

DEEP LEARNING FOR SEMANTIC SEGMENTATION OF CORAL IMAGES IN UNDERWATER PHOTOGRAMMETRY

Hanqi Zhang¹, Armin Gruen², Ming Li^{1,2*}

¹ Wuhan University, LIESMARS, 430074 Wuhan, China
Email: hqzhang@whu.edu.cn

² ETH Zurich, Institute of Theoretical Physics, 8093 Zurich, Switzerland
Email: agruen@geod.baug.ethz.ch, mingli39@phys.ethz.ch

KEY WORDS: Coral Images, Semantic Segmentation, Boundary Loss, Neural Network (NN), Deep Learning, Underwater Photogrammetry.

ABSTRACT:

Regular monitoring activities are important for assessing the influence of unfavourable factors on corals and tracking subsequent recovery or decline. Deep learning-based underwater photogrammetry provides a comprehensive solution for automatic large-scale and precise monitoring. It can quickly acquire a large range of underwater coral reef images, and extract information from these coral images through advanced image processing technology and deep learning methods. This procedure has three major components: (a) Generation of 3D models, (b) understanding of relevant corals in the images, and (c) tracking of those models over time and spatial change analysis. This paper focusses on issue (b), it applies five state-of-the-art neural networks to the semantic segmentation of coral images, compares their performance, and proposes a new coral semantic segmentation method. Finally, in order to quantitatively evaluate the performance of neural networks for semantic segmentation in these experiments, this paper uses mean class-wise Intersection over Union (mIoU), the most commonly used accuracy measure in semantic segmentation, as the standard metric. Meanwhile, considering that the coral boundary is very irregular and the evaluation index of IoU is not accurate enough, a new segmentation evaluation index based on boundary quality, Boundary IoU, is also used to evaluate the segmentation effect. The proposed trained network can accurately distinguish living from dead corals, which could reflect the health of the corals in the area of interest. The classification results show that we achieve state-of-the-art performance compared to other methods tested on the dataset provided in this paper on underwater coral images.

1. INTRODUCTION

Coral reefs represent one of the most diverse ecosystems on Earth. They are also among the most productive of all ecosystems and they have immense ecological, social and economic value. Coral reefs are undergoing rapid changes as a result of increasing ocean temperatures, acidification, eutrophication, *Acanthaster planci* (crown-of-thorn starfish) eruption, and chemical pollution. Increasingly, coral reefs worldwide are being affected by perturbations that range from short-term, localized disturbances - where return to the original state is possible - to more chronic, widespread influences of shifts in climate that may fundamentally alter the ecosystem (Knowlton, 2001; Jackson et al., 2003; Boesch et al., 2001; Cressey, 2015; Raj et al., 2021). Regular monitoring activities are important for assessing the influence of unfavourable factors on corals and tracking subsequent recovery or decline (Pavoni et al., 2021; Cai et al., 2021), Figure 1 shows the partial coral changes in Moorea Island from 2017 to 2019 close to a known underwater survey control point. It can be seen that it takes only three years for some corals to grow from health to death. Monitoring by field surveys provides accurate data but at highly localized scales and so is not cost-effective for coral reef scale monitoring at frequent time intervals. Satellite and aerial photogrammetry and remote sensing are alternative and complementary approaches, while remote sensing on coral reefs from satellites and drones cannot provide the level of detail and accuracy required. No traditional method is appropriate for large-area and short-period coral accuracy monitoring. There are also bottlenecks in data processing besides the difficulty of gathering good and suitable images. The National Oceanic and Atmosphere Administration (NOAA) reports that marine biologists are just able to analyse 1-2% of the millions of coral

images acquired each year (Pavoni et al., 2020). Deep learning-based underwater photogrammetry provides a new comprehensive solution for large scale and accurate monitoring (Yuval et al., 2021). It can quickly acquire a large range of underwater coral reef images. These images not only generate accurate 3D object models and orthophotos through advanced image processing techniques, but also through deep learning-based coral annotation and semantic segmentation techniques to enable the identification of the different corals and their distinction between living and dead corals and other objects. Through this method, we can conduct regular research on coral reefs in the monitoring area to obtain accurate information on coral growth, coverage, distribution, and recovery in a certain period of time rapidly. It cannot only improve the cost-effectiveness ratio based on underwater field observations, but also make up for the lack of accuracy of traditional photogrammetry and remote sensing.

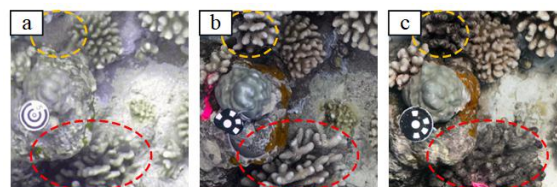


Figure 1. Coral reef images (a) 2017, (b) 2018, (c) 2019.

Coral semantic information extraction based on underwater photogrammetry and deep learning mainly includes underwater coral reef image photography, image pre-processing, image annotation and semantic image segmentation. In coral reef image acquisition and underwater photogrammetry processing, the popularity of low-cost cameras in underwater image acquisition has greatly increased the amount of coral image archived data and the frequency of observations, and methods and technologies for underwater high precision photogrammetry

* Corresponding author

have been developed. At present, the reconstruction of coral reef seabed 3D models using underwater low-cost consumer grade camera images can reach millimeter-level accuracy (Guo et al., 2016; Gruen et al., 2018; Nocerino et al., 2020). This lays the foundation for subsequent high-quality orthophoto processing. In the field of coral reef image annotation and semantic segmentation, annotation methods are developing from random point-based annotation to pixel-based annotation, and the segmentation method is changing from support vector machine method to neural network semantic segmentation (Beijbom et al., 2012; Beijbom et al., 2016; Alonso et al., 2017; Alonso et al., 2019). The use of orthophotos for coral semantic information extraction can effectively improve the semantic image segmentation of neural network caused by inconsistencies in geometric deformation and colour between images, and it is the current mainstream choice (Pavoni et al., 2019; Mizuno et al., 2020; Pavoni et al., 2021). Specific to the research work of this paper, it uses the advanced underwater photogrammetry technology proposed for orthophoto processing at first (Guo et al., 2016; Gruen et al., 2018). Then in coral annotation, the TagLab tool is used for coral annotation with biologists' participation (Pavoni et al., 2020). Compared with the traditional point-based annotation method, this method can not only save a lot of time, but also describe the specific characteristics of corals well. In the end, this paper selects five state-of-the-art neural networks suitable for semantic segmentation (RefineNet, SegNet, DenseASPP, DeepLab V3+, and U-Net) to test and evaluate the coral semantic segmentation results of different neural network models, compare their performance, and proposes a new deep learning-based coral semantic segmentation method based on our experiments, and improves the ability of coral semantic information extraction by improving coral boundary extraction and segmentation. Meanwhile, in order to quantitatively describe the performance of the semantic neural network in this experiment, this paper uses the neural network model evaluation index of mean Intersection over Union (mIoU) and Boundary IoU (Cheng et al., 2021). The proposed trained network can accurately distinguish living from dead corals, which could reflect the health of the corals in the area of interest. It is particularly suitable for periodic field studies of coral reefs through long-term ecological research programmes.

2. STUDY AREA AND IMAGE DATA

2.1 Study Area

Moorea Island is situated 17°32' South and 149°50' West, 25 km distant from Tahiti in the Society archipelago in French Polynesia. It is a high island, entirely surrounded by a coral reef rim, cut by eleven passes. Moorea island has attracted the interests of biologists, geophysicists, ecologists, environmentalists and geospatial Information Scientists due to its strategic location in the middle of the South Pacific Ocean, where it is offering some of the most complex ecosystems on earth. It can be studied to reveal the effects of natural and anthropogenic changes. The coral reefs around Moorea Island, were selected as the area observed in this study because they are flat, diverse, and representative, they have been a research hotspot for a long time and one of the important research contents of the IDEA (Island Digital Ecosystem Avatars) project. The IDEA project was recently developed as a joint initiative by an interdisciplinary group of experts to create a digital avatar of the Pacific Island of Moorea (Cressey, 2015). In the past years, many underwater coral reef data have been collected, which are very suitable for the use of underwater photogrammetry and Deep Learning techniques to study automatic semantic mapping of benthic habitats based on

underwater images, especially coral cover, individual morphology and growth. Since 1979, information has been obtained concerning the processes involved in reef complex degradation. Possible causative agents of this disturbance are: sedimentation effects, cyclones, weak bleaching events, outbreaks of *Acanthaster planci*, algal spreading, predator populations and microbial diseases. (Faurea, 1989; Adjeroud, 1997; Feeney et al., 2021). In particular, it is worth pointing out that when *Acanthaster planci* erupts, they can cause great damage to corals, and their concentrated outbreaks on the reef floor are closely related to ocean water temperature warming and nutrient enrichment caused by human development. The satellite imagery map and 3D model in Figure 2 show the distribution of coral benthic habitats in the shallow waters around Moorea Island.

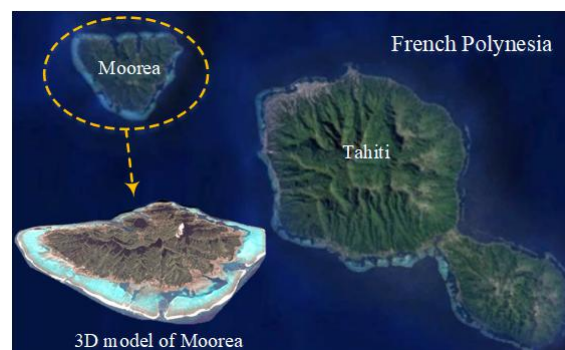


Figure 2. A 3D model and satellite imagery map of Moorea.

2.2 Coral Image Data

According to the characteristics of the underwater lighting environment, underwater images will be affected by colour absorption in certain bands, changes in clarity (turbidity in the water), geometric distortions and colour differences. Therefore, this paper uses orthophotos for image annotation. Orthophotos are geometrically corrected, have a constant pixel size and always present a vertical viewpoint, which provides convenience for semantic image segmentation. Figure 3a is the orthophoto of the approximately square study area of 5m × 5m obtained by processing 323 underwater coral reef images through Agisoft Metashape Professional 1.8.0 software. These images were taken by the IDEA project in 2019.

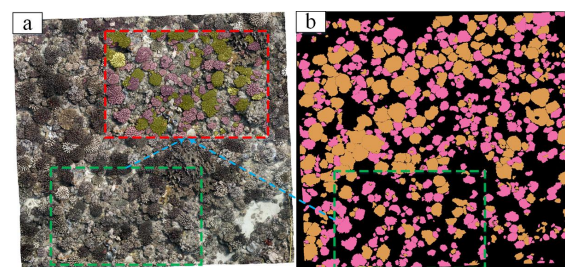


Figure 3. The picture inside the red rectangle is a partial overlay display of the corresponding parts in (a) and (b), (a) a coral reef orthophoto, (b) a corresponding expert-labelled image.

3. METHODS

3.1 Data Pre-processing

Corals have a large quantity of species, presenting the great interspecific difference and intraspecific mutability, specifically in their morphology and colour. There will be varying degrees of inconsistency in identifying benthic organisms among observers and by the same observer over time, no exception for experienced observers (Ninio et al., 2003). Therefore, the annotation of the training data set of the benthic community

requires experienced professionals to carry out. Figure 3b is the experimental annotation data of this paper. It uses a whole benthic orthographic image marked by marine experts. It is a colour RGB seabed orthographic image with a size of 11317×10773 pixels. The species marked in the picture is the staghorn cup coral *Pocillopora*, the dominant local coral species. The pink ones are live corals, and the yellow ones are dead corals. The quality of the training image will directly affect the accuracy of the semantic segmentation network. Therefore, the pre-processing of training images used for the segmentation network is very important. Figure 4 lists the key steps of image pre-processing in this paper, which is actually a data augmentation based on image processing techniques. Specifically, first of all, the experimental orthophoto is cropped into multiple 448×448 pixels coral image slices at a fixed step size of 160 pixels, and images with insufficient label data are deleted to prevent learning too much useless feature information. Of the 1967 images divided in total, 0.6 is for the training dataset, 0.2 for the validation dataset, and 0.2 for the test dataset, all of which are randomly assigned. Then, random horizontal or vertical flipping, random rotation and random translation were performed on the processed coral image slices. In order not to lose too much label information, this paper randomly rotates by ± 10 degrees and translates by ± 50 pixels to increase the richness of the image, improve the accuracy of the network, and effectively increase the robustness of the semantic segmentation network model. It should be pointed out that the above operations are a data augmentation technique for increasing the diversity of datasets in deep learning. This type of pre-processing eliminates the need to collect more real data, but still helps to improve model accuracy and prevent model overfitting. Finally, limited by the GPU memory, in order to reduce the amount of network calculations, but preserve the edge information of the image as much as possible, this paper scales the coral slices to a size of 224×224 pixels, based on the above processing. Resizing the input image size of coral image classification network to 224×224 is also based on the experience of a large number of excellent classification networks. Because in the classification task, if the size of the feature map is too small, excessive feature information will be lost. If the size is too large, the abstraction level of the information is not high enough, and the amount of calculation is also larger. Therefore, the choice of this paper is to achieve a good balance between the performance of the network and the amount of computation. After completing the above image pre-processing steps, the slice data must be normalized so that their features have the same measurement scale. Since adjacent slices that are continuously cropped have certain rules, in order to eliminate the correlation between slice data, this paper shuffles all slices before putting them into network training.

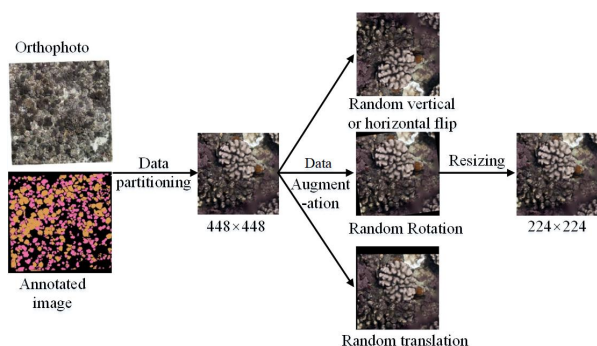


Figure 4. Data Augmentation.

3.2 Five Test CNNs for Coral Semantic Segmentation

This paper selects five mainstream neural networks (RefineNet, SegNet, DenseASPP, DeepLab V3+, and U-Net) for the evaluation and research of semantic segmentation of coral reef images. Through quantitative comparison and analysis of their performance, the new semantic image segmentation network model of this paper is finally proposed.

3.2.1 U-Net

The U-Net neural network structure was first published in 2015 (Ronneberger et al., 2015). Because of the characteristics of simple structure and suitable for small datasets, it plays a great role in semantic segmentation. It is a U-shaped network based on a Fully Convolutional Network (FCN), which adopts the symmetrical "Encoder-Decoder" idea. First, the encoder captures the context and feature of the input image, and reduces the size of the feature map. Second, the decoder performs upsampling operation to restore the details of the image lost in the encoder. The U-Net model can propagate context information to higher resolution layers through a quantity of feature channels in the up-sampled part. Its end-to-end architecture allows the input of raw data to obtain segmentation map as the final output.

3.2.2 SegNet

The SegNet neural network structure, which was proposed in 2017, is a deep fully convolutional neural network architecture for semantic pixel-level segmentation (Badrinarayanan et al., 2017). Its encoder structure is the same as the 13 convolutional layers of VGG16 in the topological structure, and the decoder part is symmetrical with the encoder part.

3.2.3 RefineNet

The RefineNet neural network structure, which was proposed in 2017, is a generic multi-path refinement network that explicitly exploits all the information available along the down-sampling process to enable high-resolution prediction using long-range residual connections (Lin et al., 2017). The deeper layers that capture high-level semantic features can be directly refined using fine-grained features from earlier convolutions. A chained residual pooling is also introduced which captures rich background context in an efficient manner.

3.2.4 DeepLab v3+

The DeepLab v3+ neural network structure was proposed in 2018 (Chen et al., 2018). It combines the aforementioned two network structures, so that the encoding and decoding structure can gradually restore spatial information to obtain clearer edge details, and the dilated convolutional pyramid pooling (ASPP) can encode multi-scale contextual information, thus combining the advantages of these two structures. Its backbone can be ResNet, MobileNet and Xception.

3.2.5 DenseASPP

The DenseASPP neural network structure was proposed in 2018 (Yang et al., 2018). It employs DenseNet as the backbone network to connect a set of atrous convolutional layers in a dense way, such that it generates multi-scale features that not only cover a larger scale range, but also cover that scale range densely, without significantly increasing the model size.

3.3 Our Semantic Image Segmentation Network

Since U-Net has the most advantage among the five potential semantic segmentation neural networks selected and tested in the semantic coral image segmentation experiments, this paper selects the U-Net network as the object of fine-tuning improvement. The accuracy comparison of these five semantic image segmentation networks will be displayed in detail in the subsequent analysis of experimental results in this paper.

In fact, the live *Pocillopora* genus and dead ones have diverse morphological characteristics: living corals have high lustre and full of shape. Although the dead ones maintain their original branches, they are completely bleached, or are eroded by sea water and coral-eating organisms, left wormholes and rust spots. There will be some algae attached to the surface as well. Both states of dead corals have the problem of low contrast with the surrounding background, making their boundaries difficult to identify and segment.

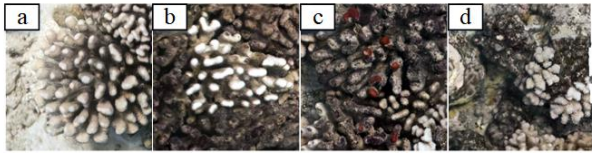


Figure 5. Different states of corals. (a) living, (b) dead with completely bleached, (c) dead with wormholes and rust spots, (d) dead covered in algae.

The RM (Refinement Module) is generally designed as a residual block, applied to the field of salient object detection, to learn the residual between the saliency map and the ground truth, attempting to refine the coarse saliency map (Islam et al., 2017; Wang et al., 2017; Deng et al., 2018). However, residual refinement module was originally proposed for semantic segmentation to refine the object boundaries (Qin et al., 2019). To refine boundary drawbacks of corals in the feature map, this paper introduces a Residual Refinement Module (Peng et al., 2017). Its main architecture contains an input layer, an encoder, a bridge, a decoder and an output layer, which is deeper than those above-mentioned modules. For the fine-tuning model we adopt is to add a residual refinement module after the output of U-Net network.

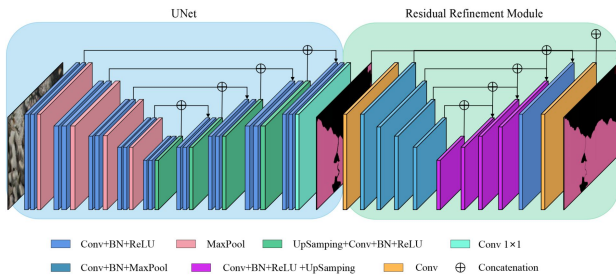


Figure 6. Architecture of our semantic segmentation network: the combination of U-Net and residual refinement module.

Loss functions commonly used in neural network classification and segmentation, such as cross-entropy or dice, are all region-based loss. In order to provide complementary information to the regional loss and obtain a clear boundary, we employ l as a hybrid loss function:

$$l = l_{ce} + l_b \quad (1)$$

where l_{ce} and l_b indicate Cross-Entropy loss and Boundary loss respectively (Kervadec, 2021).

Cross-Entropy loss is widely used in classification and segmentation, and its expression is:

$$l_{ce} = -\sum_{i=1}^N [G(i) \log S(i) + (1 - G(i)) \log(1 - S(i))] \quad (2)$$

Where $G(i) \in \{0, 1\}$ is the ground truth label, $S(i)$ is the predicted probability and N is the total number of classes.

Boundary loss is proposed for the problems associated with regional losses in highly unbalanced segmentation. As an integral approach for computing boundary variations, this paper adds it into the loss function to learn boundary information of

target corals. ∂G indicates the boundary of ground truth region G . $D_G(q)$ calculates the distance between point $q \in \Omega$ and the nearest point ∂G on contour. It is defined as:

$$l_b(\theta) = \int_{\Omega} \Phi_G(q) s_{\theta}(q) dq \quad (3)$$

where Φ_G denotes the level set representation of boundary ∂G : $\Phi_G(q) = D_G(q)$ if $q \in G$, otherwise $\Phi_G(q) = -D_G(q)$. $s_{\theta}(q)$ represents the Softmax probability outputs of the network.

The curve of cross-entropy loss will tend to be flat and the learning speed slows down due to the log characteristics. On the contrary, the boundary precision needs more attention at the later stage of the training process. The combination of standard regional loss and boundary loss ensures that the network model is able to learn enough information from different directions.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Implementation and Evaluation Metrics

This paper employed our fine-tuning network architecture without pre-trained because of weak visual similarity between the coral classes and the dataset used to train the original network. To train our model, the optimizer is Adam, with an initial learning rate of 10^{-4} and the batch size equal to 8. The implementation of our network is based on the following framework: Pytorch 1.9.0 (Paszke et al., 2019). This paper conducted the experiments on the device configured with an eight-core PC with an i7-7820HK 2.9 GHz CPU (with 32GB RAM) and a GTX 1070 GPU (with 8GB memory) for training, validating and testing.

For semantic image segmentation, in case of multiple classes, the mIoU has to be calculated rather than just calculating IoU by treating all the different classes as a single one. Moreover, the standard metric of accuracy is mIoU, which is the most commonly used methods to measure semantic image segmentation (Yang et al., 2018). It calculates (intersection ratio of true label and predicted result) IoU for each class separately, and then averages the IoU of all classes, and then averages the IoU of all classes. The expression of IoU is:

$$IoU = \frac{|G \cap P|}{|G \cup P|} \quad (4)$$

where G is ground truth, P is prediction.

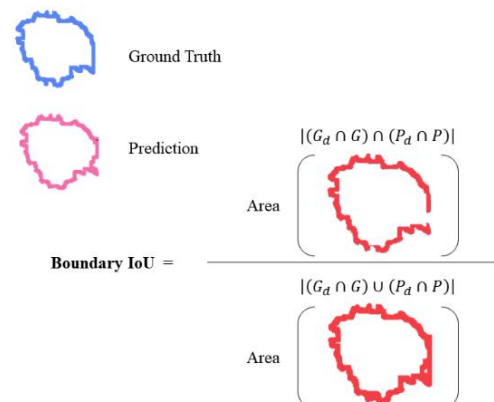


Figure 7. Boundary IoU computation illustration.

Boundary IoU measures errors from predicted and ground truth boundaries, which calculates the set of intersection-over-union for mask pixels within a certain distance from the corresponding ground truth or prediction boundary contours (Cheng et al., 2021). We introduce this metric to observe the boundary quality of coral objects. G_d and P_d are defined as set of pixels in the

boundary region of the binary mask. For multi-class Boundary IoU, it is necessary to first calculate the mask IoU of each class, and then take the average:

$$IoU_{boundary} = \frac{|(G_d \cap G) \cap (P_d \cap P)|}{|(G_d \cap G) \cup (P_d \cap P)|} \quad (5)$$

It should be pointed out that we did not add the error of boundary prediction and ground truth into the loss function optimization, which is only introduced as a new measure of experimental results.

4.2 Analysis of the Coral Segmentation Results

We first compared five kinds of image semantic segmentation convolutional neural networks. In the coral semantic segmentation experiments, we trained the model through 2019 coral images annotated by experts. The labels contained a total of 3 categories, namely live *Pocillopora* coral, dead *Pocillopora* coral and background. Then we compared the proposed semantic image segmentation network with the state of the art. Through field investigations, we found that *Pocillopora* coral dominates the study area, and it is also one of the main coral species gnawed during the outbreak of *Acanthaster planci*, which is of great significance to its dynamic monitoring and research. In our dataset, a large number of corals bleached or died, with live *Pocillopora* coral, dead *Pocillopora* coral and Background per-pixel frequencies of 13.38%, 13.87%, and 72.75%, respectively, with enough samples to avoid the class imbalance problem.

4.2.1 Performance Comparison of Different Test CNNs

The results are evaluated on the training set, validation set, and testing set of our coral images, and listed in Table 1, 2, and 3. IoU0, IoU1, and IoU2 denotes the background, the live coral and the dead coral respectively.

Methods	mIoU	IoU0	IoU1	IoU2
RefineNet	87.6%	92.8%	88.3%	80.8%
SegNet	93.7%	95.9%	93.6%	91.1%
DenseASPP	93.0%	94.7%	92.5%	91.2%
Deeplab V3+	95.0%	96.4%	94.7%	93.5%
U-Net	96.3%	97.7%	95.9%	94.9%

Table 1. Performance comparison on the training dataset.

Methods	mIoU	IoU0	IoU1	IoU2
RefineNet	87.7%	92.9%	87.7%	82.4%
SegNet	93.8%	96.3%	93.4%	91.7%
DenseASPP	91.4%	94.9%	90.6%	88.7%
Deeplab V3+	92.9%	96.0%	92.2%	90.5%
U-Net	95.5%	97.5%	94.4%	94.6%

Table 2. Performance comparison on the Validation dataset.

Methods	mIoU	IoU0	IoU1	IoU2
RefineNet	81.9%	89.1%	83.2%	72.5%
SegNet	89.8%	93.9%	90.2%	84.5%
DenseASPP	88.5%	92.5%	88.9%	83.5%
Deeplab V3+	89.8%	93.8%	90.2%	84.6%
U-Net	93.2%	96.5%	93.4%	89.1%

Table 3. Performance comparison on the test dataset.

In deep learning, the neural network model is initially fit on a training dataset first, which is a dataset of examples used during the learning process and is used to fit the parameters (e.g., weights) of, for example, a classifier. Successively, the fitted model is used to predict the responses for the observations in a second dataset called the validation data set, which provides an unbiased evaluation of a model fit on the training dataset while tuning the model's hyperparameters (i.e., the architecture) of a

classifier. Finally, the test dataset is a dataset used to provide an unbiased evaluation of a final model fit on the training dataset, which is a dataset that is independent of the training dataset, but that follows the same probability distribution as the training data set. In this paper, our original coral reef image dataset was partitioned into three subsets.

The results are illustrated in Table 1, Table 2, Table 3, and Figure 8. U-Net achieves the highest accuracy on the three sets. However, its performance in the boundary segmentation of coral individuals is not stable, and the boundary segmentation errors and incompleteness still exist. Since U-Net has the most advantages in the three-classification problem of corals, and it is also a lightweight semantic image segmentation model, we propose our own semantic segmentation network for the coral segmentation problem, based on the U-Net.

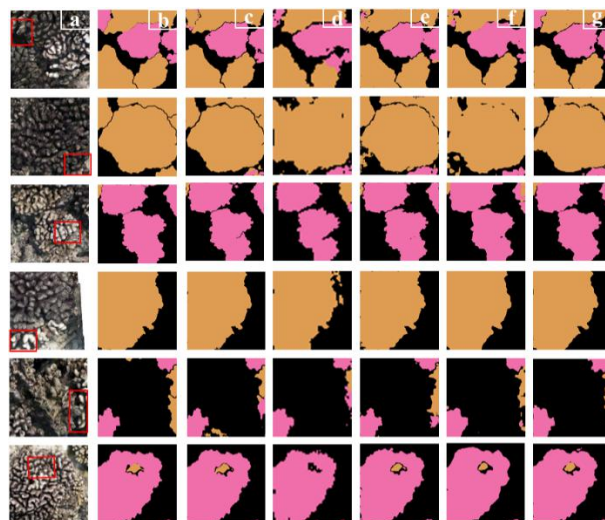


Figure 8. Comparison of the five CNNs results on the test set. (a) original coral reef images, (b) ground truth, (c) U-Net, (d) RefinNet, (e) SegNet, (f) DenseASPP, (g) Deeplab V3+.

4.2.2 Comparison of Ours and U-Net

As shown in Table 4 and Figure 9, our model performs better than the U-Net network models we listed in terms of results.

Methods	mIoU	IoU0	IoU1	IoU2
U-Net	93.2%	96.5%	93.4%	89.1%
Ours	93.6%	96.1%	93.6%	90.6%

Table 4. Performance comparison on the test dataset.

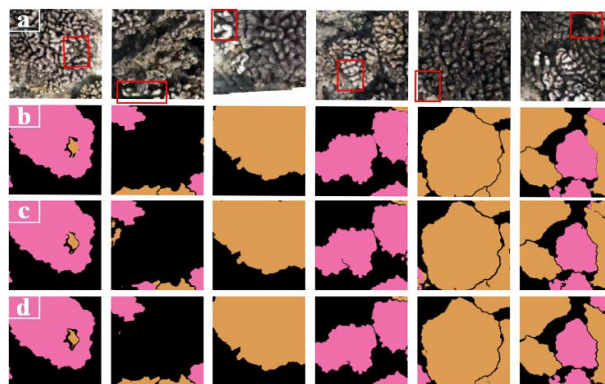


Figure 9. Comparison of our results on the test set. (a) original coral reef images, (b) ground truth, (c) U-Net, (d) ours.

Specifically, it can be seen that our improved method based on U-Net can segment more fine and clear coral individual boundaries, and our method also has a small improvement in

mIoU compared to the U-Net method, about 0.4%. These improvements have higher accuracy on each class except background, increasing by 0.2% (live coral) and 1.5% (dead coral) respectively, which demonstrates our improvements are more effective on dead corals. Yet no matter what kind of neural network, the accuracy of dead coral is about 4% lower than the live coral, and this is not only attributed to the different death states or forms of corals, but also associate with their bleaching process. It is a continuous process, for example high temperatures make corals lose their symbiont pigmentation, and becoming pale, but corals can recover if returned to optimal conditions (Jokiel and Coles, 1977). Therefore, how to define whether the corals are in an irreversible state or a sublethal state is very important, and it affects the accuracy of recognition to a certain degree. After all, the two are difficult to distinguish from a visual perspective. Furthermore, we also find that if the coral is located at the edge of the image tile, the semantic segmentation network obtains too little information and the performance of all models is not stable enough, although our proposed improved method performs better overall.

Boundary IoU can more accurately evaluate the segmentation quality of coral boundaries (Cheng et al., 2021). In addition to the common mask IoU evaluation metric, we introduce it to further evaluate our experimental results. Results of Boundary IoU on the test dataset are shown in Table 5 and Figure 10. It can be seen that our neural network model improves the boundaries of live coral and dead coral, especially the dead. The comparison results of the new Boundary IoU evaluation metric can further demonstrate the improvement of our proposed instance semantic segmentation network for coral boundary segmentation. Since the boundary of the background class has no practical significance in this paper, we do not calculate its Boundary IoU here.

Methods	Boundary mIoU	Boundary IoU1	Boundary IoU2
U-Net	82.2%	84.0%	79.8%
Ours	82.4%	84.3%	80.8%

Table 5. Boundary IoU comparison on the test dataset.

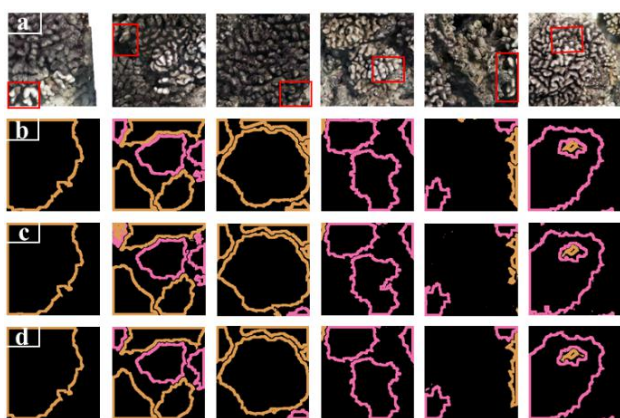


Figure 10. Comparison of our boundary results on the test set. (a) original coral reef images, (b) ground truth, (c) U-Net, (d) ours.

5. CONCLUSION

This work integrated underwater photogrammetry and deep learning to study the spatial distribution of *Pocillopora* coral populations. We propose an improved boundary-oriented U-Net model, to automatically identify and segment coral individuals from orthophoto. In addition, the network we optimized can accurately distinguish live *Pocillopora* and dead *Pocillopora*

corals, which could assess their health status. The main work includes: (1) A reasonable method to partition a whole marine coral reef map into datasets. (2) Analysis of different semantic image segmentation networks and transferability of the model on coral reef images. The potential superior semantic image segmentation network for coral semantic segmentation is screened out through experiments. (3) An improved U-Net network architecture focused on boundaries and demonstrates its superior performance. Through the research of this paper, the following aspects must be pointed out: One difficulty of coral image segmentation is that the edges of corals are relatively irregular and the information is complicated to extract. Whether it is a live coral or a dead coral, the wrong pixels basically fall on the edge of the coral, and the details are not accurate enough. Therefore, the semantic segmentation and extraction of dead corals has always been a difficult task in the current large number of coral semantic segmentation studies. Many papers often avoid discussing the accuracy of dead coral segmentation. Hence, this study can effectively improve the recognition rate of dead corals under the premise of only a slight decline in other indicators, which also illustrates the advantages and potential of this method in coral semantic segmentation. Of course, our experiment is only tested in the same study area, without considering the situation of cross-year or cross-plot in this paper. In future work, we will further conduct experiments on multiple years or multiple plots to verify the generalization ability of our network model.

ACKNOWLEDGEMENTS

The authors thank the research project Moorea IDEA (Island Digital Ecosystem Avatars) for support in underwater coral reef images obtained. Thanks to the Undergraduate on Innovation and Entrepreneurship Training Program (S202110486216, S202210486307) of Wuhan University, China. We also would like to acknowledge ETH Zurich, Switzerland for supporting this research.

REFERENCES

- Adjeroud, M., 1997: Factors influencing spatial patterns on coral reefs around Moorea, French Polynesia. *Marine Ecology Progress Series*, 159, 105-119.
- Alonso, I., Cambra, A., Munoz, A., Treibitz, T., Murillo, C., 2017: Coral-segmentation: Training dense labelling models with sparse ground truth. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2874-2882.
- Alonso, I., Yuval, M., Eyal, G., Treibitz, T., Murillo, C., 2019: CoralSeg: Learning coral segmentation from sparse annotations. *Journal of Robotic system*, 36(8), 1456-1477.
- Bourque, B., Bradbury, R., Cooke, R., Erlandson, J., Estes, J., Estes, J., Hughes, T., Kidwell, S., Lange, C., Lenihan, H., Pandolfi, J., Peterson, C., Steneck, R., Tegner, M., Warner, R., 2001: Historical overfishing and the recent collapse of coastal ecosystems. *Science*. 293(5530), 629-637.
- Boesch, D., Burreson, E., Dennison W., Houde, E., Kemp, M., Kennedy, V., Paynter, K., Orth, R., Ulanowicz, R., 2001: Factors in the decline of coastal ecosystems. *Science*, 293(5535), 1589-1591.
- Beijbom, O., Edmunds, J., Kline, I., Mitchell, G., Kriegman, D., 2012: Automated annotation of coral reef survey images. *IEEE*

- conference on computer vision and pattern recognition, 1170-1177.
- Beijbom, O., Treibitz, T., Kline, I., Eyal, G., Khen, A., Neal, B., Loya, Y., Mitchell, G., Kriegman, D., 2016: Improving automated annotation of benthic survey images using wide-band fluorescence. *Scientific reports*, 6(1), 1-11.
- Badrinarayanan, V., Kendall, A., and Cipolla, R., 2017: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- Cressey, D., 2015: Tropical paradise inspires virtual ecology lab. *Nature*, 517, 255–256.
- Chen, C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018: Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision*, 801-818.
- Cai, R., Guo, H., Amro, A., 2021: A study on the adaptation and restoration of warm water coral reef ecosystem in the context of global change. *Journal of Applied Oceanography*. 40(01),12–25.
- Cheng, B., Girshick, R., Dollár, P., Berg, A. C., Kirillov, A., 2021: Boundary IoU: Improving object-centric image segmentation evaluation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15334-15342.
- Deng, Z., Hu, X., Zhu, L., Xu, X., Qin, J., Han, G., Heng, P., 2018: R³net: Recurrent residual refinement network for saliency detection. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 684-690.
- Faurea, G., 1989: Degradation of Coral Reefs at Moorea Island (French Polynesia) by *Acanthaster planci*. *Journal of Coastal Research*, , 5(2), 295–305.
- Feeney, W., Bertucci, F., Gairin, E., Siu, G., Waqalevu, V., Antoine, M., Lecchini, D., 2021. Long term relationship between farming damselfish, predators, competitors and benthic habitat on coral reefs of Moorea Island. *Scientific Reports*, 11(1), 1-8.
- Guo, T., Capra, A., Troyer, M., Gruen, A., Brooks, A.J., Hensch, R.J., Schmitt, S.J., Holbrook, M., Dubbini, M., 2016: Accuracy Assessment of Underwater Photogrammetric Three Dimensional Modelling for Coral Reefs. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B5, 12–19.
- Gruen, A., Guo, T., Ural S., Troyer, M., 2017: DSM/DTM-related investigations of the Moorea Avatar project. *Asian Conference on Remote Sensing*, New Delhi, India, 1-10.
- Islam, A., Kalash, M., Rochan, M., Bruce, N., Wang, Y., 2017: Salient Object Detection using a Context-Aware Refinement Network. *British Machine Vision Conference*, 1-12.
- Jackson, J., Kirby, M., Berger, W., Bjorndal, K., Botsford, L., Hughes, P., Baird, A., Bellwood, R., Card, M., Connolly, S., Folke, C., Grosberg, R., Guldberg, O., Jackson, J., Kleypas, J., Lough, J., Marshall, P., Nystrom, M., Palumbi, S., Pandolfi, J., Rosen, B., Roughgarden, J., 2003: Climate change, human impacts, and the resilience of coral reefs. *Science*, 301(5635), 929-933.
- Jokiel, L., Coles, L., 1977: Effects of temperature on the mortality and growth of Hawaiian reef corals. *Marine Biology*, 43(3), 201-208.
- Knowlton, N., 2001: The future of coral reefs. *Proceedings of the National Academy of Sciences*. 98(10), 5419-5425.
- Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ayed, B., 2021: Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67, 285-296.
- Lin, G., Milan, A., Shen, C., and Reid, I., 2017: RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 5168-5177.
- Mizuno, K., Terayama, K., Hagino, S., Tabeta, S., Sakamoto, S., Ogawa, T., Fukami, H., 2020: An efficient coral survey method based on a large-scale 3-D structure model obtained by Speedy Sea Scanner and U-Net segmentation. *Scientific reports*, 10(1), 1-11.
- Ninio, R., Delean, S., Osborne, K., Sweatman, H., 2003: Estimating cover of benthic organisms from underwater video images: variability associated with multiple observers. *Marine Ecology Progress Series*, 265, 107-116.
- Nocerino, E., Menna, F., Gruen, A., Troyer, M., Capra, A., Castagnetti, C., Rossi, P., Brooks, J., Schmitt, J., Holbrook, J., 2020: Coral Reef Monitoring by Scuba Divers Using Underwater Photogrammetry and Geodetic Surveying. *Remote Sensing*, 12(18), 3036.
- Pavoni, G., Corsini, M., Callieri, M., Palma, m., Scopigno, r., 2019: Semantic Segmentation of Benthic Communities from Ortho-Mosaic Maps. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W10, 151-158.
- Pavoni, G., Corsini, M., Cignoni, P., 2020: A State of the Art Technology in Large Scale Underwater Monitoring, *ERCIM news*, 121, 17-18.
- Pavoni, G., Corsini, M., Pedersen, N., Petrovic, V., Cignon, p., 2021: Challenges in the deep learning-based semantic segmentation of benthic communities from Ortho-image. *Applied Geomatics*1, 13(1), 131-146.
- Peng, C., Zhang, X., Yu, G., Luo, g., Sun, j., 2017: Large kernel matters--improve semantic segmentation by global convolutional network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4353-4361.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Chintala, S., 2019: Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 8026-8037.
- Qin, X., Zhang, Z., Huang, C., Gao, C., Jagersand, M., 2019: Basnet: Boundary-aware salient object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7479-7489.
- Raj, D., Aeby, S., Mathews, G., Williams, G., Edward, J., 2021: Coral reef resilience differs among islands within the Gulf of Mannar, southeast India, following successive coral bleaching events. *Coral Reefs*. 40, 1029–1044.

Ronneberger, O., Fischer, P., Brox, T., 2015: U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, Cham, 234-241.

Wang, T., Borji, A., Zhang, L., Zhang, P., Lu, H., 2017: A stagewise refinement model for detecting salient objects in images. *Proceedings of the IEEE International Conference on Computer Vision*, 4019-4028.

Yang, M., Yu, K., Zhang, C., Yang, K. 2018: DenseASPP for semantic segmentation in street scenes. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3684-3692.

Yuval, M., Alonso, I., Eyal, G., Tchernov, D., Loya, Y., Murillo, A. C., Treibitz, T., 2021: Repeatable semantic reef-mapping through photogrammetry and label-augmentation. *Remote Sensing*, 13(4), 659.