

# Automatic Coral Detection using Neural Networks

Ivan Gruber<sup>1,2</sup>[0000-0003-2333-433X]\* and Jakub Straka<sup>2</sup>[0000-0002-9981-1326]\*

University of West Bohemia, Faculty of Applied Sciences,  
New Technologies for the Information Society<sup>1</sup> and Department of Cybernetics<sup>2</sup>,  
Univerzitní 8, 301 00 Plzeň, Czech Republic  
grubiv@ntis.zcu.cz, strakajk@students.zcu.cz

**Abstract.** This paper presents methods that were utilized in the ImageCLEFcoral 2020 challenge. The challenge contains two following subtasks: automatic coral reef annotation and localization, and automatic coral reef image pixel-wise parsing. In the first subtask, we tested two methods - SSD, and Mask R-CNN. In the second subtask, we tested only Mask R-CNN. Performance improvements were achieved by careful cleaning of the dataset and by both offline and online data augmentations.

**Keywords:** Object detection · Semantic segmentation · Coral localization · Convolutional neural networks · Machine learning.

## 1 Introduction

With changes in world climate in recent years, the danger of losing coral reefs and the ecosystem they support is increasing. Therefore, detailed monitoring of these ecosystems can be critical for their future. However, because of the complexity of coral images, they are very difficult for people to annotate, which opens possibilities for automatic detection.

Within the ImageCLEFcoral 2020 challenge [5, 1], authors provide 440 images with ground-truth annotations. The challenge contains two subtasks. The first task is the classic detection task, whereas the success is considered each detection with Intersection over Union (IoU) equal or bigger than 0.5. The second task is semantic parsing of the corals in the input image.

For the first subtask, we tested two detection methods, one single-shot detector - SSD [6] and one two-shot detector - Mask R-CNN [4]. In the second subtask, we utilized Mask R-CNN, because it also provides semantic parsing information.

---

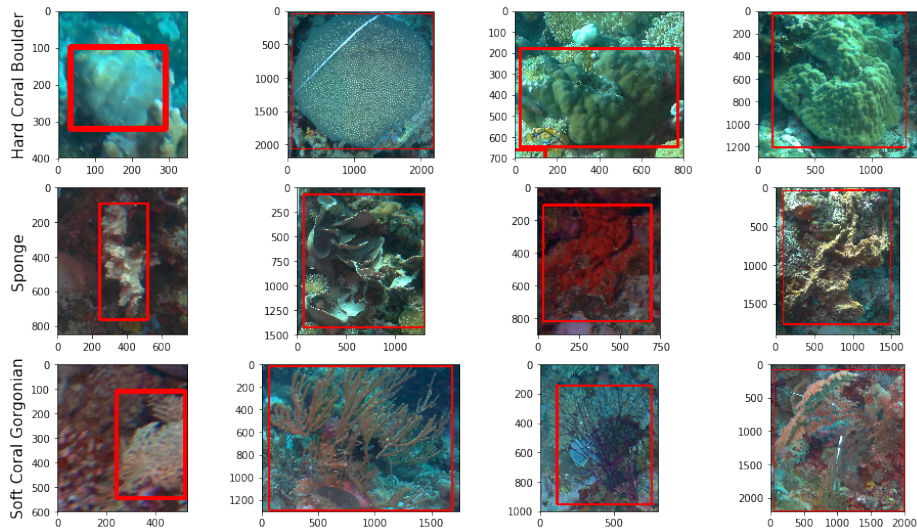
Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.

\* Both authors contributed equally

## 2 Data

The data for this task originates from a growing, large-scale collection of images 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 dataset contains 440 images in total with ground-truth annotations for 13 coral classes.

The provided dataset is challenging from many different perspectives. First, each image contains a large number of different corals, to be more specific 28 on average. Second, the large imbalance in the total number of instances among coral classes occurs. Third, there exists big intra-class variability in appearance and size, see Fig. 1.



**Fig. 1.** Examples of different class instances.

Third, the quality of images is highly inconsistent and some images are very blurry. Last but not least, during a hand-made inspection, we revealed that approximately 120 images are rotated by 180 degrees with respect to the ground-truth annotations.

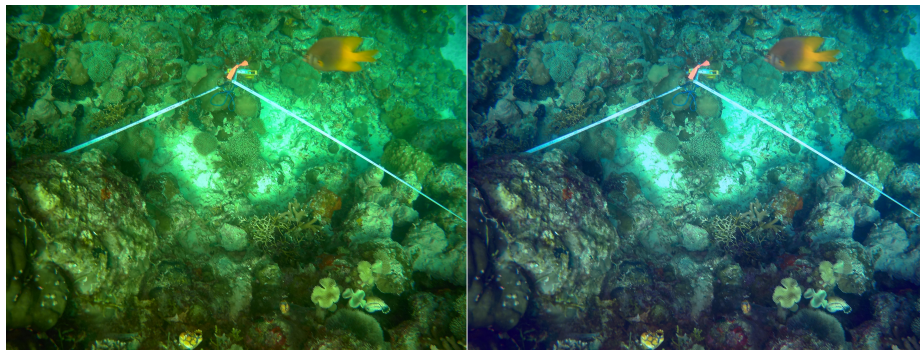
### 2.1 Data preprocessing

In the first step, we address the rotation problem mentioned above and rotate all the images to the correct orientation. In the next step, the database was split into two subsets - train, and validation set. Due to the big imbalance between classes, during the splitting, we focused to preserve class distribution across both

of the sets. Our goal was split to split the dataset in the way, that the train set contains approximately 85% of the images and validation set the rest. The final split with respect to the classes, after which the train set contains 371 images, and the validation set 69 images, can be found in Table 1.

**Table 1.** The training and the validation set split. Total number of instances with percentage information in brackets.

Class	Total number of instances	Instances in the train set (%)	Instances in the validation set (%)
Hard coral branching	1181	893 (76)	288 (24)
Hard coral submassive	198	172 (87)	26 (13)
Hard coral boulder	1642	1364 (83)	278 (17)
Hard coral encrusting	946	816 (86)	130 (14)
Hard coral table	21	18 (86)	3 (14)
Hard coral foliose	177	153 (86)	24 (14)
Hard coral mushroom	233	179 (80)	44 (20)
Soft coral	5663	4459 (79)	1204 (21)
Soft coral gorgonian	90	79 (88)	11 (12)
Sponge	1691	1514 (90)	177 (10)
Sponge barrel	139	107 (77)	32 (23)
Fire coral millepora	19	16 (84)	3 (16)
Algae macro or leaves	92	78 (85)	14 (15)



**Fig. 2.** Example of augmentation using a color filter. Original image (left) and image after augmentation using a color filter (right).

## 2.2 Data augmentation

To enrich and expand our train set, we utilize data augmentations. Primary, standard augmentations common for computer vision tasks were used, to be

more specific, random horizontal flip, random vertical flip, random crop with resize, and Gaussian blur. Moreover, during the data analysis, we noticed, that two different light 'themes' are common - the blue one, and the green one. We simulate this effect by using a color filter, see Fig 2.

We tested both offline (before the training) and also online (during the training) augmentations, however, we reached better results using online augmentations during the experiments. A comparison of the influence of distinct online data augmentation on mean average precision (mAP) on our validation set can be found in Table 2.

**Table 2.** The comparison of augmentations for both models on our validation set (localization task).

Method	Mask R-CNN mAP	SSD mAP
W/O augmentations	6.94	4.44
Flip, Color filter	8.12	-
Flip, Color filter, Blur	9.84	-
Flip, Color filter, Blur, Random crop	10.18	14.94

### 3 Methods and experimental setup

We tested two detection models - SSD [6], and Mask R-CNN [4]. Both models were pretrained on Pascal VOC 2007 dataset [3]. For both models, we used standard implementation in Keras [2].

All the images were resized to the resolution  $1024 \times 1024$  pixels during the Mask R-CNN training and to  $512 \times 512$  pixels during the SSD training. Both tested methods were trained with a batch size of 1 during 200 epochs using SGD optimizer with an initial learning rate  $l = 0.0001$  and step decay  $d = 0.1$  after 100 epochs. For the purpose of challenge, we choose the model with best mAP on the validation set.

### 4 Results

We evaluate both trained models on the validation set. Exemplary results on the validation set can be found in Fig. 3. Mask R-CNN detects much more bounding boxes than SSD, however, the majority of them are false positives. To be more specific, on the validation set, Mask R-CNN detects 2148 bounding boxes, but only 44.7% are true positives. On the other hand, SSD detects only 1029, but 71.3% are true positives.

Detailed comparison of average precision in the localization subtask for individual classes can be found in Table 3. It should be noted that neither of the models were able to learn to detect five of the low frequent class. On the



**Table 3.** Comparison of average precision (AP) for individual classes on the validation set in the localization subtask.

Class	Mask R-CNN	SSD
Hard coral branching	19.56	23.13
Hard coral submassive	0.00	0.00
Hard coral boulder	25.76	33.29
Hard coral encrusting	2.35	5.48
Hard coral tablet	0.00	0.00
Hard coral foliose	8.33	14.24
Hard coral mushroom	5.91	15.91
Soft coral	44.97	34.41
Soft coral gorgonian	0.00	0.00
Sponge	3.52	5.55
Sponge barrel	21.98	62.17
Fire coral millepora	0.00	0.00
Algae macro or leaves	0.00	0.00
<b>mAP</b>	10.18	14.94

**Table 4.** Results comparison for both models on our validation set and the challenge test set.

Method	Task	Validation mAP	Test mAP
Mask R-CNN	localization	10.18	24.3
Mask R-CNN	segmentation	10.29	30.4
SSD	localization	14.94	49.0

## 5 Conclusion

In this paper, we present working notes from ImageCLEFcoral 2020 challenge. We employed two detection methods, both of them based on neural networks, SSD, and Mask R-CNN. In the localization subtask, SSD reached better mAP by a large margin. Despite the fact reached results were mediocre, we believe that with more advanced state-of-the-art detection methods, it is possible to reach satisfactory results. Other future improvements we see in an expansion of the training set, or utilizing knowledge distillation approach.

**Acknowledgement:** The work has been supported by the grant of the University of West Bohemia, project No. SGS-2019-027. Moreover, access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

## References

1. Chamberlain, J., Campello, A., Wright, J.P., Clift, L.G., Clark, A., García Seco de Herrera, A.: Overview of the ImageCLEFcoral 2020 task: Automated coral reef

- image annotation. In: CLEF2020 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org (2020)
2. Chollet, F., et al.: Keras. <https://keras.io> (2015)
  3. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>
  4. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
  5. Ionescu, B., Müller, H., Péteri, R., Abacha, A.B., Datla, V., Hasan, S.A., Demner-Fushman, D., Kozlovski, S., Liauchuk, V., Cid, Y.D., Kovalev, V., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Ninh, V.T., Le, T.K., Zhou, L., Piras, L., Riegler, M., Halvorsen, P., Tran, M.T., Lux, M., Gurrin, C., Dang-Nguyen, D.T., Chamberlain, J., Clark, A., Campello, A., Fichou, D., Berari, R., Brie, P., Dogariu, M., Ștefan, L.D., Constantin, M.G.: Overview of the ImageCLEF 2020: Multimedia retrieval in lifelogging, medical, nature, and internet applications. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020), vol. 12260. LNCS Lecture Notes in Computer Science, Springer, Thessaloniki, Greece (September 22-25 2020)
  6. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European conference on computer vision. pp. 21–37. Springer (2016)