Automatic Coral Reef Annotation, Localization and Pixel-wise Parsing Using Mask R-CNN

Lukáš Soukup[1]

[1]*University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics*
Univerzitní 8, 301 00, Plzeň, Czech Republic

Abstract

This paper describes the methods that were used for annotation, localization and pixel-wise parsing of the coral reefs from underwater images. The proposed system achieved competitive results in the third edition of ImageCLEFcoral 2021 challenge. Specifically, in case of annotation and localization task achieved mean average precision with Intersection over Union (IoU) greater that 0.5 (mAP@0.5) 0.121 and in case of pixel-wise parsing task achieved mAP@0.5 0.075 on the test set. The proposed method is based on Mask R-CNN object detection and segmentation framework with online data augmentations.

Keywords

Object detection, Semantic segmentation, Neural networks, Deep learning, Machine learning, Coral reefs detection, Coral reefs segmentation

1. Introduction

The ImageCLEFcoral 2021 challenge [1] is motivated by the impact of recent climate changes on the coral reefs and the ecosystem they support. Coral reefs are in danger of being lost within next 30 years which would lead to not only extinction of many marine species but also to a humanitarian crisis on a global scale for people who rely on the reef services. By monitoring the changes of reef we could help with prioritizing conservation efforts.

The goal of the challenge is to create an automatic system for monitoring the coral reefs using the provided dataset. The challenge consist of the tasks: (1) annotation and localization, and (2) pixel-wise parsing.

The proposed solution is using state-of-the-art object detection model Mask R-CNN [2] which provides both detection and semantic segmentation information.

Provided dataset [1] consists of 1052 train images and 485 test 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. Each coral object in the training set was annotated by an expert including a bounding box, segmentation polygon and a class representing one of 13 substrate types. In total 21,749 objects were annotated in the dataset.

The dataset is very challenging from different perspectives. Each image contain many different coral objects, on average there are over 20 corals in a single image. The dataset

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania
lsoukup@kky.zcu.cz (L. Soukup)
0000-0001-7140-6604 (L. Soukup)

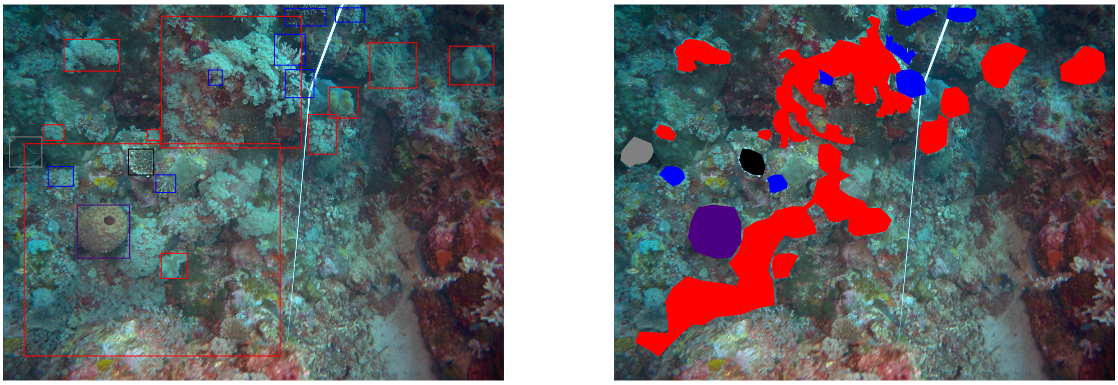

Figure 1: Example image with drawn (a) detection annotation, (b) pixel-wise parsing annotation

is highly unbalanced having 33.5% of all objects from class *c_soft_coral* and only 0.12% of all objects from class *c_fire_coral_millepora*. Additionally, the quality of the images is very inconsistent, some images are heavily blurred and there are noticeable color variations. Fig. 1 show example from the dataset with drawn annotations for both tasks.

1.1. Dataset split

For the purpose of optimizing network parameters the provided training dataset [1] was divided into train set and validation set. To correctly evaluate the performance on the validation set it is crucial that the validation set has the same data distribution as the training set. To preserve the distribution we decided to make the split with respect to location where the image was taken. From each location 80% of images were added to the train set and rest to the validation set.

1.2. Data Preprocessing

The implementation of the CNN [2] used for the experiments expects the target data to be in the specific format. For the subtask annotation and localization the expected target data are the bounding boxes specified by 4 numbers - coordinates of upper left corner, width and height of the bounding box. The provided bounding box annotations for this subtask were given by coordinates of the upper left corner and bottom right corner. Thus, the preprocessing of target data was simple.

The preprocessing of target data for the second subtask (pixel-wise parsing) was more interesting. The CNN expects the target data for the segmentation to be binary segmentation masks. The annotation provided in the challenge were marking every segmentation object as a set of points making a polygons around the coral (as shown in Fig. 1).

To create a submission to the challenge, the segmentation masks had to be transferred to the set of points again. Several methods of creating polygons were tested. The only one not creating self-intersecting polygons turned out to be searching for the contours in every binary mask and then creating a convex hull of the contour.

2. Method

The proposed object detection and pixel-wise parsing method is state-of-the-art convolutional neural network Mask R-CNN [2] pretrained on ImageNet dataset [3]. Specifically, PyTorch [4] implementation of this network was used in the experiments. The model provides predictions useful for both subtasks - bounding boxes for annotation and localization, and binary segmentation masks for pixel-wise parsing.

2.1. Experimental Setup

Even though resolution of input images is crucial for the task of object detection since some of the objects are relatively small, all the training images were resized to 1000×1000 due GPU memory limitation. The model was trained with batch size 2 and accumulated gradient 4 using SGD optimizer with an initial learning rate 0.005 step decay 0.0005 after 3 epochs. The best model was chosen by an early stopping method over last 5 epoch evaluated on the validation set.

2.2. Augmentations

To enrich the training set data augmentation were applied to the training images (online - during the training process). When loading the image, first, a random horizontal flip was applied with probability 0.5. Second, random brightness and contrast variations were used to simulate color inconsistency in the training data. Specifically, brightness variation with delta of 0.15 with probability 0.6 and contrast and saturation variations scaled by random value in range from 0.85 to 1.15 with probability 0.8.

2.3. Evaluation

The trained models were evaluated on the validation set and the best models used for the prediction on the test set. The evaluation criteria for both subtasks is mean average precision with Intersection over Union (IoU) greater that 0.5 (mAP@0.5). The model chosen for the the prediction on test set achieved mAP@0.5 of 0.18 in case of object detection and mAP@0.5 of 0.35 in case of instance segmentation on the validation set.

2.4. Submissions

The submissions to the challenge were created as the prediction of the proposed method on the test set provided in the challenge [1]. For evaluation of participants submission, the AICrowd platform was used. Each team was allowed to submit up to 10 runs. I have used 2 runs for annotation and localization task and only 1 run for pixel-wise parsing task.

Tables 1 and 2 show the description and result of each submission to both subtasks of the challenge. Fig. 2 shows example of object detection on one of the images from test set.

Table 1
Annotation and localization results on the test set.

Setup	mAP@0.5	recall
Mask R-CNN	0.105	0.055
Mask R-CNN + augmentations	**0.121**	**0.059**

Table 2
Pixel-wise parsing results on the test set.

Setup	mAP@0.5	recall
Mask R-CNN + augmentations	**0.075**	**0.048**

3. Competition results

The official competition results are shown in Table 3 for pixel-wise parsing task and in Table 4 for annotation and localization task. The proposed method achieved the best score in both tasks of the ImageCLEFcoral 2021 competition. Specifically, achieved mAP@0.5 of **0.075** in case of pixel-wise parsing (run id 139084) and mAP@0.5 of **0.121** in case of annotation and localization (run id 138115) of coral reefs.

Table 3
Comparison with other participants in pixel-wise parsing task.

Group	mAP@0.5
MTRU	0.011
MTRU	0.017
MTRU	0.018
MTRU	0.021
University of West Bohemia	**0.075**

Table 4
Comparison with other participants in annotation and localization task.

Group	mAP@0.5
UAlbany	0.001
University of West Bohemia	0.105
University of West Bohemia	**0.121**

4. Conclusion

This paper presents automatic system for annotation, localization and segmentation of coral reefs which was used in ImageCLEFcoral 2021 challenge. The detection method based on

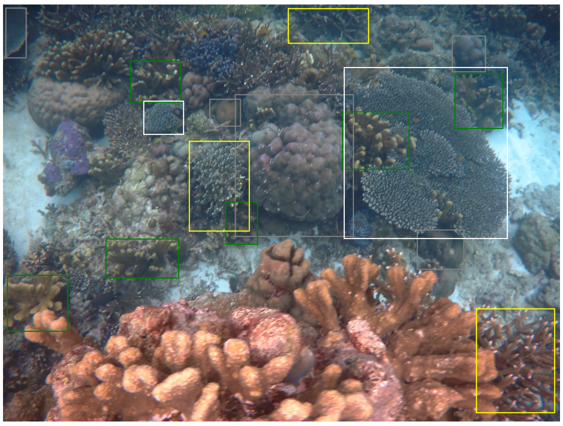

Figure 2: Example result of object detection on the test set.

Mask R-CNN achieved mAP@0.5 of 0.121 in case of annotation and localization task and 0.075 in case of pixel-wise parsing task. Despite the unsatisfying results, I believe that more advanced methods and utilization of knowledge distillation could significantly improve the results in the future.

Acknowledgments

This work was supported by the grant of University of West Bohemia, project No. SGS-2019-027. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

References

[1] J. Chamberlain, A. García Seco de Herrera, A. Campello, A. Clark, T. A. Oliver, H. Moustahfid, Overview of the ImageCLEFcoral 20201task: Coral reef image annotation of a 3d environment, in: CLEF2021 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Bucharest, Romania, 2021.

[2] K. He, G. Gkioxari, P. Dollár, R. B. Girshick, Mask R-CNN, CoRR abs/1703.06870 (2017). URL: http://arxiv.org/abs/1703.06870. arXiv:1703.06870.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.

[4] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Sys-

tems 32, Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.