NOVEL APPROACHES TO ENHANCE CORAL REEFS MONITORING WITH UNDERWATER IMAGE SEGMENTATION

Hanqi Zhang [1], Ming Li [1,2] *, Xiaotian Pan [3], Xinlin Zhang [3], Jiageng Zhong [1], Jiangying Qin [1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - hqzhang@whu.edu.cn, zhongjiageng@whu.edu.cn, jy_qin@whu.edu.cn

[2] Institute of Theoretical Physics, ETH Zurich, Zurich, Switzerland - mingli39@ethz.ch

[3] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China - hanse8023@whu.edu.cn, nanako@whu.edu.cn

Commission III, WG III/1

KEY WORDS: Deep Learning, Underwater Image Segmentation, Coral Reefs, Monitoring, Image Processing, Review.

ABSTRACT:

Coral reefs not only inhabiting millions of species that are primarily or completely associated with them, but also produce economic and cultural benefits to coastal societies around the world. In recent years, affected by climate change and human factors, coral reef ecosystem has been experiencing accelerated degradation. Coral reef monitoring activities are therefore required to assess the impact of adverse factors on corals and to track subsequent recovery or decline. The collection of image data has become a common approach in the field of underwater monitoring, but traditional coral image data analysis mainly has high time and labor costs. We need to investigate the spatial distribution of different coral populations in the study area through image segmentation methods to help oceanographers develop effective management and conservation strategies. In fact, deep learning has shown better prediction performance than traditional image processing or traditional machine learning algorithms in coral image segmentation tasks. Starting from classification of random point annotations, segmentation of sparsely labeled data, and segmentation of densely labeled data, this paper summarizes state-of-the-art techniques of segmentation in deep learning applied to underwater images. Then, we discuss the problems and challenges of CNN-based underwater image segmentation of coral reefs, and make corresponding solutions or possible directions for future.

1. INTRODUCTION

As an important part of marine ecosystems, coral reefs provide habitat and sufficient nutrition for multiple organisms, they also produce enormous economic and cultural benefits to coastal societies around the world. However, direct and indirect anthropogenic pressures are being increasingly endured by coral reefs, including climate change, ocean acidification, overfishing and pollution. These stress factors work together to cause the large-scale degradation of coral reef ecosystems gradually. Status of Coral Reefs of the World 2020 report that 14 percent of the world's coral reefs were damaged between 2009 and 2018 (Souter D, 2021). Monitoring activities on coral reefs are required to assess the impact of adverse factors on corals and track subsequent recovery or decline in order to develop more effective management and conservation strategies to maintain the integrity and resilience of coral reef ecosystems (Obura et al., 2019). With the development of camera systems, image data collection has become a common approach in the field of underwater monitoring. Traditional analysis of coral image data has largely relied on experts with specialized knowledge, they distinguish corals by comparing morphological features. However, due to the repetitiveness of the labeling task and the required experience, this method has high time and labor costs, and there are certain degree of subjectivity of the level of recognition accuracy (Beijbom et al., 2015). How to knowledge mining efficiently and automatically from underwater video and image sequences is a new challenge. Because of the complex underwater imaging environment and lighting conditions, flaws in underwater images undoubtedly make the task of coral identification difficult: 1) The energy loss of light propagating in water leads to low contrast and blurred texture of images. 2) The uneven spectral propagation causes color distortion in underwater imaging. 3) Impurities and suspended particles in water usually lead to large noise in underwater images. In addition, underwater images have complicated backgrounds. Varied benthic assemblages result significant change in different image backgrounds (Hedley et al., 2016). For example, dead corals are covered in algae, similar to their surroundings, such as reefs. Free-swimming fish often block the target, making accurate identification of their features challenging.

Traditional computer vision techniques often capture visual features through textures, spots, or edges (Shihavuddin et al., 2013; Mary et al., 2017). After enough features are extracted from the image, these features are able to generate the definition of each target class for the purpose of simple classification of different coral classes, such as SIFT, histogram, Local Binary Patterns (LBP), etc. Typically, traditional computer vision and machine learning algorithms employ pre-designed features and classifiers for image recognition, these features they extracted lack adaptability, and cannot be generalized to other classes or datasets (Mittal et al., 2022). Determining which features better describe the target class requires the researcher's own judgment because of coral's numerous and complicated branches. In general, neither the speed nor the accuracy of traditional techniques can meet the needs of current coral reef monitoring tasks.

Nowadays, the field of deep learning has made significant progress in image processing, providing new methods for coral

* Corresponding author

image analysis. Image segmentation can automatically extract the most representative and salient semantic features from a large number of collected images, and understand the population and distribution of benthic communities in the study area. In fact, deep learning has shown better predictive performance than traditional image processing or traditional machine learning algorithms in coral image segmentation tasks (King et al., 2018). In this review, we mainly summarize deep learning techniques for underwater coral image segmentation. The following is divided into two parts: the first part discusses the segmentation method of coral images, then the second part looks forward to future research directions.

2. UNDERWATER IMAGE SEGMENTATION

2.1 Classification of Random Point Annotations

In the early identification and classification and image processing of marine organisms, it was very time-consuming to achieve complete manual segmentation, therefore most benthic communities were annotated by random point sampling (Beijbom et al., 2012). The label of dataset image for network training input is to expand each annotated point into a square area around a point. Patch-based CNNs generate predictions for a single class for each patch. Following will present some approaches for coral image segmentation based on patch classification.

The authors applied the pretrained VGG-Net to the coral reef classification task (Mahmood et al., 2016). The dataset is a subset of the Benthoz15 dataset (Bewley et al., 2015), with only two classes of corals and non-corals. For the training images input to the network, the authors introduce the idea of Spatial Pyramid Pooling (SPP) to extract patches with multiple scales centered on labeled pixels, which scales are 28×28, 224×224, and 448×448, respectively. This approach can improve classification accuracy and make the resulting features scale-invariant. The preprocessing step adopts color channel stretching. Experimental results show that the best performance is achieved with the training and test datasets consist of images from the same year. Performance degrades when employing experimental data across multiple years. Ultimately, to demonstrate that this study could investigate growth trends in coral populations, they generated coral coverage maps using the best performing classification network and calculated coverage percentage for each site and every year.

Likewise, Beijbom et al chose a low-cost modified consumer camera FluorIS to capture wide-band wide-field-of-view fluorescence images (Beijbom et al., 2016). FluorIS can be used for underwater fluorescence imaging of green and red. The wide-band components of the fluorescence spectra correspond to the emission spectra of green fluorescent protein (GFP) and chlorophyll-a, respectively. The combined fluorescence and reflectance image data has both high-contrast coral and provides background and information about non-fluorescent substrates. These reflectance and fluorescence image pairs, acquired by standard SLR and FluorIS cameras, were annotated by coral ecologists at 200 random point locations. Then they introduced a supervised automatic annotation algorithm based on a convolutional neural network (CNN) to automatically classify the five major coral genera.

Compared with the traditional automatic annotation approach, there is no difference in the accuracy of training on the reflection image. However, when inputting FluorIS images, the proposed method based on the CNN is higher than the accuracy of 81.0% obtained by the traditional method, which achieves an accuracy of 85.5%. The CNNs were trained using only reflectance per pixel, reflectance and fluorescence mean, fluorescence only, and on 5-channel images. The results displayed that FlourIS image intensities were more effective than reflectance image intensities in distinguishing corals from non-corals, with the highest accuracy (90.5% ± 0.8%) achieved with a three-channel network that includes mean values of fluorescence and reflectance. Moreover, they also released the web-based point annotation platform CoralNet (Chen et al., 2021).

González-Rivero et al. applied deep learning as a tool to support coral reef monitoring, employing a convolutional neural network architecture VGG-D 16 pretrained on the large dataset ImageNet to classify random point annotations (Gonzalez-Rivero et al., 2020). The patch area cropped around each point is set to 224*224 pixels to align with the predefined image input size of the VGG architecture. The authors use absolute error to represent the overall difference in abundance estimates for each label class between machines and observers. Among the major functional groups of underwater benthic organisms, the differences in error about abundance estimates were most significant between categories but smaller between study areas. Algae showed the greatest variation, with a 3%–5% error, followed by hard and soft corals.

On average, there was 97% high agreement between expert and automated observations. This consistency also exists at the community level. The authors compare the performance of machine and observer estimations in Australia, Central Indian Ocean, and Central Pacific Regions. The similarity between automated estimates of benthic composition and manual observations varied across taxonomic groups, ranging between 84% and 94%. Provided the errors of network estimates are comparable to those by multiple observers, the estimated cost of single image annotation of $5.41 by experts is much higher compared to just $0.07 using machine learning. These results demonstrate the practicality of deep learning methods for underwater monitoring tasks, and its application in automatic image classification reduces costs while accelerating image analysis. In addition, they discussed the advantages of machine learning techniques to facilitate coral reef monitoring: combining underwater image monitoring with automatic image annotation greatly improves the efficiency of coral reef measurement and monitoring, and provides new ideas for future region management and data integration.

2.2 Segmentation of Sparsely Labeled Data

In the case of environmental monitoring of coral reefs, data labeling needs relevant expertise and is expensive. Therefore, datasets of benthic communities are mostly sparsely annotated, such as random point annotations. This approach assumes that the surrounding pixels of an annotated point are also of the same class, and makes the choice of patch size difficult: patches must be large enough to describe the structure of benthic organisms, but small enough to prevent other classes. Patch-based classification can only roughly discriminate objects in underwater images, it lacks per-pixel accuracy, and the result does not fit the coral region contours well. Enhancing sparse annotations can reduce the labeling work but acquire densely labeled segmentation data.

(Aloso et al., 2017) proposed a CNN-based strategy for semantic segmentation of coral reef images with available sparsely labeled data. They recommended simple augmentation of ground-truth with superpixel approaches to generate approximate but dense labels. Moreover, they evaluated different combinations of fluorescence data and RGB image data. It was concluded that the best combination was a three-channel input through combining two channels of fluorescence images with the mean of the RGB channels. The network model adopted was a fine-tuned SegNet, and used the cross-entropy loss with the median frequency balancing as the objective function for training to eliminate the impact of the difference in class sample numbers on the network. The encoder-decoder CNN model trained on both data types provides the highest average precision and recall for coral class. They also proposed a dense score based on superpixel labels that serves as a quality assessment for segmentation. Compared with patch-based classification methods, the overall accuracy of both are comparable, but superpixel-based methods produce more coral-like morphological characteristics in segmentation and follow object contours better.

In order to balance the fitting accuracy of the single-level superpixel strategy and unlabeled region numbers, they proposed a multilevel superpixel segmentation Algorithm in (Alonso et al., 2019), which iteratively applies superpixel segmentation, and gradually reducing iterate of each generated superpixels to obtain more accurate labels for per-pixel classification. 15 superpixels is used as the default setting for multi-level methods in most experiments because the accuracy starts to converge. It improves mIoU of 11% on the average by contrast with the previous single-stage method. The modified cross-entropy loss, i.e. give more importance to pixels whose neighboring pixel predictions are not the same, which considering neighboring pixels without adding more computation. The authors compare three semantic segmentation models, Deeplab v3, Deeplab v3+, and symmetric Deeplab v3+. The results show that the symmetric architecture performs better, but increases inference time. Besides, CoralNet pretrained encoder they published is applied for all coral reef semantic segmentation models, and the proposed sparse label augmentation method is also applicable to other datasets in the non-coral field (Alonso et al., 2018).

King et al. suggested an image annotation tool based on superpixel generation and computation of graph cuts in their research (King et al., 2018). Further, the annotation tool uses RGB histogram and Gabor filter features to measure region similarity and performs k-means clustering based on the similarity measure. The authors compare patch-based network models and semantic segmentation models. Then they performed normalization as a preprocessing step before the images are fed into the segmentation model. Experiments demonstrate that Resnet152 achieves the highest patch accuracy of 90.03%, and DeepLab v2 achieves the highest pixel accuracy of 67.7%.

They also introduced stereo disparity in (King et al., 2019) to provide multi-view information: after the image was corrected using camera calibration parameters, a disparity map was created with a semi-global block matching-based disparity estimation algorithm, then they used an inpainting algorithm to interpolate missing disparity data. The resulting disparity map is fed into FCNN as an additional fourth channel concatenated with RGB channels. In addition, they also designed TwinNet, which is a stereo FCNN architecture, the basic architecture is the front-end of the Dilation8 architecture, and the input data is a stereo image pair, weights are shared in the backbone. The left and right outputs by Dilation8 are fed to different Siamese sub-networks, and the weights of the sub-modules are learned dividually. Then it is fed into a stereo module consisting of three convolutional layers, and its output is finally sent to the collapse module after concatenating the channel axes. The experiments compare the performance of a standard FCNN with three channels as input, FCNN with four channels as input, TwinNet with only left-view images as input, and TwinNet with stereo image pairs as input. The results show that the TwinNet architecture using stereo image pairs achieves the highest accuracy of 66.44% on 10 categories.

For the algorithm of generating dense labels, (Pierce et al., 2020) proposed an improved version of the multi-level superpixel segmentation (MSS) algorithm, namely Fast MSS, which can convert existing sparse labels into dense labels required for training FCN, and greatly reduce execution time for generating dense labels. Fast MSS modifies how to combine labels created during each iteration and adopts Fast-SLIC as the over-segmentation algorithm. After each iteration, the class labels of the MSS propagated to adjacent pixels are stored in a 2D array, piling a 3D array of size (H x W x I), where I refers the total number of iterations. Starting from the array created in the first iteration, MSS joins each 2D array in the stack, and Fast-MSS makes dense labels by computing the statistical pattern of class labels on the 3-rd dimension of the stack. Fast-SLIC is optimized on the basis of SLIC, such as color quantization, subsampling, parallelization and integer-based arithmetic.

In (Raine et al., 2022), the authors propose a new point-label-aware superpixel method for dense label generation, which uses a custom loss function to optimize the superpixel center position to maximize the similarity of pixels within each region , while reducing class labels containing conflicts. The loss function is weighted by the two loss terms: distortion loss and conflict loss. The distortion loss clusters similar pixels to generate superpixels that perfectly fit the coral boundaries. The weighted distortion loss is obtained by weighting the distance between each pixel F_p and the superpixel feature $S_i = [T_i, X_i]$.

$$\mathcal{L}_{distortion}(p) = \sum_{i \in I} \left\| F_p - S_i \right\|^2 \mu_{p,i} \tag{1}$$

where the vector $F_p = [T_p, X_p]$ consists of the output T_p of the feature extractor and the scaled (x, y) coordinates. They use the Superpixel Sampling Network (SSN) to extract dense feature vectors representing the area around each image pixel. The fuzzy membership $\mu_{p,i}$, which computes the normalized distance from a pixel to each superpixel.

The conflict loss only considers the labeled pixels, and ensures that only one class exists in each superpixel. The conflict loss sums the inner products of the fuzzy memberships of the conflict markers. To maximize the possibility that a superpixel region contains point labels, they output three results using three models with different hyperparameters, and combine the three ground truth masks by adopting a specific pattern at each pixel. The data augmentation on the UCSD Mosaics dataset includes random horizontal or vertical flips, gain, and gamma. In enhancing coral image labels, the superpixel point label perception method has a larger increase in accuracy and mIoU than CoralSeg, Fast Multilevel Algorithm and SSN, and also significantly improves computational speed of each image.

(Yu et al., 2021) proposed an iterative strategy that only relies on sparse labels to generate pseudo-labels to improve coral image segmentation performance: First, use wide Residual Networks (WRN) which trained by sparse labels to extract features, then use Dirichlet assignment (LDA) with spatial coherence in network embeddings to infer class distributions on discrete features and generate pseudo-labels. Finally, train the network by adding some pseudo-labels inferred from prior knowledge to improve the learning of the model. Pseudo-labeled samples are weighted by sample and class. In the first stage of training, the objective function consists of supervised loss and mutual information constraint loss. The supervised loss is the cross-entropy between the prediction of the labeled samples and their corresponding true label, the mutual information constraint loss is the mutual information between input space and latent feature space. In the second stage of training with pseudo-labels, the objective function adds the cross-entropy loss function between the predictions of unlabeled samples and their corresponding pseudo-labels. The results indicate that LDA with spatial coherence and training with weighted confidence samples has the highest segmentation accuracy, with mIoU of 73.5%. Adding spatial information can improve performance, and considering the confidence level of pseudo-labels can improve the generalization ability of the network and reduce the bad effects of mistake pseudo-labels. If adding virtual adversarial training (VAT), it has the best performance of 75.1%. Yet this approach doesn't work well for classes with a low percentage of overall pixels.

2.3 Segmentation of Densely Labeled Data

As described, one of the factors limiting the accuracy of image segmentation is the lack of images classified by pixel-based labels, and the application of approximate labels will increase the error of network predictions (Pavoni et al., 2021a). How to efficiently obtain accurate ground truth always be a block to image segmentation performance.

(Pavoni et al., 2021b) open-sourced TagLab, a semi-automatic labeling tool based on artificial intelligence (AI), which can accomplish pixel-level labeling of benthic organisms. With a wealth of data, they employed underwater ortho maps to segment benthic communities (Pavoni et al., 2019a). Regions were divided as training, validation, and test datasets by biologically-inspired dataset partition methods, which weighted the calculated values of three ecological metrics describing the spatial patterns of benthic communities as similarity scores (S), selected with triplet of best S-values. To address the class imbalance problem, a simple oversampling strategy is also proposed based on the actual area coverage.

Ortho-mosaic need to be divided into tiles by overlapping sliding windows into the input network due to memory constraints. Such a partition method will get different predicted class values on boundary of each tile, so there requires a reassembling strategy. They employed Bayesian fusion to obtain the final scores for smoother boundaries at tile connection. The network selects the pretrained CNN architecture Bayesian SegNet for semantic segmentation tasks. They also used a PyQt-based interactive validation tool to examine the classification of the network and correct wrong predictions, possibly exporting more correct labels to retrain the network (Pavoni et al., 2019b). Through these works, the new accuracy of the network is 96.7%.

In (Pavoni et al., 2020), the dataset used in the study are manually labeled orthos of 10 plots from the 100 Island Challenge Project. A 100-square-meter parcel was imaged with a Nikon d7000 camera on each island, then they use Agisoft Metashape software to create a single continuous 3D model of each parcel. The dense cloud is then put into the Viscore platform to create an orthographic projection. The proposed oversampling strategy is to use Poisson disk sampling to select the center of the tile, and generate non-overlapping random samples not smaller than the specified minimum distance 2r through the Dart Throwing algorithm with class-related radius. The goal is to crop patches with different local densities. Similarly, they provide several image augmentation techniques: colorimetric enhancement (simulates changes associated with underwater image formation) and geometric enhancement (random rotation and translation), weight decay and batch normalization (subtracting each channel average).

The network chooses the pre-trained DeepLab V3+. They compared performance of the four loss functions, Weighted Cross-Entropy, Generalized Dice loss, Focal Tversky and GDL+Boundary in terms of choosing the loss function. As can be seen that although the Focal Tversky loss proved to be the best loss for training coral segmentation models, it was unable to handle datasets with overly imbalanced proportions. Beside conducting experiments on the same area, they also applied the same model to other test areas to evaluate its generalization ability. The mIoU ranged from 0.69 to 0.97, with accuracy depending on species richness and abundance.

In addition to introducing deep learning for coral image segmentation, some strategies for dealing with overall class imbalance and differences in class distribution among images are described in detail (Steffens et al., 2019). To balance the distributions of datasets, they created N fixed-size training and validation sets with randomly selected images, and applied cosine distances weighted by the overall class distribution to compare the train/validation distributions. The training input images are cropped at random square sizes, then scaled to a fixed size of 256*256, and randomly flipped vertically or horizontally. Data augmentation can enhance the robustness of the network model and prevent the network from overfitting. Correspondingly, DeeplabV3 also uses a weighted cross-entropy loss function based on the class distribution measurement.

Mizuno et al.'s research focuses on how to efficiently and accurately estimate coral cover (Mizuno et al., 2020). They described the construction process of the 3D model in detail: First, use a towed optical camera array system (Speedy Sea Scanner) for data acquisition. Continuous still images were obtained from the video data, and 3-D point clouds were reconstructed from them using Agisoft Metashape, resulting in a large-scale 3-D structural model with a resolution of 0.01 m. The Structure from Motion technique is then employed to generate digital elevation models (DEMs) and 2-D orthophotos based on the 3-D structural model. Furthermore, they constructed a U-Net-based deep neural network for experiments. To evaluate the prediction results, they also used five-fold cross-validation. The results show that U-Net with color correction and data-augmentation preprocessing achieved the highest accuracy (0.910) and F-measure (0.772). They concluded that not only UNet may have better performance than pixel-level CNN, but it also has low computational cost. Finally, they constructed a predictive model using 200 images and the network structure with the best performance and estimated coral cover for the entire survey area, with the percentage of coral cover distributed between 0-35%.

(Song et al., 2021) published the CoralS dataset for coral semantic segmentation, which includes spectral images at different wavelengths and RGB images of several coral species under different illuminations. Especially, they classified coral classes with modified DeepLab model DeeperLabC, distinguishing coral and non-coral areas. The DeeperLabC is shown in Figure 1. The structure of the DeeperLabC model includes encoder and decoder, and the adjusted pre-trained ResNet34 is used as the skeleton network. The channel-combined features are fed into the decoder, including low-level features that are further downsampled via convolution and Space2Depth operations, and high-level features extracted via Atrous Spatial Pyramid Pooling (ASPP). Feature maps are upsampled to the original size using depth-to-space and linear interpolation in the decoder. Image preprocessing is to crop and scale all RGB images, spectral images and labels into 512*512 single-channel images. And training set are randomly augmented, rotated, and flipped horizontally or vertically. A class activation map (CAM) module is added to the tail of the DeeperLabC model to visualize feature maps for semantic segmentation. Compared with FCN, UNet and DeepLabV3+, this way achieves the best performance with mIoU of 93.90%.

Although there are mature techniques for semantic segmentation of benthic communities, most of the existing annotation are related to specific tasks, such as coral reef classification and cover statistics tasks as summarized in this paper, and there is no general object classification dataset like ImageNet to facilitate large-scale training and benchmark evaluation of segmentation models. Therefore, (Islam et al., 2020a) proposed the first large-scale dataset SUIM for general multi-object semantic segmentat-

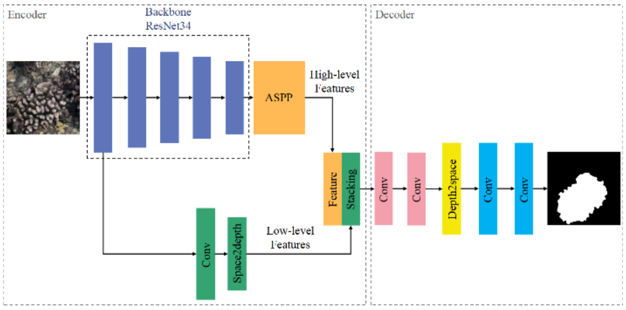

Figure 1. The structure of DeeperLabC for single-channel coral segmentation.

-ion of underwater scenes. It contains 1525 natural underwater images, corresponding ground truth and 110 images for test. The SUIM dataset considers the following eight object categories: fish (vertebrates), coral reefs (invertebrates), aquatic plants, wrecks/ruins, human divers, robots, and seafloor, annotated by humans. It also utilizes some data from large datasets of EUVP (Islam et al., 2020b), USR-248 (Islam et al., 2020c) and UFO-120 (Islam et al., 2020d). Further, they proposed SUIM-Net, which follows an encoder-decoder architecture with optional hierarchical skip connections on its core building blocks, called residual skip block (RSB), which is shown in Figure 2. Skip connections are fed through intermediate convolutional layers when skip=0, and directly from the input when skip=1 for local residual learning. The network structure is shown in Figure 3:

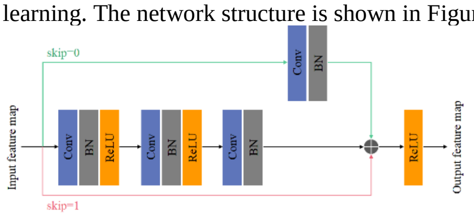

Figure 2. The architecture of Residual Skip Block.

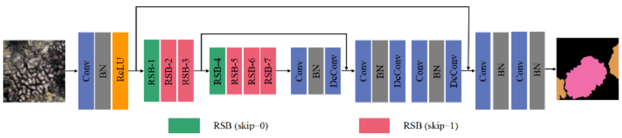

Figure 3. The end-to-end network architecture.

Combining skip connections and residual learning, SUIM-Net not only is comparable to current state-of-the-art architectures, but also has a light-weight architecture for real-time inference. They use state-of-the-art network models and SUIM-Net to evaluate performance on the SUIM dataset through two tasks, semantic segmentation and single-channel saliency detection for five main class. The results indicate that UNet$_{RGB}$, FCN8$_{VGG}$ and DeepLabV3 generally perform the best in two tasks. The evaluation metrics F and mIoU scores of SUIM-Net are within 5% of their respective top scores. In fact, SUIM-Net has higher speed than other models, and its memory efficiency is more than 10 times higher than the above three networks. In general, these important cases of the proposed SUIM dataset provide ideas for realizing real-time annotation and processing of underwater imagery, and also help to explore the feasibility of deep learning models for other specific application tasks. There is another study complementing DeepLabV3+ experiments on the SUIM dataset (Chicchon et al., 2022).

2.4 Conclusion

The deep learning based coral image segmentation methods are summarized in Table 1. Given the above, we overview a set of pipelines for analysing underwater coral images:

Image Preprocessing. The processing of underwater images normally includes the following two points: image enhancement and image restoration. For images with color decay, an enhanced model of deep learning is used to improve image quality (Islam et al., 2020; Yang et al., 2021). It is also a popular practice to generate orthophotos from continuous video for subsequent studies on coral cover estimation.

Image Annotation. At early stages, due to scarce resources, researchers usually chose the method of random point annotation to distinguish corals. To achieve higher annotation quality, there are related researches on generating dense labels from sparse labels. The current requirement for coral accuracy has led professionals to conduct pixel-wise annotation, with semi-automatic annotation tools available.

Segmentation Network. From patch-based classification networks to pixel-based segmentation networks, improvements in computing power of computer have enabled deeper network models to extract more semantic information.

Application Research. After segmenting coral individuals automatically and precisely, researchers often apply it for estimating coral cover in study area, contributing to environmental protection.

3. DISCUSSION AND CONCLUSION

We provided a comprehensive survey of the recent development in coral reefs monitoring using deep learning techniques, including classification at random point annotation, semantic segmentation of sparsely labels and semantic segmentation of densely labels. We presented a comprehensive performance and advantage of various methods in each field. Nonetheless, Coral image segmentation still faces unprecedented and important challenges, and there are many issues that need further thinking and research. Below, we outline some outstanding issues and identify potential research directions.

3.1 Coral Growth Trends and Coverage

Most studies at this stage focus on monitoring the overall coverage of coral reefs in a study area, ignoring each coral genus. Even though classifications of multiple categories were carried out in the study, there was no further analysis and estimation of each species. The dominant species in each marine region are different, and the coverage of different coral reef genus needs to be considered, which can better assess the ecology and resilience of coral reefs. In the future, the segmentation of coral images will improve towards the direction of multi-class segmentation, and how to deal with the class imbalance of different coral species is also a point issue in the future. The growth of corals can be considered from multiple perspectives such as coral size, coral color, and growth environment, which requires pixel-wise labeling of corals in different instance areas, that is, instance segmentation.

3.2 Insufficient Training Data in Corals

Sufficient data is the key for training CNN models for underwater image classification. Due to the high time and equipment cost of situ survey, the automatic acquisition of underwater images will be completed by unmanned underwater vehicles in the future. While preparing the ground truth of the dataset still requires manual annotation of human experts, and preparing the remote

	Method[Refer.]	Data augmentation	Feature extractor	Loss	Contribution
Classification of Random Point Annotations	Mahmood et al.	-	VGG	-	Estimating coral cover
	Beijbom et al.	-	CNN	-	Combines RGB and fluorescence images
	González-Rivero et al.	-	VGG-D 16	-	Measuring machine and observer error
Segmentation of Sparsely Labeled Data	Aloso et al.	-	Fine-tuned SegNet	Weighted	Augmenting sparse labels with superpixels.
	Aloso et al.		-	Weighted	Augmenting sparse labels with multi-level superpixels.
	King et al.	√	-	-	Comparing patch-based and pixel-wise models.
	King et al.	-	TwinNet	-	Design a network that takes stereo images as input.
	Pierce et al.	√	-	-	Fast-MSS.
	Raine et al.	√	SSN	Hybrid	A new point-label-aware superpixel method.
	Yu et al.	-	WRN		Iterative strategy for generating pseudo-labels.
Segmentation of Densely Labeled Data	Pavoni et al.	√	Bayesian SegNet	Weighted	Biologically-inspired dataset partition methods..
	Pavoni et al.	√	DeepLab V3+	Hybrid	Poisson Disk Oversampling Strategy.
	Steffens et al.	√	DeeplabV3	Weighted	Techniques to deal with class imbalance.
	Mizuno et al.	√	UNet	-	Efficient and accurate estimation of coral cover.
	Song et al.	√	DeeperLabC	-	New single-channel segmentation network.
	Islam et al.	√	SUIM-Net	-	Lightweight segmentation network and general dataset.

Table 1. Summary of coal image segmentation methods.

sensing dataset of orthophotos also takes more time. How to develop a large amount of labeled data on the basis of reducing labor cost is a serious challenge. However, some research is devoted to image enhancement techniques that address insufficient training data, they are insufficient when applied to complex or large deep learning models. Large-scale datasets of coral imagery need the attention of the global coral conservation community, and they may drive ecological experts or researchers studying underwater robotics to develop a generic pixel-wise labeling dataset for underwater coral reefs scenes.

3.3 Network Design

From the current research, it can be seen that most deep learning-based coral image segmentation research focuses more on the accuracy rather than on the time efficiency. At the same time, few reasearch specifically design a lightweight network model with applicability and mobility in the field of underwater benthic communities. Future research may design a coral-specific network model that can be employed not only in automatic annotation and coral cover estimation, but also embedded in visual-based unmanned underwater vehicles for real-time data collection and identification.

ACKNOWLEDGEMENTS

This research was funded by the National Key R&D Program of China, grant numbers 2018YFB0505400, the National Natural Science Foundation of China (NSFC), grant number 41901407 and the College Students' Innovative Entrepreneurial Training Plan Program, Application of intelligent photogrammetry and deep learning in dynamic monitoring of coral reefs, grant numbers S202110486216.

REFERENCES

Souter D., Planes S., Wicquart J., et al., 2021: Status of Coral Reefs of the World 2020. *Global Coral Reef Monitoring Network.* gcrmn.net/2020-report.

Obura D. O., Aeby G., Amornthammarong N., et al., 2019. Coral reef monitoring, reef assessment technologies, and ecosystem-based management. *Frontiers in Marine Science*, 580. doi.org/10.3389/fmars.2019.00580.

Beijbom O., Edmunds P. J., Roelfsema C., et al., 2015. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. *PloS one*, 10(7), e0130312. doi.org/10.1371/journal.pone.0130312.

Shihavuddin A. S. M., Gracias N., Garcia R., et al., 2013. Image-based coral reef classification and thematic mapping. *Remote Sensing*, 5(4), 1809-1841. doi.org/10.3390/rs5041809.

Mary N. A. B., harma D., 2017. Coral reef image classification employing improved LDP for feature extraction. *Journal of Visual Communication and Image Representation*, 49, 225-242. doi.org/10.1016/j.jvcir.2017.09.008.

Pavoni G., Corsini M., Pedersen N., Petrovic V., Cignoni P., 2021a. Challenges in the deep learning-based semantic segmentation of benthic communities from Ortho-images. *Applied Geomatics*, 13(1), 131-146. doi.org/10.1007/s12518-020-00331-6.

Mittal S., Srivastava S., Jayanth J. P., 2022. A Survey of Deep Learning Techniques for Underwater Image Classification. *IEEE Transactions on Neural Networks and Learning Systems.* doi.org/10.1109/TNNLS.2022.3143887.

King A., Bhandarkar S. M., Hopkinson B.M., 2018. A comparison of deep learning methods for semantic segmentation of coral reef survey images. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. doi.org/1394-1402.10.1109/CVPRW.2018.00188.

Beijbom O., Edmunds P. J., Kline D. I., et al., 2012. Automated annotation of coral reef survey images. *IEEE conference on computer vision and pattern recognition.* doi.org/1170-1177. IEEE.10.1109/CVPR.2012.6247798.

Bewley M., Friedman A., Ferrari R., et al., 2015. Australian sea-floor survey data, with images and expert annotations. *Scientific data*, 2(1), 1-13. doi.org/10.1038/sdata.2016.113.

Mahmood A., Bennamoun M., An S., Sohel F., et al., 2016. Automatic annotation of coral reefs using deep learning. Oceans 2016 mts/IEEE monterey, 1-5.

Chen Q., Beijbom O., Chan S., et al., 2021. A New Deep Learning Engine for CoralNet. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3693-3702. doi.org/10.1109/ICCVW54120.2021.00412.

Gonzalez-Rivero M., Beijbom O., Rodriguez-Ramirez A., et al., 2020. Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. *Remote Sensing*, 12(3), 489. doi.org/10.3390/rs12030489.

Alonso I., Yuval M., Eyal G., et al., 2019. CoralSeg: Learning coral segmentation from sparse annotations. *Journal of Field Robotics*, 36(8), 1456-1477. doi.org/10.1002/rob.21915.

King A., Bhandarkar S. M., Hopkinson B. M., 2018. A comparison of deep learning methods for semantic segmentation of coral reef survey images. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops.* 1394-1402. doi.org/10.1109/CVPRW.2018.00188.

Beijbom O., Treibitz T., Kline D. I., et al., 2016. Improving automated annotation of benthic survey images using wide-band fluorescence. *Scientific reports*, 6(1), 1-11. doi.org/10.1038/srep23166.

Alonso I., Cambra A., Munoz A., et al., 2017. Coral-segmentation: Training dense labeling models with sparse ground truth. *Proceedings of the IEEE international conference on computer vision workshops.* 2874-2882. doi.org/10.1109/ICCVW.2017.339.

Alonso I. Murillo A. C., 2018. Semantic segmentation from sparse labeling using multi-level superpixels. *IEEE/RSJ International Conference on Intelligent robots and Systems IROS*, 5785-5792. doi.org/10.1109/IROS.2018.8594185.

King A., M Bhandarkar S., Hopkinson B. M., 2019. Deep learning for semantic segmentation of coral reef images using multi-view information. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.* 1-10.

Pierce J. P., Rzhanov Y., Lowell K., Dijkstra J. A., 2020. Reducing Annotation Times: Semantic Segmentation of Coral Reef Survey Images. *Global Oceans 2020: Singapore–US Gulf Coast*, 1-9. doi.org/10.1109/IEEECONF38699.2020.9389163.

Raine S., Marchant R., Kusy B., et al., 2022. Point Label Aware Superpixels for Multi-species Segmentation of Underwater Imagery. arXiv preprint arXiv:2202.13487.

Yu X., Ouyang B., Principe J. C., 2021. Coral Image Segmentation with Point-Supervision via Latent Dirichlet Allocation with Spatial Coherence. *Journal of Marine Science and Engineering*, 9(2), 157. doi.org/10.3390/jmse9020157.

Pavoni G., Corsini M., Ponchio F et al., 2021b. TagLab: AI-assisted annotation for the fast and accurate semantic segmentation of coral reef ortho images. *Journal of Field Robotics*. doi.org/10.1002/rob.22049.

Pavoni G., Corsini M., Callieri M., et al., 2019. Semantic segmentation of benthic communities from ortho-mosaic maps. *Copernicus GmbH. 2019 Underwater 3D Recording and Modelling" A Tool for Modern Alications and CH Recording"* Vol. 42, No.2. 151-158. 10.5194/isprs-archives-XLII-2-W10-151-2019.

Pavoni G., Corsini M., Palma M., Scopigno R., 2019. A Validation Tool for Improving Semantic Segmentation of Complex Natural Structures. *Eurographics Short Papers*, 57-60. doi.org/10.2312/egs.20191014.

Pavoni G., Corsini M., Callieri M et al., 2020. On improving the training of models for the semantic segmentation of benthic communities from orthographic imagery. *Remote Sensing*, 12(18), 3106. doi.org/10.3390/rs12183106.

Steffens A., Campello A., Ravenscroft J., et al., 2019. Deep Segmentation: using Deep Convolutional Networks for Coral Reef Pixel-wise Parsing. *CLEF Working Notes*.

Mizuno K., Terayama K., Hagino S., et al., 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. doi.org/10.1038/s41598-020-69400-5.

Song H., Mehdi S. R., Zhang Y., et al., 2021. Development of coral investigation system based on semantic segmentation of single-channel images. *Sensors*, 21(5), 1848. doi.org/10.3390/s21051848.

Islam M. J., Edge C., Xiao Y., et al., 2020a. Semantic segmentation of underwater imagery: Dataset and benchmark. *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS*, 1769-1776. doi.org/10.1109/IROS45743.2020.9340821.

Islam M. J., Enan S. S., Luo P., Sattar J., 2020c. Underwater image super-resolution using deep residual multipliers. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 900-906. doi.org/10.1109/ICRA40945.2020.9197213.

Islam M. J., Luo P., Sattar J., 2020d. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. arXiv preprint arXiv:2002.01155.

Chicchon M., Bedon H., 2022. Semantic Segmentation of Underwater Environments Using DeepLabv3+ and Transfer Learning. *Springer, Singapore*. *Smart Trends in Computing and Communications*, 301-309.

Hedley J. D., Roelfsema C. M., Chollett I., et al., 2016. Remote sensing of coral reefs for monitoring and management: a review. *Remote Sensing*, 8(2), 118. doi.org/10.3390/rs8020118.

Islam M. J., Xia Y., Sattar J., 2020b. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2), 3227-3234.

Yang H. H., Huang K. C., Chen W. T., 2021. Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 685-692.