



Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks

Gómez-Ríos, Anabel; Tabik, S.; Luengo, Julián; Shihavuddin, A. S.M.; Herrera, Francisco

Published in:
Knowledge-Based Systems

Link to article, DOI:
[10.1016/j.knosys.2019.104891](https://doi.org/10.1016/j.knosys.2019.104891)

Publication date:
2019

Document Version
Early version, also known as pre-print

[Link back to DTU Orbit](#)

Citation (APA):
Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A. S. M., & Herrera, F. (2019). Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks. *Knowledge-Based Systems*, 184, Article 104891. <https://doi.org/10.1016/j.knosys.2019.104891>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

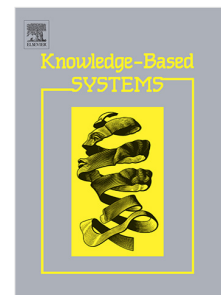
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Journal Pre-proof

Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks

Anabel Gómez-Ríos, Siham Tabik, Julián Luengo,
A.S.M. Shihavuddin, Francisco Herrera



PII: S0950-7051(19)30356-9
DOI: <https://doi.org/10.1016/j.knosys.2019.104891>
Reference: KNOSYS 104891

To appear in: *Knowledge-Based Systems*

Received date : 28 January 2019
Revised date : 2 July 2019
Accepted date : 27 July 2019

Please cite this article as: A. Gómez-Ríos, S. Tabik, J. Luengo et al., Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks, *Knowledge-Based Systems* (2019), doi: <https://doi.org/10.1016/j.knosys.2019.104891>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier B.V.

Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks

Anabel Gómez-Ríos^{a,*}, Siham Tabik^a, Julián Luengo^a, ASM Shihavuddin^b,
Francisco Herrera^a

^aAndalusian Research Institute in Data Science and Computational Intelligence, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18071, Granada, Spain

^bDept. of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

Abstract

Corals are crucial animals as they support a large part of marine life. The automatic classification of corals species based on underwater images is important as it can help experts to track and detect threatened and vulnerable coral species. However, this classification is complicated due to the nature of coral underwater images and the fact that current underwater coral datasets are unrealistic as they contain only texture images, while the images taken by autonomous underwater vehicles show the complete coral structure. The objective of this paper is two-fold. The first is to build a dataset that is representative of the problem of classifying underwater coral images, the StructureRSMAS dataset. The second is to build a classifier capable of resolving the real problem of classifying corals, based either on texture or structure images. We have achieved this by using a two-level classifier composed of three ResNet models. The first level recognizes whether the input image is a texture or a structure image. Then, the second level identifies the coral species. To do this, we have used a known texture dataset, RSMAS, and StructureRSMAS.

Keywords: Coral Images Classification, Structure Coral Images, Deep

*Corresponding author

Email addresses: anabelgrios@decsai.ugr.es (Anabel Gómez-Ríos), siham@ugr.es (Siham Tabik), julianlm@decsai.ugr.es (Julián Luengo), shihav@dtu.dk (ASM Shihavuddin), herrera@decsai.ugr.es (Francisco Herrera)

Learning, Convolutional Neural Networks, Inception, ResNet, DenseNet.

1. Introduction

Coral reefs are extremely valuable ecosystems for marine life and humans. A recent evaluation of 40% of the total number of coral species has shown that more than 200 species are threatened [1]. This is a direct consequence of air and water pollution and the changes in ocean temperatures due to climate change [2].

The automatic classification of corals based on images is a hard task. There are thousands of coral species and their taxonomy is continuously updated as more information is obtained. In addition, some coral species look really similar externally whereas their differences are based on internal characteristics. This makes difficult to keep a constant record of all the species, even more their extension rates. Nowadays, thousands of images of coral reefs and other benthic habitats are being captured regularly by Autonomous Underwater Vehicles (AUVs). However, analysing that huge amount of data and acquiring useful information out of it is still a bottleneck, as many hours of experts manual work are involved in such tasks. Having a rigorous accurate automatic coral classifier can potentially help in analysing the large amount of data, thereby progressing in the understanding of coral reefs. Nevertheless, the available coral datasets contain, in general, texture images, as opposed to the structure images obtained by AUVs. This is a consequence of using classical machine learning models, since they need to extract textures as a previous feature extraction step. Texture images show only a small part of the coral and do not include any information of the whole structure of the coral body, while structure images contain the whole coral or a large part of it. Figure 1 shows an example of (a) a coral texture or local information and (b) a coral structure or global information. The available public coral datasets, such as EILAT (which contains images taken near Eilat) RSMAS (Rosenstiel School of Marine and Atmospheric Sciences) or MLC (Moorea Labelled Corals), among others, only contain texture images. As far

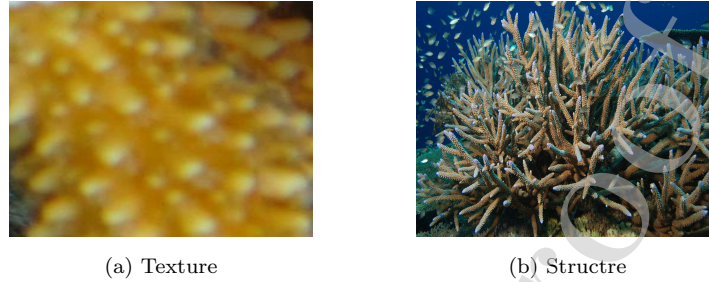


Figure 1: Difference between (a) coral texture and (b) coral structure.

as we know, there is no publicly available datasets containing coral structure
 30 images.

Making the coral classification automatic has been tackled in previous works, using either normal machine learning models combined with feature extraction methods [3, 4, 5, 6] or Convolutional Neural Networks (CNNs) [7, 8, 9, 10]. Recently, CNNs are providing outstanding performance in pattern recognition
 35 in many fields, particularly in Computer Vision [11, 12, 13]. A clear example is the ImageNet Large Scale Visual Recognition Challenge(ILSVRC) competition [14], whose top ten models have been CNNs since 2012. A key to its success is that they are capable to extract simple and complex features as the network goes deeper [15, 16]. In fact, since the last few years, they are capable to use
 40 the simple features in the deeper layers by adding some connections that skip layers [17, 18]. Furthermore, they are able to overcome the limitation of large datasets requirements in the training phase by increasing the size of the training set artificially (data augmentation) or by starting the training using the weights from the pre-trained network on another dataset (transfer learning).

45 However, none of the previous works address the problem of classifying coral texture and structure images together, so it is still an unsolved issue. Most CNN works analyse classical architectures such as LeNet [19] and VGGnet [20]. In [10], the authors analysed more recent CNNs, like Inception [16], ResNet [17] and DenseNet [18] on two texture datasets, RSMAS and EILAT. They achieved
 50 the state-of-the-art accuracies in both datasets. In general, CNNs outperform

classical feature extraction in the classification of coral texture images.

In this work, we aim at identifying coral species based on their texture or structure images. To the best of our knowledge, this is the first work classifying corals based on texture and structure images. We propose to use recent
 55 CNNs to classify any coral image, either texture or structure. We want to provide a classifier that could be used with pictures provided by AUVs, irrespective of the portion of coral that is contained in each image. In order to achieve this, we have used a known texture dataset, RSMAS [21], and a new structure dataset, more realistic, that we have built in this work, called
 60 StructureRSMAS. This dataset is available through the following link: <http://sci2s.ugr.es/CNN-coral-image-classification>. We have taken advantage of the costly experimentation that it has been done in [10] to choose the best model and configuration for RSMAS. For StructureRSMAS, we have evaluated Inception, ResNet and DenseNet, and we have chosen the best model and
 65 its best configuration. We have also tested the influence of some image enhancement techniques on both datasets, RSMAS and StructureRSMAS, in order to improve the images before using the CNN models. Finally, we have built a two-level classifier, whose first level is to decide whether the input coral image is a texture or a structure and whose second level is to identify the coral species.
 70 Similar strategies have also been used in other works [22]. To obtain the model in the first level, we have also evaluated the three CNN architectures mentioned before. To the best of our knowledge, this work is the first identifying coral species based on its local or global information.

The rest of the paper is organized as follows. In Section 2, we give an
 75 overview on CNNs and the three architectures we have used. In Section 3, a summary on previous works for classifying underwater coral images is given. In Section 4, we present the structure dataset we have created. In Section 5, we describe our proposal to classify coral images, either texture or structure and in Section 6 we explain the experiments we have carried out to obtain such a
 80 classifier. Finally, in Section 7 we state the conclusions of this work.

2. Convolutional Neural Networks (CNNs) and improvement techniques

CNNs are a widely used type of artificial neural network. They constitute the state-of-the-art in object recognition in images. Using CNNs we do not have to extract the features from the images with a previous algorithm. This is possible thanks to its principal operation: the *convolution*, which is defined, for a point (i, j) , as [23]:

$$\sum_m \sum_n I(i - m, j - n) K(m, n) , \quad (1)$$

where I is the input of the convolution and K is the kernel of the convolution with size $m \times n$. Each convolutional layer has different numbers of kernels. As we can see in Figure 2, the input of the convolutional layer is convoluted with each kernel, resulting in an output feature map per kernel. The values of these kernels, called weights, are learned by the CNN autonomously during the training process. At the end of the training, these values are the selected features of the images.

To increase the non-linearity of the models, every convolutional layer is followed by a non-linear operation, typically the Rectified Linear Unit (ReLU) operation.

To increase the abstraction level of the extracted features, it is usual to reduce the size of the feature maps from the convolutional layer using a *pooling* layer, which takes a neighbourhood of size $m \times n$ and performs an operation to it. There are several variants, depending on the operation used: maximum pooling, minimum pooling, average pooling, etc. The size is reduced, as we can see in Figure 2, by using a stride between one group of pixels and the next one. Among other things, pooling allows the network to extract simple features at the beginning and complex features as the network goes deeper. This is why the first improvement to obtain better results with CNNs was to make the networks deeper. However, it has the problem that the deeper the network, the harder it is to train it. On the other hand, it is usual that, at the same

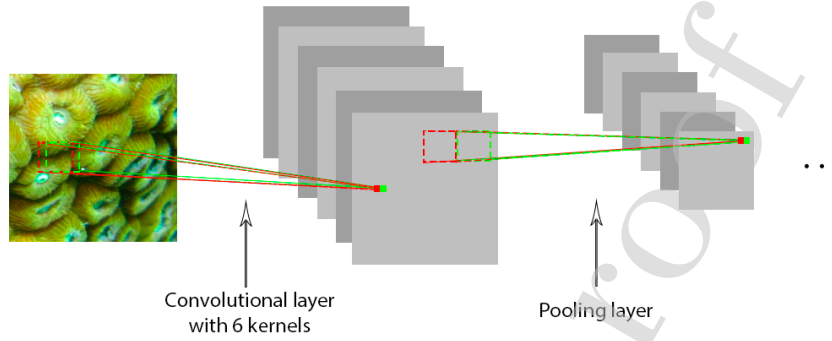


Figure 2: An example of a convolutional layer and a pooling layer in a CNN.

level, the network has various convolutional layers, increasing the width of the
 110 network and extracting different features. This is clearly seen in Inception v3
 architecture [16].

In the last years, some architectures, like ResNet [17] and DenseNet [18], have
 explored other ways to improve the performance of the networks. In particular,
 they added connections that skip layers, making the network to use the simple
 115 features extracted at the beginning in deeper layers. This has been proved to
 work well: ResNet won the ILSVRC competition [14] in 2015 and DenseNet
 beat its results in 2016.

In order to obtain good results from a CNN, it is necessary to train it with
 large datasets. However, there are two improvement techniques that help to
 120 overcome this constraint: transfer learning and data augmentation.

By using *transfer learning* we can start the training of the network from the
 pre-trained weights in another problem, although it is recommended that this
 other problem is somehow related with the problem that we want to resolve.
 Then, we can choose between retraining only the last layer of the network, which
 125 classifies the images into our classes, or to retrain all the weights in the network
 (or some of them), which is also called fine-tuning.

Data augmentation allow us to artificially increase the size of the training
 set, so we have more images to train the network. The increase in the training

set is done by applying several distortions to the original images, like zooming
 130 them, flipping them horizontally or vertically, rotating them, shifting them, etc.
 They can be applied alone or combined, and in most of them we can choose how
 much distortion we want to apply: rotate an image 15 degrees, zoom 25% of the
 image, etc. Depending on the original images, we need to be careful with the
 data augmentation techniques we apply as the images may lose their meaning:
 135 for example, flipping the images in hand-writing digit classification.

In this work, we will use Inception v3, ResNet and DenseNet. These three
 architectures are based on the repetition of a block, different in each case. The
 composition of these networks is shown in Table 1. As it is seen in the table,
 Inception v3 has a fixed number of layers. That is because it has a fixed number
 140 of blocks, although the base inception module is only used at the beginning of
 the network and it has several modifications as the network goes deeper. On the
 other hand, the number of layers in ResNet and DenseNet is a hyperparameter
 and it depends on the number of times their blocks are repeated. There are two
 more hyperparameters in the three architectures: the number of epochs and the
 145 batch size we used to train them.

3. Related work on coral classification

In this section we first present the available coral datasets that have been
 used in the literature. Second, we present the previous works on the classifica-
 tion of coral images and the results obtained for RSMAS, the texture dataset
 150 we have used.

3.1. Coral datasets

The current available eight open datasets are KTH-TIPS (Textures under
 varying Illumination, Pose and Scale), CURET (Columbia-Utrecht Reflectance
 and Texture), UIUCtex (University of Illinois at Urbana-Champaign texture
 155 dataset), MLC (Moorea Labelled Corals), EILAT, EILAT2, RSMAS (Rosenstiel
 School of Marine and Atmospheric Sciences) and the Red Sea Mosaic. Only

Table 1: Description of the composition of Inception v3, ResNet and DenseNet. BN stands for Batch Normalization.

Architecture	Name of the block	Composition of the block	Same block along the network?	Number of layers of the network
Inception v3	Base inception module	Three 1×1 convolutions, four 3×3 convolutions and one pooling. Some of these operations are made in parallel, so it has a concatenation filter at the end. All the convolutions are followed by BN and a ReLU.	No	42
ResNet	Building block	Three consecutive operations: 1×1 conv., 3×3 conv., and 1×1 conv. It has an additional connection between the input of the first 1×1 conv. and the output of the second 1×1 conv that performs a 1×1 conv. or an identity. All the convolutions are followed by BN and a ReLU.	Yes	It depends on the number of times the block is repeated.
DenseNet	Dense block	A repetition of the sequence: BN, ReLU, 1×1 conv., BN, ReLU and 3×3 conv. The output of all the layers inside the block is connected with the input of all the following layers in the block.	Yes	It depends on the number of times the block is repeated and the repetitions inside each block.

MLC, EILAT, EILAT2, RSMAS and the Red Sea Mosaic contain RGB images. From these, RSMAS and EILAT are the ones with more number of classes, 14 and 8, respectively, and the most recent. EILAT2 is a subset of EILAT, MLC
 160 only contains five coral classes and the Red Sea Mosaic is actually a large image containing different coral species.

The images in all these datasets share some characteristics inherent to underwater images: the water movement cause lightning variations between images taken at the same time; the water causes the images to be blurry; and it is very
 165 common that animals, like fish, cover part of the corals when the images are being taken. On the other hand, the images share some characteristics inherent to coral images, like the occurrence of various coral species in the same image or the subjective classification of the images by different experts.

Both RSMAS and EILAT are texture datasets, but we choose RSMAS as
 170 our texture dataset because the labelling is the scientific Latin name of the coral species, which allow us to obtain structure images from these species and to create StructureRSMAS, a coral structure dataset that contains images from the same species that RSMAS. RMSAS contains 766 images that are patches that contain specific and small parts of the corals, not the entire structure of
 175 them. They are close-up images, so they are sometimes blurry and small: each one of them have 256×256 pixels in size. Some images from RSMAS can be seen in Figure 3.

3.2. Previous works

Before recent advances in deep learning, most previous approaches on the
 180 automatic coral reef classification with underwater images combined classical machine learning models with feature extraction algorithms [24, 6, 3, 5]. They used, in general, an algorithm to extract color features and another one to extract texture features. Then, they used them to train a classical machine learning model, like Support Vector Machines in [3] or a three layer neural
 185 network in [24]. Most of them used a single dataset, usually small and containing several non-coral classes, except for [5], where the authors designed an algorithm

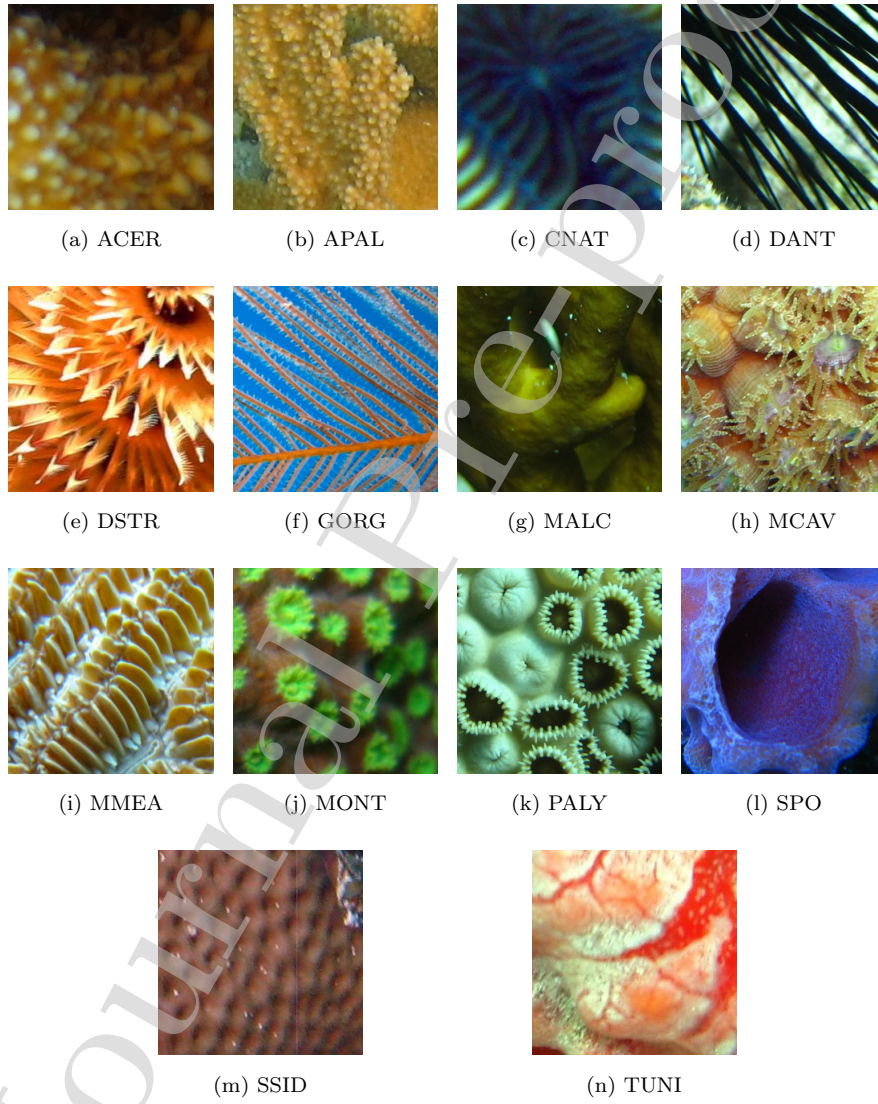


Figure 3: One texture image from each RSMAS class.

that could be used with different datasets. They divided their algorithm into steps and at each step several sub-algorithms could be chosen in accordance with the dataset that was going to be classified.

190 The authors in [25, 26, 27] proposed three different modifications of the local binary pattern feature descriptor, and they used the extracted features to train machine learning models in order to classify several datasets.

In 2015, the author in [7] used CNNs for the first time to resolve this task. He used a LeNet-5 model, but he still combined it with feature extraction algorithms. Since then, most works classifying coral images have used CNNs, as 195 they provide better results and do not need a previous step of feature extraction. The authors in [8] and [9] used a VGGnet model and in [9] the authors also combined the features extracted by the CNN model with hand-crafted features, although the improvement from such custom features was small. In [8] they 200 used the Benthos15 dataset and in [9] they used the MLC dataset.

In [10], the authors proposed the use of more powerful CNNs to classify EILAT and RSMAS [21], and they found that ResNet was the best CNN architecture in both datasets, improving the state-of-the-art results obtained by [5, 25, 26] and [27], which were all the works that used EILAT and RSMAS.

205 The works that used RSMAS and the results they obtained can be seen in Table 2. The highest result is a recall of 99.34%, obtained by [27], but using a held out test of 10%. The authors in [10] obtained an accuracy of 98.63% using a five fold cross validation technique, which means that they were using tests sets of 20% of the dataset in each partition. Then, they took the mean of 210 the test accuracy on the five partitions. As a consequence, this result is more stable. Because of this, the best model for RSMAS, and the one we choose to use in this work, is the one obtained in [10].

An extensive review on the classification of coral images using deep learning can be found in [28]. However, most of these works used texture datasets whose 215 images show close-up and very specific patches of the corals, not the whole structure of them. The present work is different from all the previously cited works in that it develops an automatic model capable of identifying the species

Table 2: Results from previous works on RSMAS. The results of Shihavuddin et al. using a 5 fold cross validation can be found in [10].

Authors	Ref.	Metric	Result	Test method
Brown Mary et al.	[25]	Recall	98.87%	10% held out test
Brown Mary et al.	[25]	Recall	84.9%	25% held out test
Brown Mary et al.	[26]	Recall	98.1%	10% held out test
Brown Mary et al.	[26]	Recall	85.72%	25% held out test
Brown Mary et al.	[27]	Recall	99.34%	10% held out test
Brown Mary et al.	[27]	Recall	85.8%	25% held out test
Shihavuddin et al.	[5]	Accuracy	92.74%	5 fold cross validation
Gómez-Ríos et al.	[10]	Accuracy	98.63%	5 fold cross validation

of a coral based on an input image of either texture or structure.

4. StructureRSMAS: a new coral structure dataset

220 In this section we present the new coral dataset of coral structure images we have built: StructureRSMAS. We have considered the same coral species, nomenclature and number of classes as in RSMAS. We have downloaded the images from official scientific websites, e.g., the Encyclopedia of Life [29], the IUCN Red List of Threatened Species [30] or the coralpedia of the University
225 of Warwick [31]. Few images were available per class.

We have built StructureRSMAS because typical coral pictures often capture the whole coral body and structure and do not focus only on the texture of its parts. An ideal classifier should be able to recognize the coral based either on its texture or structure and with images taken under different conditions,
230 cameras, etc. To develop such a model we need images that show both texture and structure of the coral species.

StructureRSMAS contains 409 coral images of variable size but larger than the ones in RSMAS. Some examples can be seen in Figure 4. The dataset, along with a list of the sources of the images, is available through the following link:
235 <http://sci2s.ugr.es/CNN-coral-image-classification>. In Figure 3 and Figure 4 we can see the difference between texture and structure images from

Table 3: Characteristics of RSMAS and StructureRSMAS. #imgs refers to the number of images in that class.

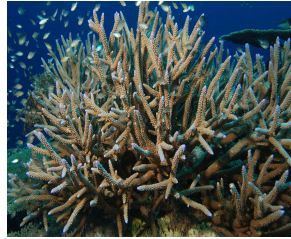
Classes	#imgs in RSMAS	#imgs in StructureRSMAS
Acropora Cervicornis (ACER).	109	44
Acropora Palmata (APAL).	77	41
Colpophyllia Natans (CNAT).	57	34
Diadema Antillarum (DANT).	63	20
Diploria Strigosa (DSTR).	24	16
Gorgonians (GORG).	60	18
Millepora Alaicornis (MALC).	22	33
Montastraea Cavernosa (MCAV).	79	38
Meandrina Meandrites (MMEA).	54	30
Montipora spp. (MONT).	28	21
Palythoa Palythoa (PALY).	32	32
Sponge Fungus (SPO).	88	23
Siderastrea Sidera (SSID).	37	36
Tunicates (TUNI).	36	23

the same coral species.

The images in RSMAS and StructureRSMAS were taken with different cameras and under different conditions. They are both imbalanced, as we show in Table 3, since there are classes that contain more images than others, like ACER in RSMAS, which has 109 images, while MALC has 22 images. The differences are smaller in StructureRSMAS. The largest difference in this dataset is between ACER, with 44 images, and DSTR, with 16 images.

5. A two-level classifier for coral classification using a texture model and a structure model

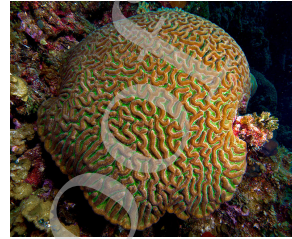
We propose the use of a two-level classifier to address the automatic classification of corals based on either texture or structure images. This classifier is composed of three models, one used in the first level and the other two used in the second level. As we show in Figure 5, the first level is to determine if



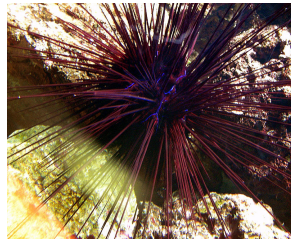
(a) ACER



(b) APAL



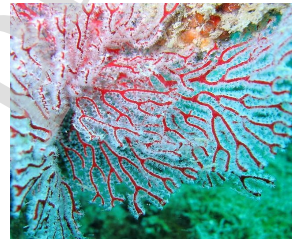
(c) CNAT



(d) DANT



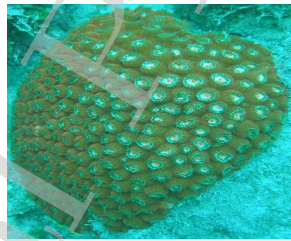
(e) DSTR



(f) GORG



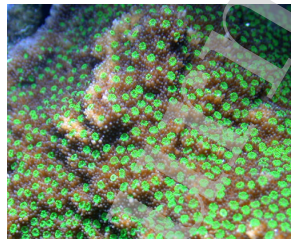
(g) MALC



(h) MCAV



(i) MMEA



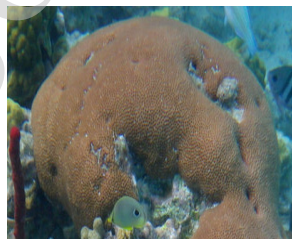
(j) MONT



(k) PALY



(l) SPO



(m) SSID



(n) TUNI

Figure 4: One structure image from each StructureRSMAS class.

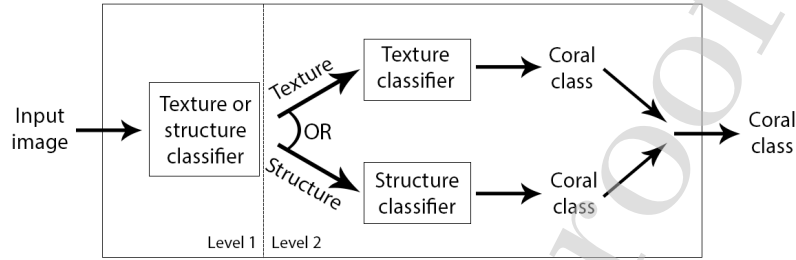


Figure 5: The two-level classifier we have developed to classify any coral image, either texture or structure.

the image is a texture or a structure. Therefore, the model used in this level is a binary classifier trained over the images of RSMAS (texture images) and the images of StructureRSMAS (structure images) as two separated classes. In the second level, we decide which coral species the image belongs to using a texture model, trained over RSMAS, or a structure model, trained over StructureRSMAS, depending on the output obtained in the first level. To build our two-level model, at the first level we have developed a binary texture or structure classification model. At the second level, we have considered the most accurate texture model from the literature and developed a new structure classification model using StructureRSMAS.

In particular, for the texture dataset, RSMAS, we have used the state-of-the-art model proposed in [10], a ResNet-152 model that used transfer learning from ImageNet and data augmentation techniques.

For the structure dataset, StructureRSMAS, and the texture or structure binary classifier, we have evaluated different CNN architectures and configurations, as we will see in the following section.

In both datasets, we have evaluated the improvement of some image enhancement techniques before the application of the data augmentation techniques.

Table 4: The set of hyperparameters we have test in the three architectures.

	Number of layers	Number of epochs	Batch size
Inception v3	42	100, 300, 500, 700, 1000, 1300	32, 64, 128
ResNet	50, 152		
DenseNet	121, 161		

6. Experimental Analysis

In this section we describe the process we have used to build the two-level classifier. First, we describe the experimental framework we have used in all the experiments. Second, we evaluate different image enhancement techniques in the RSMAS dataset. Third, we analyse and compare different CNN architectures for StructureRSMAS and choose the best of them. We also evaluate the image enhancement and data augmentation techniques in this dataset, and choose the best configuration for it. Fourth, we evaluate and compare the same CNN architectures, and data augmentation techniques to classify an image into texture or structure using the image enhancement techniques for both datasets that we found the best for them. Finally, we analyse and evaluate the performance of the proposed two-level classifier.

6.1. Experimental framework

The results shown in this section are obtained using a 5 fold cross validation technique. The known accuracy metric has been used to compare the performance of the CNNs and the data augmentation techniques.

For the implementation of Inception, ResNet and DenseNet, we have used Keras [32] with Tensorflow [33] as backend. For Inception, we have used the model already available in Keras 2.0.4, and we have adapted the code by Yu Felix in GitHub [34] for ResNet and DenseNet.

To search for the best model for StructureRSMAS and for the decision between texture and structure, we have evaluated the hyperparameters that we show in Table 4 in a grid search.

Table 5: Description of the evaluated data augmentation techniques.

Data augmentation technique	Parameter	Description
Shift	A float number x	To shift vertically and horizontally the images by a random fraction of the width or length, respectively, in $[0, x]$.
Zoom	A float number x	To zoom the images so their width and length are a random number in $[1 - x, 1 + x]$.
Rotation	An integer number x	To rotate the images by a random degree in $[0, x]$.
Flip	True/False	If true, randomly choose if the image is horizontally flipped or not.

Since RSMAS and StructureRSMAS are very small to train the CNNs from scratch, as we can see in Table 3, we have used transfer learning from ImageNet [35]. That way, we have all the networks with its weights pre-trained on ImageNet. Then, we have removed the last Fully Connected (FC) layer with 1000 neurons, which classifies the inputs into ImageNet classes, and we have added two FC layers, the first with 512 neurons and a ReLU activation and the second with as many neurons as classes in the dataset we are classifying (14 for StructureRSMAS and 2 for the classifier between RSMAS and StructureRSMAS) and a softmax activation. Lastly, we have only trained the two FC layers we have added.

Once we have chosen the best CNN model and its best parameters, we test the following image enhancement techniques: contrast and brightness enhancement (referred to later as CBE) [36, 37], saliency detection [38, 39] and deblurring [40, 41]. Then, we choose the best combination of them for each dataset and with it, we evaluate data augmentation techniques. We have evaluated the performance of data augmentation by carrying out experiments with and without the use of data augmentation techniques. The description of the data augmentation techniques we have used is in Table 5.

Table 6: Results obtained for RSMAS using ResNet-152 for each image enhancement technique. The best accuracy is stressed in bold.

	Without enhancement	CBE	Deblur	Saliency	Deblur + CBE	Deblur + CBE + Saliency
Accuracy	96.710	96.575	96.438	79.726	95.342	72.740

6.2. Second level: texture model

In this section we evaluate the image enhancement techniques in the RSMAS dataset. We use the best model and hyperparameters found in [10] for this dataset: a ResNet152 model with batch size 32 and 300 epochs for the training process. The results can be seen in Table 6, where CBE refers to contrast and brightness enhancement. As we can observe in this table, none of the image enhancement techniques helps to improve the accuracy of the model in RSMAS, though the differences, if we exclude the saliency method, are similar. We argue that this is happening because the images in this dataset are already preprocessed. In the case of the saliency method, we think this is normal as we are losing all the information in the background of the images.

As a result, we are not using any image enhancement techniques in this dataset and thus we are using the best data augmentation technique reported in [10], a random zoom of 0.4, which give us an accuracy of 98.356%.

6.3. Second level: structure model

To automatically classify any coral image from AUVs, it is necessary to classify images of entire corals. At this level, we focus on the problem of classifying structure images before solving the classification of texture and structure images together. These images are generally larger than the images in RSMAS. To simplify the number of experiments, we first have chosen the best model to classify the complete images in StructureRSMAS, among several CNNs models, without using data augmentation. After that, we have chosen the best model to evaluate the image enhancement methods and we have chosen the best of them. Once we have the best model and the best image enhancement method, we evaluate the data augmentation techniques.

Table 7: Best results obtained for StructureRSMAS using Inception, ResNet-50, ResNet-152, DenseNet-121 and DenseNet-161, and the set of hyperparameters used to obtain them, without data augmentation. The best accuracy is stressed in bold.

	Inception	ResNet-50	ResNet-152	DenseNet-121	DenseNet-161
Accuracy	81.316	83.158	83.158	54.737	55.526
Best batch size	32	32	32	64	32
Best number of epochs	700	300	1300	700	700

The results without data augmentation and without image enhancement, and the hyperparameters we have used to obtain those results, are shown in Table 7. Similarly to RSMAS, ResNet is the best classifier. Particularly, ResNet-50 achieves the highest accuracy, although Inception also obtains a competitive accuracy. DenseNet is the only CNN that does not show a good performance in this problem. As stated in [10], this is probably due to the large number of connections in DenseNet, which makes it more adapted to ImageNet than the rest of the architectures. As a consequence, it is more difficult to classify a new dataset just by training the last two layers of the network.

We observe that the accuracy obtained for StructureRSMAS is lower than the one obtained for RSMAS. This can be explained by the fact that StructureRSMAS contains fewer examples than RSMAS, and the images that belong to the same class have very different characteristics e.g., resolution, angle of view or distance from which the images are captured. Despite all of this, we achieve a good accuracy, 83.158%.

Next, we have used the ResNet-50 model with its best parameters to test the image enhancement methods in this dataset. The results we have obtained are showed in Table 8, and we can see that for this dataset it is better to first preprocess the images by using contrast and brightness enhancement and deblurring. The highest accuracy with image enhancement, 2.6% higher than without applying any image enhancement, was obtained by applying the deblurring method alone. While applying the contrast and brightness enhancement improved the accuracy by only 1.3%. As a result, we have preprocessed the

Table 8: Results obtained for StructureRSMAS using ResNet-50 for each image enhancement technique. The best accuracy is stressed in bold.

	Without enhancement	CBE	Deblur	Saliency	Deblur + CBE	Deblur + CBE + Saliency
Accuracy	82.368	83.684	85.000	52.632	83.158	51.358

Table 9: Best results obtained for StructureRSMAS using ResNet-50 for each data augmentation technique. The best accuracy is stressed in bold.

	shift = 0.4	flip	zoom = 0.4	rotation = 6	flip + rotation = 8
Accuracy	84.211	85.000	83.421	84.737	83.684

images in StructureRSMAS with the deblurring method.

Finally, we have evaluated different data augmentation techniques with the ResNet-50 classifier, and, for the sake of brevity, the best results can be seen in Table 9. They show that there is no improvement from using data augmentation in this dataset. We argue that this can be due to the very small size of the dataset, which also affects even when using data augmentation, as there are few images from which to obtain new images.

6.4. First level: texture or structure binary model

In this section we develop a classifier to distinguish whether an image is a texture or a structure. To do this, we have joined in one dataset all the images in RSMAS in one class and all the images in StructureRSMAS, preprocessed with the deblurring method, in another class. We have followed the same scheme: first, we have chosen the best model without data augmentation, and then we have evaluated several data augmentation techniques using the best model.

The results without data augmentation, along with the best hyperparameters, can be found in Table 10. We can see that all the models provide good results distinguishing textures and structures. This is actually the expected output as the images in both classes are very dissimilar. In this case the best model is ResNet-152, achieving an accuracy of 99.730%.

We have evaluated ResNet-152 using different data augmentation techniques. The ones that have given the best results can be seen in Table 11. The improve-

Table 10: Best results obtained for the texture or structure binary classifier using Inception, ResNet-50, ResNet-152, DenseNet-121 and DenseNet-161 and the set of hyperparameters used to obtained them, without data augmentation. The best accuracy is stressed in bold.

	Inception	ResNet-50	ResNet-152	DenseNet-121	DenseNet-161
Accuracy	95.495	99.640	99.730	98.920	99.009
Best batch size	64	32	64	32	64
Best number of epochs	300	700	1000	1000	1300

Table 11: Best results obtained for the texture or structure binary classifier using ResNet-152 and data augmentation techniques. The best accuracy is stressed in bold.

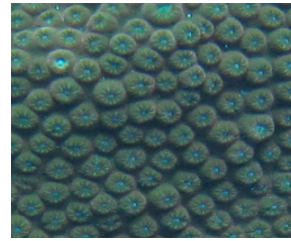
	shift = 0.4	zoom = 0.3	rotation = 4	flip
Accuracy	99.640	99.820	99.460	99.550

ment is small as we already had a very good result. Even so, we have obtained an accuracy of 99.820%. This means that the model is only wrong in two images that can be seen in Figure 6. The first one (a) is a structure image classified as a texture image, and the second one (b) is a texture classified as a structure.

The high classification accuracy obtained by this binary classifier opens up the possibility of using it as a first level model to decide whether the image is a texture or a structure without losing accuracy in the second level.



(a) MMEA



(b) MCAV

Figure 6: Coral images misclassified by the texture or structure binary classifier.

Table 12: Results obtained using the two-level classifier over the test set from RSMAS \cup StructureRSMAS, RSMAS and StructureRSMAS.

	RSMAS \cup StructureRSMAS	RSMAS	StructureRSMAS
Accuracy	93.874	98.356	85.263

385 6.5. Two-level classifier: identification of coral species based on texture or structure images

At this point, we already have the three models needed to build the two-level classifier for the classification of coral species based on either texture images or structure images. The first level of the classifier is to use the model that distinguishes between textures and structures, which we built in the previous section, to classify the input image in a texture image or a structure image. The second level of the two-level classifier is to classify the image into one of the coral species. To do this, the image is given to the RSMAS classifier or the StructureRSMAS classifier depending on the output of the first level.

395 We have evaluated the two-level classifier individually on structure images, on texture images and on structure images and texture images. The results can be seen in Table 12. It is important to note that, in each partition of the 5 fold cross validation, we are testing the two-level classifier with the same images that we used to test the StructureRSMAS model, the RSMAS model and the texture or structure binary classifier. That is, to test the two-level classifier we are not using images that were used to train any of its components.

Thanks to the high accuracy of the classifier in the first level, the accuracies obtained by the two-level classifier when we evaluate it only with texture images and only with structure images are very similar to the ones obtained by the RSMAS classifier and the StructureRSMAS classifier, respectively. This means that we classify textures and structures separately without decreasing the classification accuracy obtained by the texture classifier alone and the structure classifier alone. In fact, the accuracy for StructureRSMAS is now slightly better, which means that the RSMAS classifier in the second level is classifying correctly the image that are misclassified in the first level as a texture.

When we test the two-level classifier with texture and structure images, the accuracy is 93.874%, which is higher than the weighted arithmetic mean, taking into account the number of images in each dataset, between the obtained accuracy using the RSMAS model and the obtained accuracy using the StructureRSMAS model, which is 93.707%. Therefore, we have built a robust classifier that obtains a very good accuracy classifying any type of coral image.

7. Conclusions

The problem of classifying together structure and texture underwater coral images is a complicated task for three reasons: 1) the underwater images involve lightning problems and camera focusing problems due to the water, along with partial occlusion of marine animals; 2) different coral species look very similar and some species coexist together; and 3) there are not available coral structure datasets, and coral texture and structure images are very different from each other, no matter that they belong to the same class.

We have resolved this last problem by creating a coral structure dataset called StructureRSMAS, and we have tackled the classification of any coral image by using CNNs. Particularly, we have used one of the newest and most powerful CNNs, ResNet, which have been the one that have obtained better accuracies in the texture dataset, RSMAS, and the structure dataset, StructureRSMAS. We have also used image enhancement methods in the two datasets, data augmentation techniques and transfer learning from ImageNet to improve our results. We have resolved this classification building a two-level classifier composed of three models: the best model known model for RSMAS, the best model we have developed in this work for StructureRSMAS and a model to distinguish between a texture image and a structure image, also developed in this work.

We have observed that data augmentation does not bring much benefit when classifying structure images alone, and we argue that this happens because StructureRSMAS is small.

440 The two-level classifier we have developed in this work first identifies if the input image is a texture or a structure and then uses one of two specialized CNNs, depending on whether the image is a texture or a structure. It is able to correctly classify 93.874% of the images in RSMAS \cup StructureRSMAS.

Acknowledgments

445 This work was partially supported by the Spanish Ministry of Science and Technology under the project TIN2017-89517-P. Siham Tabik was supported by the Ramón y Cajal Program (RYC-2015-18136). Anabel Gómez-Ríos was supported by the FPU Program 998758-2016. The NVIDIA Titan Xp used for this research was donated by the NVIDIA Corporation.

References

- 450 [1] Iucn red list table of number of threatened species by major groups of organisms, Available from: http://cmsdocs.s3.amazonaws.com/summarystats/2017-3_Summary_Stats_Page_Documents/2017_3_RL_Stats_Table_1.pdf, accessed on 22-01-2019.
- 455 [2] J. H. Seinfeld, S. N. Pandis, Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, 2016.
- [3] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, D. Kriegman, Automated annotation of coral reef survey images, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 1170–1177. doi:10.1109/CVPR.2012.6247798.
- 460 [4] O. Pizarro, P. Rigby, M. Johnson-Roberson, S. B. Williams, J. Colquhoun, Towards image-based marine habitat classification, in: OCEANS 2008, IEEE, 2008, pp. 1–7. doi:10.1109/OCEANS.2008.5152075.
- [5] A. Shihavuddin, N. Gracias, R. Garcia, A. C. Gleason, B. Gintert, Image-based coral reef classification and thematic mapping, Remote Sensing 5 (4) 465 (2013) 1809–1841. doi:10.3390/rs5041809.

- [6] M. D. Stokes, G. B. Deane, Automated processing of coral reef benthic images, *Limnol. Oceanogr.: Methods* 7 (157) (2009) 157–168. doi:10.4319/lom.2009.7.157.
- 470 [7] M. Elawady, Sparse coral classification using deep convolutional neural networks, arXiv preprint arXiv:1511.09067.
- [8] A. Mahmood, M. Bennamoun, S. An, F. Sohel, F. Boussaid, R. Hovey, G. Kendrick, R. Fisher, Automatic annotation of coral reefs using deep learning, in: *OCEANS 2016 MTS/IEEE Monterey*, IEEE, 2016, pp. 1–5. doi:10.1109/OCEANS.2016.7761105.
- 475 [9] A. Mahmood, M. Bennamoun, S. An, F. Sohel, F. Boussaid, R. Hovey, G. Kendrick, R. Fisher, Coral classification with hybrid feature representations, in: *Image Processing (ICIP)*, 2016 IEEE International Conference on, IEEE, 2016, pp. 519–523. doi:10.1109/ICIP.2016.7532411.
- 480 [10] A. Gómez-Ríos, S. Tabik, J. Luengo, A. Shihavuddin, B. Krawczyk, F. Herrera, Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation, *Expert Systems with Applications* 118 (2019) 315 – 328. doi:10.1016/j.eswa.2018.10.010.
- 485 [11] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* 25, Curran Associates, Inc., 2012, pp. 1097–1105.
- 490 [12] E. Basaeed, H. Bhaskar, M. Al-Mualla, Supervised remote sensing image segmentation using boosted convolutional neural networks, *Knowledge-Based Systems* 99 (2016) 19 – 27. doi:https://doi.org/10.1016/j.knosys.2016.01.028.
- [13] G. Liu, Z. Yin, Y. Jia, Y. Xie, Passenger flow estimation based on convolutional neural network in public transportation system, *Knowledge-Based*

- 495 Systems 123 (2017) 102 – 115. doi:<https://doi.org/10.1016/j.knosys.2017.02.016>.
- [14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision (IJCV) 115 (3) (2015) 211–252. doi:[10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y).
500
- [15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9. doi:[10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
505
- [16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. doi:[10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308).
- 510 [17] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. doi:[10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [18] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 2261–2269. doi:[10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
515
- [19] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324. doi:[10.1109/5.726791](https://doi.org/10.1109/5.726791).
- 520 [20] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).

- [21] Coral reef dataset, v2. shihavuddin, asm, accessed on 22-01-2019. doi: 10.17632/86y667257h.2.
URL <https://data.mendeley.com/datasets/86y667257h/2>
- 525 [22] B. Xiao, K. Wang, X. Bi, W. Li, J. Han, 2d-lbp: an enhanced local binary feature for texture image classification, IEEE Transactions on Circuits and Systems for Video Technology doi:10.1109/TCSVT.2018.2869841.
- [23] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, <http://www.deeplearningbook.org>.
- 530 [24] M. S. A. C. Marcos, M. N. Soriano, C. A. Saloma, Classification of coral reef images from underwater video using neural networks, Optics express 13 (22) (2005) 8766–8771.
- [25] N. Ani Brown Mary, D. Dharma, Coral reef image classification employing improved ldp for feature extraction, Journal of Visual Communication and
535 Image Representation 49 (2017) 225 – 242. doi:10.1016/j.jvcir.2017.09.008.
- [26] N. Ani Brown Mary, D. Deje, Classification of coral reef submarine images and videos using a novel z with tilted z local binary pattern, Wireless Personal Communications 98 (3) (2018) 2427–2459. doi:10.1007/s11277-017-4981-x.
540
- [27] N. Ani Brown Mary, D. Deje, Coral reef image/video classification employing novel octa-angled pattern for triangular sub region and pulse coupled convolutional neural network (pccnn), Multimedia Tools and Applications doi:10.1007/s11042-018-6148-5.
- 545 [28] A. Mahmood, M. Bennamoun, S. An, F. Sohel, F. Boussaid, R. Hovey, G. Kendrick, R. B. Fisher, Chapter 21 - deep learning for coral classification, in: P. Samui, S. Sekhar, V. E. Balas (Eds.), Handbook of Neural Computation, Academic Press, 2017, pp. 383 – 401. doi:<https://doi.org/10.1016/B978-0-12-811318-9.00021-1>.

- 550 [29] Encyclopedia of life, accessed on 23-01-2019.
URL <http://eol.org/>
- [30] Iucn red list of threatened species, accessed on 23-01-2019.
URL <http://www.iucnredlist.org/>
- [31] Coralpedia of the univerisity of warwick, accessed on 23-01-2019.
555 URL <http://coralpedia.bio.warwick.ac.uk/en>
- [32] F. Chollet, et al., Keras, accessed on 23-01-2019 (2015).
URL <https://github.com/keras-team/keras>
- [33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Cor-
rado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp,
560 G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Lev-
enberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster,
J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke,
V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke,
Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heteroge-
565 neous systems (2015).
URL <https://www.tensorflow.org/>
- [34] F. Yu, Resnet and densenet cnns in keras, https://github.com/flyyufelix/cnn_finetune, accessed on 23-01-2019 (2017).
- [35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A
570 large-scale hierarchical image database, in: Computer Vision and Pattern
Recognition (CVPR), 2009. IEEE Conference on, IEEE, 2009, pp. 248–255.
doi:10.1109/CVPR.2009.5206848.
- [36] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz,
T. Greer, B. ter Haar Romeny, J. B. Zimmerman, K. Zuiderveld, Adaptive
575 histogram equalization and its variations, Computer vision, graphics, and
image processing 39 (3) (1987) 355–368. doi:[https://doi.org/10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X).

- [37] B. Xiao, H. Tang, Y. Jiang, W. Li, G. Wang, Brightness and contrast controllable image enhancement based on histogram specification, *Neurocomputing* 275 (2018) 2798–2809. doi:<https://doi.org/10.1016/j.neucom.2017.11.057>.
580
- [38] D. Zhang, D. Meng, J. Han, Co-saliency detection via a self-paced multiple-instance learning framework, *IEEE transactions on pattern analysis and machine intelligence* 39 (5) (2016) 865–878. doi:10.1109/TPAMI.2016.2567393.
585
- [39] J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for simplicity: The all convolutional net, *arXiv preprint arXiv:1412.6806*.
- [40] H. Tang, B. Xiao, W. Li, G. Wang, Pixel convolutional neural network for multi-focus image fusion, *Information Sciences* 433 (2018) 125–141. doi:<https://doi.org/10.1016/j.ins.2017.12.043>.
590
- [41] X. Tao, H. Gao, X. Shen, J. Wang, J. Jia, Scale-recurrent network for deep image deblurring, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8174–8182.

June 30, 2019

Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks

Anabel Gómez-Ríos, Siham Tabik, Julián Luengo, ASM Shihavuddin and Francisco Herrera.

Research highlights

- Creation of a new coral structure dataset.
- Study the performance of promising CNNs in the classification of coral images.
- Analyse data augmentation on the performance of the coral classification models.
- Evaluate image enhancement techniques on the performance of the coral classification models.
- Creation of a two-level model able to classify any coral image (texture or structure).

AUTHOR DECLARATION

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). She is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from eesserver@eesmail.elsevier.com

Signed by all authors as follows:

2nd July 2019

Anabel Gómez-Ríos



Siham Tabik



Julián Luengo



ASM Shihavuddin



Francisco Herrera

