

Coral Reef Annotation and Localization using Faster R-CNN

S M Jaisakthi¹, P Mirunalini², and Chandrabose Aravindan²

¹ Vellore Institute of Technology, Vellore, India

jaisakthi.murugaiyan@vit.ac.in

² SSN College of Engineering, Kalavakkam, Chennai, India

{miruna,aravindanc}@ssn.edu.in

Abstract. Coral reefs are the most diverse and valuable ecosystems in the world. It is also called as rainforests of the sea as they are so diverse. Coral reefs are important since it provide shelter and food to many marine species and also act as the source of nitrogen and other essential nutrients for marine food chains. Recent studies show that coral reefs ecosystems are extremely threatened due to pollution, sedimentation, unviable fishing practices, and climate change. So, coral reefs should be protected and monitored to save marine ecosystem. Hence, to monitor coral reef a task was introduced in ImageCLEF 2019, to automatically identify and label different types of benthic substrate with bounding boxes in a given image. This paper presents a Convolutional Neural Network (CNN) based method to locate and detect different types of benthic substrate. We have used faster RCNN architecture to detect the substrate since this method is much faster and accurate in detecting the objects.

Keywords: Coal Reef · Object Detection · Faster R-CNN · Convolutional Neural Network (CNN).

1 Introduction

Coral reefs [1], [3] are large underwater structures which are composed of the skeletons of colonial marine invertebrates called coral. These colonies are groups of individual animals called polyps. The reef structures are formed by the polyps secretion, upon which they live, which is made up of a substance called calcium carbonate. These coral make significant contributions to the well-being of people, animals, and plants in marine and coastal environments. They protect the coastal land from erosion that is caused by waves and storms. Coral reefs are not only important in terms of worldwide tourism, but it also serve as an important indicator to evaluate the health of our planet. In addition, they are an essential source of food and protein for millions of people throughout the world and also provide medical benefits to us. But today, we are the ones threatening reefs.

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

Around the world, roughly 50 percent of coral reefs have died in just the past few decades. The Great Barrier Reef was even declared dead last year. Coral reefs should be protected and many organisations are working hard to protect coral reefs. So an automatic system to locate and detect the coral reef in the sea will be helpful to conserve coral reefs. In the ImageCLEFcoral 2019 task, coral reefs are localised and annotated automatically using the CNN based object detection methods in an given image.

Predicting the location of the object along with the class label is called as object detection. This can be achieved with deep learning or computer vision techniques by localizing the objects along with image classification in each image. The traditional object detection methods involve detection based on block-wise orientation of histogram features. These methods use low level characteristics of the object features and hence, not able to discriminate objects of different labels well. But deep learning based methods construct a representation in hierarchical manner using low to high level features extracted from neural networks which improves the detection accuracy much better.

In deep learning object detection problem can be considered as a classification problem by classifying the image patches extracted from the images. In general the CNNs used for classification were too slow and computationally expensive because of running on so many patches generated by sliding window detector. This problem can be solved using R-CNN, which uses selective search that reduces number of bounding boxes to the classifier. Selective search uses local cues like texture, intensity, color and/or a measure of insiderness etc. to generate all the possible locations of the objects. The selected objects regions are wrapped to a fixed size pixels and are fed to a classifier which gives the individual probability of the region belonging to background and classes. So to locate and detect the coral reefs in an input image we have used faster R-CNN technique to achieve good accuracy.

2 Proposed Methodology

Our method for substrate detection is based on the Faster R-CNN [8] architecture. Faster R-CNN uses Region Proposal Network (RPN) using CNN to generate region proposals. This architecture consist of 3 layers, namely convolutional layer to extract feature maps, RPN to obtain region proposals with the help of anchor boxes and detection network for predicting object classes and bounding boxes. The proposed method consists of 3 stages namely preprocessing, training and substrate detection. In this work we have used five variants of pretrained COCO object detection model namely (1)Faster RCNN with NasNet (with augmentation), (2) Faster RCNN with NasNet (without augmentation) (3) Faster RCNN with inception V2 (with augmentation) (4) Faster RCNN with inception V2 (without augmentation) and (5) Faster RCNN with resnet101 (with augmentation).

2.1 Dataset

The dataset for this task is taken from coral reefs around the world as part of a coral reef monitoring project with the Marine Technology Research Unit at the University of Essex. The images contains the following 13 types of substrates: Hard Coral – Branching, Hard Coral – Submassive, Hard Coral – Boulder, Hard Coral – Encrusting, Hard Coral – Table, Hard Coral – Foliose, Hard Coral – Mushroom, Soft Coral, Soft Coral – Gorgonian, Sponge, Sponge – Barrel, Fire Coral – Millepora and Algae - Macro or Leaves. The data set contains 240 training images with 6670 annotated substrates along with ground truth annotations as bounding boxes.

2.2 Preprocessing

To reduce the computational complexity we have scaled down the input images. To build a strong object detector we have applied image augmentation [7] to improve the accuracy. We have created more training images using augmentation by applying horizontal and vertical flips, rotating by 90 degrees and randomly adjusting the contrast and brightness of the images.

2.3 Training

We have used five variants of Faster R-CNN architecture that comes with the Tensorflow Object Detection API [4]. The architectures were pre-trained using the COCO dataset [6] that contains 300k images from 80 categories of animals, furniture, vehicles, etc. for general object detection. In order to make the pre-trained models to learn the characteristics of benthic substrates we have fine-tuned it using the data set provided by the task ImageCLEFcoral 2019 [2] [5].

2.4 Coral Reef Image Annotation and Localisation

To localize the coral reef we have trained the models using the dataset provided by the organizers using the hyper-parameters recommended in the Tensorflow Object Detection API. The different models used in this task are discussed in the following sections.

Faster RCNN with NasNet In this model we have trained Faster R-CNN with NasNet as backbone. To study the performance of this model we have conducted two different experiments, one with image augmentation and the other without augmentation. This architecture used NasNet to extract the features in the first stage with l2 regularizer. Since NasNet utilizes very large memory space we experienced resource allocation problem during the training phase. So we have down scaled the input images size as 300×300 and trained the architecture using dataset for 120000 epochs.

Faster R-CNN with Inception V2 Faster R-CNN with inception V2 model extracts the features from the input images using inception resnet v2 during the first stage. To reduce the computational complexity the input images are reduced to the size of 600*1024. The model is trained for 100000 epochs with l2 regularizer and truncated normal initializer in the first stage. Anchors are generated for 4 scales with 3 different aspect ratio. For box predictor, the model is trained with l2 regularizer and variance scale initializer is used. The model performance is evaluated and analysed with both with and without image augmentation techniques.

Faster R-CNN with Resnet101 In this architecture we have used Faster RCNN Resnet101 to extract features in the first stage along with the image augmentation technique. The model is trained with the coral reef dataset for 150000 epochs.

2.5 Results and Discussion

The proposed methods were evaluated using intersection over union (IoU), the area of intersection between the foreground in the output segmentation and the foreground in the ground-truth segmentation, divided by the area of their union. The final results were calculated using average performance over all images of all concepts, and also per concept performance over all images. The following table shows the result of the proposed method presented in this paper.

Table 1. Results of the proposed Method

Method	MAP_50	R_50	MAP_0
Faster R-CNN with NasNet (with augmentation)	0.139962	0.068156	0.43097514
Faster R-CNN with NasNet (without augmentation)	0.134396	0.072253	0.42396952
Faster R-CNN with inception V2 (with augmentation)	0.084863	0.045624	0.42396952
Faster R-CNN with inception V2 (without augmentation)	0.048321	0.028678	0.28710386
Faster R-CNN with resnet101 (with augmentation)	0.040993	0.027374	0.27161182

MAP_50 is the localised Mean average precision (MAP) for each submitted method for using the performance measure of $IoU \geq 50$ of the ground truth, R_50 is the localised mean recall for each submitted method for using the performance measure of $IoU \geq 50$ of the ground truth and MAP_0 is the image annotation average for each method with success if the concept is simply detected in the image without any localisation. From the above table it is clear that all the methods with image augmentation techniques produced a good mean average precision when compared with the other methods trained without augmentation. But mean average recall is slightly higher for Faster R-CNN with NasNet without augmentation when compared to NasNet with augmentation. This variation may be due to which we have downscaled the input images too much. So further we need to conduct experiments by increasing the size of the

input images. It is also found that among three different architectures Faster R-CNN with NasNet produced a good result in terms of both precision and recall.

In terms of per substrate accuracy, Faster R-CNN with NasNet produced a good accuracy when compared to the methods presented by the other participants. Table 2 shows the results of per substrate accuracy presented by other teams.

Table 2. Accuracy per substrate obtained by all the participating teams

Run	Group	hard coral branching	hard coral sub-massive	hard coral boulder	hard coral encrusting	hard coral table	hard coral foliose	hard coral mushroom	soft coral	soft coral gorgonian	sponge	sponge barrel	fire coral millepora	algae macro or leaves
27115	VIT	0.0436	0	0.0809	0.0168	0	0.0128	0.0664	0.0722	0	0.0349	0.0526	0	0
27347	VIT	0.0456	0	0.0374	0.0055	0	0	0.0204	0.0918	0	0.0239	0.0498	0	0
27348	VIT	0.0548	0	0.0956	0.0171	0	0.0129	0.119	0.0782	0	0.0365	0.0579	0	0
27349	VIT	0.0637	0	0.1012	0.0195	0	0.0028	0.0758	0.0804	0	0.0329	0.0619	0	0.0004
27350	VIT	0.0597	0	0.0305	0.0141	0	0	0.0422	0.0808	0	0.0299	0.0598	0	0
27398	HHUD	0.0013	0	0.0116	0	0	0	0	0.0702	0	0.002	0	0	0
27413	HHUD	0.0068	0.0021	0.0063	0.0014	0	0.0022	0	0.0523	0	0.0063	0	0	0
27414	HHUD	0.0089	0	0.016	0.0015	0	0	0	0.0562	0	0.0104	0.0054	0	0
27415	HHUD	0	0	0	0	0	0	0	0.0731	0	0	0	0	0
27416	HHUD	0.0346	0	0.0343	0.0064	0	0	0.0437	0.055	0.0008	0.0094	0.0222	0	0.0158
27417	HHUD	0.0356	0	0.033	0.0069	0	0	0.0406	0.0505	0.0008	0.0094	0.0213	0	0.0157
27418	HHUD	0.0246	0	0.0321	0.0038	0	0	0.0269	0.0447	0.0005	0.0086	0.0318	0	0.0012
27419	HHUD	0.0261	0	0.0315	0.0042	0	0.0023	0.0228	0.0423	0.0005	0.0088	0.0304	0	0.0012
27421	HHUD	0.0007	0	0.0167	0.0048	0	0.0006	0.0106	0.0571	0	0.0072	0	0	0
27497	ISEC	0.0198	0	0.0007	0.0121	0	0	0	0.0079	0	0.0277	0	0	0

From Table 2 it is evident that our method produced better accuracy in identifying many substrate types.

3 Acknowledgements

The authors would like to thank the management of VIT university, Vellore, India and SSN College of Engineering, Chennai India for funding the respective research labs where the research work is being carried out. One of the authors S M Jaisakthi would like to thank NVIDIA for providing a GPU grant in support of this research work and similarly P Mirunalini and Chandrabose Aravindan would like to thank the management for providing the GPU machine where this research is carried out.

References

1. Introduction to Coral Reefs. <http://www.deepbluediscoveries.com/introduction-to-coral-reefs>
2. Chamberlain, J., Campello, A., Wright, J.P., Clift, L.G., Clark, A., García Seco de Herrera, A.: Overview of ImageCLEFcoral 2019 task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org (2019)
3. Gray, C.: Coral Reefs: An introduction. <https://www.edgeofexistence.org/blog/coral-reefs-an-introduction/> (2012)
4. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., Murphy, K.: Speed/accuracy trade-offs for modern convolutional object detectors. CoRR **abs/1611.10012** (2016), <http://arxiv.org/abs/1611.10012>
5. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Ben Abacha, A., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasilopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 9-12 2019)
6. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
7. Mikołajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW). pp. 117–122 (May 2018). <https://doi.org/10.1109/IIPHDW.2018.8388338>
8. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. CoRR **abs/1506.01497** (2015), <http://arxiv.org/abs/1506.01497>