

Machine Learning and Deep Learning Based Coral Species Classification: A Survey

¹Ms. Zaheda Parveen Mohammad, ²Dr. L.R. Sudha, ³Dr. R. Santhoshkumar,

¹Research Scholar,

Department of CSE,

Annamalai University, Chidambaram, India

²Associate Professor,

Department of CSE,

Annamalai University, Chidambaram, India

³Associate Professor,

Department of CSE,

St. Martin's Engineering College, Secunderabad,

Abstract:

Marine ecosystems cannot function without coral reefs. There have been reports of coral reefs declining in health and population recently. The amount of information on coral reefs available for study has expanded dramatically because to underwater imaging tools like autonomous underwater vehicles (AUV) and towed diver sleds. Variations within classes are large, class boundaries are complex, and picture clarity varies greatly in coral reefs. Because of this, categorizing corals is difficult. In order to better understand coral reefs, this study provides a comprehensive overview of how machine learning and deep learning have been applied to underwater picture processing. Image categorization, object detection, and scene understanding are just a few of the many computer vision problems where deep learning approaches have produced state-of-the-art results. From a computer vision standpoint, marine ecosystems are challenging because of the scene complexity involved. Deep learning has the ability to effectively solve the difficult task of classifying coral species.

Keywords:Corals, machine learning, classification, deep learning, marine images, convolutional neural networks

1 Introduction

Coral reefs play a crucial role in the health of marine environments. Many marine organisms depend on them as a safe haven and a nutrient-rich home. They supply a lot of nitrogen and other nutrients that benthic organisms need. In addition to safeguarding shorelines from waves and storms, they are crucial in the recycling of nutrients. The proximity of many fish and other species to coral reefs aids in maintaining a thriving fishing business. The tourism business also benefits from shallow water coral reefs like Australia's Great Barrier Reef. Researchers in the ocean have noticed a widespread decline in coral reefs. A study conducted in 2011 found that 19% of coral reefs have been gone, with another 75% now under threat [1]. The beneficial and bad effects of global warming, urbanization, human population, and

heavy use of the sea for shipping, mineral exploration, recreational uses like boating, and industrial commerce and activity on coral reefs have been substantial [2]. Coral bleaching and mortality are caused by rising ocean temperatures [3, 4]. Because of this, marine life around the world has been rapidly disappearing [5]. Marine ecosystems need to be closely watched so that damage to the ocean can be kept to a minimum. That's why we need optical imaging that works underwater. Most modern management plans need extensive long-term monitoring, remote sensing, and tracking of marine species and their associated habitats. Therefore, management applications that rely heavily on automatically annotated marine data are at the forefront of the field [6]. Standard methodologies for analyzing and reducing the negative impacts on seawater environmental sustainability can be developed with the advent of

underwater optical imaging tools. The use of digital cameras and videos is also expected to skyrocket,

which will boost the demand for data storage and processing automation.



Fig. 1. (a) Fringing reef off the coast of Eilat, Israel. (b) The Great Barrier Reef, Australia. (c) Lighthouse Atoll Reef, Belize. (d) Coral Bleaching

Marine biologists have amassed a substantial collection of coral reef images, numbering in the thousands; however, these photos have not yet been through the process of tagging. Every year, the Integrated Marine Observation System (IMOS) and comparable monitoring programs capture a substantial number of photographs of coral reefs located around Australia. Nevertheless, it is worth noting that a relatively small proportion of these photographs, typically around 5%, get scrutiny and analysis by experts in the field of marine biology. In addition, manual annotation is a time-consuming and labor-intensive process that requires a lot of people. Continuous monitoring of marine ecosystems is impossible without automated technology for doing so and receiving feedback from human specialists. Given these conditions, computerized annotation of underwater photographs can produce understandable visuals. The proposed findings may contribute to reducing a worldwide danger. The fact that this area of study is receiving so much interest is a strong indicator of its potential relevance. In this study, we investigate

the potential of deep learning for the automatic annotation of coral reef photographs, with the goal of mitigating some of the difficulties associated with the automatic interpretation of marine data.

A.Annotation of Coral Reefs: Methods and Challenges

In terms of biodiversity and economic value, coral reefs are among the planet's most significant ecosystems. Fringing reefs, barrier reefs, and atolls are the three primary forms of coral reefs. Nearer to the shore, fringe reefs (like the reefs off the coast of Eilat, Israel) often border islands and coasts. Barrier reefs, like Australia's Great Barrier Reef, are located offshore from their respective coasts. As a result, there is a body of water known as a lagoon between the coast and the reef. Atoll reefs (like the Lighthouse Atoll Reefs in Belize) are circular or oval in shape and can be found hundreds of feet below sea level. The primary reason to keep tabs on coral reefs throughout time is to learn more about the effects of coral bleaching. This inquiry spans both regional and international scopes.

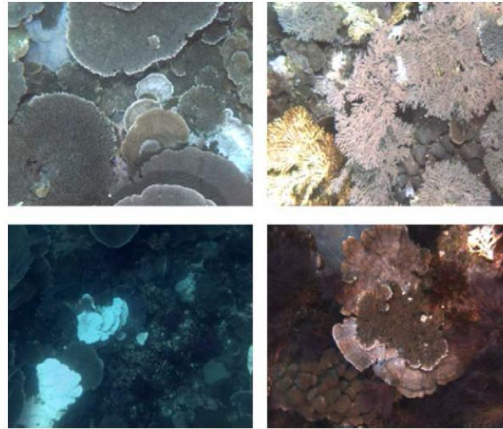


Fig. 2. Sample marine images

A wide variety of coral species exist. The Great Barrier Reef in Australia is home to a diverse array of over 600 distinct species. The species exhibits a wide range of sizes, forms, and colors, indicating a high level of diversity. The two main categories of corals are hard corals and soft corals. Soft corals lack a skeleton and are more pliable than their hard coral counterparts, leading to frequent misidentification as plants. The hard corals in a reef are the best barometer of its overall condition. The most frequent way of estimating the coral reef population is by calculating their percentage of coverage. Healthy corals bleach and die due to climate change impacts including pollution and rising sea surface temperatures.

B. Methods for Conventional Annotation

Autonomous underwater vehicles (AUVs) and other forms of underwater imaging have dramatically expanded the volume of marine data available for study. Manually annotating this data, however, is a time-consuming and ineffective operation [7]. In fact, marine scientists frequently use random point annotations, in which anywhere from 20 to 200 randomly chosen points are annotated on each image. It may take up to 30 minutes to finish annotating a single image. Class borders are unclear and difficult to describe in terms of color, form, or texture [7], making it tedious and challenging to repeat the same technique for millions of photos. Coral Point Count [9] is one piece of software that can help with this annotation system. The National Coral Reef Institute (NCRI) has released free software for the management and monitoring of coral reefs, aimed at researchers and professionals in the field. Random pixels are superimposed on a

picture using CPCe. These random pixels are then given a category classification by a marine specialist.

C. Challenges

The utilization of seafloor exploration and imaging techniques has enabled the comprehensive examination of the expansive and intricate marine ecosystem. The acquisition of data pertaining to the sea bottom is of paramount importance for the advancement of scientific understanding about the intricate ecosystems present therein. However, the logistical challenges associated with underwater environments can pose significant obstacles to the collection of such data. The impetus for the progress in imaging techniques arose from the advent of superior underwater cameras and an increasing captivation with the exploration of submerged surroundings. Optical imaging is widely used in the fields of maritime archaeology, marine geology, marine biology, and marine biodiversity [1, 10, 11]. An accurate automatic annotation method, rather than relying on manual labeling, is required to make the maritime data usable for analysis.

The field of digital photography in relation to marine life and the automated identification of species poses significant technical obstacles. Underwater classes are employed for the purpose of generating training datasets, exhibiting variations in attributes like as form, color, texture, size, rotation, illumination, view angle, camera distance, and lighting conditions. The primary obstacles encompass:

- The images exhibit notable variability within each class.

- The spatial borders between classes are intricate and intricate.
- The annotation of the images differs across experts.
- There are variations in the limitations of spatial and spectral resolution.
- Partial or total obstruction of things of interest
- The marine bottom undergoes gradual structural modifications over time.
- The presence of lighting artifacts can be attributed to the phenomenon of wave refraction.

li Literature Survey

A. Automatic Coral Classification using Machine Learning

1)Coral Classification with Hand-crafted Features

The primary criteria for categorizing corals are their color and texture. As a result, the extraction of color and texture-based, human-created elements for image representation has been the subject of substantial research. Given the corals' irregular forms and murky class boundaries, shape-encoding features are less than ideal for usage in this context. It is common practice to favor aesthetic details that combine color and texture. There are no guaranteed feature sets that can be applied to any given coral dataset and guarantee success. The features are often chosen according to the characteristics that can be used to distinguish corals from non-corals within a particular dataset.

In his work, Marcos [13] employed the color space of Normalized Chromaticity Coordinate (NCC) and the texture space of Local Binary Pattern (LBP). Five groups—living corals, dead corals, corals with algae, abiotic, and algae—were separated using a three-layer feed-forward back propagation neural network. It was suggested that LBP is resistant to variations in brightness, while NCC color properties are invariant under different lighting situations. Unfortunately, complicated underwater photos are beyond the capabilities of the NCC and LBP features. Then, using only 300 photos from each of the three coral classes, this strategy was put to the test. Combining LBP and color information was reported to have better results than combining LBP with NCC.

Normalized colour histograms were employed by Stokes and Deane [14] as colour descriptors, and a discrete cosine transform (DCT)-based feature vector was utilised to characterize the texture.

Their training set included 3000 different photos that were sorted into 18 different categories. For the purpose of categorization, a brand new method known as "probability density-weighted mean distance (PDWMD)" was presented. This strategy can be put into action quickly and easily. During the process of feature extraction, however, the weights of the colour and texture features are determined manually by the user. Additionally, DCT descriptors are not particularly robust or precise when it comes to describing textures.

Pizarro [15] employed an NCC histogram-based feature vector, bag of words (BoW), for scale-invariant feature transform (SIFT) and hue-histograms. A BoW is created from a subset of training samples, and the test picture is characterized using this vocabulary. Their approach classifies images into one of eight types. Training and vocabulary learning used 453 images. Pixel annotations with this annotation method often miss important details. This study ignores sub-image classification. BoW texturing on SIFT features is inefficient in complex underwater environments. Beijbom [8] provided the Moorea Labelled Coral (MLC) dataset with four non-coral and five coral classifications and employed a Maximum Response (MR) filter bank followed by texon maps for feature extraction at several scales. K-means clustering and a selection of training photos created a texture dictionary. They also found that L*a*b pre-processing outperforms RGB. They classified by employing an SVM classifier with an RBF kernel. Coral maps from three years were automatically annotated.

In [16], the best classification accuracy across several benthic datasets was investigated using a mixture of hand-crafted features and various classifiers. Completed Local Binary Patterns (CLBP), grey level co-occurrence matrix (GLCM), Gabor feature, and opponent angle and hue channel color histograms were among the descriptors employed. This work used scale-invariant feature vectors that were color distortion- and low-light-resistant. Classifiers included SVM, KNN, neural networks, and PDWMD. The six test datasets performed best with different feature and classifier combinations. This work did not address concerns like patch extraction scale and class overlap identification.

2) Coral Classification with Learned Features

Deep neural networks are sophisticated machine learning algorithms that stack neural networks along the depth and width of lesser structures. Deep networks have shown discriminative and representation learning capabilities in many applications recently. ML researchers are exploring new applications for deep learning. Marine scene classification is upcoming. Training deep networks requires lots of labeled data. Deep neural networks can separate millions of tagged images with effective training techniques. The trained network may also learn effective picture representations for other benthic datasets. Next, we explore deep learning and its state-of-the-art architectures before considering coral classification applications.

BAutomatic Coral Classification using Deep Learning Classification, object detection, and scene understanding need the extraction of discriminative features or picture representations from input data. Computer vision has used domain-specific hand-crafted picture representations for decades. Recently, representation learning algorithms have outperformed hand-crafted representations. Deep learning techniques use complex, deep neural networks. Many-layered neural networks can abstract data well [17]. Many cutting-edge computer vision systems succeed by extracting high-level abstractions. Support vector machines outperformed neural networks in the 2000s. The fundamental work in [18] popularized deep neural networks in computer vision.

1) Convolutional Neural Networks

One such essential class of neural networks for solving a wide variety of computer vision issues is the convolutional neural network (CNN) [18]. In particular, deep CNNs are made up of several interconnected layers of linear and non-linear processes that are learned all at once, from beginning to finish. The parameters of these layers are learned over multiple iterations in order to accomplish a certain objective. In recent years, CNN-based approaches have gained interest for extracting features from visual data.

Alternating convolutional and pooling layers make up a convolutional neural network (CNN). CNNs are robust against transformations, scaling, and distortions because of their sparse connectivity, parameter sharing, sub-sampling, and local receptive fields. When the input picture is less than

the kernel size, fewer connections are made between the input and output layers, creating sparse connectivity. Redundancy between channels and within a channel can be used to increase sparsity. In addition, fewer operations and less memory are needed to store the weights throughout the computation of the output. One weight element is multiplied by the input only once in a non-convolutional neural network. In contrast, a convolutional layer repeatedly convolves the input image with each component of the kernel matrix. In the convolutional layers, the input is convolved with stacks of filters of fixed sizes. Convolutional layers' parameter sharing is more space- and time-efficient than, say, a dense matrix multiplication. Convolutional layers can be made equivariant to linear translations (meaning that any change in the input will result in an equivalent change in the output) by sharing their parameters. But convolutional layers are not scale- and rotation-invariant.

2) Representation Learning

Data-driven learning of discriminative picture representations has emerged as a fruitful topic of study. Learning the image's characteristics allows for an accurate depiction of the data's past distributions. As a result of the hierarchical nature of these features (low and high level features), the picture representations are trained to define higher-level concepts in terms of lower-level concepts. A good learnt representation needs to be easy to understand (often dependent in a linear fashion), sparse, and consistent in both space and time. One other factor that plays a role in representation learning is the network's depth. Data representations learned in deeper layers of neural networks encode more abstract properties. Extracted image representations from CNNs have proven to be state-of-the-art in many image classification tasks [20], thanks to their training on huge datasets like ImageNet [18] and fine-tuning on domain-specific datasets. In computer vision tasks including image classification, object detection, fine-grained recognition, attribute detection, and instance retrieval, these learnt features have shown to be extremely useful as universal picture representations. Most studies have used CNN activations in the first completely linked layer. However, equivalent performances have also been

observed in the activations of intermediate convolutional layers. In the 'local feature' mode of [21], subarrays of activations from convolutional layers are retrieved and used as region descriptors. The local features extracted from two consecutive convolutional layers are then combined to form a single feature vector. Significant performance gains were seen in scene categorization tasks using this method, which is known as "cross-convolutional layer pooling" [11].

The field of computer vision is become increasingly interested in visualization techniques that might shed light on learned deep representations of images. CNN-based picture representations were previously treated as black boxes for deep feature extraction before the discovery of these techniques. Recently, in [13], a novel visualization approach was presented. Pre-images (an image created by the inverse transform of the learned representation) with prominent image representations form the basis of this technique. The term "natural pre-images" is used to describe pictures like this. The efficacy of both traditional hand-crafted representations and CNN representations was studied via the lens of three different visual representations: inversion, activation maximization, and caricaturization. It was shown that inversions using representations like HOG are more exact than those using CNN features. Different layers of CNN, however, remember important details about the input image, including differences in stance and lighting. The CNN's deep layers were able to maintain both object-level detail and global variation. Additionally, substantial differences in object arrangements were caught by completely connected layers. Convolutional layers in the middle appeared to retain local variations and structures like lines, edges, curves, and components. These findings significantly advanced our knowledge of deep generic features.

CGoing Deeper with Neural Nets

Both [17] and [8] give extensive research explaining why deep learning performs better than other shallow networks. To address the difficulty of a function encoded by a deep network, the number of unique linear regions was used as a critical parameter in [17]. It was determined that the capacity to encode pieces of input information was exponential in nature for any given layer of a deep

network. The complexity of the functions computed by the later layers increased, but they retained some of the rigidity introduced by the earlier levels' replication. The ability of deep networks to generalize to input samples they have never seen gives them an advantage over shallow models. In [8], a new approach was described that uses computational geometry for piecewise linear functions to comprehend a model's expressiveness. More linearity zones were produced by deep and narrow rectifier MLPs compared to shallow networks of the same size.

Training deep neural networks incurs a substantial computational cost as their depth and width grow larger (resulting in a greater number of weight parameters to change). Increases in dataset size, model depth, processing speed, and novel algorithms for optimization and efficient training of deeper networks have all contributed to the phenomenal success of CNNs during the past five years. Partitioning and abstraction are the two main features of traditional CNNs. By starting with extremely small filters and then expanding them as we travel deeper into the network, we can enhance partitioning. To generate an abstraction from the input patch, a typical CNN uses a linear classifier and a non-linear activation function. These simplifications lack sensitivity. [9] suggests a new structure named "Network in network (NIN)" to improve the efficacy of such abstractions. Multilayer perceptrons and micro neural networks both undergo an initialization phase.

This micro network can be thought of as an extra 11 convolutional layer that comes after the rectified linear activation layer in a standard neural network architecture. The 1x1 convolutions (small filters) increase the network's width while decreasing its computing cost by decreasing the input vector's dimension. When applied together, these two components outperform a linear classifier followed by a non-linear activation function at approximating a given function. "Network in network" refers to the process in which these smaller networks are convolved with the input image inside of a bigger network. When these structures are stacked repeatedly, deep NINs are produced. Additionally, overfitting-prone fully-connected layers are swapped out for global-average pooling layers. State-of-the-art performance was shown by deep

NINs on the CIFAR-10, CIFAR-100, and SVHN datasets.

The computational difficulty of training a neural network has been shown to be related to the depth of the network. If we keep adding layers on top of each other beyond a certain depth, the accuracy of a deep network will plateau. However, if the computations used in training are tuned properly, greater depth can improve performance. In [34], the author describes one such method, which he calls residual networks (ResNets). Many mini-CNNs, or "residual blocks," make up a residual network. Each of these residual blocks is not only layered on top of the other; rather, it is directly connected to the outputs of the following block. There are 3.6 billion multiply-add operations in a 34-layer ResNet, but there are 19.6 billion in a 19-layer VGGnet. As a result, ResNets are more conducive to training, and their accuracy throughout training does not plateau. A following study [15] demonstrated better performance with a 1001-layer deep ResNet on CIFAR-10.

lii Hybrid And Ensemble Deep Learning For Coral Classification

Variations within classes are large, class boundaries are complex, and picture clarity varies greatly in coral reefs. The ability of the derived features to reliably distinguish between similar images is crucial to the success of any classification method. One or two data characteristics, such color, form, or texture, are typically encoded via hand-crafted features. It's a tall order to come up with a fresh, hand-crafted feature representation for marine photographs that overcomes all of the obstacles that can arise. It is far more practical to use pre-trained deep network CNN features derived from a huge image dataset. When applied to a new domain, CNN features have proven to be highly discriminatory [10].

A. Hybrid and Quantized Features

Due to the pixel (instead of bounding box) annotations found in benthic datasets, CNN features cannot be directly applied in coral picture classification. Up until recently, the topic of classifying coral reefs has not been investigated using deep features. To take advantage of the complimentary benefits of different representation methods, [11] proposes applying generic deep

features extracted from VGGnet in conjunction with hand-crafted features for coral reef categorization. The dataset was too little to train a CNN with random seed inputs. Therefore, a local variation of SPP was applied to make the picture representations scale-invariant, and pre-trained CNN-based features were recovered from patches centered at labelled pixels across several scales. The CNN characteristics were supplemented by texture and color based hand-crafted features taken from the same patches. To cut down on storage space needs by a factor of 16, researchers looked into using a 2-bit feature representation approach.

On the MLC coral benchmark dataset, the suggested technique achieved a greater classification accuracy than the state-of-the-art methods. The best results were achieved by a combination of both hand-crafted and learned characteristics. It's also suggested that the CNN features and the hybrid features did a better job of addressing the issue of class disparity. When coral patches are collected at a single scale, the most common type dominates the other types. The max-pooled feature vectors give greater prominence to the less common classes because they were collected from patches of varying sizes. The outcomes of the experiments proved this. Because of this, the classifier is better able to deal with the issue of class imbalance.

B. Coral Population Analysis

Benthos15 dataset [12] was also used to test the proposed classification algorithm of [14]. Experts analysed a collection of geo-referenced photos of benthic organisms taken by an autonomous underwater vehicle (AUV) at various locations across Australia. Totaling 407,968 expert-labeled points, the collection consists of 9,874 unique photos captured at varying depths across nine sites in Australia during the previous few years. This dataset is notoriously difficult to automatically annotate due to the presence of over 40 unique class labels. In [13], the classifier was trained on a subset of this data consisting of photos from Western Australia (WA). A deep network was used to extract the multi-scale features. Automatic annotation of the unlabeled mosaics using our best classifier was also used to analyze the coral population of the Abrolhos Islands (situated off the

west coast of Western Australia). Due to a lack of ground-truth labeling, a marine specialist was brought in to create and verify coral cover maps. The coral population in the area was found to be declining using this method. It was a significant advance in our automated quest to learn how long-term environmental change affects marine ecosystems' ability to thrive.

C. Cost-sensitive Learning for Corals

Class imbalance is a problem in marine datasets just as it is in most real-world computer vision datasets. The class distribution is heavily weighted toward coral classes despite the existence of numerous non-coral classes. Due to difficulties in learning clear class borders, classification performance suffers when classes are not distributed evenly. In [12], the authors propose a cost-sensitive deep network to solve this problem. The VGG-net (16-layer variant) was used as the foundation for this network's design. A cost learning layer was introduced before to the soft-max layer of the classifier in order to prevent any tampering with the initial class distributions (which may have been caused by methods such as over-sampling and under-sampling). It was proposed that an optimisation strategy be utilised in order to achieve optimisation of the network parameters as well as the cost-sensitive layer parameters. This approach was evaluated using a large number of datasets, one of which being a coral dataset called MLC. These datasets had class imbalances. Their approach produced results that were superior to those obtained by MLC, as stated in [8]. This performance, on the other hand, does not live up to the standards set out in [11].

D. CNN with Fluorescent Images

Since most color images are processed by deep networks, the input layer typically contains three separate channels (R, G, and B). Nonetheless, in theory, a CNN can contain as many input channels as needed to encode the necessary data. In [15], for example, one such strategy is proposed. There was a mashup of reflectance and fluorescence photos with RGB ones. The final picture was made by using a pixel-by-pixel average function. The fluorescent photographs showed the corals with striking contrast, while the reflectance images gave insight into the non-fluorescent backgrounds. A convolutional neural network (CNN) was trained

using an input image with 5 channels after registration. The architecture of this CNN was quite similar to the Caffe-defined CIFAR10 architecture. Following extraction and downsampling to 32x32, patches of 128x128 were fed into three consecutive rounds of convolutional layers. For coral categorization, the 5-channel CNN outperformed its 2-channel counterpart. This 5-channel CNN was also evaluated against the gold standard of [8]. Using combined reflectance and fluorescence photos resulted in a 22% lower classification error-rate compared to using simply reflectance images.

Traditional methods of coral categorization have involved the manual extraction of traits based on color and texture. Marcos et al. [16] employed a 3-layer feed-forward back propagation neural network after employing Normalized Chromaticity Coordinate (NCC) for color and Local Binary Pattern (LBP) for texture. However, it was shown that the NCC traits were insufficiently discriminatory between certain groups of coral. Images of benthic coral were classified with a k-nearest neighbor classifier developed by Stokes and Deane [17] using an RGB histogram and a discrete cosine transform based feature vector. Although this technique is quick, it requires manual adjustment of the weights of color and texture elements. NCC histogram, SIFT bag of words, and Gabor filter response were all employed to create a feature vector that was put to use by Pizaro et al. [18]. The best pairs were voted on to determine their classification. Each image is assigned to a single category in their system, and the classification of individual pixels is ignored. The MLC dataset was first presented by Beijbom et al. [4], who extracted features using a Maximum Response (MR) filter bank and then texton maps. They also demonstrated that the performance improves when the features are extracted in the LAB color space as opposed to the RGB color space. For this, they relied on a Support Vector Machine (SVM) classifier trained with a Radial Basis Function (RBF) kernel.

Recent work by DenseNet [8] utilizing fewer hyper-parameters has showed encouraging results in the classification of common object (non-coral) datasets like CIFAR-10, CIFAR-100, SVHN, and ImageNet. The increased information flow and quicker gradient propagation during back-propagation made possible by extensive

connections between layers was demonstrated by Huang et al. [8]. Dense connections mean that every layer is linked to every other layer below it, including the loss layer. The network is able to reduce over-fitting in smaller size training sets thanks to the deep supervised learning that occurs as a result of the connection between the loss layer and each preceding layer.

Iv future Prospects

Applications of deep learning to ecological research have the potential to introduce a genuinely objective metric for the detection, discrimination, and identification of species, as well as their behavior and morphology. This will lessen the influence of observer bias and subjective interpretation, both major problems in human observational studies. Such automated processing tools will guarantee that research findings are shared openly and analysts use consistent processes. It will also allow for more systematic and objective cross-study comparisons of individuals, populations, and species. Data sets can be processed at speeds much exceeding those achievable by human experts. This is especially important to remember when performing boring, repetitive work. In studies with limited funds and extensive data, freeing up human resources for more complex tasks is becoming increasingly vital. Additional ecological game-changers include (i) a speedy quantitative survey of the vast majority (95%) of the obtained underwater images that has not yet been processed. This will make it possible to create image baselines of unprecedented size and scope for maritime environments. A quantitative assessment of climate change's effects could be made with such information; (ii) the growth, mortality, recruitment rates, and competitive abilities of marine species (such as coral reef, lobsters, kelps) associated with warming and acidification could be tracked; and (iii) our understanding of marine ecosystems, about which relatively little is known, could be greatly enhanced.

V Conclusion

We provided a brief history of deep learning and an overview of current deep neural network topologies in this survey study. We went about the importance of exploring the ocean floor and the

difficulties of doing so. Following this, we provided a brief literature review of methods for classifying maritime images. By addressing the most current work completed by our team and others, we delved deeper into the possible uses of deep learning for benthic image classification. We also shared some ideas on where to take deep learning and underwater scene understanding in the near future. We hope that by publishing this report, we might inspire many academics in the fields of computer vision and marine societies to work together on similar long-term projects.

References

- [1] P. Fairweather, "Statistical power & design requirements for environmental monitoring," *Australian Journal of Marine and Freshwater Research*, vol. 42, pp. 555-567, 1991.
- [2] A. Underwood, "Experiments in ecology," in Cambridge University Press, Cambridge, 1997.
- [3] E. Graham, M. Dethier, S. Cohen and L. Tear, "Visual versus random-point percent cover estimations: objective is not always better.," *Marine Ecology Progress Series*, vol. 96, p. 93-100, 1993.
- [4] A. Wojcicka and Z. Wrobel, "The panoramic visualization of metallic materials in macro- and microstructure of surface analysis using Microsoft Image Composite Editor (ICE)," pp. 358-368, 2012.
- [5] C. N. Rooper and M. Zimmermann, "A bottom-up methodology for integrating underwater video and acoustic mapping for seafloor substrate classification," *Continental Shelf Research*, vol. 27 (7), pp. 947-957, Apr 2007.
- [6] R. Hummel, "Image enhancement by histogram transformation, computer graphics and image processing," vol. 6, pp. 184-195, 1997.
- [7] C. E. Woodcock, "The factor of scale in remote sensing," *Remote Sensing of Environment*, vol. 21, no. 3, pp. 311-332, April 1987.
- [8] "What is spectral signature in remote sensing?," 2013. <https://articles.extension.org/pages/40071/what-is-a-spectral-signature-in-remote-sensing> (last accessed: Nov 3, 2017)
- [9] P. J. Mumby, C. D. Clark, E. P. Green and A. J. Edwards, "Benefits of water column correction

- and contextual editing for mapping coral reefs," *international Journal of Remote sensing*, vol. 19 (1), 1998.
- [10] H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa and H. Kim, "Underwater optical image processing: a comprehensive review," *Proc. CVPR*, 2017.
- [11] R. Coles, L. McKenzie, M. Rashed, J. Mellors, H. Taylor, K. Dew, S. McKenna, T. Sankey, A. Carter and A. Grech, "Status and trend of seagrass habitats in the Great Barrier Reef world heritage area," Report to the Marine and Tropical Sciences Research Facility, Reef and Rainforest Research Centre on behalf of the Australian Government's Marine and Tropical Sciences Research Facility, Cairns, Australia, 2007.
- [12] F. T. Shorta, B. Polidoro, S. R. Livingstone, K. E. Carpenter, S. Bandeira, J. S. Bujang, H. P. Calumpong and T. J. Carruthers, "Extinction risk assessment of the world's seagrass species," *Biological Conservation*, vol. 144, pp. 1961-1971, 2011.
- [13] D. I. Walker and R. I. T. Prince, "Distribution and biography of seagrass species on the northwest coast of Australia," *Aquatic Botany*, vol. 29(1), pp.19-32, 1987.
- [14] M. Vanderklift, D. Bearham, M. Haywood, H. Lozano-Montes, R. McCallum, J. McLaughlin, K. McMahan, N. Mortimer and P. Lavery, "Natural dynamics: understanding natural dynamics of seagrasses in north-western Australia," WAMSI Dredging Science Node Report Perth, Oct 2016.
- [15] M. I. SAUNDERS et al., "Coastal retreat and improved water quality mitigate losses of seagrass from sea level rise," *global change Biology*, vol. 19, p. 2569–2583, 2013.
- [16] R. Pu and S. Bell, "A protocol for improving mapping and assessing of seagrass abundance along the West Central Coast of Florida using Landsat TM and EO-1 ALI/Hyperion image," *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 116-129, Jul 2013.
- [17] C.-F. Chen, V.-K. Lau, N.-B. Chang, N.-T. Son, P.-H.-S. Tong and S.-H. Chiang, "Multi-temporal change detection of seagrass beds using integrated Landsat TM/ETM+/OLI imageries in Cam Ranh Bay, Vietnam," *Ecological Informatics*, pp. 43-54, Jul 2016.