CNet: A Novel Seabed Coral Reef Image Segmentation Approach Based on Deep Learning

Hanqi Zhang
WHU
hqzhang@whu.edu.cn

Ming Li*
WHU, ETH Zurich
mingli39@ethz.ch

Jiageng Zhong
WHU
zhongjiageng@whu.edu.cn

Jiangying Qin
WHU
jy_qin@whu.edu.cn

Abstract

Achieving underwater coral seabed image segmentation involves dividing an image into meaningful regions or segments, which, in this case, could represent different types of corals, substrate, or other features in the underwater habitat. We introduce an innovative network architecture, CNet, designed for the segmentation of coral seabed images. This architecture incorporates a three-branch parallel encoder structure, employing an RGB encoder based on the ResNet block, a Depth encoder based on the VGG block, and a ShapeConv block-based Fusion encoder. The study conducts comprehensive performance comparisons and ablation experiments to evaluate the efficacy of CNet in comparison to state-of-the-art (SOTA) methods. The results demonstrate an impressive mIoU of 81.83% on the coral dataset, with the IoU of the minority class, Acropora, reaching 73.61%. This is of crucial significance in the fields of marine biology and environmental monitoring, playing a pivotal role in the comprehensive understanding of coral reef ecosystems. By automatically and accurately identifying different coral classes, scientists can gain insights into threatened corals and their growth in different environments, providing crucial data for developing targeted conservation plans to promote coral recovery.

1. Introduction

Coral reefs, known as the "rainforests of the ocean", are a vital part of the marine ecosystem, providing refuge for approximately 25% of ocean fish species and bestowing significant economic and cultural benefits upon coastal global communities [8]. However, human activities such as coastal development, land reclamation, and overfishing have led to the disappearance of about 14% of coral reefs [32]. With the development of underwater remote sensing [7, 25], the long-term monitoring and tracking of coral reefs growth and health status through underwater coral seabed visual images have become imperative for predicting early warnings, evaluating adverse events, and ensuring ecosystem resilience. This monitoring process involves the analysis and extraction of valuable information from extensive underwater image data to evaluate various key reef indicators, including coral species, area, and abundance. Unlike satellite and aerial photogrammetry and remote sensing technologies vulnerable to adverse weather and air-water interface effects, underwater videography enables large-scale detailed imaging with the benefits of high imaging resolution and low cost, thus enabling granular coral monitoring [11]. Despite addressing the challenges of fine data collection, automating the process of analyzing high spatial resolution images presents its own set of complexities. Manual annotation of corals is a time-consuming task for machine learning-based image segmentation prone to annotation errors over time and among individuals due to the repetitive nature of labeling [22]. Therefore, quantifying coral growth and degradation, as well as identifying coral individuals, urgent needs advanced accurate automated image processing tools.

To meet this need, we propose an automated processing method for multimodal data that integrates effective data partitioning techniques and novel deep learning-based neural network structure. It first performs reasonable dataset partitioning from limited underwater remote sensing images and then uses a designed multi-modal deep neural network for fast and accurate coral segmentation. This method can predict and analyze multi-year coral maps. Specifically, our research focusses on the application of the multi-modalities semantic segmentation to coral orthophotos, a conventional data product extensively used by terrestrial ecological monitoring scientists for extracting ecological information from remote sensing imagery [21]. Orthophotos provide a top-down perspective and accurate geographical reference, facilitating the generation of coral coverage maps. Our proposed neural network, which integrates high-resolution orthophotos and corresponding depth information, can more accurately identify and classify corals at the genus level, making it reliable for dynamic coral reef monitoring tasks

[*] Corresponding author

in in situ survey.

2. Related work

Accurate coral monitoring is essential for documenting and comprehending the long-term dynamics in coral reef decline and recovery on a spatial scale through a variety of meta-analysis integrating different data sources. The widely adopted metric for assessing coral reef condition, such as the estimation of "percent live coral cover", can be efficiently and rapidly computed using photogrammetric computer vision and deep learning-based image segmentation algorithms [18, 29]. Despite the capability of underwater remote sensing technology to cover extensive coral reef substrates, the heterogeneous nature of corals necessitates the involvement of knowledgeable experts for initial data annotation, resulting in limitations in the dataset available for deep learning-based coral monitoring tasks. In dealing with remote sensing images for training of deep neural network model, researchers employ fixed-stride sliding windows [4, 28] and Poisson sampling [20] to effectively utilize limited data. Compared to sliding windows, Poisson sampling offers the advantage of controlling the number of patches of both the majority and minority classes through the manipulation of the sampling radius, thereby achieving category balance. In addition, the high heterogeneity of coral reef systems [13] presents a challenge in providing a comprehensive description of these organisms using singular features of a specific type, hierarchy, or scale. Consequently, there is an urgent need to establish generalized features that can effectively describe corals in their underwater environments. Thus, careful design, selection, and optimization of suitable network models hold significant importance. For conventional network models in other fields, several network architectures such as DeepLab v3+, U-Net, and MultiResUNet, have been successfully applied to underwater coral image segmentation [10, 17, 20]. Researchers often introduce new modules or structures to enhance these existing networks, aiming to improve the accuracy and performance of segmentation models when dealing with complex objects such as corals [15, 31, 35, 36]. These techniques facilitate the fusion of fine-grained features, thereby enhancing the accuracy and performance of segmentation models when dealing with complex objects like corals. Moreover, SUIM-Net [14], tailored specifically for underwater image segmentation tasks, includes a core component called Residual Skip Block (RSB). This feature enables optional hierarchical skip connections within its core building blocks, effectively capturing contextual information. In addition to model optimization, techniques such as transfer learning and ensemble networks find extensive usage in this domain [20, 33].These collective efforts and advancements lead to highly precise segmentation analysis of coral images acquired through underwater measurement methodologies, thereby opening up new avenues for the study and comprehension of complex coral ecosystems.

3. Method

3.1. Data collection & pre-processing

We concentrate on a dataset comprising human-labeled orthophotos provided by the Moorea IDEA (https://mooreaidea.ethz.ch) project, encompassing coral imagery from two monitoring sites on Moorea Island. The dataset spans a three-year period and was taken around August. The camera system was positioned approximately 2 meters from the reef, enabling clear identification of individual coral colonies and intricate coral branches within the images. To minimize the influence of the underwater lighting environment on image quality, we conduct radiation correction on the acquired images. Specifically, for chromatic aberration, we employ Adobe Camera Raw for precise color correction. Additionally, the Wiener filter algorithm [24] is applied to resolve image quality issues stemming from motion errors. Subsequently, we utilize the photogrammetry program in Agisoft Metashape software to complete key processes such as marking control points, image matching, and the generation of dense point clouds. This process culminates in the production of high-resolution (1mm) orthophoto and depth image products. The manual annotation process is overseen by experienced professionals, augmented by the application of Taglab [23] tools for semi-automatic annotation. This dual approach ensures the acquisition of high-quality pixel-level annotation data, critical for our coral image semantic segmentation analysis.

3.2. Dataset partition

The dominant coral classes on Moorea Island primarily consist of the Pocillopora genus, followed by the Acropora genus, totally constituting 85% of our coral monitoring area. Considering that deep learning models require sufficient data for model training and parameter optimization, we have focused on these two dominant genera, setting aside the consideration of other less common coral genera.

Despite this, the substantial numerical disparity between these two coral genera results in a pronounced imbalance in their representation. This imbalance can lead to biases in the train progress of deep learning models, particularly in capturing minority classes. The current basic and commonly used sliding window extraction technique, which involves employing a fixed-stride sliding window to segment images into training, validation, and test datasets. However, this straightforward implementation method proves insufficient in handling class imbalance and often falls short in providing adequate contextual information through smaller image patches. On the other hand, the dataset partition method based on Poisson disk sampling controls the ra-

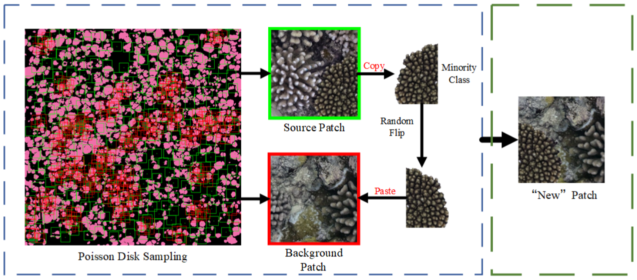

Figure 1. Improved Copy-paste partition methods. (1) Utilize Poisson disk sampling to partition the background patch (shown in the green box) and the source patch containing minority classes (shown in the red box); (2) Implement the random flipping operation to merge the minority class objects from the source patches onto the background patches.

dius of the sampling disk to sample the image multiple times without excessive repetition. Nonetheless, the resulting cropped image retains the same spatial distribution in the overlapping areas, thus limiting the capacity to enhance feature representation. Taking inspiration from the concept of copy-paste [9], a simple mechanism for randomly pasting objects, we design a copy-paste for class imbalance technique, as shown in Figure 1. Specifically, we utilize the image obtained through the Poisson disk sampling method as the background and randomly paste only minority class objects. This approach not only facilitate the segmentation of the limited remote sensing data into an adequate number of sections but also ensure class balance by altering the spatial distribution of various corals.

3.3. Network Structure

We propose CNet, a model designed for coral image semantic segmentation with two symmetric input branches and a fused input branch, which is based on ACNet [5]. The proposed network structure is depicted in Figure 2. We employ a pretrained ResNet50 encoder [12] for the RGB branch, while the VGG encoder [30] is utilized for the depth branch. The fused input branch also employs a pretrained ResNet50 encoder. In this branch, the 3×3 convolution used for feature extraction in the Bottleneck is replaced by ShapeConv [1]. We utilize this branch to integrate and refine semantic information from the two symmetric input branches. In the RGB branch, the original ResNet50 model is used for RGB images as the baseline network with five encoding modules. In comparison, depth images typically contain less feature information. However, they are adept at extracting detailed edge contour information, given their representation of the depth value for each pixel in the RGB image. In the context of semantic segmentation, when the encoder has a deeper number of layers in the network, there is a higher likelihood of losing edge contour information due to the lack of high-level features within the edge contour information. Thus, for the depth images, we advocate the application of a shallow encoder to extract auxiliary feature information, thereby preventing the loss of valuable coral edge contour information. We replace the Bottleneck block in ResNet50 with the shallower VGG block in the depth branch. Based on the original Bottleneck block, we substitute the 3×3 convolution in the middle with ShapeConv. ShapeConv represents a distinctive convolutional layer that dynamically adjusts the weights of convolution kernels based on the shape information of the input data, facilitating the seamless integration of shape features from RGB images and depth images. This modification empowers the network to effectively capture the intricate morphological structure and fine-grained textural patterns of corals without a substantial increase in computational overhead. Finally, we integrate the feature information output from the three branches and employ convolutional layers to optimize the fused features. Subsequently, we utilize four transposed convolution blocks to restore the spatial resolution of the optimized feature maps to match the size of the RGB image. Drawing inspiration from UNet's skip connections [26], we connect the feature map output of each transposed convolution block with the output of the corresponding layer in the encoder through a convolutional layer. This process aims to leverage the underlying low-level features, which is crucial for the restoration of detailed information within the image. The final layer of each decoder stage is supervised by ground truth.

To further enhance the model's performance, we employ a hybrid loss function. This chosen approach combines a weighted cross-entropy loss function and an intersection-over-union (IoU) loss function [34]. The weighted cross-entropy loss function is tailored to assign greater weight to minority classes, thus effectively addressing the challenge of class imbalance. On the other hand, the IoU Loss function serves to measure the similarity between the predicted results and the actual results. Maximizing the IoU loss value encourages the model to produce more precise and accurate predictions. The equation for the weighted cross-entropy loss is defined as follows:

$$L_{wce} = -\frac{1}{N}\sum_{i=1}^{N}[\omega G(i)\log S(i) + (1 - G(i))\log(1 - S(i))], \tag{1}$$

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. IoU Loss is a metric learning method like Dice Loss, and the formula is defined as follows:

$$L_{iou} = 1 - \frac{X \cap Y}{X \cup Y}, \tag{2}$$

where X is the ground truth label, Y is the predicted probability map.

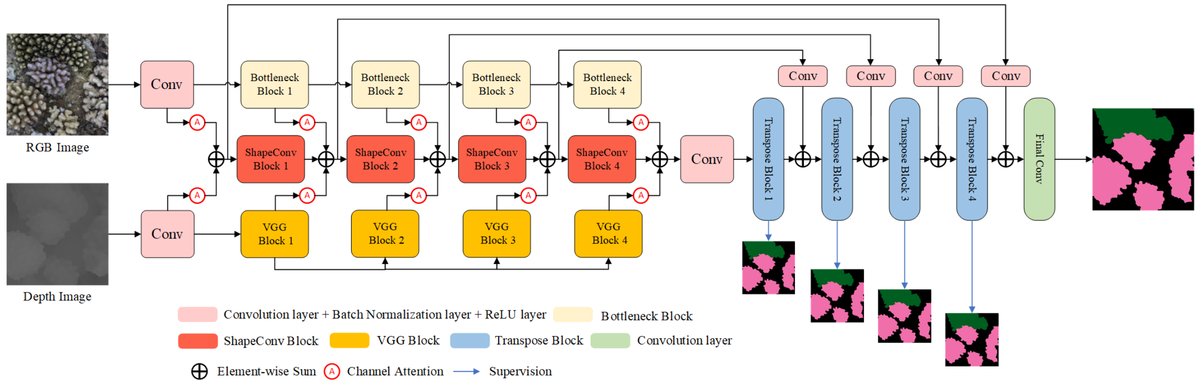

Figure 2. The network architecture of CNet.

Through the combination of these loss functions, we derive a comprehensive loss function that effectively mitigates class imbalance issues to achieve precise model predictions. In subsequent experiments, we employ various weighting coefficients to balance the impact of the two loss functions, with the optimal combination as follows:

$$L = L_{wce} + \lambda \cdot L_{iou}, \tag{3}$$

the value of λ is taken to be 0.4.

4. Experimental results and analysis

We first assess the improvement in network performance resulting from our proposed data partitioning method. Subsequently, we evaluate our method using our custom partitioned coral dataset, as outlined in Section 4.3. In Section 4.4, we conduct a series of ablation studies to confirm the effectiveness of component selection in our CNet.

4.1. Implementation details & Datasets

We trained our network using PyTorch [19] for 120 epochs with batches of size 16. Our optimization process involved using SGD with a momentum of 0.9 and Adam [16] with an initial learning rate of 1e-4, coupled with a minor weight decay of 5e-4. We adopted PyTorch's poly learning rate scheduler to dynamically adjust the learning rate. Moreover, to expand our training samples, we employed random scaling, translation, and flipping techniques for image augmentation. In the case of RGB images, we also implemented random adjustments for brightness, contrast, and blur. Our evaluation used three representative common metrics: average accuracy (mAcc), average union over intersection (mIoU), and F1 score (F1).

The underwater remote sensing images collected between 2017 and 2019 in area A are partitioned into 1230 coral RGB-D images of size 300*300. The training and validation datasets are randomly divided in a 4:1 ratio and the labels were set to three classes: background, Pocillopora, and Acropora. For the test dataset, we utilize underwater remote sensing images from area B in 2018 and 2019. The test dataset employs a basic sliding window division to facilitate the stitching of predicted patches, resulting in 5421 images. Despite the images in area A produce fewer patches, utilizing these patches for network training enable the network to learn the color and shape characteristics of corals across different years, which is beneficial to enhancing the generalization ability of the network. It should be pointed out that the mosaic orthophotos we used for data partitioning were obtained through photogrammetric processing using approximately 400 underwater coral seabed images with millimeter-level spatial resolution.

4.2. Results with different dataset partition methods

We apply the proposed data partitioning method to three-year orthophotos of area A, which were used for training the neural network. We sample the orthophotos to acquire non-overlapping background patches and multi-sampled source patches containing the minority class, Acropora. The number of oversampled patches depend on the coral's area and distribution. Subsequently, we randomly paste the Acropora objects from the source patches into the background patches. To illustrate the effectiveness of this sampling method, we compare it with both the sliding window and Poisson disk sampling methods. And we divide the orthophotos to yield approximately 1,200 image patches due to ensure fairness among these three methods. We count the pixel frequencies of each class in the coral remote sensing image dataset derived from the various division methods, as outlined in Table 1. In comparison with the sliding window and Poisson disk sampling, our improved copy-paste method achieved a more balanced class frequency in the dataset, resulting in the frequency of the minority class, Acropora, increasing from 1.47% to 9.33%. Moreover, the baseline network model, ACNet, trained on each orthophoto dataset, demonstrate that our data partitioning method facil-

Methods	BGCf	PCf	ACf	mIoU	BGIoU	PIoU	AIoU
Sliding Window	59.67%	38.86%	1.47%	75.58%	**88.73%**	**83.02%**	55.01%
Poisson Sampling	67.32%	29.73%	2.95%	80.36%	88.45%	82.81%	69.83%
Copy-paste (Ours)	56.56%	34.11%	9.33%	**80.77%**	88.70%	82.90%	**70.72%**

Table 1. Class frequencies (Cf) and network performance on training area using our over-sampling strategy and other methods. BG, P, and A respectively refer to background, Pocillopora, and Acropora.

itate the comprehensive comprehension of the characteristics of each class by the network. Notably, the minority class display a remarkable IoU of 70.72%.

4.3. Performance comparison with SOTA methods

In this section, we present a performance evaluation of the coral semantic segmentation network. Specifically, we compare our proposed underwater coral image segmentation method with SOTA networks such as SUIM-Net [14], DeepLab v3+ [2], UNet [26], SA-Gate [3], ESANet [27], and BBSNet [6]. Additionally, we compare our proposed network structure with our previous work [36]. All experiments are conducted using the same dataset and hyperparameter settings. For SUIM-Net, DeepLab v3+, and UNet, which lack depth map input channels, we integrate the RGB channel with the depth channel, generating the images with four channels for network input. Table 2 displays the segmentation performance achieved by our proposed CNet in comparison to the aforementioned methods. CNet demonstrates the highest overall performance, with mAcc, IoU, and F1-Score of 92.52%, 81.83%, and 89.87% respectively. Our proposed method exhibits superior accuracy when compared to other well-known methods designed solely for RGB images, such as DeepLab v3+ and UNet. Additionally, our fusion method outperforms existing RGB-D image fusion techniques, surpassing the performance of SA-Gate, ESANet, and BBSNet. It is worth noting that our network outperforms the previously designed improved DeepLab V3+, which involves the simple operation of replacing all vanilla convolutions with ShapeConv. We extract different features from RGB images and depth images, respectively, and then use a smaller but effective number of shape convolutions to fuse RGB and depth features, achieving better network performance without increasing computational overhead. Furthermore, the network performance of SUIM-Net, tailored specifically for underwater image segmentation, also falls short in comparison to our proposed CNet. These results indicate the effectiveness of our proposed improvements in fusing the multi-modal features of corals, thereby yielding higher quality predictive images segmentation results.

Methods	mAcc	mIoU	F1-score
SUIM-Net [14]	88.12%	72.29%	82.90%
DeepLab V3+ [2]	88.65%	72.84%	83.89%
Zhong et al. [36]	89.92%	73.30%	84.06%
UNet [26]	91.01%	77.12%	86.64%
SA-Gate [3]	91.43%	72.93%	83.33%
ESANet [27]	91.14%	77.07%	86.85%
BBSNet [6]	92.05%	78.68%	87.82%
CNet (ours)	**92.52%**	**81.83%**	**89.87%**

Table 2. Quantitative comparison results with SOTA methods.

Figure 3 depicts the visualization results acquired from various semantic segmentation network models. It is evident that the segmentation results produced by CNet exhibit finer boundaries and more precise coral outlines compared to other methods, which only roughly identify the overall coral outline. Moreover, distinguishing between dead and living corals is challenging due to their similar morphologies. However, CNet demonstrates the capability to effectively exclude dead corals, as evidenced in the third and sixth columns. It is essential to highlight a minor inconsistency in the sixth column, where the newly generated coral in the upper left corner lacks significant height. Consequently, the depth information in the associated depth image is not prominent, resulting in challenges for some network models in accurately recognizing it. Despite its exceptional performance, CNet is not completely immune to errors. For instance, in the first column of 2019, while most networks successfully identify the Acropora region, CNet erroneously categorizes it as Pocillopora.

4.4. Ablation study

Table 3 presents the results of the ablation study conducted on CNet, where it is compared with the baseline ACNet, and various components are gradually integrated to assess their impact on segmentation performance. ACNet is used as the baseline. The network, which incorporates the ShapeConv fusion branch, is labeled as S. Each encoder block of the depth branch is substituted with a VGG Block represent as V, while the hybrid loss function used is marked as H. The ablation experiments reveal that the baseline ACNet achieve an 80.77% mIoU, surpassing all SOTA networks compared in Section 4.3. Following the integration of ShapeConv to merge multi-model coral fea-

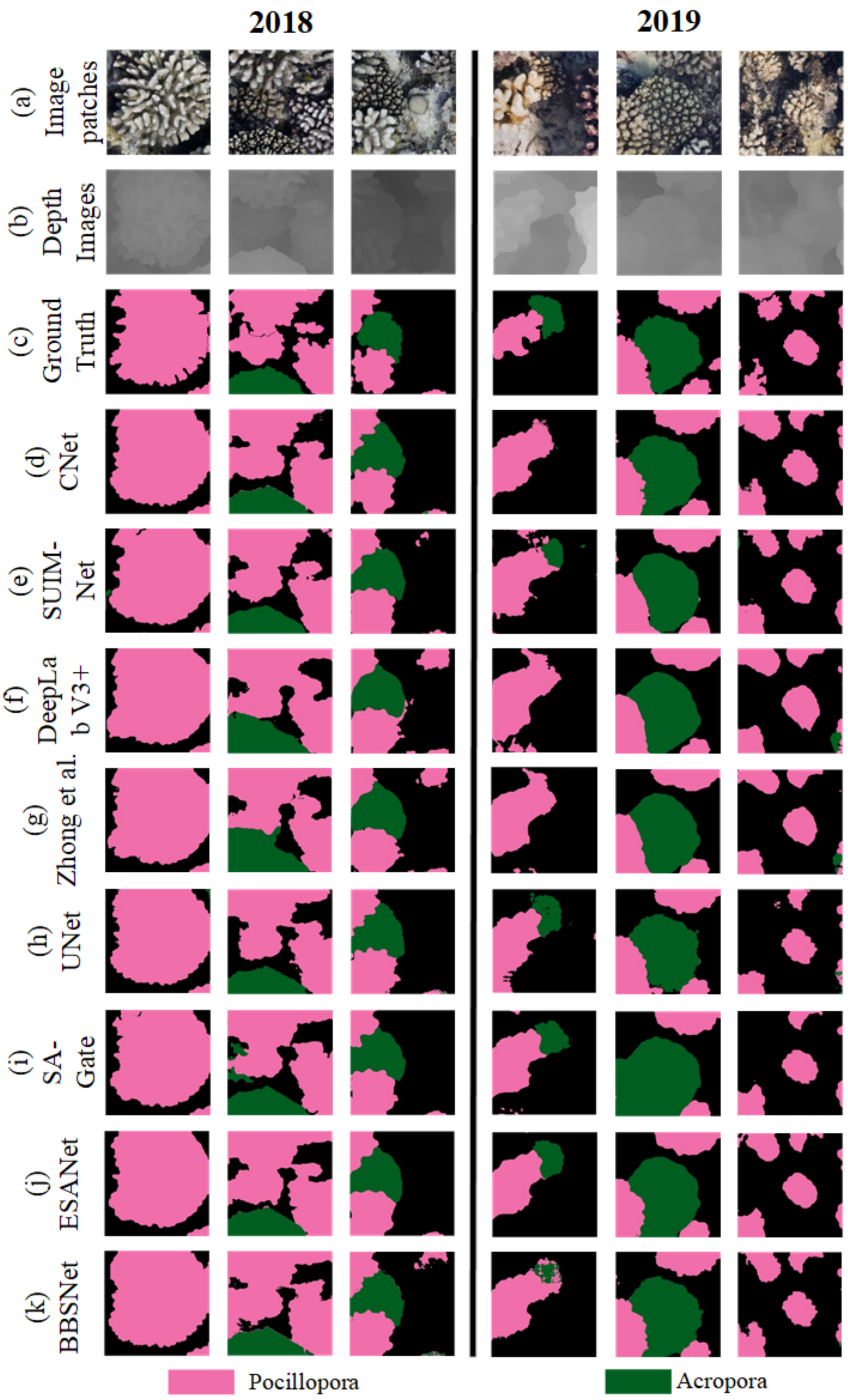

Figure 3. Examples of segmentation patches of CNet and other comparison methods. Masks of Pocillopora coral is in pink and Acropora coral is in green.

tures, the IoU of the minority class Acropora significantly increase by 1.76%, while the IoU of other classes showed a slight decline. This demonstrates the superior ability of ShapeConv to effectively capture and fuse local geometries encoded by deep features compared to traditional convolutions. Additionally, compared with the original ResNet50 encoder, employing shallow encoders to extract the morphological information of corals in depth images improves mIoU from 80.77% to 81.00%. This suggests that for depth information with fewer features, shallow encoding is more beneficial for feature extraction, although the improvement effect on minority classes is not substantial. The use of hybrid functions also enhances accuracy, especially on the minority class, reaching 72.94%. The combination of the depth branch based on the VGG block and the fusion branch adding ShapeConv elevates the IoU of the three classes to an impressive level, particularly for the background and Pocillopora, which experience a growth of approximately 1%, reaching 89.22% and 83.43%, respectively. The implementation of the hybrid loss function further boosts the mIoU to the highest 81.83%, significantly enhancing the accuracy of the minority class to 73.61%. This showcases how the IoU loss effectively adjusts the overall loss bias to achieve a balanced segmentation outcome. Our refinements result in a 1.06% mIoU improvement for the final CNet compared to baseline ACNet, without a substantial increase in parameters or additional computational overhead.

Methods	mIoU	BGIoU	PIoU	AIoU
Baseline	80.77%	88.70%	82.90%	70.72%
Baseline+S	81.12%	88.39%	82.48%	72.48%
Baseline+V	81.00%	88.78%	83.18%	71.02%
Baseline+H	81.32%	88.53%	82.51%	72.94%
Baseline+S+V	81.58%	89.22%	83.43%	72.08%
Baseline+S+V+H	81.83%	88.76%	83.10%	73.61%

Table 3. An ablation study conducted on CNet. BG, P, and A respectively refer to background, Pocillopora, and Acropora.

5. Discussions

The predicted map we generated demonstrates the alarming rate of coral mortality within a single year, emphasizing the critical urgency of implementing effective coral protection measures. While our expertise lies predominantly in the realm of photogrammetric computer vision, we recognize the multi-dimensional complexity of coral degradation within the broader context of marine biology and ecological dynamics in visual images. Utilizing quantitative analysis, we conduct a thorough examination of the pixel frequency within the two-year timeframe, serving as a rudimentary yet critical metric for assessing changes in coral coverage. Our findings indicate a substantial decline in the coverage of Pocillopora and Acropora by 53.29% and 71.40% respectively, mirroring the reduction trends derived from Ground Truth measurements (49.27% and 74.92%) with an error margin within 5%. This observation implies that Acropora experiences a more substantial decline in coverage than Pocillopora, indicating higher sensitivity to environmental changes. However, given the smaller original coverage of Acropora, additional regional data may be essential to further validate this conclusion. It is important to note that comprehensive coral monitoring necessitates the integration of various ecological indicators, each playing a crucial role in elucidating the intricate dynamics of coral reef ecosystems. We only take the simplest example of coverage rate to show the potential for automated processing of underwater benthic images. We emphasize the capacity of this approach to expedite the analysis of extensive coral remote sensing data, thereby facilitating comprehensive monitoring conducted by marine experts.

Moreover, we recognize that our examination primarily pertains to the distribution of live corals on orthophotos, highlighting the imperative need to account for the presence of dead and bleached corals. Of particular significance is the dynamic nature of bleached corals, as they respond to environmental fluctuations by either regenerating or experiencing further deterioration, ultimately leading to either coral recovery or continued bleaching and eventual mortality. Real-time monitoring of bleached corals serves as a pivotal tool in enabling ecologists to discern the temporal variations in coral health, allowing for the identification of key environmental stressors and the formulation of timely, data-driven conservation strategies. The design of our network, tailored explicitly for multi-modal data fusion and morphological characterization of corals, underscores its potential in bolstering the efficacy of coral reef management initiatives. Looking ahead, we envisage the expansion of our network's capabilities to encompass a wider array of coral monitoring tasks, thereby fostering a more holistic and adaptive approach to marine conservation efforts.

6. Conclusions

This paper presents a cost-effective approach for automated monitoring of changes in coral reefs, utilizing underwater remote sensing technology. To address the challenge of limited data and class imbalance, we introduce a Copy-paste-based data partitioning technique, significantly enhancing the performance of minority classes within the network. Additionally, we design a novel multi-modal network model, CNet, specifically tailored for semantic segmentation of underwater coral remote sensing images. CNet accurately identifies different coral genera and is characterized by the use of two encoders with layer asymmetry, the fusion branch for effective fusion of RGB and D features and the skip connections for deep and shallow feature fusion. Our experimental results demonstrate the effectiveness of these strategies in enhancing the network's performance and generalization capabilities. Moreover, our proposed framework holds significant potential for practical applications in fine-grained coral ecological monitoring tasks. It may have the potential to enhance the accuracy and dependability of underwater visual image segmentation technology in the foreseeable future.

Acknowledgments

This study was funded by the NSFC (41901407). The author thanks Wuhan University and ETH Zurich for their support in providing the computing environment. Special appreciation goes to the Moorea IDEA project for its data contributions.

References

[1] Jinming Cao, Hanchao Leng, Dani Lischinski, Daniel Cohen-Or, Changhe Tu, and Yangyan Li. Shapeconv: Shape-aware convolutional layer for indoor rgb-d semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7088–7097, 2021. 3

[2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 5

[3] Xiaokang Chen, Kwan-Yee Lin, Jingbo Wang, Wayne Wu, Chen Qian, Hongsheng Li, and Gang Zeng. Bi-directional cross-modality feature propagation with separation-and-aggregation gate for rgb-d semantic segmentation. In *European Conference on Computer Vision*, pages 561–577. Springer, 2020. 5

[4] Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, 2020. 2

[5] Xiaohan Ding, Yuchen Guo, Guiguang Ding, and Jungong Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1911–1920, 2019. 3

[6] Deng-Ping Fan, Yingjie Zhai, Ali Borji, Jufeng Yang, and Ling Shao. Bbs-net: Rgb-d salient object detection with a bifurcated backbone strategy network. In *European conference on computer vision*, pages 275–292. Springer, 2020. 5

[7] Aidan Fitzpatrick, Ajay Singhvi, and Amin Arbabian. An airborne sonar system for underwater remote sensing and imaging. *IEEE Access*, 8:189945–189959, 2020. 1

[8] Mark F Forst. The convergence of integrated coastal zone management and the ecosystems approach. *Ocean & Coastal Management*, 52(6):294–306, 2009. 1

[9] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2918–2928, 2021. 3

[10] Anna Barbara Giles, Keven Ren, James Edward Davies, David Abrego, and Brendan Kelaher. Combining drones and deep learning to automate coral reef assessment with rgb imagery. *Remote Sensing*, 15(9):2238, 2023. 2

[11] Manuel González-Rivero, Pim Bongaerts, Oscar Beijbom, Oscar Pizarro, Ariell Friedman, Alberto Rodriguez-Ramirez, Ben Upcroft, Dan Laffoley, David Kline, Christophe Bailhache, et al. The catlin seaview survey–kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 24(S2):184–198, 2014. 1

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[13] TP Hughes, AH Baird, EA Dinsdale, NA Moltschaniwskyj, MS Pratchett, JE Tanner, and BL Willis. Patterns of recruitment and abundance of corals along the great barrier reef. *Nature*, 397(6714):59–63, 1999. 2

[14] Md Jahidul Islam, Chelsey Edge, Yuyang Xiao, Peigen Luo, Muntaqim Mehtaz, Christopher Morse, Sadman Sakib Enan, and Junaed Sattar. Semantic segmentation of underwater imagery: Dataset and benchmark. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1769–1776. IEEE, 2020. 2, 5

[15] Andrew King, Suchendra M Bhandarkar, and Brian M Hopkinson. Deep learning for semantic segmentation of coral reef images using multi-view information. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–10, 2019. 2

[16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[17] Katsunori Mizuno, Kei Terayama, Seiichiro Hagino, Shigeru Tabeta, Shingo Sakamoto, Toshihiro Ogawa, Kenichi Sugimoto, and Hironobu Fukami. 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):12416, 2020. 2

[18] Hassan Mohamed, Kazuo Nadaoka, and Takashi Nakamura. Assessment of machine learning algorithms for automatic benthic cover monitoring and mapping using towed underwater video camera and high-resolution satellite images. *Remote Sensing*, 10(5):773, 2018. 2

[19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 4

[20] Gaia Pavoni, Massimiliano Corsini, Marco Callieri, Giuseppe Fiameni, Clinton Edwards, and Paolo Cignoni. On

improving the training of models for the semantic segmentation of benthic communities from orthographic imagery. *Remote Sensing*, 12(18):3106, 2020. 2

[21] G Pavoni, M Corsini, M Callieri, M Palma, and R Scopigno. Semantic segmentation of benthic communities from orthomosaic maps. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:151–158, 2019. 1

[22] G Pavoni, M Corsini, N Pedersen, V Petrovic, and P Cignoni. Challenges in the deep learning-based semantic segmentation of benthic communities from ortho-images. *Applied Geomatics*, 13:131–146, 2021. 1

[23] Gaia Pavoni, Massimiliano Corsini, Federico Ponchio, Alessandro Muntoni, Clinton Edwards, Nicole Pedersen, Stuart Sandin, and Paolo Cignoni. Taglab: Ai-assisted annotation for the fast and accurate semantic segmentation of coral reef orthoimages. *Journal of field robotics*, 39(3):246–262, 2022. 2

[24] William K Pratt. Generalized wiener filtering computation techniques. *IEEE Transactions on Computers*, 100(7):636–641, 1972. 2

[25] Victor Quintino, Rosa Freitas, Renato Mamede, Fernando Ricardo, Ana Maria Rodrigues, Jorge Mota, Angel Pérez-Ruzafa, and Concepción Marcos. Remote sensing of underwater vegetation using single-beam acoustics. *ICES Journal of Marine Science*, 67(3):594–605, 2010. 1

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 3, 5

[27] Daniel Seichter, Mona Köhler, Benjamin Lewandowski, Tim Wengefeld, and Horst-Michael Gross. Efficient rgb-d semantic segmentation for indoor scene analysis. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 13525–13531. IEEE, 2021. 5

[28] Ronghua Shang, Jiyu Zhang, Licheng Jiao, Yangyang Li, Naresh Marturi, and Rustam Stolkin. Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images. *Remote Sensing*, 12(5):872, 2020. 2

[29] ASM Shihavuddin, Nuno Gracias, Rafael Garcia, Arthur CR Gleason, and Brooke Gintert. Image-based coral reef classification and thematic mapping. *Remote Sensing*, 5(4):1809–1841, 2013. 2

[30] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 3

[31] Hong Song, Syed Raza Mehdi, Yangfan Zhang, Yichun Shentu, Qixin Wan, Wenxin Wang, Kazim Raza, and Hui Huang. Development of coral investigation system based on semantic segmentation of single-channel images. *Sensors*, 21(5):1848, 2021. 2

[32] David Souter, Serge Planes, Jérémy Wicquart, Murray Logan, David Obura, and Francis Staub. Status of coral reefs of the world: 2020, 2021. 1

[33] Mathew Wyatt, Ben Radford, Nikolaus Callow, Mohammed Bennamoun, and Sharyn Hickey. Using ensemble methods to improve the robustness of deep learning for image classification in marine environments. *Methods in Ecology and Evolution*, 13(6):1317–1328, 2022. 2

[34] Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang. Unitbox: An advanced object detection network. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 516–520, 2016. 3

[35] Hanqi Zhang, Armin Grün, and Ming Li. Deep learning for semantic segmentation of coral images in underwater photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:343–350, 2022. 2

[36] Jiageng Zhong, Ming Li, Hanqi Zhang, and Jiangying Qin. Combining photogrammetric computer vision and semantic segmentation for fine-grained understanding of coral reef growth under climate change. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 186–195, 2023. 2, 5