

Pixelwise annotation of coral reef substrates

Jessica Wright¹, Ioana-Lia Palosanu¹, Louis Clift¹, Alba García Seco de Herrera¹ and Jon Chamberlain¹

¹*School of Computer Science and Electronic Engineering, University of Essex, UK*

Abstract

Coral reef substrate composition is regularly surveyed for ecosystem health monitoring. The current method of visual assessment is slow and limited in scale. ImageCLEFcoral aims to identify reef areas of interest and annotate them appropriately. We present an adaptation of a submission to the 2019 ImageCLEFcoral task that uses a semantic segmentation model, DeepLabV3, with a ResNet-101 backbone. We implemented pre-training image colour enhancement and supplemented the available training data with that of NOAA NCEI for specific runs. Our runs had no overall improvement from the 2019 code, though did predict submassive corals and table corals with greater accuracy (+3.022% and +0.353%). Though none of our model runs had the highest precision or accuracy, we did best predict submassive corals (3.022%), boulder corals (12.787%), table corals (0.353%), foliose corals (0.097%), gorgonian soft corals (0.002%) and algae (0.027%) across 3 of our 4 runs. Image colour enhancement benefited the prediction accuracy of boulder corals (+1.209 – 5.026%), encrusting corals (+1.7 – 2.578%) and algae (+0.027%), most likely by making them more distinct from their surroundings. Adding NOAA data enhanced the precision of encrusting coral, soft coral and gorgonian predictions despite only providing additional annotations for encrusting and foliose corals. Our results suggest that a more balanced approach to data augmentation combined with image-specific colour improvements may provide a more desirable outcome, particularly when paired with a model that is fine-tuned to the data set used.

Keywords

Image segmentation, automatic annotation, coral reef annotation, semantic segmentation

1. Introduction

Coral reefs are vital marine systems that are known to provide many ecosystem functions and services [1] while supporting one third of marine species [2]. Their decline has been widely reported and tracked [3]. Current monitoring of coral reefs benthic communities relies on *in-situ* data collection, sometimes followed by *ex-situ* video analysis, requiring time and expertise to analyse correctly [4]. Automatic annotation from video stills or photographs would greatly increase the speed and scale of feasible monitoring, and could free up reef experts to focus on other areas to gain a wider view of shifting coral reef dynamics.

Deep learning algorithms provide an answer to automatic annotation [5]. The underlying architecture of most are Convolutional Neural Networks, often used for image and pattern recognition [6]. Image segmentation models have been the most successful in the ImageCLEFcoral pixel-wise parsing task [7, 8], where each pixel is predicted as a particular class.

This is the third iteration of an annual ImageCLEF task [9, 10, 11] which has subtasks looking

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania

✉ jpwriga@essex.ac.uk (J. Wright); ip17484@essex.ac.uk (I. Palosanu)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

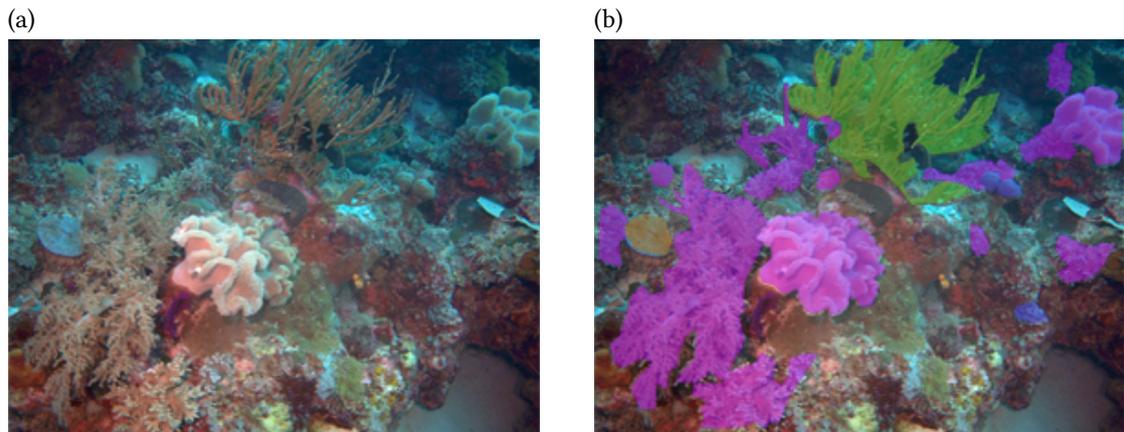


Figure 1: Examples of reef images (a) without and (b) with morphological substrate annotations.

into (1) Coral reef image annotation and localisation and (2) Coral reef image pixel-wise parsing. Considering the value of each subtask in terms of practical use in monitoring reef systems accurately, we focused on subtask 2.

2. Data

The initial data provided were split into a training and test images of coral reef systems. The training set was annotated (Fig. 1) with the morphological substrate classes set in the task [9] and the test set was not annotated. The training set was then provided to build and train a network, with the test set given later to get submission runs. More details about the dataset can be found at Chamberlain et al. [11].

2.1. The training set

879 images with a combined set of 21,748 annotations were provided as the training data. The annotations were not evenly split across classes, likely as some are more prevalent than others in reef systems (Fig. 2).

Each substrate morphology can be indistinct from others due to the variation in that class' species. This is particularly true of classes that are not broken down into morphological groups, i.e. "soft coral", and less of an issue with classes that are split, i.e. each "hard coral" group.

The use of NOAA NCEI¹ and/or CoralNet² data was recommended for the task. We chose to utilise the NOAA data set for some experiments. 3032 NOAA images were downloaded of a possible 15,019, due to time limitations on our machines. The NOAA data set contains a greater number of classification labels than the ImageCLEFcoral classes. These classifications are also of a single pixel (10 pixels per image) so did not provide enough information for our image analysis and recognition algorithms. We developed a NOAA Translation processor to capture

¹<https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0211063>

²<https://coralnet.ucsd.edu/>

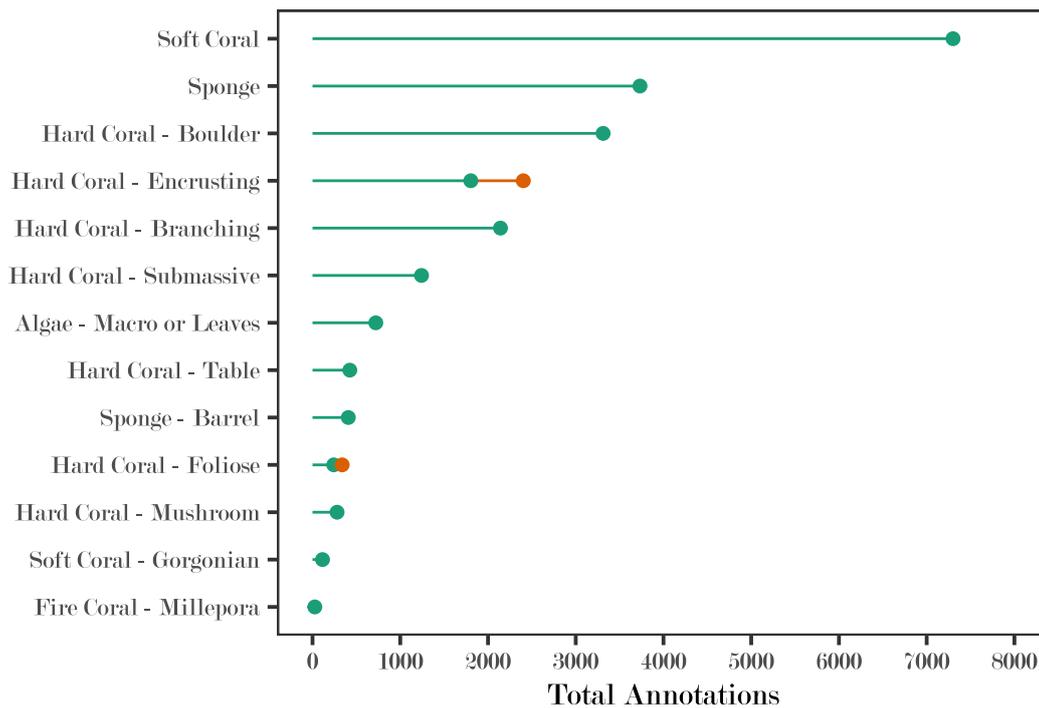


Figure 2: Substrate annotations in the ImageCLEF training set of 879 images ($n = 21,748$; green), and in the training set when combined with an additional 502 NOAA images ($n = 22,403$; orange).

the classification types within the data set and translate them via an expert defined translation matrix into the ImageCLEFcoral classes which we made available through the ImageCLEFcoral website for other participants. The processor then created an adjustable Region Of Interest (ROI) around the same pixel to provide an image patch, typically a 10x10 pixel area, that enabled our machine learning routines to adapt to the NOAA data sets.

5 substrate classes were then selected to refine the number of images to a more manageable amount: Fire Coral - Millepora, Hard Coral - Foliose, Hard Coral - Table, Hard Coral - Sub-Massive, and Hard Coral - Encrusting. These classes had a lower number of annotations than others and were chosen to increase accuracy. Despite low incidence, Soft Coral - Gorgonian, Hard Coral - Mushroom, and Sponge - Barrel were not chosen from the NOAA data set as they have more distinct morphologies than the selected classes so were more likely to be predicted despite relatively few occurrences. Algae - Macro or Leaves were also not selected from the NOAA data set despite low incidence. Algae classification of the ImageCLEF set only accounted for large leaf macroalgae, whereas the NOAA data set also included other types such as turf and CCA, so conflicting annotations could have hampered the model predictions.

502 viable NOAA images were found, within which 2 of the 5 selected classes were found: Hard Coral - Encrusting and Hard Coral - Foliose. This almost doubled the processing time per epoch and pushed the entire model training time from 10 hours to 17.5 hours (10 epochs total), and increased the total number of substrate annotations from 21,748 to 22,403 (Fig. 2).

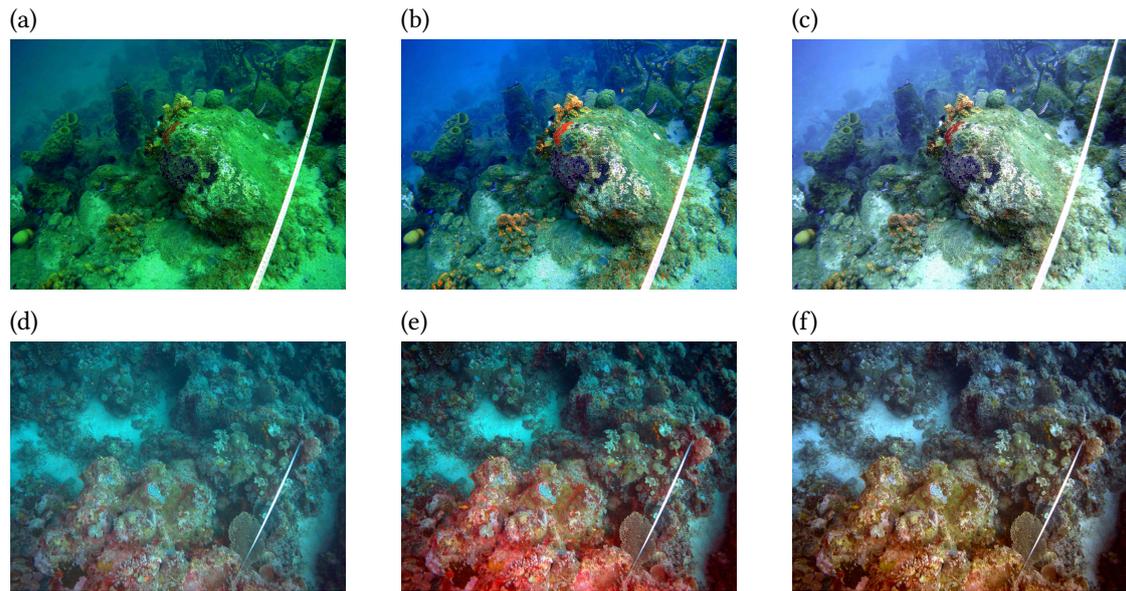


Figure 3: Transformation of (a) a green and (d) blue image through 2 stages to an image used in training. Each image had (b,e) balanced RGB levels, then (c,f) when through a generalised channel mixing process to balance the colors while maintaining image contrast. The leveling and mixing were selected to optimise substrate color and contrast with less focus on the background and water colouring.

2.2. Image enhancement

Underwater imagery is often lower quality than that taken on land. Light attenuation distorts colour detection and water turbidity can reduce image quality, and with all underwater imagery there is a greater chance for blurred or unfocused photographs. Taking steps to investigate, process and augment the provided data is expected to improve the data quality and subsequent network results [12, 7].

Images were visually assessed and split into those with accurate colouring and contrast, those with a heavy green tint and those with a heavy blue tint. Accurate images were not altered in any way. Green and blue images were passed through an RGB histogram leveller followed by an RGB channel mixer, generalised to green or blue images for speed (Fig. 3). This would allow all the images to be processed easily but would not allow for image-specific editing.

2.3. Data augmentation

Before training the model, each image was cropped into 12 squares which were each then cropped at a random point to a 480px square. Random horizontal flips were also utilized due to the limited amount of data. These pre-processing techniques are used to present the model with different iterations of the same images, increasing the size of the data set available.

2.4. The test set

The provided test set included 485 unannotated images from 4 different regions:

- Region 1: The training set location.
- Region 2: A geographically and biologically similar region.
- Region 3: A geographically distinct but biologically similar region.
- Region 4: A region that is both geographically and biologically distinct.

The test images were also cropped into 12 squares to match the training images used on the model. Each test image was then resized to a 520px square, which allowed us to predict all test images despite system limitations.

The predicted pixel array of each test image had to be resized to its original dimensions before submission to match the ground truth annotation mask. This was carried out using spline interpolation through the zoom function in SciPy³.

3. The model

We used the DeepLabV3 model based on a previous submission [8]. It makes use of a ResNet-101 backbone and the application of both atrous convolution and Atrous Spatial Pyramid Pooling (ASPP). ResNet-101 is used for feature extraction before atrous convolution and ASPP are applied. Atrous convolution increases the field of view in the last layer of ResNet-101 by inserting 0-values between filter values used in the network layer [13]. The atrous rate utilised corresponds to the amount of 0-values inserted - the higher the rate, the bigger the field of view becomes. ASPP is then applied to assign a label to each pixel using 4 atrous convolution rates. This enables the model to utilise different aspects of the objects it is identifying, ensuring that when pixels are labelled the network has seen multiple perspectives of field of view.

The model was evaluated using different crop and batch sizes during training. Batch size 4 was used in each run as it had the best performance within our system limitations. A crop size of 480px was selected as, when combined with batch size 4, it provided the greatest overall accuracy (per mAP0.0 and mAP0.5) of all tested crop size combinations.

4. Submission

Each team in the competition were allowed to submit up to 10 runs per task using the collaboration platform AICrowd⁴. We chose to submit 4 files to the pixelwise-parsing subtask only, representing 4 individual runs:

- MTRU1:** the “baseline” run, using the 2019 submission [8] that was rewritten and finetuned by experimenting on crop × batch size combinations. Batch size 4 with crop size 480 were found to give the best results and were used in this run.

³<https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.zoom.html>

⁴<https://www.aicrowd.com/>

- MTRU2:** the edited ImageCLEF run, using the same settings as MTRU1. Poorly coloured training images were enhanced to represent more accurate coloring of the coral reefs.
- MTRU3:** the NOAA run, using additional data from NOAA in three different substrates. The images were not enhanced or edited in any way, and the same settings from MTRU1 were used.
- MUTR4:** the fully edited run, using same settings as MTRU1, with both the additional NOAA data and image colour enhancements where needed.

4.1. Self-intersecting polygons

All 4 runs predicted some images containing self-intersecting polygons. These polygons invalidate a run and are not permitted in the submission file so must be removed. The evaluation script was used to identify any images with self-intersecting polygons and the substrate type of the polygon. This process involved removing each polygon of the relevant substrate type one by one, re-running the evaluation script each time to check if the error was resolved.

Initial images were checked polygon by polygon in this manner to minimise any impact on model accuracy but the time constraints of the challenge required faster processing of the latter runs. Images in these runs were checked in polygon "batches", where several at a time would be deleted before running the evaluation script. While this did increase the speed of evaluation before submission, it is likely that a significant proportion of the deleted polygons were not self-intersecting and as such the mean average precision (mAP) of the runs would be both lower and less accurate.

4.2. Blank predictions

3 of the 4 runs (MTRU2, MTRU3 and MTRU4) predicted images with no substrate classes. While clearly an error as all images were of coral reef substratum, these predictions were a part of our model outcome and therefore our submitted runs. The evaluation script used upon submission blocks these images and deem runs with them a failure so each had to be altered. As all images from the test set must be used, the blank images could not be removed. Our solution to this was to include a small square annotation in the center of the blank images and label it as Fire Coral - Millepora. This class was used as it had the lowest number of annotations and had no additional images added from NOAA images so was likely to be the least accurate class, limiting the effect on overall accuracy as much as possible.

5. Results and discussion

Results provided by ImageCLEFcoral after submission used 2 metrics. mAP_{0.5} showed the localised mean average precision using $\text{IoU} \geq 0.5$. Accuracy per substrate calculated the segmentation accuracy as the number of correctly labelled pixels of class over the number of pixels labeled with class in ground truth.

Overall results of the pixel-wise parsing subtask (Table 1) show that our runs were less accurate and precise than the other participant team. When considering the accuracy per class,

however, there were some substrate categories that were better predicted by our model.

Across the MTRU runs, we saw the highest accuracy of submassive coral, table coral and foliose coral predictions when images run unedited and without additional NOAA data. The greatest prediction accuracy of boulder corals and algae across the subtask occurred when images were colour corrected, and of gorgonian soft corals occurred when unedited ImageCLEF data and NOAA data were used. MTRU3 was the only instance of gorgonian predictions with positive accuracy (0.002%) across all submissions. Similarly, MTRU1 was the only instance of positive accuracy in foliose coral prediction (0.097%). None of our runs predicted mushroom corals, sponges, barrel sponges or fire coral accurately.

Of our runs, the greatest precision was seen in MTRU1 ($mAP_{0.5} = 0.021$), though it did not have the highest accuracy (2.767%). MTRU4 was most accurate (2.951%) despite having the lowest overall precision ($mAP_{0.5} = 0.011$).

Table 1

Overall precision ($mAP_{0.5}$), average accuracy (%) and substrate class accuracy (%) of pixel-wise parsing subtask submissions to ImageCLEFcoral 2021 (4 MTRU runs, 1 run from team Pilsen Eyes that performed best overall). The best scores for each category are shown in orange.

Category	MTRU1	MTRU2	MTRU3	MTRU4	Pilsen Eyes
$mAP_{0.5}$	0.021	0.018	0.017	0.011	0.075
Average accuracy	2.767	2.531	1.942	2.951	6.147
Hard Coral - Branching	1.090	2.299	0.536	5.562	11.095
Hard Coral - Submassive	3.022	0.279	1.036	0.039	2.704
Hard Coral - Boulder	9.607	12.787	7.601	8.827	5.385
Hard Coral - Encrusting	0.017	2.595	0.729	2.429	2.615
Hard Coral - Table	0.353	0	0	0	0.008
Hard Coral - Foliose	0.097	0	0	0	0
Hard Coral - Mushroom	0	0	0	0	0
Soft Coral	0	0	0.228	0	50.433
Gorgonian	0	0	0.002	0	0
Sponge	0	0	0	0	1.625
Barrel Sponge	0	0	0	0	0.329
Fire Coral - Millepora	0	0	0	0	0
Algae	0	0.027	0	0	$1.0e^{-4}$

Overall precision and average accuracy were also lower than the 2019 run of this model [8], however we did show improvement in the prediction of submassive corals ($MTRU1 = 0.030221$, 2019 = 0) and table corals ($MTRU1 = 0.003534$, 2019 = 0), neither of which were predicted with any accuracy in 2019.

5.1. Image colour enhancement

The colour adjustments made to the images increased the prediction accuracy of boulder corals, encrusting corals and algae (Table 1). For boulder corals, colour enhancement may have distinguished them from other reef substrates and enabled greater recognition of the coral over rocks and other substratum that they can easily resemble. Encrusting corals would benefit for

the same reasons. Algae would likely show improvement with colour enhancement due to the removal of green image tints, which would allow the natural green of the algae to become more defined and clear. Brown and red algae would also benefit from the red channel correction to make them more distinct from surrounding substrate.

Submassive corals were less accurately predicted with image enhancement, as well as table corals, foliose corals, soft corals and gorgonians. Any loss in predictive power is likely due to the general nature of the colour correction performed. While some images would improve with the balancing and mixing at the levels set, others may have had colour blow outs or excessive input from one or more RGB channels. This could have a blur-like effect, wherein neighbouring substrate categories look indistinct from each other due to a lack of colour definition.

5.2. Augmenting annotations with NOAA data

The NOAA data used was from a different location than the ImageCLEF data which could greatly impact mAP0.5 and prediction accuracy as substrates from different geographic regions can show vastly different morphologies. Of the 2 categories with increased annotations from the NOAA data set, encrusting corals saw a greater accuracy while foliose corals had less prediction accuracy. Encrusting corals are very similar globally despite varying conditions, so increasing the number of annotations would likely improve the models predictive power by adding distinctive pixels to train on. This is not the case with foliose corals, which are more likely to show differing morphologies as they are not flat to the substrate. Foliose corals also have extensive structures that often appear layered and often appear to have many shadows that could hamper the training capabilities of the model. Any shadows would look like black, probably with a flat texture, regions of the image. These would provide no benefit to the model and may cause it to relate any dark spots to foliose corals or to fail to recognise them at all.

Adding NOAA data had a detrimental effect on the accuracy of most other substrate categories. Where a prediction accuracy > 0 without NOAA data (MTRU1 and MTRU2), adding NOAA annotations reduced the prediction accuracy of submassive, boulder, and table corals as well as algae. This could occur if the additional NOAA annotations skewed the models perception of each category and altered the predictions made as a result. Accuracy also decreased for branching corals between MTRU1 and MTRU3 (unedited images), but increased between the colour enhanced runs (MTRU2 and MTRU4) by 5.026%. Predictions were also more accurate for soft coral (+0.228%) and gorgonians (+0.002%) when NOAA data was added but no colour enhancement was performed. These substrate categories can form more distinct morphologies across all locations that may have become more distinct with an increasingly balanced data set at the expense of the other classes. Although the soft coral category encompasses several distinct organisms with different morphologies, the abundance of annotations likely compensated by providing many examples of each structure.

6. Conclusions

Image colour enhancements can increase the accuracy of coral reef substrate predictions when those substrates are otherwise difficult to distinguish from the surrounding environment. It can also be detrimental when the editing performed is generalised instead of image specific.

Similarly, augmenting the training data set with NOAA annotations can improve the predictions of substrates that are either morphologically general across different geographical regions or those that form distinct structures despite changing geography. Large increases in the number of annotations should be reflected in a subsequent increase of accuracy in the represented classes. When this does not occur, the abundance of data can impair the predictive power of the model by blurring the line between substrate categories through incorrect annotation or by skewing the predictions made as a result of an imbalanced data set.

A combination of an augmented data set with distinct image enhancement pathways for either different geographic locations or substrate categories may provide a more accurate and precise prediction array. Combining these steps with improved hyperparameters would enhance model performance and provide a coral reef substrate prediction tool that would be applicable to reefs across the globe.

6.1. Limitations of the model

The use of a dedicated GPU greatly increases the computational power of machine learning models. Training time can then be diminished and hyperparameters can be improved. The machine we used to run our model was affected by a lack of GPU memory, which can only be rectified by changing the graphics card to a more powerful one. The memory limitation heavily impacted batch sizes testing, limiting tests to batch size 4 at most. DeepLabV3 works best with a batch size of 16 (demonstrated on the PASCAL VOC data set [13]). Using a computer with a better GPU would allow for a greater batch size to be used which would improve the model parameters and strengthen the power of the predictions.

In the future we plan to include a greater volume of NOAA data when training the model. This would both increase the number annotations per class across the training data. More specific pixel expansion would have also enabled us to be more precise in training and may have provided more pixels per class than otherwise achieved. A potential method could have different expansion shapes set by class (i.e. boulder coral expands as a circle) and a pixel selection/rejection threshold based on annotated pixel value.

6.2. Improving the approach

Images and predictions would likely benefit from a more tailored colour correction approach. This could be performed with the commonly used Rayleigh distribution [12, 14, 15] or with a different approach such as red channel weighted compensations [16] that leverage the other colour input channels to colour balance an image with accuracy.

Leveraging the results from this approach, developing a staggered pipeline may improve prediction accuracy in the future. A bounding box approach to gain a generalised location of each substrate could be used to then send images through different processing steps, such as colour correction, blur reduction, contrast changes, etc, based on the class found. This could then feed into a pixel-wise prediction model to find precise location of substrate classes within an image.

Acknowledgments

We would like to thank the team that developed the 2019 base code that we used [8], particularly Antonio Campello for his support and advice throughout this process. We would also like to thank NOAA and the MTRU team of participants at the 2020 NOAA hackathon <https://www.gpuhackathons.org/event/noaa-gpu-hackathon>, when we began working on this project.

References

- [1] F. Moberg, C. Folke, Ecological goods and services of coral reef systems, *Ecological Economics* 29 (1999) 215–233.
- [2] B. Bowen, L. Rocha, R. Toonen, S. Karl, The origins of tropical marine biodiversity, *Trends in Ecology and Evolution* 28 (2013) 359–366.
- [3] L. Jones, P. Mannion, A. Farnsworth, P. Valdes, S. Kelland, P. A. Allison, Coupling of palaeontological and neontological reef coral data improves forecasts of biodiversity responses under global climatic change, *Royal Society Open Science* 6 (2019).
- [4] J. Hill, C. Wilkinson, *Methods for Ecological Monitoring of Coral Reefs*, 1 ed., Australian Institute of Marine Science, Townsville, Australia, 2004.
- [5] 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*, 2016.
- [6] K. O’Shea, R. Nash, *An Introduction to Convolutional Neural Networks*, 2015. [arXiv:1511.08458](https://arxiv.org/abs/1511.08458).
- [7] L. Pícek, A. Říha, A. Zita, Coral reef annotation, localisation and pixel-wise classification using Mask R-CNN and Bag of Tricks, in: *CLEF2020 Working Notes*, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020.
- [8] A. Steffens, A. Campello, J. Ravenscroft, A. Clark, H. Hagraš, Deep segmentation: Using deep convolutional networks for coral reef pixel-wise parsing, in: *CLEF2019 Working Notes*, volume 2380 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019.
- [9] J. Chamberlain, A. Campello, J. P. Wright, L. G. Clift, A. Clark, A. García Seco de Herrera, Overview of ImageCLEFcoral 2019 task, in: *CLEF2019 Working Notes*, volume 2380 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019.
- [10] J. Chamberlain, A. Campello, J. P. Wright, L. G. Clift, A. Clark, A. García Seco de Herrera, Overview of the ImageCLEFcoral 2020 task: Automated coral reef image annotation, in: *CLEF2020 Working Notes*, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020.
- [11] J. Chamberlain, A. García Seco de Herrera, A. Campello, A. Clark, T. A. Oliver, H. Moustahfid, Overview of the ImageCLEFcoral 2021task: Coral reef image annotation of a 3d environment, in: *CLEF2021 Working Notes*, *CEUR Workshop Proceedings*, CEUR-WS.org, Bucharest, Romania, 2021.
- [12] M. Arendt, J. Rückert, R. Brüngel, C. Brumann, C. Friedrich, The effects of colour enhancement and IoU optimisation on object detection and segmentation of coral reef structures,

in: CLEF2020 Working Notes, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020.

- [13] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking Atrous Convolution for Semantic Image Segmentation, 2017. [arXiv:1706.05587](https://arxiv.org/abs/1706.05587).
- [14] A. Abdul Ghani, N. Mat Isa, Underwater image quality enhancement through composition of dual-intensity images and rayleigh-stretching, *SpringerPlus* 3 (2014) 757.
- [15] A. Abdul Ghani, N. Mat Isa, Underwater image quality enhancement through integrated color model with rayleigh distribution, *Applied Soft Computing* 27 (2014) 219–230.
- [16] W. Xiang, P. Yang, S. Wang, B. Xu, H. Liu, Underwater image enhancement based on red channel weighted compensation and gamma correction model, *Opto-Electronic Advances* 1 (2018) 180024.