Monitoring Coral Reefs Using Faster R-CNN

Felix Kerlin[1], Kirill Bogomasov[1] and Stefan Conrad[1]

[1]Heinrich-Heine-Universität Düsseldorf (Germany), Universitätsstraße 1, 40225 Düsseldorf, Germany

Abstract
Monitoring coral reefs is an important procedure to protect the persistence of many marine species. The imageCLEFcoral 2022 Challenge aims to identify and annotate corals on underwarter images. These images vary in terms of quality and are therefore of a high complexity. While our investigation, we focused on the data set and searched for ways to improve the image quality. To be specific, we minimized the impact of color casts, and erratic annotations by a color balancing strategy, as well as combining the prediction results of the trained deep learning architectures on preprocessed and original images. Object detection was handled by deep learning entirely. In particular, faster R-CNN with a ResNet+FPN backbone network was the architecture of the choice. The merging strategy is based by a Non-maximum Suppression (NMS) and reduces therefore overlapping predictions. Additionally, we analyzed the impact of the depth of the chosen backbone network. We have identified a connection between increasing network depth and increasing accuracy for underwater imaging. Overall, our best approach achieved a $\text{MAP}_{0.5}$ value of 0.396.

Keywords
Computer Vision, Object Detection, Neural Networks, Coral Reefs Detection, Deep Learning

1. Introduction

The CLEF Initiative's [1] imageCLEFcoral 2022 Challenge [2] addresses the issue of the destruction of coral reefs due to climate change and human activities. The reefs and the entire surrounding ecosystems are threatened with extinction within the next 30 years [3]. This would lead not only to the end of many marine species, but also to a humanitarian crisis of global proportions, as many regions depend on coral reefs. A quick change in the near future is essential. By this reason, an invention is indispensable. An appropriate intervention, in terms of environmental protection, can only take place if it is known which steps need to be taken. These are depending on the current state of the coral landscape, i.e. coral distribution, stocks and many more. For this reason, the entire area of coral reefs needs to be analyzed and regularly monitored subsequently. Manual monitoring by experts such as marine biologists is expensive and not feasible at all, keeping in mind the total area of 255 000 km^2 covered by corals [4]. Therefore automation is necessary. Our aim is to investigate how well we can locate and classify corals. For this purpose, we use efficient technologies from the field of deep learning. In the following chapters we will describe procedures used for the submissions to the challenge in detail.

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy
kerlin@hhu.de (F. Kerlin); bogomasov@hhu.de (K. Bogomasov); stefan.conrad@hhu.de (S. Conrad)

2. Related Work

The annually occurring Coral reef image annotation and localization task is taking place for the fourth time in a row. The results in the recent past have always shown potential for future developments. Last years' winning team achieved a $MAP_{0.5}$ value of just 0.121 in 2021 [5]. Even though the data sets have been revised this year and therefore the results cannot be directly compared, these are the results that can best serve as a benchmark for our work. Over the years, we have seen different approaches in submissions to the challenge, both classic feature engineering [6], as was commonly used in computer vision, and newer deep learning methods [7, 8, 9]. Our preliminary work, compared and combined both [10, 11]. A key lesson learned from previous investigation is the need of balancing the highly unbalanced data, which is not trivial and necessity of improving the quality of underwater images according to typical issues such as cloudiness and shifted color distribution. This year we want to benefit from these findings once again. Additionally, as proposed in [10], we will build our investigations mainly around regions with CNN features. Although, [9] already experimented with Faster R-CNN and achieved a $MAP_{0.5}$ value of 0.13996, we see potentials for improvement.

3. Data set

The data set was provided by the CLEF initiative. The training data set consisted of 1,374 images from a total of four different locations with a total of 31,517 annotations and 13 different classes. Additionally, the evaluation data for the submissions consisted of 200 images from one location and was not available for own investigations until the final submissions.

Regarding the quality of the data, it was varying and therefore challenging. Many images had a severe color shift, some images were blurry. Another complexity in a multi-label classification task is the number of objects. The number of corals in an image varied from 1 to 116. In addition, in groups of multiple corals of the same coral species, the corals are sometimes annotated in one bounding box and sometimes in multiple boxes as shown in Figure 2.

As can be seen in Table 1, the individual classes are distributed very unbalanced, i.e. the substrate type "c_soft_coral" alone comprising 24.65% of the annotations. The three most frequent classes account for 67.37% of the annotations, while the three least frequent classes account for just 1.28% of the annotations.

To give an example: Figure 1 illustrates four images of the data set with visualized annotations. Noticeable are, the color and quality differences of the images that can be easily observed. In particular, while images b) and c) have a strong blue cast, image a) is blurred. Especially in figure d), the problem of delineating identical corals within an assemblage becomes clear. To be specific: the corals of the type "c_hard_coral_branching" are divided into a total of three annotations. Contrarily, in the same image the three large corals of type "c_hard_coral_table" in the lower right corner are combined into one annotation, although they are clearly separable.

Another example is shown in Figure 2. In image a), many corals of type "c_soft_coral" are annotated by a single bounding box per coral. In image b) a group of corals of type "c_soft_coral" is annotated by one large bounding box. Remembering the main evaluation metric $MAP_{0.5}$, the impact of varying strategies while annotating becomes clear. For example, splitting a group

Table 1
Distribution of the individual classes in the training data set

Class	absolute occurrence	relative occurrence
c_soft_coral	7769	24.65%
c_hard_coral_boulder	7373	23.39%
c_sponge	6091	19.33%
c_hard_coral_branching	3132	9.94%
c_hard_coral_submassive	2637	8.37%
c_algae_macro_or_leaves	1870	5.93%
c_hard_coral_table	920	2.92%
c_sponge_barrel	606	1.92%
c_hard_coral_encrusting	380	1.21%
c_hard_coral_mushroom	335	1.06%
c_hard_coral_foliose	233	0.74%
c_soft_coral_gorgonian	171	0.54%
c_fire_coral_millepora	0	0.00%

of the same coral species among more or fewer annotations in the submission would have a negative impact on the score.

4. Approach

The core strategy used for object detection uses the state-of-the-art convolutional neural network Faster R-CNN [12]. It is built with different ResNet backbone networks and the framework detectron2 [13] for PyTorch [14]. We will first observe the effect of network depth on coral detection and secondly try to compensate for the previously mentioned weaknesses of the dataset through image enhancement.

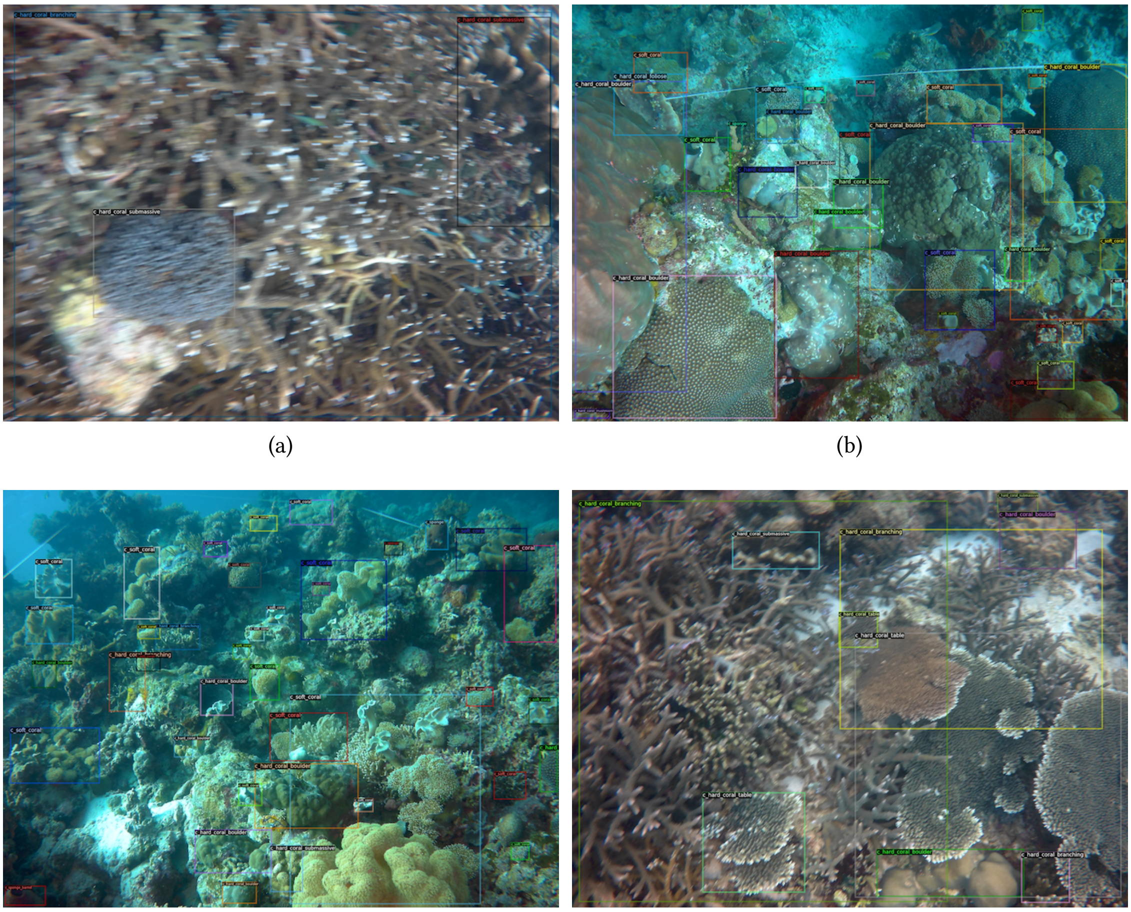
4.1. Network architecture

For the network, we chose Faster R-CNN as described in the related work section. As a backbone network, we have chosen ResNet+FPN [15]. This approach achieved the best results on the COCO-dataset [16] in the FPN paper [15] and in detectron2's Model Zoo baseline [13]. In addition to the commonly used ResNet-50 and ResNet-101 and according to He et al. [17] who showed that residual networks gain precision by increasing depth, we included ResNet-152.

4.2. Training hyperparameter

Because of the small data set, we divided it into 90% training data and 10% validation data. To monitor overfitting, we calculated the evaluation metric of the challenge $\text{MAP}_{0.5}$ in small intervals of 250 iterations (≈ 104 epochs).

Figure 7 shows the total loss of the training process and the $\text{MAP}_{0.5}$ of the checkpoints as an example. Although the training loss decreased over the complete 100.000 iterations (≈ 41727



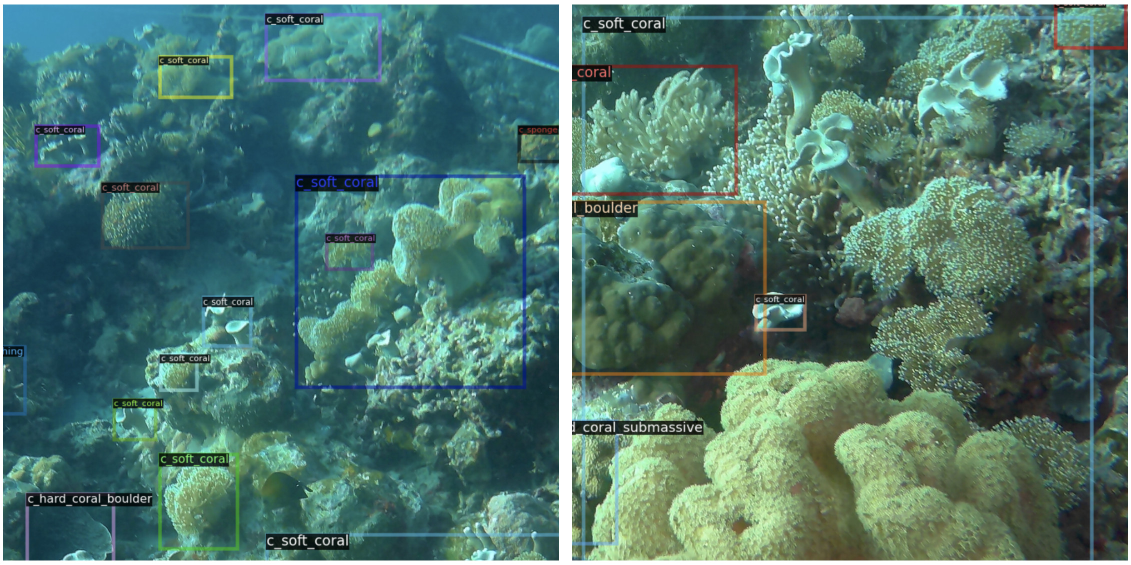
(a) (b) (c) (d)

Figure 1: Different images from the data set with visualized annotations

epochs), the network started to overfit at 70.000 iterations (≈ 29209 epochs) and the $\mathrm{MAP}_{0.5}$ decreased from there on. Therefore, in the end, we chose the checkpoint with the best $\mathrm{MAP}_{0.5}$ on the evaluation data for the final submission.

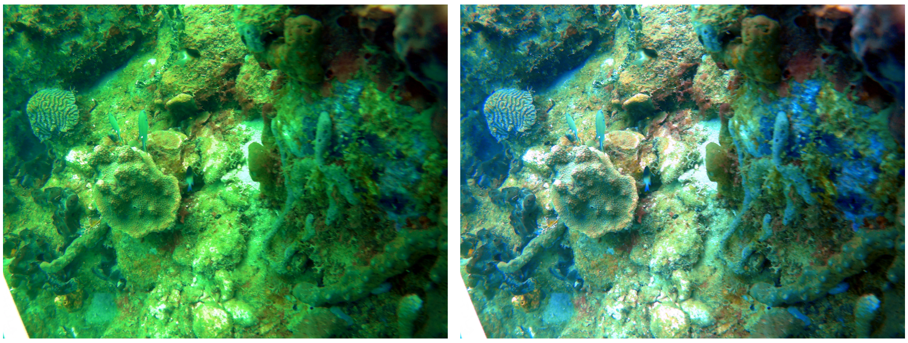
For the learning process, we chose a base learning rate of 0.0005 for the first 25,000 iterations (≈ 10431 epochs). After that, we lowered the learning rate to 0.0001 for the next 25,000 iterations (≈ 10431 epochs) and to 0.00005 after 50,000 iterations (≈ 20863 epochs).

Figure 8 shows a comparison of the different batch sizes. The ResNet-50 and ResNet-101 networks were trained with batch sizes 32, 64, 128, 256 and 512 each. In both cases, the networks with larger batch sizes performed better than those with smaller batch sizes. Therefore, we chose a batch size of 512 for the final submissions.

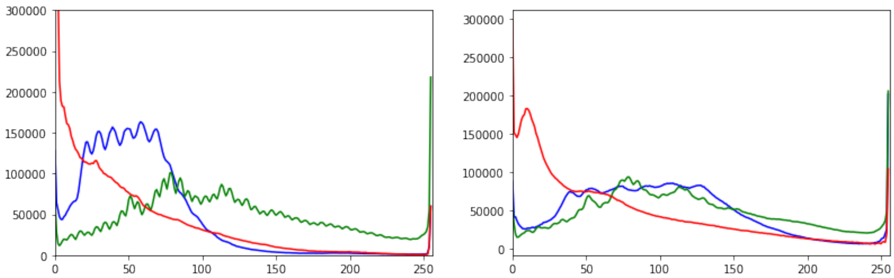


(a) (b)

Figure 2: Different annotation strategies for assemblages of the same coral species



(a) original image (b) enhanced image



(c) Histogram of original image (d) Histogram of enhanced image

Figure 3: Training image before and after image enhancement

4.3. Color balancing

To counteract the problem of color casts in the images, a function was written that removes blue and green casts. For this purpose, the average values for red, green and blue were calculated for each image. Then the image was shifted slightly into the red range until the average red value reached a certain threshold. Since the termination criterion depends only on the red component of the image, both images with a blue cast and images with a green cast can be improved in this way. [18]

A comparison of a random image before and after color balancing, including the corresponding histograms, is shown in Figure 3. Contrary to the histogram of the original image, the green channel value of the postprocessed image is much less dominant. Its value was shifted by the image enhancement. Because of that, the histogram of the processed image shows a much more balanced distribution for all three channels.

4.4. Dual network approach

Both image variations are used for training subsequently. For a better comparison, all tested networks share the same settings and hyper-parameters. The setup for the training is explained in Figure 4. The finally submitted predictions were then calculated for both types of data and combined to to a common result. This process is illustrated in Figure 5.

All potential duplicates were removed using Non-maximum Suppression. NMS iteratively removes boxes with a lower confidence for all overlapping boxes that have an IoU greater than 0.8 and keeps the box with the highest confidence. If two boxes of different classes with an IoU greater than 0.8 overlap, the box with smaller confidence is also discarded.

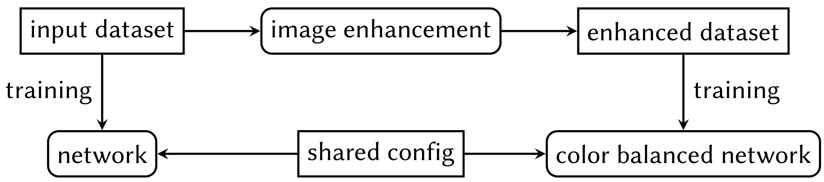


Figure 4: Training process of the combined Network

5. Submissions

Nine runs were submitted to the "Coral reef image annotation and localisation" task. These consist of a combination of the backbone construction i.e. depth variation and the image type that was used for training. Each run configuration is explained in Table 2. According to the challenge organization its main evaluation metric is mean average precision. Furthermore we added the submissions' mean average recall for a better comparison.

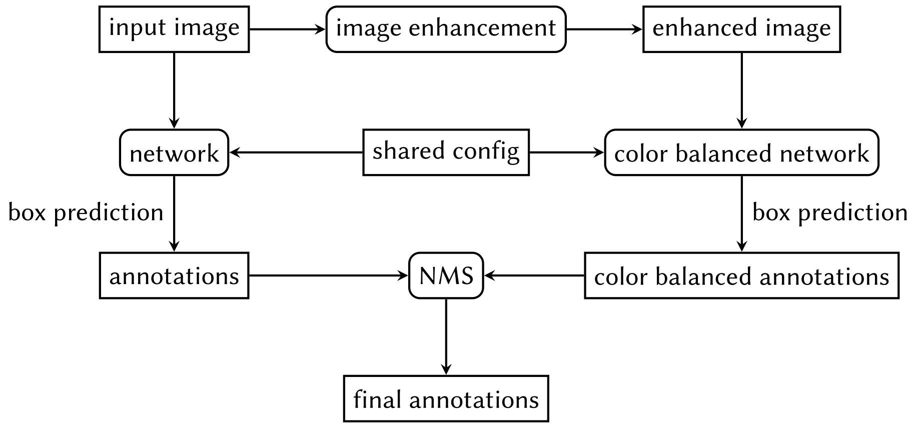


Figure 5: Use of the combined Network

Table 2
Results of the submission runs

Run-ID	Backbone	Image type	**Precision**	Recall
183911		default	0.365	0.269
183912	ResNet-50+FPN	color balanced	0.318	0.256
183913		combined	0.297	0.337
183914		default	0.371	0.246
183916	ResNet-101+FPN	color balanced	0.305	0.270
183918		combined	0.291	0.344
183919		default	**0.396**	0.292
183920	ResNet-152+FPN	color balanced	0.366	0.292
183922		combined	0.336	**0.393**

6. Results and discussion

The results of the submitted runs are shown in Table 2. The best result according to the Challenge's evaluation metric was the run with the ID 183919. It achieved an $MAP_{0.5}$ of 0.396.

In view of increasing precision in connection with increasing depth, we can assume, that precision can be improved using an even deeper backbone. The same applies to recall.

With the combined method using both the original and color balanced images, the recall could be improved significantly. For example the $MAR_{0.5}$ for the ResNet-152+FPN network with combined images was 0.393 while the $MAR_{0.5}$ of the network with the original images was 0.292.

The results of all three image-type approaches used can be seen in Figure 6. In total, 5 corals could be identified in the exemplary image section by combining using NMS. In comparison, the network on the original images found only 3 corals and the network on the enhanced images found only 4 corals. The duplicate annotations found on both images were correctly sorted out by NMS. Using the data for the $MAR_{0.5}$ from Table 2, it can be seen that this method was able to find more corals for all 3 network architectures by combining the predictions on the original images and the predictions on the enhanced images.

(a) predictions on original image (b) predictions on enhanced image (c) combined predictions

Figure 6: Predictions for original image, enhanced image and combined predictions

The average precision per substrate for the ResNet-152 approach for the original images and the enhanced images given in Table 3. For most coral species the difference is less than 0.01, but the species "c_hard_coral_boulder", "c_hard_coral_mushroom" and "c_hard_coral_foliose" could be detected much better by the network with the enhanced images. In contrast, it was significantly worse especially with the species "c_soft_coral_gorgonian". Because both networks had strengths and weaknesses for certain coral species, the combined network was able to benefit from the strengths of both.

Table 3
Average precision per substrate on validation data

Class	AP original images	AP enhanced images	difference	difference(%)
c_soft_coral	0.237	0.233	-0.004	-1,7%
c_hard_coral_boulder	0.220	0.242	+0.022	+10%
c_sponge	0.121	0.117	-0.004	-3,3%
c_hard_coral_branching	0.208	0.199	-0.009	-4,3%
c_hard_coral_submassive	0.197	0.197	0	0%
c_algae_macro_or_leaves	0.049	0.057	+0.008	+16,3%
c_hard_coral_table	0.318	0.316	-0.002	-0,6%
c_sponge_barrel	0.354	0.326	-0.028	-7,9%
c_hard_coral_encrusting	0.461	0.467	+0.006	+1,3%
c_hard_coral_mushroom	0.251	0.316	+0.065	+25,9%
c_hard_coral_foliose	0.103	0.184	+0.081	+78,6%
c_soft_coral_gorgonian	0.254	0.142	-0.112	-44,1%
c_fire_coral_millepora	0	0	0	0

7. Conclusion and Perspective

Overall, the results of the challenge are satisfying and the improvements we made to the models had the desired effects. Some image quality issues, as described in section 2, were improved. However, the quality of the annotation data should be addressed in future versions of the challenge. For instance, it is a bad starting point to have inaccurate bounding boxes containing in some cases a set of individual corals, and in other cases grouping these objects to a single annotation. It appears to make much more sense to have more precise annotations, similar to the "coral reef image pixel-wise parsing" subtask.

Furthermore, dense coverage of ground surface as well as fluctuating image quality, makes it difficult to distinguish the different substrate types. However, a deeper network seems to be more capable of handling these difficulties. The image enhancements made have not produced better results on their own, but have found different corals than the network with the original images. By combining the bounding boxes of both approaches and applying non-maximum suppression, the best overall $\mathrm{MAR}_{0.5}$ was obtained.

References

[1] B. Ionescu, H. Müller, R. Péteri, J. Rückert, A. Ben Abacha, A. G. S. de Herrera, C. M. Friedrich, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, S. Kozlovski, Y. D. Cid, V. Kovalev, L.-D. Ştefan, M. G. Constantin, M. Dogariu, A. Popescu, J. Deshayes-Chossart, H. Schindler, J. Chamberlain, A. Campello, A. Clark, Overview of the ImageCLEF 2022: Multimedia retrieval in medical, social media and nature applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 13th International Conference of the CLEF Association (CLEF 2022), LNCS Lecture Notes in Computer Science, Springer, Bologna, Italy, 2022.

[2] J. Chamberlain, A. Garcia Seco de Herrera, A. Campello, A. Clark, ImageCLEFcoral task: Coral reef image annotation and localisation, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 13th International Conference of the CLEF Association (CLEF 2022), LNCS Lecture Notes in Computer Science, Springer, Bologna, Italy, 2022.

[3] O. Hoegh-Guldberg, The Impact of Climate Change on Coral Reef Ecosystems, Springer Netherlands, Dordrecht, 2011, pp. 391–403. URL: https://doi.org/10.1007/978-94-007-0114-4_22. doi:10.1007/978-94-007-0114-4_22.

[4] M. Spalding, A. Grenfell, New estimates of global and regional coral reef areas, Coral reefs 16 (1997) 225–230.

[5] L. Soukup, Automatic coral reef annotation, localization and pixel-wise parsing using mask r-cnn (2021).

[6] C. M. Caridade, A. R. Marçal, Automatic classification of coral images using color and textures., in: CLEF (Working Notes), 2019.

[7] L. Picek, A. Ríha, A. Zita, Coral reef annotation, localisation and pixel-wise classification using mask-rcnn and bag of tricks., in: CLEF (Working Notes), 2020.

[8] A. Steffens, A. Campello, J. Ravenscroft, A. Clark, H. Hagras, Deep segmentation: using

deep convolutional networks for coral reef pixel-wise parsing., in: CLEF (Working Notes), 2019.

[9] S. Jaisakthi, P. Mirunalini, C. Aravindan, Coral reef annotation and localization using faster r-cnn., in: CLEF (Working Notes), 2019.

[10] K. Bogomasov, P. Grawe, S. Conrad, A two-staged approach for localization and classification of coral reef structures and compositions, in: CLEF (Working Notes), 2019.

[11] K. Bogomasov, P. Grawe, S. Conrad, Enhanced localization and classification of coral reef structures and compositions, in: CLEF (Working Notes), 2020.

[12] S. Ren, K. He, R. B. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, CoRR abs/1506.01497 (2015). URL: http://arxiv.org/abs/1506.01497. arXiv:1506.01497.

[13] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, R. Girshick, Detectron2, https://github.com/facebookresearch/detectron2, 2019.

[14] 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 Systems 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.

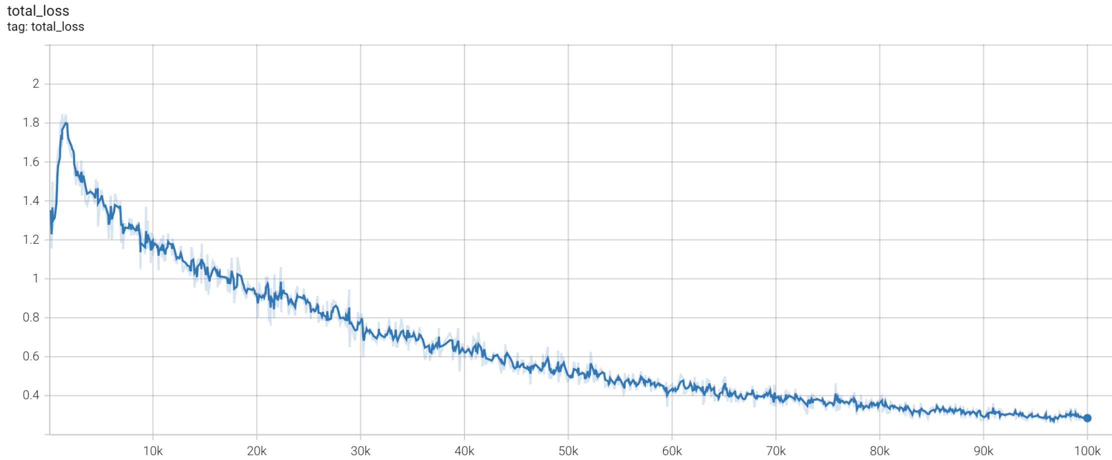
[15] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, 2016. URL: https://arxiv.org/abs/1612.03144. doi:10.48550/ARXIV.1612.03144.

[16] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, Microsoft coco: Common objects in context, 2014. URL: https://arxiv.org/abs/1405.0312. doi:10.48550/ARXIV.1405.0312.

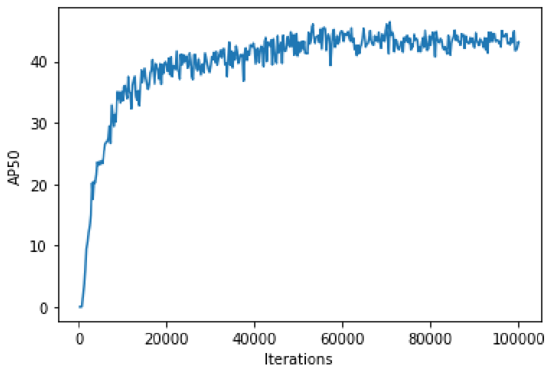
[17] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015. URL: https://arxiv.org/abs/1512.03385. doi:10.48550/ARXIV.1512.03385.

[18] N. B. Andersen, underwater-image-color-correction, https://github.com/nikolajbech/underwater-image-color-correction, 2020.

A. Figures

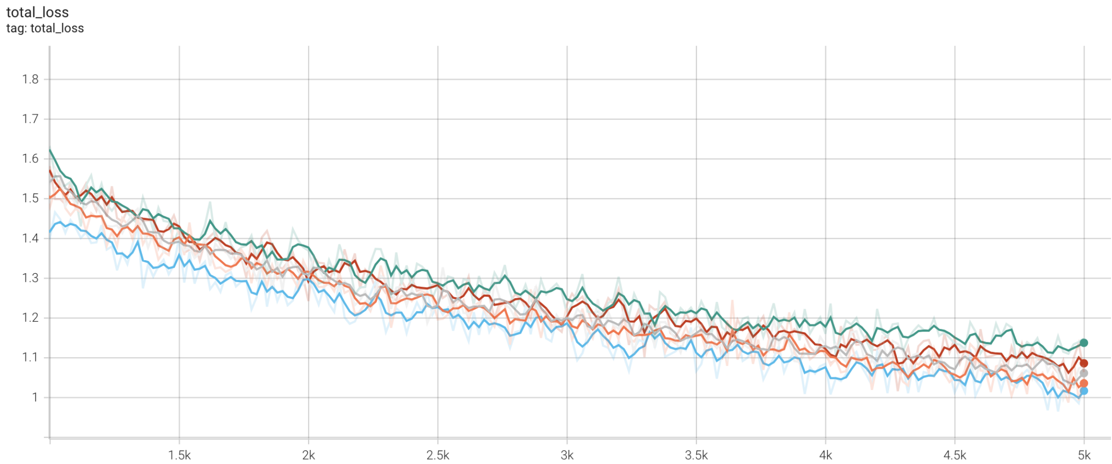


(a) total_loss

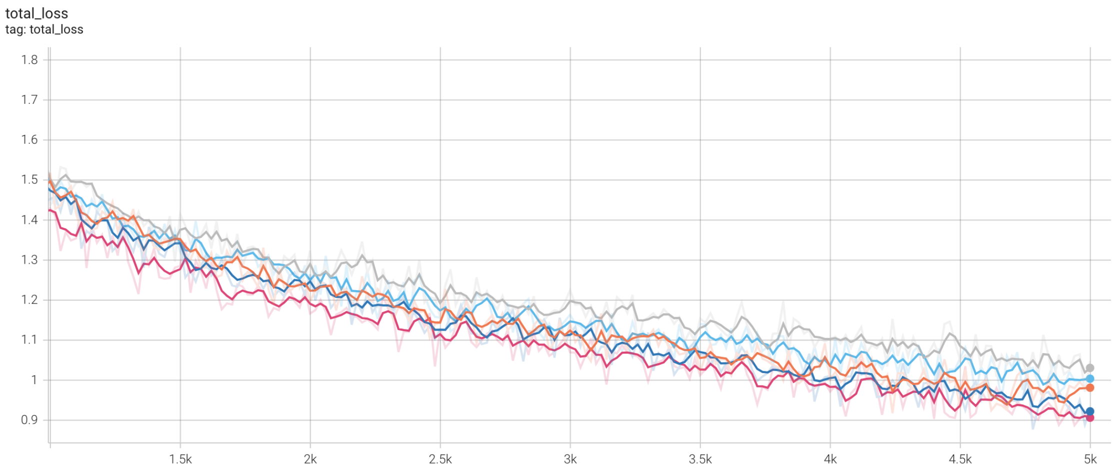


(b) $\mathrm{MAP}_{0.5}$

Figure 7: Total training loss and $\mathrm{MAP}_{0.5}$



(a) training loss for ResNet-50 with different batch sizes
32, 64, 128, 256, 512



(b) training loss for ResNet-101 with different batch sizes
32, 64, 128, 256, 512

Figure 8: Total training loss for different batch sizes