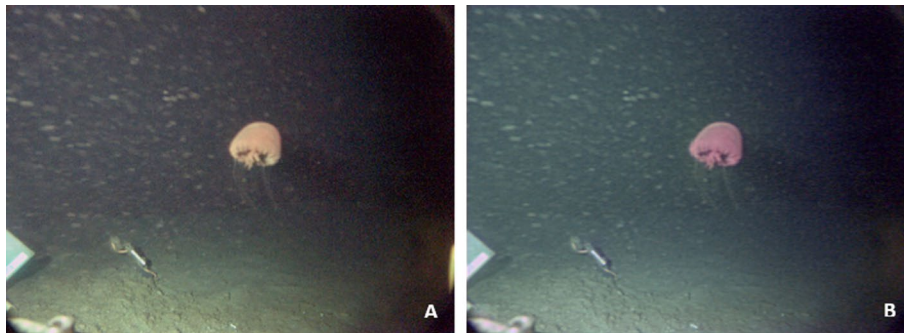
Fig. 8 Comparison of CNN input and output images. **A** Shows an input image after WB, GC and CLAHE processes (from Dataset4). **B** Shows an image generated by the CNN (from Dataset7)

Table 6 shows the results obtained by the classic classifiers on the different datasets. The dataset that obtained the lowest values overall during this classification was the Dataset1. The remaining datasets obtained fairly equal results. The algorithm that obtained the highest performance on all the datasets (both accuracy and AUROC values) was the RF2. The highest accuracy value of 0.6147 was achieved with the Dataset5 dataset, while the highest AUROC value of 0.8041 was achieved with the Dataset7 dataset.

The results obtained by the deep learning techniques can be seen in Table 7. The dataset that obtained the lowest values overall during classification was the so-called

Table 3 UIQM and UCIQE mean values for the different datasets

	UIQM	UCIQE
Dataset0	1.845	0.2719
Dataset1	1.936	**3.515**
Dataset2	2.421	2.778
Dataset3	2.299	2.475
Dataset4	2.023	3.096
Dataset5	**2.585**	2.164
Dataset6	2.454	1.892
Dataset7	1.908	2.406

The bold values refer to the highest (and best) value

Table 4 SSIM mean values between the different datasets

SSIM	Dataset1	Dataset2	Dataset3	Dataset4	Dataset5	Dataset6	Dataset7
Dataset0	$0.5464\pm$ 0.0663	$0.5298\pm$ 0.0668	$0.6465\pm$ 0.0496	$0.5199\pm$ 0.0723	$0.5355\pm$ 0.0585	**$0.6496\pm$** **0.0382**	$0.5348\pm$ 0.0620

The bold values are the highest (and therefore best) value

Table 5 PSNR mean values between the different datasets

PSNR	Dataset1	Dataset2	Dataset3	Dataset4	Dataset5	Dataset6	Dataset7
Dataset0	$15.3835\pm$ 0.6807	$15.4722\pm$ 0.6855	$19.5634\pm$ 0.9484	$16.3320\pm$ 0.7462	$15.7494\pm$ 0.6848	**$20.1172\pm$** **0.9509**	$17.0195\pm$ 0.7969

The bold values are the highest (and therefore best) value

Table 6 Test accuracy and AUROC values obtained by the classical algorithms with the 7 classes datasets

Classifier	*Dataset1*		*Dataset5*		*Dataset6*		*Dataset7*	
	Test AUROC	Test accuracy	Test AUROC	Test accuracy	Test AUROC	Test accuracy	Test AUROC	Test accuracy
LSVM	0.6443	0.3390	0.7365	0.4976	0.7484	0.5141	0.7506	0.5171
SVM_SGD	0.6098	0.2868	0.6457	0.3911	0.6773	0.4534	0.7220	0.4850
K-NN1	0.6557	0.3579	0.7077	0.4083	0.6999	0.4050	0.7179	0.4282
K-NN2	0.6136	0.2999	0.6715	0.3462	0.6590	0.3446	0.6763	0.3558
DT1	0.6780	0.4135	0.7047	0.4634	0.6972	0.4570	0.7065	0.4744
DT2	0.6743	0.4112	0.6998	0.4559	0.6966	0.4571	0.7020	0.4648
RF1	0.6744	0.4067	0.7033	0.4571	0.7050	0.4665	0.7062	0.4721
RF2	**0.7524**	**0.5214**	**0.8009**	**0.6147**	**0.7965**	**0.6034**	**0.8041**	**0.6110**

The bold values refer to the highest (and best) value

Dataset1. Several of the networks obtained high values of accuracy and AUROC, and low values of loss (also during test).

The 10-classes classification results

Three more classes were added to represent floating particles and sand suspended in the water, objects placed on the ground and unclassified species and species that contained

Table 7 Training accuracy, training loss, test AUROC and test accuracy and loss values obtained by the deep learning approaches with the 7 classes datasets

| Classifier	Train accuracy	Train loss	Test AUROC	Test accuracy	Test loss
Dataset1					
CNN-1	0.6097	1.2148	0.5839	0.2508	2.4334
CNN-2	0.6568	1.0570	0.5983	0.2779	2.3739
CNN-3	0.9704	0.1289	0.7960	0.6106	3.4402
CNN-4	**0.9875**	**0.0490**	0.7918	0.6043	4.3183
DNN-1	0.6330	0.7024	**0.8264**	0.6046	2.3056
DNN-2	0.6311	0.7548	0.8181	0.6042	3.6272
DNN-3	0.5799	0.9233	0.8097	**0.6373**	**1.1579**
DNN-4	0.5693	0.9631	0.8090	0.6358	1.1581
Dataset5					
CNN-1	0.6345	1.1534	0.6602	0.4595	1.4518
CNN-2	0.6603	1.0562	0.6509	0.4374	1.4733
CNN-3	0.9807	0.1263	0.7910	0.6075	3.8420
CNN-4	**0.9883**	**0.0723**	0.7923	0.6083	4.2763
DNN-1	0.6395	0.6841	0.8208	0.6563	2.4390
DNN-2	0.6297	0.7835	0.8169	**0.6573**	4.1019
DNN-3	0.5952	0.8647	**0.8227**	0.6531	**1.1528**
DNN-4	0.5978	0.8466	0.8215	0.6525	1.2127
Dataset6					
CNN-1	0.6236	1.1459	0.6679	0.4296	1.5100
CNN-2	0.6709	1.0069	0.6721	0.4348	1.5241
CNN-3	0.9632	0.1421	0.7846	0.5925	3.5784
CNN-4	**0.9809**	**0.0775**	0.7837	0.5851	4.3230
DNN-1	0.6377	0.6915	0.7987	0.6193	2.7051
DNN-2	0.6300	0.7649	0.8026	0.6218	4.2045
DNN-3	0.5653	0.9716	0.7990	0.6106	**1.2135**
DNN-4	0.5715	0.9392	**0.8076**	**0.6264**	1.2195
Dataset7					
CNN-1	0.6375	1.1002	0.6668	0.4738	1.4635
CNN-2	0.7033	0.9094	0.6584	0.4335	1.4909
CNN-3	0.9741	0.1291	0.7889	0.5950	3.8693
CNN-4	**0.9866**	**0.0735**	0.7970	0.6144	3.9761
DNN-1	0.6371	0.6880	0.8210	0.6520	2.6627
DNN-2	0.6312	0.7615	0.8246	0.6616	3.8673
DNN-3	0.5908	0.8806	**0.8291**	0.6616	**1.1330**
DNN-4	0.6018	0.8376	0.8277	**0.6644**	1.2163

The bold values refer to the best value

very few records, in order to improve the results obtained at the time of detection. In fact, elements within these three categories were detected but were not classified in initially available classes. This dataset was only considered in its Dataset7 processing trials, where it delivered the best result. As can be seen in Table 8, RF2 obtained an accuracy value of more than 0.75 and an AUROC value of almost 0.87.

The confusion matrix (Fig. 9) corresponds to the results obtained by RF2, as the best performing algorithm. It classified quite correctly several classes, like Asteroidea, *Bathyagonus nigripinnis*, Human-made objects and Floating particles, but it frequently

Table 8 Test accuracy and AUROC values obtained by the classical algorithms with the 10 classes for the images from Dataset7

Classifier	Test AUROC	Test accuracy
LSVM	0.7549	0.5458
SVM_SGD	0.7089	0.4719
K-NN1	0.7787	0.5802
K-NN2	0.7207	0.4727
DT1	0.7587	0.5610
DT2	0.7582	0.5592
RF1	0.7608	0.5637
RF2	**0.8691**	**0.7568**

The bold values refer to the highest (and best) value
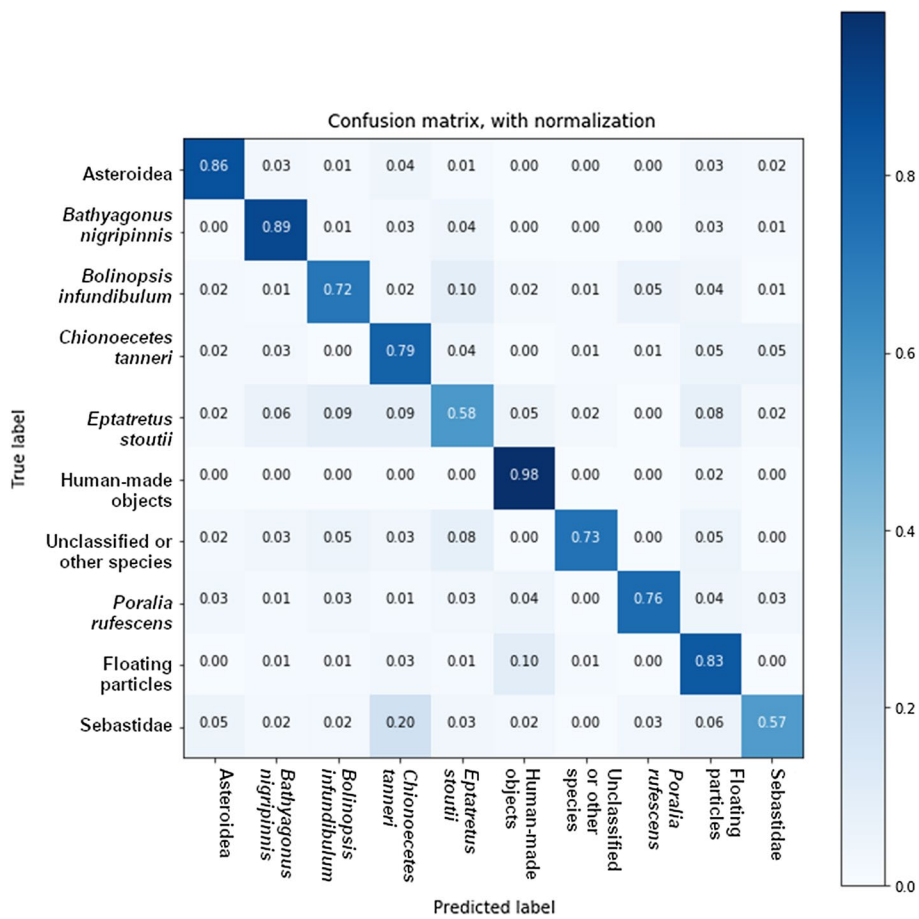
Fig. 9 Confusion matrix for the classification results (accuracy) obtained by RF2

misclassified two classes (*Eptatretus stoutii* and Sebastidae), for which it achieved a 60% of success rate.

As for neural networks, training accuracy and loss did not show major differences compared to those obtained for 7 classes datasets. However, other metrics such as the AUROC value, the test accuracy and test loss values improved (Table 9). The highest AUROC value was 0.8864, achieved by DNN-2. The best test accuracy value was

Table 9 Training accuracy, training loss and test AUROC, accuracy and loss values obtained by the deep learning approaches for the 10 classes Dataset7

Classifier	Train accuracy	Train loss	Test AUROC	Test accuracy	Test loss
CNN-1	0.5860	1.4096	0.6170	0.3124	2.4164
CNN-2	0.5314	1.6716	0.6370	0.3410	3.5452
CNN-3	0.9224	0.2387	0.8242	0.6749	1.9160
CNN-4	**0.9552**	**0.1326**	0.8136	0.6359	2.8140
DNN-1	0.5929	0.8642	0.8859	**0.7944**	1.2662
DNN-2	0.5813	0.9618	**0.8864**	0.7888	1.4685
DNN-3	0.5594	1.0097	0.8479	0.7299	0.8745
DNN-4	0.5626	0.9914	0.8760	0.7578	**0.8389**

The bold values refer to the best value
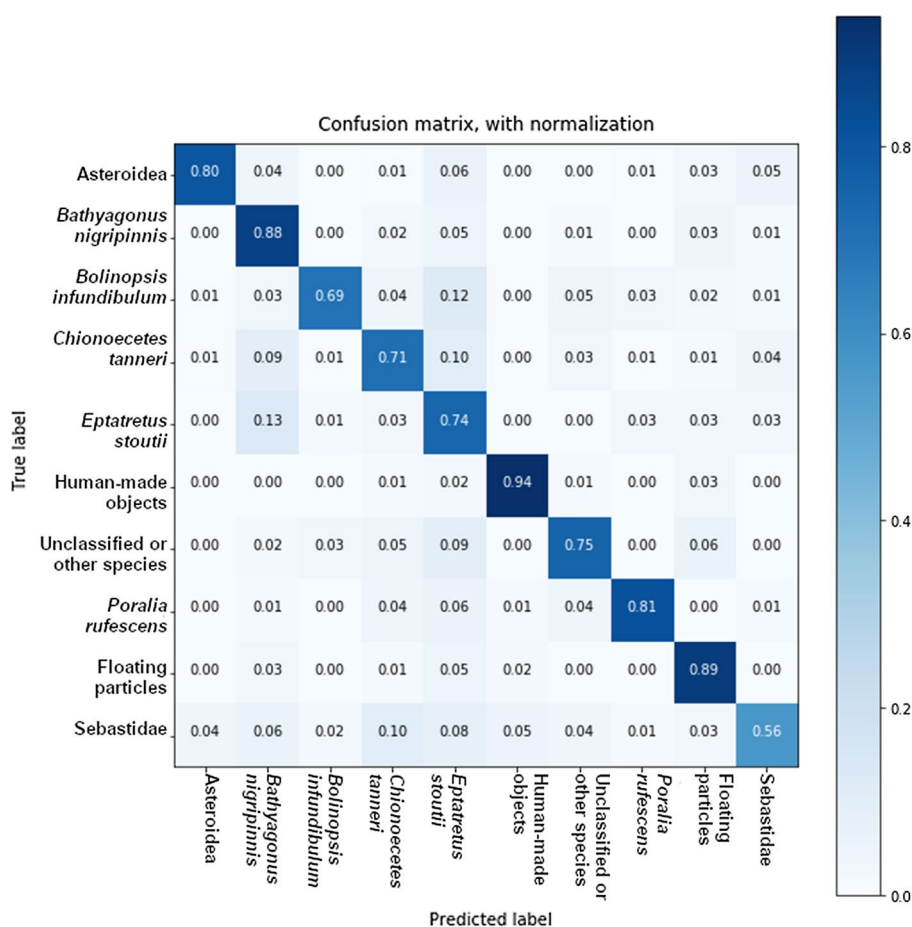
Fig. 10 Confusion matrix for the classification results (accuracy) obtained by the DNN-4

0.7944, much higher than in any of the previous datasets of 7 classes (see Table 7). Finally, the test loss also decreased to 0.8389 in the case of DNN-4.

Figure 10 shows the confusion matrix for the classification results obtained by DNN-4, which achieved good results for almost every class. In this case, five classes (Asteroidea, *Bathyagonus nigripinnis*, human-made objects, *Poralia rufescens* and Floating Particles)

were correctly classified with a rate above 80%, and the worst ranked class (Sebastidae) had 56% correctly labeled.

The performance during the training of the DNN-4 can be seen in Fig. 11. This network was trained during 889 epochs. Figure 11A shows the progress of the accuracy value, while Fig. 11B shows the decreasing of the loss value.

Some examples of detection and classification by DNN-4 are shown in Figs. 12 and 13. Figure 12 shows cases of correct detection and classification, while Fig. 13 shows cases where the algorithm confused the classes, leading to a wrong labelling.

It is possible that the misclassification of floating particles as ctenophore shown in Fig. 13 is due to their similarity in colour, as this species body is transparent and has some white spots. The objects classified as species may have been confused because of their location, as these are common areas where such species are found, and probably also because of their shape.

Discussion

In this study, we presented a novel pipeline for the enhancement of dark deep-sea images and the automated classification of visible fauna, in footage taken by a crawler as a moving benthic platform on a changing background. We elaborated an enhancement procedure that allowed to improve the animals classification capability, hence the functionalities previously achieved with static cameras at cable observatories [57]. For this purpose, different image enhancement techniques were first investigated and then applied to generate different datasets. Then, a residual network was modelled and trained with these datasets in order to generate a new set of enhanced images. Although the evaluation metrics of the image sets generated by the residual network could be improved, the best values of test accuracy, loss and AUROC in classification were achieved with one of the datasets generated by the neural network, which is the principal objective.
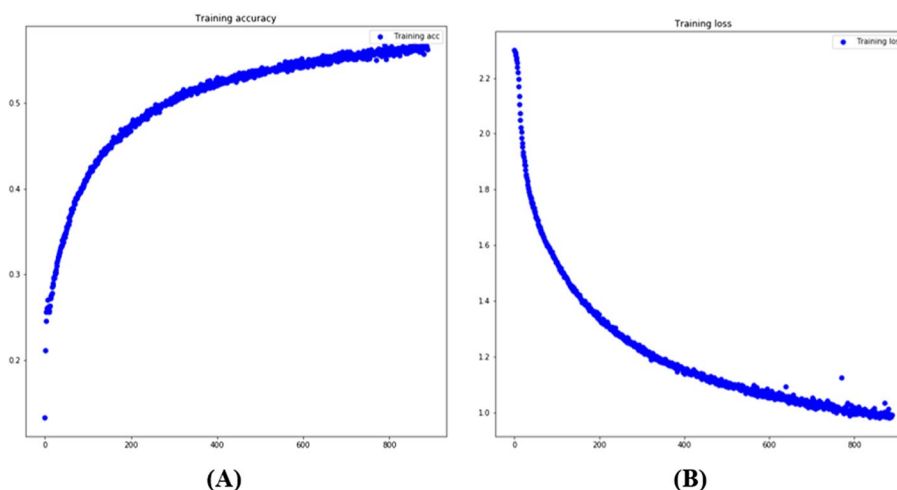
Fig. 11 Training accuracy and loss plots of the DNN-4. The X axis of the plots shows the number of epochs, while the Y axis shows the accuracy and the loss value respectively that was reached during the training. **A** Accuracy values obtained in every epoch at training time and **B** Loss values obtained in every epoch at training time
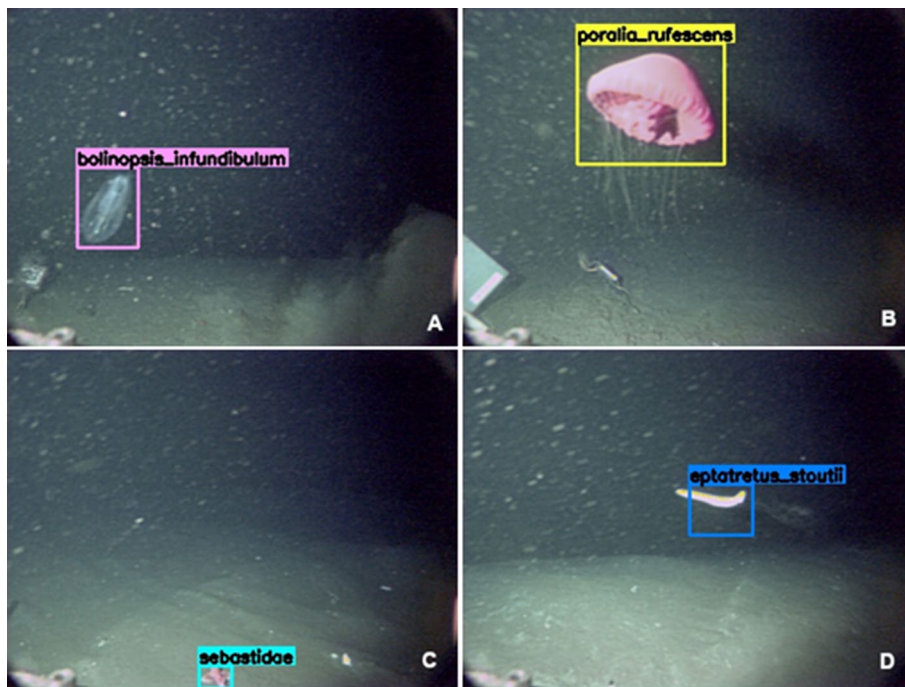
Fig. 12 Some examples of detections and correct classification. **A** Shows a correctly labelled ctenophore, **B** shows a correctly labelled jellyfish, **C** shows a correctly labelled rockfish and **D** shows a correctly labelled hagfish
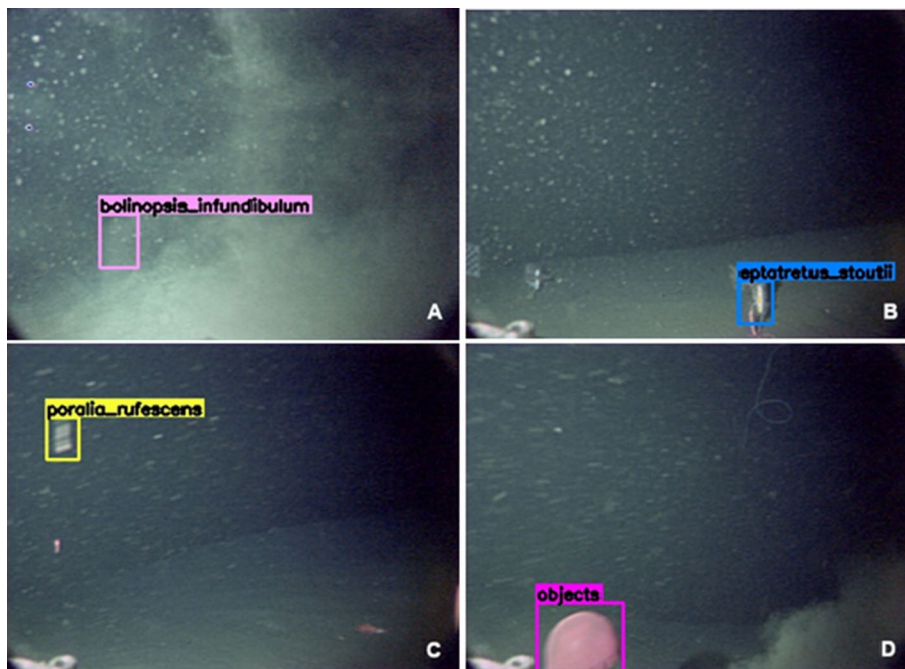
Fig. 13 Some examples of detections and incorrect classification. **A** Shows floating particles incorrectly labelled as ctenophore, **B** shows ground object incorrectly labelled as hagfish, **C** shows floating particles incorrectly labelled as jellyfish and **D** shows a jellyfish incorrectly labelled as objects

The residual convolutional network shows some problems with some hues when generating new images. For example, orange colours have been generated with bluish hue, transforming them into pinks. This is probably because these colours do not appear very often in the whole set of images. The UIQM and UCIQE values were slightly higher for the input images of the network. This may be because the images generated by the network are more blurred than the input images, as are the images transformed by applying techniques such as white balance, gamma correction and CLAHE. Similar studies, e.g. [88, 89], applied similar methods to pre-process the images, such as CLAHE, in order to obtain a mask on the Norway lobster (*Nephrops norvegicus*) detection, and then apply CV techniques and a Mask-RCNN for detection and segmentation comparison. Other studies in which the dataset was also obtained by an ROV, as in [69], we can observe that although the images are not as dark as those in the present paper, they do have that characteristic blue-green colour of the water. The method proposed in [16] provides colour enhancement and restoration to marine images, and although they tested it on not very dark images to which turbidity has been added, it is intended for AUVs and ROVs. They obtained a PSNR value of 21.840 dB, while ours was 20.117 dB. In [78] authors present an image enhancement method in which they also apply CLAHE, as well as other techniques, such as gray-level co-occurrence matrix (GLCM) feature extraction. However, the images they used were obtained from a dataset whose characteristics are totally different from the one used here, since it is collected by static cameras and at a shallower depth, since the images have natural light.

For the classification of marine species, two different types of methods were used in this study, i.e., classical algorithms and DL techniques. Data augmentation techniques were applied to the species with the most elements, and on the other hand, classes with insufficient number of elements were discarded. Similar studies also detected the advantages of DL over ML methods in marine environments (X. [19, 80, 83, 95]. However, the datasets used by these studies were obtained in coral reef areas, where there is still some sunlight, while the dataset we used was obtained at depths of more than 800 m, where visibility is low. For deeper water applications mimicking environmental conditions similar to those where the crawler is deployed, [88, 88, 89, 89] evidenced that advanced DL techniques, such as segmentation networks, can be an efficient tool for monitoring catches in pelagic fishery. In addition, the crawler generated clouds of sand and prevented the observation of species and objects on several occasions, which would not happen with a fixed camera.

For the test values, there are notable differences both among datasets and among neural networks. Regarding the classical algorithms, the RF2 was clearly the model that obtained the best results on all the datasets. With regard to the neural networks, it can be seen that, in the case of the CNNs, good results were not obtained, since CNN-1 and CNN-2 networks obtained quite low validation accuracy and AUC values, while CNN-3 and CNN-4, due to the loss values, were probably over trained. The deep neural networks (DNN-1 to DNN-4) achieved better AUROC, accuracy and loss test values than other algorithms. However, the sequential networks DNN-1 and DNN-2 have performed rather poorly for datasets of 7 and 10 classes, reaching loss values as high as 1.2662 and 1.4685 respectively. On the other hand, the other two deep networks (DNN-3 and DNN-4), have had a good performance and result, obtaining the best value for the

test accuracy (0.6644) and the best value for the test loss (1.1330) for 7 classes datasets, while for the 10 classes dataset DNN-4 obtained a test accuracy value of 0.7578 and a test loss value of 0.8389.

In [88, 89] authors also used CV techniques for the classification of marine species, which they compared with the results obtained by a Mask R-CNN. In their case, they obtained higher results with CV techniques, although in a later work they improved their classification with the segmentation network in a dataset of four classes [88, 89]. In [78] authors performed classification on the enhanced images using SVM, DT and k-NN, among others. The SVM achieved an accuracy value of 79.66%, while the k-NN obtained a value of 72.96% and the DT of 64.03%. They also used a backpropagation neural network (BPNN), which achieved an accuracy of 93.73%. The values achieved in the present study were quite close to those, despite the fact that the dataset is totally different, more complex and darker.

If we compare the results obtained for the 7 classes dataset and the 10 classes dataset, we can see that the best results were obtained for the dataset with more classes. This may be due to the fact that the classification pipeline in [57] detects the elements that move along the different images, and that these elements were not correctly classified because they did not correspond to any class. In the 10 classes dataset, these extra classes have been added and those elements can be assigned to a class and then be correctly classified.

Compared to the results of [57], the results here obtained achieved better metrics. As for the ML methods, RF2 was the algorithm that obtained the best test values for accuracy and AUROC in both investigations. In this paper the test values of 0.7568 for accuracy and 0.8691 for AUROC were reached, while in [57] the accuracy value for this algorithm was 0.6527 and that for AUROC was 0.8210. Regarding DL techniques, in [57] the network that obtained the highest values was the DNN-4, with test accuracy value of 0.7140 and an AUROC value of 0.8503, whereas here higher values have been achieved with several networks. DNN-4 obtained a test accuracy value of 0.7578 and an AUROC value of 0.8389. DNN-3 also obtained, anyway, higher values than the previous paper. DNN-1 and DNN-2 also outperformed the previous results but obtained high error values. We can state definitely that the results obtained in this paper outperform those of [57].

The next technological application scenario

Marine robotics is creating platforms that can be transformed into intelligent tools for autonomous ecosystem monitoring needs [4], as is nowadays required for monitoring an increasing number of marine activities, e.g., oil extraction and mining [46, 87, 91]. Implementing routines for automated individual tracking and classification and later integrate all those component routines into an operational hardware and software product, is a key aspect to improve the ecological monitoring functionality of all mobile platforms, including the crawler [8, 35, 63, 98, 104]. The proposed approaches and results represent a first step toward the establishment of an autonomous software focused on image processing to be installed on-board of the crawler. This represents a critical bottleneck for full autonomous monitoring of deep-sea ecosystem functions and services, by this class of IOVs.

Extracting the essential information on species presence, counts (as an indicator for abundance) and derived spatiotemporal changes, picturing community dynamism, is a time-consuming manual process. The tasks proposed in this research with the state-of-the-art of CNN algorithms indicate the possibility to allow embedded pre-processing of acquired images for object tracking via image quality enhancing/rendering with CNN approaches. At the same time, a refinement of species classification procedure is available with the post-processing of imaging products on a server with the help of newly added morphological descriptors [1, 2]. Moreover, the integration of all the detection and classification connected with the video capture processes would allow to transmit and store only the frames where the algorithms detect some kind of labelled species.

Conclusions and future work

The designed neural network, in combination with the detection and classification pipeline, generated enhanced underwater images leading to a more accurate classification process. The improvement and enhancement of underwater images also play an important role in feature detection, since a clear improvement of the images could reduce the subsequent work of feature detection and obtains better classification rates. We demonstrated that a neural network is a good option for generating enhanced images automatically, without the need to apply multiple techniques to an image. Due to their particular characteristics, enhancement of underwater images prior to detection and classification is indispensable for the improvement of classification results, regardless of the use of traditional classifiers or DL approaches.

As future work in this line of research, the current developed CNN for image enhancement could be modified by adding or removing layers, modifying the number of units in each layer, or applying different parameter settings, i.e., modifying the number of epochs, the batch size or using different activation functions. Another step which would be of interest for practical applications would be the optimization of image quality *vs.* computational cost when applying these procedures to the original-sized (1600×1200 pixel) images, in order to minimize the processing time without compromising the extracted amount of valuable information. As for classification, to improve the results, other strategies like transfer learning, or even object detection networks and segmentation networks could be used. However, the amount of floating particles in some images, and the small size of some species, could hinder the performance and results of this type of networks.

Availability of data and materials
All raw footage used in this study were archived and are available in the Oceans 2.0 database (data.oceannetworks.ca/DataSearch).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Aguzzi J, Chatzievangelou D, Company J, Thomsen L, Marini S, Bonofiglio F, Juanes F, Rountree R, Berry A, Chumbinho R, et al. The potential of video imagery from worldwide cabled observatory networks to provide information supporting fish-stock and biodiversity assessment. ICES J Mar Sci. 2020;77(7–8):2396–410.
2. Aguzzi J, Chatzievangelou D, Francescangeli M, Marini S, Bonofiglio F, del Rio J, Danovaro R. The hierarchic treatment of marine ecological information from spatial networks of benthic platforms. Sensors. 2020;20(6):1751.
3. Aguzzi J, Chatzievangelou D, Marini S, Fanelli E, Danovaro R, Flögel S, Lebris N, Juanes F, De Leo FC, Del Rio J, et al. New high-tech flexible networks for the monitoring of deep-sea ecosystems. Environ Sci Technol. 2019;53(12):6616–31.
4. Aguzzi J, Costa C, Calisti M, Funari V, Stefanni S, Danovaro R, Gomes HI, Vecchi F, Dartnell LR, Weiss P, et al. Research trends and future perspectives in marine biomimicking robotics. Sensors. 2021;21(11):3778.
5. Aguzzi J, Costa C, Matabos M, Azzurro E, Lázaro A, Menesatti P, Sarda F, Canals M, Delory E, Cline D, Favali P, Juniper S, Furushima Y, Fujiwara Y, Chiesa J, Marotta L, Bahamón N, Priede I. Challenges to the assessment of benthic populations and biodiversity as a result of rhythmic behaviour: video solutions from cabled observatories. Oceanogr Mar Biol. 2012;50:235–86.
6. Aguzzi J, Costa C, Menesatti P, García JA, Bahamon N, Puig P, Sarda F, et al. Activity rhythms in the deep-sea: a chronobiological approach. Front Biosci (Landmark Edition). 2011;16:131–50.
7. Aguzzi J, Costa C, Robert K, Matabos M, Antonucci F, Juniper SK, Menesatti P. Automated image analysis for the detection of benthic crustaceans and bacterial mat coverage using the VENUS undersea cabled network. Sensors. 2011;11(11):10534–56.
8. Aguzzi J, Flögel S, Marini S, Thomsen L, Albiez J, Weiss P, Picardi G, Calisti M, Stefanni S, Mirimin L, et al. Developing technological synergies between deep-sea and space research. Elementa-Sci Anthropocene. 2022;10(1):1–9.
9. Ancuti CO, Ancuti C, De Vleeschouwer C, Bekaert P. Color balance and fusion for underwater image enhancement. IEEE Trans Image Process. 2018;27(1):379–93.
10. Anh DH, Pao S, Wataru K. Fish detection by LBP cascade classifier with optimized processing pipeline. 2013.
11. Ballard DH. Modular learning in neural networks. AAAI, 1987;279–284.
12. Bellingham JG, Rajan K. Robotics in remote and hostile environments. Science. 2007;318(5853):1098–102.
13. Beyan C, Browman HI. Setting the stage for the machine intelligence era in marine science. ICES J Mar Sci. 2020;77(4):1267–73.
14. Bicknell AW, Godley BJ, Sheehan EV, Votier SC, Witt MJ. Camera technology for monitoring marine biodiversity and human impact. Front Ecol Environ. 2016;14(8):424–32.
15. Bjerring JC, Busch J. Artificial intelligence and patient-centered decision-making. Philos Technol. 2021;34(2):349–71.
16. Boudhane M, Balcers O. Underwater image enhancement method using color channel regularization and histogram distribution for underwater vehicles AUVs and ROVs. Int J Circuits. 2019;13:571–8.
17. Boudhane M, Nsiri B. Underwater image processing method for fish localization and detection in submarine environment. J Vis Commun Image Represent. 2016;39:226–38.
18. Cao S, Zhao D, Liu X, Sun Y. Real-time robust detector for underwater live crabs based on deep learning. Comput Electron Agric. 2020;172: 105339.
19. Cao X, Zhang X, Yu Y, Niu L. Deep learning-based recognition of underwater target. IEEE Int Conf Digital Signal Proc (DSP). 2016;2016:89–93.
20. Chatzievangelou D, Aguzzi J, Ogston A, Suárez A, Thomsen L. Visual monitoring of key deep-sea megafauna with an Internet Operated crawler as a tool for ecological status assessment. Prog Oceanogr. 2020;184: 102321.
21. Chatzievangelou D, Aguzzi J, Scherwath M, Thomsen L. Quality control and pre-analysis treatment of the environmental datasets collected by an internet operated deep-sea crawler during its entire 7-year long deployment (2009–2016). Sensors. 2020;20(10):2991.
22. Chatzievangelou D, Bahamon N, Martini S, del Rio Fernandez J, Riccobene G, Tangherlini M, Roberto D, Cabrera De Leo F, Pirenne B, Aguzzi J. Integrating diel vertical migrations of bioluminescent deep scattering layers into monitoring programs. Front Mar Sci. 2021;8:615.

23. Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Comput Electron Agric. 2018;151:61–9.

24. Corrigan D, Sooknanan K, Doyle J, Lordan C, Kokaram A. A low-complexity mosaicing algorithm for stock assessment of seabed-burrowing species. IEEE J Oceanic Eng. 2018;44(2):386–400.

25. Costello MJ, Cheung A, De Hauwere N. Surface area and the seabed area, volume, depth, slope, and topographic variation for the world's seas, oceans, and countries. Environ Sci Technol. 2010;44(23):8821–8.

26. Cutter G, Stierhoff K, Zeng J. Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: Labeled fishes in the wild. Applications and Computer Vision Workshops (WACVW), 2015 IEEE Winter, 2015;57–62.

27. Danovaro R, Aguzzi J, Fanelli E, Billet D, Gjerde K, Jamieson A, Ramirez-Llodra E, Smith C, Snelgrove P, Thomsen L, et al. A new international ecosystem-based strategy for the global deep ocean. Science. 2017;355:452–4.

28. Danovaro R, Fanelli E, Aguzzi J, Billett D, Carugati L, Corinaldesi C, Dell'Anno A, Gjerde K, Jamieson AJ, Kark S, et al. Ecological variables for developing a global deep-ocean monitoring and conservation strategy. Nat Ecol Evol. 2020;4(2):181–92.

29. Death G, Fabricius KE, Sweatman H, Puotinen M. The 27–year decline of coral cover on the Great Barrier Reef and its causes. Proc Natl Acad Sci. 2012;109(44):17995–9.

30. Del Río J, Aguzzi J, Costa C, Menesatti P, Sbragaglia V, Nogueras M, Sarda F, Manuèl A. A new colorimetrically-calibrated automated video-imaging protocol for day-night fish counting at the OBSEA coastal cabled observatory. Sensors. 2014;13(11):14740–53.

31. Del-Rio J, Nogueras M, Toma DM, Martínez E, Artero-Delgado C, Bghiel I, Martinez M, Cadena J, Garcia-Benadi A, Sarria D, et al. Obsea: a decadal balance for a cabled observatory deployment. IEEE Access. 2020;8:33163–77.

32. Doya C, Chatzievangelou D, Bahamon N, Purser A, De Leo F, Juniper K, Thomsen L, Aguzzi J. Seasonal monitoring of deep-sea cold-seep benthic communities using an Internet Operated Vehicle (IOV). PLoS ONE. 2017;12: e0176917.

33. Favali P, Chierici F, Marinaro G, Giovanetti G, Azzarone A, Beranzoli L, De Santis A, Embriaco D, Monna S, Bue NL, et al. NEMO-SN1 abyssal cabled observatory in the Western Ionian Sea. IEEE J Oceanic Eng. 2013;38(2):358–74.

34. Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett. 2006;27(8):861–74.

35. Flögel S, Ahrns I, Nuber C, Hildebrandt M, Duda A, Schwendner J, Wilde D. A new deep-sea crawler system-MANSIO-VIATOR. OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). 2018;2018:1–10.

36. Forczmański P, Nowosielski A, PawełMarczeski. Video stream analysis for fish detection and classification. 2015; (pp. 157–169). Springer.

37. Garcia JA, Sbragaglia V, Masip D, Aguzzi J. Long-term video tracking of cohoused aquatic animals: a case study of the daily locomotor activity of the Norway lobster (Nephrops norvegicus). 2019.

38. Gervais F, Juniper S, Matabos M, Spicer A. Marine Life Field Guide. NEPTUNE-Canada Publications. 2012. https://www.oceannetworks.ca/science/publications/general-interest/marine-life-field-guide.

39. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning (Vol. 1). MIT press Cambridge. 2016.

40. Haddock SH, Christianson LM, Francis WR, Martini S, Dunn CW, Pugh PR, Mills CE, Osborn KJ, Seibel BA, Choy CA, et al. Insights into the biodiversity, behavior, and bioluminescence of deep-sea organisms using molecular and maritime technology. Oceanography. 2017;30(4):38–47.

41. Hays GC, Ferreira LC, Sequeira AM, Meekan MG, Duarte CM, Bailey H, Bailleul F, Bowen WD, Caley MJ, Costa DP, et al. Key questions in marine megafauna movement ecology. Trends Ecol Evol. 2016;31(6):463–75.

42. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016;770–778.

43. Heidemann J, Ye W, Wills J, Syed A, Li Y. Research challenges and applications for underwater sensor networking. IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006; 1: 228–235.

44. Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. 2010 20th International Conference on Pattern Recognition, 2010;2366–2369.

45. Hossain E, Alam SS, Ali AA, Amin MA. Fish activity tracking and species identification in underwater video. 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016;62–66.

46. Jones DO, Gates AR, Huvenne VA, Phillips AB, Bett BJ. Autonomous marine environmental monitoring: application in decommissioned oil fields. Sci Total Environ. 2019;668:835–53.

47. Juniper SK, Matabos M, Mihaly SF, Ajayamohan RS, Gervais F, Bui AOV. A year in Barkley Canyon: a time-series observatory study of mid-slope benthos and habitat dynamics using the NEPTUNE Canada network. Deep Sea Res Part II. 2013;92:114–23.

48. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. ArXiv Preprint ArXiv: 1412.6980.

49. Kratzert F, Mader H. Fish species classification in underwater video monitoring using Convolutional Neural Networks. OpenKratzert, Frederik, and Helmut Mader. "Fish Species Classification in Underwater Video Monitoring Using Convolutional Neural Networks". EarthArXiv, 2018;15.

50. Lantéri N, Legrand J, Moreau B, Lagadec JR, Rolin JF. The EGIM, a generic instrumental module to equip EMSO observatories. OCEANS 2017-Aberdeen, 2017;1–5.

51. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324.

52. LeCun Y, Jackel LD, Boser B, Denker JS, Graf HP, Guyon I, Henderson D, Howard RE, Hubbard W. Handwritten digit recognition: applications of neural network chips and automatic learning. IEEE Commun Mag. 1989;27(11):41–6.

53. Lee J, Davari H, Singh J, Pandhare V. Industrial artificial intelligence for industry 4.0-based manufacturing systems. Manuf Lett. 2018;18:20–3.

54. Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, Tao D. An underwater image enhancement benchmark dataset and beyond. IEEE Trans Image Process. 2019;29:4376–89.

55. Li J-PO, Liu H, Ting DS, Jeon S, Chan RP, Kim JE, Sim DA, Thomas PB, Lin H, Chen Y, et al. Digital technology, telemedicine and artificial intelligence in ophthalmology: a global perspective. Progress in Retinal and Eye Research, 2020;100900.

56. Liang J, Fu Z, Lei X, Dai X, Lv B. Recognition and Classification of Ornamental Fish Image Based on Machine Vision. 2020 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), 2020;910–913.

57. Lopez-Vazquez V, Lopez-Guede JM, Marini S, Fanelli E, Johnsen E, Aguzzi J. Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. Sensors. 2020;20(3):726. https://doi.org/10.3390/s20030726.

58. Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F. ResFeats: residual network based features for underwater image classification. Image Vis Comput. 2020;93: 103811.

59. Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, Fisher RB. Automatic annotation of coral reefs using deep learning. Oceans 2016 Mts/Ieee Monterey, 2016;1–5.

60. Marini S, Corgnati L, Mantovani C, Bastianini M, Ottaviani E, Fanelli E, Aguzzi J, Griffa A, Poulain P-M. Automated estimate of fish abundance through the autonomous imaging device GUARD1. Measurement. 2018;126:72–5.

61. Marini S, Fanelli E, Sbragaglia V, Azzurro E, Fernandez JDR, Aguzzi J. Tracking fish abundance by underwater image recognition. Sci Rep. 2018;8(1):1–12.

62. Martin-Abadal M, Ruiz-Frau A, Hinz H, Gonzalez-Cid Y. Jellytoring: real-time jellyfish monitoring based on deep learning object detection. Sensors. 2020;20(6):1708.

63. Mason JC, Branch A, Xu G, Jakuba MV, German CR, Chien S, Bowen AD, Hand KP, Seewald JS. Evaluation of AUV search strategies for the localization of hydrothermal venting. 2020.

64. McLean CN. United Nations Decade of Ocean Science for Sustainable Development. AGU Fall Meeting Abstracts, 2018, PA54B-10.

65. McLean DL, Parsons MJ, Gates AR, Benfield MC, Bond T, Booth DJ, Bunce M, Fowler AM, Harvey ES, Macreadie PI, et al. Enhancing the scientific value of industry remotely operated vehicles (ROVs) in our oceans. Front Mar Sci. 2020;7:220.

66. Migliore DA, Matteucci M, Naccari M. A revaluation of frame difference in fast and robust motion detection. Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks, 2006;215–218.

67. Milligan R, Morris K, Bett B, Durden J, Jones D, Robert K, Ruhl H, Bailey D. High resolution study of the spatial distributions of abyssal fishes by autonomous underwater vehicle. Sci Rep. 2016;6(1):1–12.

68. Milligan R, Scott E, Jones D, Bett B, Jamieson A, O'Brien R, Costa S, Rowe G, Ruhl H, Smith K, Susanne P, Vardaro M, Bailey D. Evidence for seasonal cycles in deep-sea fish abundances: a great migration in the deep SE Atlantic? J Anim Ecol. 2020;89:1593–603.

69. Naddaf-Sh M, Myler H, Zargarzadeh H, et al. Design and implementation of an assistive real-time red lionfish detection system for AUV/ROVs. Complexity, 2018.

70. Nair V, Hinton G. Rectified linear units improve restricted Boltzmann machines. Proc ICML. 2010;27:807–14.

71. Osterloff J, Nilssen I, Järnegren J, Van Engeland T, Buhl-Mortensen P, Nattkemper TW. Computer vision enables short-and long-term analysis of Lophelia pertusa polyp behaviour and colour from an underwater observatory. Sci Rep. 2020;9(1):1–12.

72. Panetta K, Gao C, Agaian S. Human-visual-system-inspired underwater image quality measures. IEEE J Oceanic Eng. 2016;41(3):541–51.

73. Panetta K, Zhou Y, Agaian S, Jia H. Nonlinear unsharp masking for mammogram enhancement. IEEE Trans Inf Technol Biomed. 2011;15(6):918–28.

74. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

75. Piechaud N, Hunt C, Culverhouse PF, Foster NL, Howell KL. Automated identification of benthic epifauna with computer vision. Mar Ecol Prog Ser. 2019;615:15–30.

76. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, Romeny BH, Zimmerman JB, Zuiderveld K. Adaptive histogram equalization and its variations. Comput Vision Graphics Image Proc. 1987;39:355–68.

77. Poynton C. Digital video and HD: Algorithms and Interfaces. Elsevier. 2012.

78. Pramunendar RA, Wibirama S, Santosa PI, Andono PN, Soeleman MA. A robust image enhancement techniques for underwater fish classification in marine environment. Int J Intell Eng Syst. 2019;12(5):116.

79. Ramirez-Llodra E, Brandt A, Danovaro R, Mol BD, Escobar E, German CR, Levin LA, Martinez Arbizu P, Menot L, Buhl-Mortensen P, et al. Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. Biogeosciences. 2010;7(9):2851–99.

80. Rathi D, Jain S, Indu DS. Underwater fish species classification using convolutional neural network and deep learning. 2018. ArXiv Preprint ArXiv: 1805.10106.

81. Rimavicius T, Gelzinis A. A comparison of the deep learning methods for solving seafloor image classification task. International Conference on Information and Software Technologies, 2017;442–453.

82. Roelfsema C, Kovacs EM, Vercelloni J, Markey K, Rodriguez-Ramirez A, Lopez-Marcano S, Gonzalez-Rivero M, Hoegh-Guldberg O, Phinn SR. Fine-scale time series surveys reveal new insights into spatio-temporal trends in coral cover (2002–2018), of a coral reef on the Southern Great Barrier Reef. Coral Reefs, 2021;1–13.

83. Salman A, Jalal A, Shafait F, Mian A, Shortis M, Seager J, Harvey E. Fish species classification in unconstrained underwater environments based on deep learning. Limnol Oceanogr Methods. 2016;14:570–85.

84. Sanila K, Balakrishnan AA, Supriya M. Underwater image enhancement using white balance, USM and CLHE. Int Symposium Ocean Technol (SYMPOL). 2019;2019:106–16.

85. Schoening T, Bergmann M, Ontrup J, Taylor J, Dannheim J, Gutt J, Purser A, Nattkemper TW. Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN. PLoS ONE. 2012;7(6):e38179.

86. Schoening T, Purser A, Langenkämper D, Suck I, Taylor J, Cuvelier D, Lins L, Simon-Lledó E, Marcon Y, Jones DO, et al. Megafauna community assessment of polymetallic-nodule fields with cameras: platform and methodology comparison. Biogeosciences. 2020;17(12):3115–33.

87. Simon-Lledó E, Bett BJ, Huvenne VA, Köser K, Schoening T, Greinert J, Jones DO. Biological effects 26 years after simulated deep-sea mining. Sci Rep. 2019;9(1):1–13.

88. Sokolova M, Mompó Alepuz A, Thompson F, Mariani P, Galeazzi R, Krag LA. A deep learning approach to assist sustainability of demersal trawling operations. Sustainability. 2021;13(22):12362.
89. Sokolova M, Thompson F, Mariani P, Krag LA. Towards sustainable demersal fisheries: NepCon image acquisition system for automatic Nephrops norvegicus detection. PLoS ONE. 2021;16(6): e0252824.
90. Spampinato C, Giordano D, Salvo RD, Chen-Burger Y-HJ, Fisher RB, Nadarajan G. Automatic fish classification for underwater species behavior understanding. Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, 2010;45–50.
91. Sutton TT, Frank T, Judkins H, Romero IC. As Gulf oil extraction goes deeper, who is at risk? Community structure, distribution, and connectivity of the deep-pelagic fauna. In Scenarios and Responses to Future Deep Oil Spills. 2020; (pp. 403–418). Springer.
92. Sweetman AK, Thurber AR, Smith CR, Levin LA, Mora C, Wei CL, Gooday AJ, Jones DO, Rex M, Yasuhara M, et al. Major impacts of climate change on deep-sea benthic ecosystems. Elementa Sci Anthropocene. 2017; 5.
93. Thomsen L, Aguzzi J, Costa C, De Leo F, Ogston A, Purser A. The oceanic biological pump: rapid carbon transfer to depth at continental margins during winter. Sci Rep. 2017;7(1):1–10.
94. Thomsen L, Barnes C, Best M, Chapman R, Pirenne B, Thomson R, Vogt J. Ocean circulation promotes methane release from gas hydrate outcrops at the NEPTUNE Canada Barkley Canyon node. Geophys Res Lett, 2012;39(16).
95. Villon S, Chaumont M, Subsol G, Villéger S, Claverie T, Mouillot D. Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between Deep Learning and HOG+ SVM methods. International Conference on Advanced Concepts for Intelligent Vision Systems, 2016;160–171.
96. Wang Y, Zhang J, Cao Y, Wang Z. A deep CNN method for underwater image enhancement. 2017 IEEE International Conference on Image Processing (ICIP), 2017;1382–1386.
97. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12.
98. Wedler A, Wilde M, Dömel A, Müller MG, Reill J, Schuster M, Stürzl W, Triebel R, Gmeiner H, Vodermayer B, et al. From single autonomous robots toward cooperative robotic interactions for future planetary exploration missions. Proceedings of the International Astronautical Congress, IAC. 2018.
99. Willis BL, Page CA, Dinsdale EA. Coral disease on the great barrier reef. In Coral health and disease. 2004;(pp. 69–104). Springer.
100. Wu D, Yuan F, Cheng E. Underwater no-reference image quality assessment for display module of ROV. Scientific Programming, 2020; 8856640:1–8856640:15.
101. Wu H, He S, Deng Z, Kou L, Huang K, Suo F, Cao Z. Fishery monitoring system with AUV based on YOLO and SGBM. 2019 Chinese Control Conference (CCC), 2019;4726–4731.
102. Yang M, Sowmya A. An underwater color image quality evaluation metric. IEEE Trans Image Process. 2015;24(12):6062–71.
103. Yao H, Duan Q, Li D, Wang J. An improved K-means clustering algorithm for fish image segmentation. Math Comput Model. 2013;58(3–4):790–8.
104. Zhang Y, Ryan JP, Hobson BW, Kieft B, Romano A, Barone B, Preston CM, Roman B, Raanan B-Y, Pargett D, et al. A system of coordinated autonomous robots for Lagrangian studies of microbes in the oceanic deep chlorophyll maximum. Sci Robot. 2021;6(50):eabb9138.
105. Zuazo A, Grinyó J, López-Vázquez V, Rodríguez E, Costa C, Ortenzi L, Flögel S, Valencia J, Marini S, Zhang G, et al. An automated pipeline for image processing and data treatment to track activity rhythms of paragorgia arborea in relation to hydrographic conditions. Sensors. 2020;20(21):6281.
106. Zuiderveld K. Contrast limited adaptive histogram equalization. Graphics Gems IV. 1994;474–485.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.