Article

Automatic Hierarchical Classification of Kelps Using Deep Residual Features

Ammar Mahmood [1,*], **Ana Giraldo Ospina** [2], **Mohammed Bennamoun** [1], **Senjian An** [3], **Ferdous Sohel** [4], **Farid Boussaid** [5], **Renae Hovey** [2], **Robert B. Fisher** [6] **and Gary A. Kendrick** [2]

1   Computer Science and Software Engineering, The University of Western Australia, Crawley, WA 6009, Australia; mohammed.bennamoun@uwa.edu.au
2   School of Biological Sciences and Oceans Institute, The University of Western Australia, Crawley, WA 6009, Australia; ana.giraldoospina@research.uwa.edu.au (A.G.O.); renae.hovey@uwa.edu.au (R.H.); gary.kendrick@uwa.edu.au (G.A.K.)
3   School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, WA 6845, Australia; s.an@curtin.edu.au
4   College of Science, Health, Engineering and Education Murdoch University, Murdoch, WA 6150, Australia; F.Sohel@murdoch.edu.au
5   Electrical, Electronic and Computer Engineering, The University of Western Australia, Crawley, WA 6009, Australia; farid.boussaid@uwa.edu.au
6   School of Informatics, University of Edinburgh, Edinburgh EH8 9YL, UK; rbf@inf.ed.ac.uk
*   Correspondence: ammarmahmood@live.com

check for updates

Abstract: Across the globe, remote image data is rapidly being collected for the assessment of benthic communities from shallow to extremely deep waters on continental slopes to the abyssal seas. Exploiting this data is presently limited by the time it takes for experts to identify organisms found in these images. With this limitation in mind, a large effort has been made globally to introduce automation and machine learning algorithms to accelerate both classification and assessment of marine benthic biota. One major issue lies with organisms that move with swell and currents, such as kelps. This paper presents an automatic hierarchical classification method local binary classification as opposed to the conventional flat classification to classify kelps in images collected by autonomous underwater vehicles. The proposed kelp classification approach exploits learned feature representations extracted from deep residual networks. We show that these generic features outperform the traditional off-the-shelf CNN features and the conventional hand-crafted features. Experiments also demonstrate that the hierarchical classification method outperforms the traditional parallel multi-class classifications by a significant margin (90.0% vs. 57.6% and 77.2% vs. 59.0%) on Benthoz15 and Rottnest datasets respectively. Furthermore, we compare different hierarchical classification approaches and experimentally show that the sibling hierarchical training approach outperforms the inclusive hierarchical approach by a significant margin. We also report an application of our proposed method to study the change in kelp cover over time for annually repeated AUV surveys.

Keywords: deep learning; hierarchical classification; kelp cover; kelps; manual annotation; benthic marine population analysis

1. Introduction

Kelp forests support diverse and productive ecological communities throughout temperate and arctic regions worldwide. Environmental anomalies such as cyclones, storms, marine heat waves and climate change have a detrimental effect on benthic marine life including kelps [1]. Significant declines

in kelp bed were observed around the globe in recent decades, with the main drivers identified as eutrophication and climate change related environmental stressors. For instance, large-scale disappearance of kelp was observed in 2002 in the southern coast of Norway [2]. In Spain, large scale reductions in two main species of kelp have also been observed since the 1980's [3].

Similarly, kelp populations in Australia have decreased as a consequence of climate change driven environmental stressors. In the east coast of Tasmania, the coverage of giant kelp *Macrocystis pyrifera* in the present decade is around 9% of the coverage in the 1940's [4]. This decline is consistent with the intrusion of warmer, nutrient poor water from the East Australian Current, which now extends 350 km further south than in the 1940's [5]. Wernberg et al. [6] reported a rapid climate-driven transition of kelp forests to seaweed turfs in the Australian temperate reef communities with kelp forests showing a 100 km poleward contraction from their pre-heatwave distribution on the Western Australia coast. This trend is alarming for the numerous endemic species that rely on kelp forests for support. Loss of kelp forests is also a major threat for Australia's fishing and tourism industries, which generate more than 10 billion Australian dollars per annum [7]. There is thus a pressing and immediate need for monitoring programs to document changes in kelp dominated habitats along coastlines worldwide and especially in temperate Australia.

Autonomous underwater vehicles (AUVs) are emerging as highly effective tools for monitoring changes in benthic marine environments, because (i) they can autonomously conduct non-destructive sampling in remote marine habitats; (ii) they can repeatedly survey the same spatial region to detect change over time; and (iii) they are fitted with a range of instrumentation to acquire both physical and biological data. AUVs were used to monitor the marine benthos across temperate and tropical environments in Australia [8,9]; to survey invasive pest species [10]; to document rapid loss of corals associated with warming events [9,11]; to describe benthic community structure at depths greater than 1000 m [12]; and assess environmental impacts of the Deepwater Horizon oil spill [13]. In a large-scale study of deep waters, the distribution patterns of kelp forests were investigated to provide useful insights on the effect of environmental changes on the kelp population [14]. The survey took an extremely long time to complete as marine biologists had to manually classify images and to identify kelp from imagery.

AUV driven monitoring can generate large quantities of imagery. For example, an AUV deployed in Western Australia collected more than 15,000 stereo image pairs each day and was deployed between 10 and 12 days each year [9]. Manual analysis of such a large number of images per deployment (150,000 to 200,000 stereo image pairs) takes a significant amount of time and effort and is the major bottleneck in data acquisition from AUV surveys. To promptly identify changes in benthic species, especially dominant habitat formers (such as kelps and corals), it is necessary to match image-analysis time to surveying time so data can be analyzed rapidly and identification of change patterns can be accomplished. Automatic classification is critical to speed up image analysis and consequently automatic classification of benthic species has raised interest in ecologists and computer scientists (such as [15–19]). Nonetheless, automated classification of AUV collected imagery is challenging because images are captured in dynamic shallow water with little to no control on lighting and significant variations in what is visible and how it is perceived.

In this paper, we tackle the challenge of automatically annotating underwater imagery for the presence of kelp to detect changes in the coverage of Australian kelp forests. The common practice is to study the distribution and density of benthic species, which involves manually annotating a smaller dataset and then extrapolating these results to make inferences about the sites under study. Automating the process of determining kelp coverage will significantly decrease image processing times and will allow for large scale analysis of datasets and for early identification of changes in kelp cover. To automate this process, it is paramount to select appropriate features. In computer vision tasks, the general trend has shifted from conventional hand-crafted features to off-the-shelf deep features [20]. Hand-crafted features which usually encode one aspect of data (i.e., color, shape or texture) were a popular choice as image representations for benthic marine species recognition tasks

in the works of [15,18,21,22]. Moreover, given that hand-crafted features are designed specifically for a current task at hand, they generally do not perform well when applied on a different task. Recently, Convolutional Neural Networks (CNNs) and features extracted from pre-trained CNNs have become the preferred choice for benthic marine image classification tasks, e.g., [19,23–25]. These off-the-shelf features are image representations learned by a deep network trained on a larger dataset such as ImageNet. Off-the-shelf CNN features are generic and have shown better performance as compared to hand-crafted features on a variety of image recognition tasks [20]. In this paper, we propose to apply image representations extracted from deep residual networks (ResNets) to further improve the automatic annotation of benthic species. Besides better performance, one big advantage of ResNets is their faster training time and ease of optimization. Figure 1 depicts the evolution of classification pipelines for automatic benthic marine species annotation.
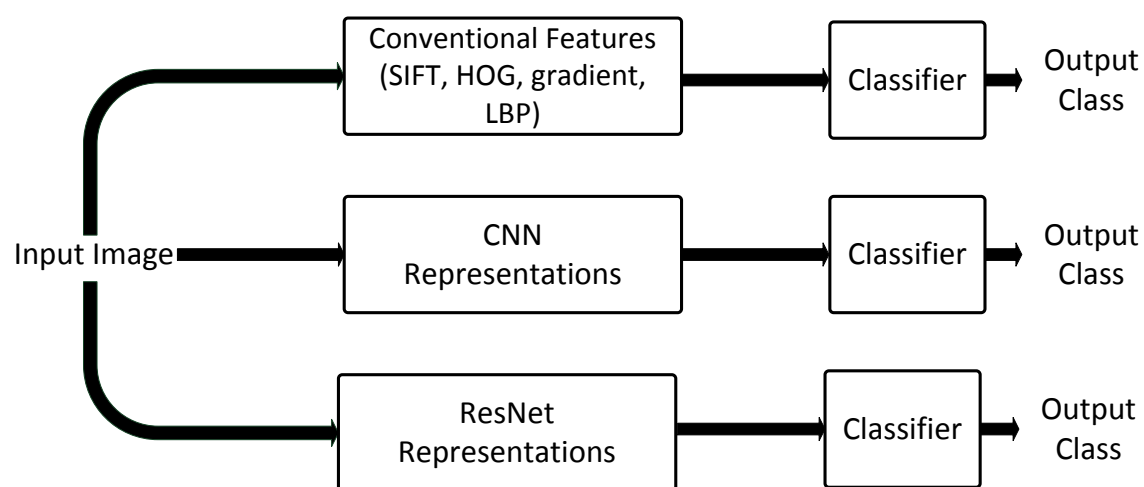
Figure 1. Evolution of classification pipelines (the most recent at the bottom). Off-the-shelf deep residual features have the potential to replace the previous classification pipelines and improve performance for benthic marine image classification tasks. (SIFT: scale invariant feature transform, HOG: histograms of gradient, LBP: local binary patterns, CNN: convolutional neural networks, ResNet: residual networks).

The main motivation for using ResNet as a base network to extract features for kelp classification is its superior performance over previous deep networks [26]. Moreover, the feature extraction is fast due to the low computational complexity of ResNets and the reduced number of floating point operations (FLOPs). Also, the feature extracted from ResNet is 2048-dimensional, which is half of the traditional 4096-dimensional feature vector of previous networks such as VGG16 [27]. These compact features result in reduced memory requirements for storing the features of large benthic marine datasets.

The main contributions of this paper are:

1. The first application of deep learning for automated kelp coverage analysis.
2. A supervised kelp image classification method based on features extracted from deep residual networks, termed as Deep Residual Features (DRF).
3. A comparison of the classification performance of the DRF with the widely used off-the-shelf CNN features for automatic annotation of kelps.
4. Experiments demonstrating DRF's superior classification accuracy compared to previous methods for kelp classification.
5. We compare hierarchical image classification with multi-class image classification and report the accuracies and mean f1-scores for two large datasets.

6.   An application of our proposed method to automatically analyze kelp coverage across five regions of Rottnest Island in Western Australia.

7.   We demonstrate the performance of the proposed kelp coverage analysis technique using ground truth data provided by marine experts and show a high correlation with previously conducted manual surveys.

The paper is organized as follows. In Section 2, we will briefly review related work. In Section 3, we present our proposed approach and explain the features extracted from deep networks. We then report the experimental results and kelp coverage analysis. In Section 4, we discuss the next steps required to implement our proposed method to a platform to rapidly analyze benthic images. Section 5 concludes this paper.

2. Related Work

2.1. Kelp Classification

Previous studies on automatic classification and segmentation of kelps in benthic marine imagery were based on hand-crafted features (Table 1). To the best of our knowledge, deep networks or features extracted from deep networks have not yet been applied to solve this problem. Here we briefly summarize a few of the prominent studies focused on automating kelp identification.

Table 1. A brief summary of methods for benthic image classification.

Authors	Methods	Classes	Main Species
Marcos et al. [15]	Color histograms, local binary pattern (LBP) and a 3-layer neural network	3	Corals
Stokes and Deane [21]	Color histograms, discrete cosine transform and probability density-based classifier	18	Corals, Macroalgae
Pizarro et al. [22]	Color histograms, Gabor filter response, scale-invariant feature transform (SIFT) and a voting-based classifier	8	Corals, Macroalgae
Beijbom et al. [18]	Maximum response filter bank with SVM classifier	9	Corals, Macroalgae
Denuelle and Dunbabin [16] *	Haralick texture features with Mahalanobis distance classifier	2	Kelp
Bewley et al. [17] *	Principal Component Analysis (PCA) and LBP descriptors with SVM classifier	19	Corals, Algae and Kelp
Bewley et al. [28] *	Hierarchical classification with PCA and LBP features	19	Corals, Algae and Kelp
Beijbom et al. [23] •	Deep neural network with reflectance and fluorescence images	10	Corals, Macrolagae
Mahmood et al. [19] •	Hybrid (CNN + handcrafted) features with a multilayer perceptron (MLP) network	9	Corals, Macrolagae
Mahmood et al. [24] •	Off-the-shelf CNN features with SVM classifier	2	Corals, Macroalgae

Key: * have reported results on kelps and • have used methods based on deep learning.

Denuelle and Dunbabin [16] utilized a technique that employed generation of kelp probability maps using Haralick texture features across an entire image. They reported that supervised and unsupervised segmentation yielded similar results. Color imbalance resulted in a significant number of false positives thus implying that the images collected must be diversified to cater for the various possible underwater lighting and visibility conditions. When compared to manual segmentation by experts, the results show good agreement.

Bewley et al. [17] presented a technique for the automatic detection of kelps using AUV gathered images. The proposed method used local image features which are fed to Support Vector Machines (SVM) [29] to identify whether kelp is present in the image under examination. Comparison of several

descriptors such as Local Binary Patterns (LBP) and Principal Component Analysis was carried out across multiple scales. This algorithm was tested on benthic data (collected from Tasmania in 2008), which contained 1258 images with 62,900 labels and 19 classes. The f1-score, which is the harmonic mean of precision and recall was used to evaluate the performance of their proposed method:

$$f1 = 2 \times \frac{precision \times recall}{precision + recall}$$

A maximum f1-score of 0.69 was reported for kelps. It was also suggested that practical systems can be built to assist scientists with automatic identification of kelps. They also concluded that results could be improved by using combinations at multiple scales, finding superior descriptors and by using more supplementary AUV data. The study concluded that for a local geographical region, and for a particular species, sufficient generalization is possible.

This work was extended in [28] for a multi-class classification problem in the presence of a taxonomical hierarchy. A local classifier was trained for each node of the hierarchy tree for LBP features and the classification results were compared through multiple hierarchy training methods. This algorithm achieved an f1-score of 0.75 for kelps and an overall mean f1-score of 0.197 for all 19 classes present in the dataset.

2.2. Deep Learning for Benthic Marine Species Recognition

In recent years, deep networks and off-the-shelf CNN features have become the first choice to tackle computer vision tasks. Only a handful of studies have developed benthic marine species recognition methods based on deep learning. Beijbom et al. [23] trained three and five-channel deep CNNs based on the CIFAR10 LeNet architecture [30] to improve the classification performance for coral and non-coral species. Reflectance and fluorescence images were registered together to obtain a five-channel image, which improved the classification performance by a significant margin. This was the first reported study to employ training of deep networks (from scratch) for benthic marine species recognition.

Off-the-shelf CNN features [20] along with multi-scale pooling were first used for coral classification in [19] on the Moorea Labelled Coral (MLC) dataset, which is a challenging dataset introduced in [18]. This paper also explored a hybrid feature approach, combining CNN features with texton maps to further improve the classification accuracy on this dataset. Class imbalance is an additional problem which refers to the disproportionate difference in the amount of points allocated to some classes compared to others. This is a common issue in benthic marine datasets, as some species are significantly more abundant than others. To address the class imbalance, a cost-sensitive learning approach was studied in [31] using off-the-shelf CNN features for MLC dataset. In another study, features extracted from pre-trained deep networks were used to generate coral population maps for the Abrolhos Islands in Western Australia [24]. This study reported a trend of decreasing live coral cover in this region. This is consistent with the manual analysis of AUV images conducted by marine researchers [9,11].

Deep residual networks (ResNets) are a special class of CNNs and are deeper, faster to train and easier to optimize than previous CNN architectures [26]. ResNets employ techniques such as residual learning and identity mapping for shortcut connections [32], which enables them to overcome the limitations of traditional CNNs and outperform them in training speed and accuracy. ResFeats, features extracted from the output of convolutional layers of a 50-layer ResNet (ResNet-50), were reported to improve the performance of different image classification tasks in [33], including coral classification on the MLC dataset. Although these features are computationally expensive large arrays, we chose to use the image representations extracted from the layers closer to the output end of ResNet-50 to reduce computation cost and alleviate the need for dimensionality reduction.

3. Methods and Results

In this section, we outline the key components of our proposed method (Figure 2) and present the adopted experimental protocols.

Figure 2. The block diagram of our proposed framework.

3.1. Datasets

3.1.1. Benthoz15 Dataset

This Australian benthic data set (Benthoz15) [34] consists of an expert-annotated set of geo-referenced benthic images and associated sensor data. These images were captured by AUV Sirius during Australia's integrated marine observation system (IMOS) benthic monitoring program at multiple temperate locations (Table 2) around Australia [8]. Marine experts manually annotated each of these images according to the Collaborative and Automation Tools for Analysis of Marine Imagery and Video (CATAMI) classification scheme. For each image, up to 50 randomly selected pixels were hand labelled using the Coral Point Count with Excel Extensions (CPCe) software package [35]. For each labelled pixel (point), a square patch of 224×224, centered at the labelled pixel is extracted. This patch is then used as an input for feature extraction. These pixels were randomly selected using CPCe for manual annotations. Several of these pixels can be found on class boundaries, making the classification problem more challenging. The whole dataset contains 407,968 expert labelled points, taken from 9,874 distinct images collected at different depths and sites over the past few years. There are 145 distinct class labels in this dataset, with pixel labels ranging from 2 to 98,380 per class. 33 out of these 145 classes belong to macroalgae (MA) species. 63,722 labelled points out of the total belong to the kelp class. Further details on the labeling methodology can be found in [34].

Table 2. Benthoz15 data.

Site	Survey Year	# of Pixel Labels	# of Images
Abrolhos Islands	2011, 2012, 2013	119,273	2377
Tasmania	2008, 2009	88,900	1778
Rottnest Island	2011	63,600	1272
Jurien Bay	2011	55,050	1101
Solitary Islands	2012	30,700	1228
Batemans Bay	2010, 2012	24,825	993
Port Stevens	2010, 2012	15,600	624
South East Queensland	2010	10,020	501
Total	-	407,968	9874

3.1.2. Rottnest Island Dataset

The Rottnest Island dataset was also collected by AUV Sirius and contains 297,800 expert labelled points, taken from 5956 distinct images collected at different depths from five sites around Rottnest Island from 2010 to 2013 (Table 3). Three out of the five sites are labelled north (15 m, 25 m and 40 m depth) and two as south (15m and 25 m depth). There are 78 distinct class labels in this dataset, with pixel labels ranging from 2 to 155,776 per class (Table A1). This makes the classification quite challenging. 25 out of these 78 classes belong to macroalgae species. 156,000 labelled points out of the total belongs to the kelp class.

Table 3. Rottnest Island data.

Survey Year	# of Images	# of Pixel Labels	# of Classes
2010	1680	84,000	61
2011	1680	84,000	55
2012	1033	51,650	44
2013	1563	78,150	55
Total	5956	297,800	78

3.2. Classification Methods

Deep residual features are extracted from the output of the last convolutional block of a 50-layer deep residual network (ResNet-50) [26] that is pre-trained on ImageNet. Figure 3 shows the architecture of the ResNet-50 deep network which we have used for feature extraction. The ResNet-50 is made up of five convolutional blocks stacked on top of each other (Figure 3). The convolutional blocks of a ResNet are different from those of the traditional CNNs because of the introduction of a shortcut connection between the input and output of each block. Identity mappings when used as shortcut connections in ResNets [32], can lead to better optimization and reduced complexity. This in turn allows one to use deeper ResNets which are faster to train and are computationally less expensive than the conventional CNNs i.e., VGGnet [27].
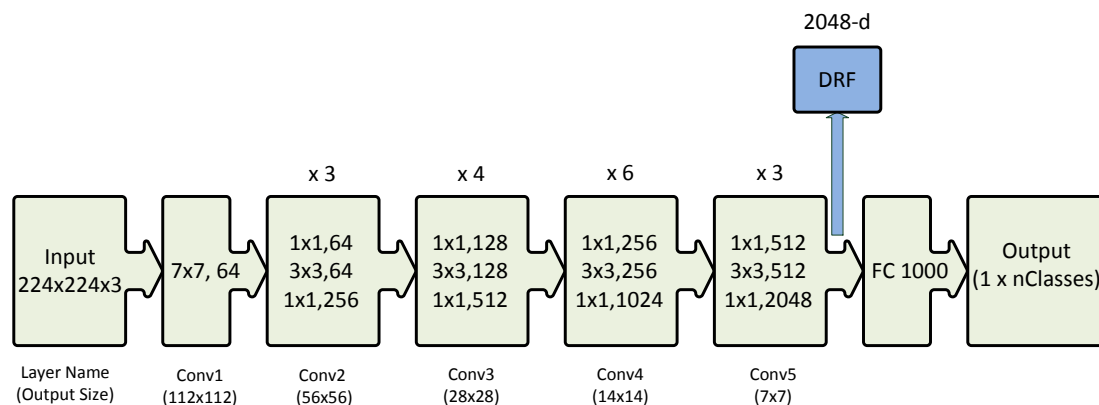
Figure 3. ResNet-50 architecture [26] shown with the residual units, the size of the filters and the outputs of each convolutional layer. DRF extracted from the last convolutional layer of this network is also shown. *Key: The notation $k \times k, n$ in the convolutional layer block denotes a filter of size k and n channels. FC 1000 denotes the fully connected layer with 1000 neurons. The number on the top of the convolutional layer block represents the repetition of each unit. nClasses represents the number of output classes.*

The image representations extracted from the fully connected layers of deep networks pre-trained on ImageNet [20] capture the overall shape of the object contained in the region of interest. The features extracted from the deeper layers encode class specific properties (i.e., shape, texture and color) and give superior classification performance as compared to features from shallower layers [36]. Hence, we propose to extract the features from the output of the last convolutional block of ResNet-50 (Figure 3). The output of the Conv5 block is a $7 \times 7 \times 2048$ dimensional array and is used as input of the FC-1000 layer. This large array is however, first converted to a 2048-dimensional vector by using a max-pool layer. We extract this 2048-dimensional vector and name it DRF. We do not use the FC-1000 layer for feature extraction because it is used as an output layer to classify the 1000 classes of the ImageNet dataset, which was used to pre-train this network. Our feature extraction method is different from the conventional method employed in previous deep networks such as VGGnet. The presence of multiple fully connected layers in the VGGnet makes the feature extraction straightforward. The only

fully connected layer in ResNet is class specific to the ImageNet dataset. Therefore, we proposed to use the output of the last convolution block for DRF extraction.

There are three different approaches described in [37] to deal with the hierarchical classification problem:

1. **Flat Classification:** This approach ignores the hierarchy and treats the problem as a parallel multi-class classification problem.
2. **Local Binary Classification:** A binary classifier is trained for every node in the hierarchical tree of the given problem.
3. **Global Classification:** A single classifier is trained for all classes and the hierarchical information is encoded in the data.

We have used the local binary classification technique in this paper to identify kelps from other taxa. This approach is easier to implement and more useful when all the nodes in the hierarchy are not labeled to a specific leaf node level. For example, some macroalgae are not labeled to the species level in the Benthoz15 dataset [34]. Moreover, this approach also allows for the use of different features, training sets and classifiers for each node of the hierarchy tree. The hierarchy tree for kelps is shown in Figure 4.
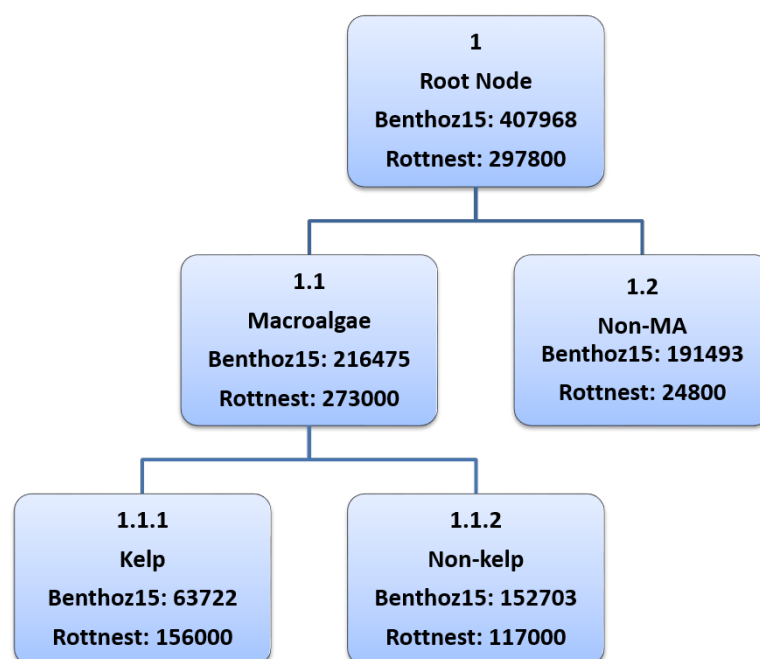
Figure 4. Hierarchy tree for kelps in our benthic data. In each node, the first line shows the node number, 2nd line shows the name of the specie, and 3rd and 4th lines show the number of labels belonging to that particular species in Benthoz15 and Rottnest Island data respectively.

3.3. Training and Testing Protocols

In this paper, two training approaches are used, namely *inclusive training* and *sibling training*. In the inclusive training method, all the non-kelp samples from the entire dataset are treated as negative samples i.e., nodes 1.2 and 1.1.2 in Figure 4. However in the sibling training method, only those non-kelp samples are considered to be negative which comes under the macroalgae node i.e., node 1.1.2 in Figure 4. We use a linear Support Vector Machines (SVM) [29] classifier because it has shown excellent performance with features extracted from deep networks [20]. We use the SVM classifier in a one-vs-all configuration with a linear kernel. We perform 3-fold cross validation within the training set to optimize the SVM parameters and mean performance are reported in Section 3.

3.4. Image Enhancement and Implementation Details

We applied color channel stretch on each image in the dataset to reduce the effect of underwater color distortion phenomenon. We calculated the averages of the lowest 1% and the highest 99% of the intensities for each color channel. The average of the lowest 1% intensities was subtracted from all the intensities in each respective channel and the negative values were set to zero. These intensities were then divided by the average of the highest 99% of the intensities. This process enhanced the color information of benthic marine images.

For feature extraction, we used a pre-trained ResNet-50 [26] deep network architecture in our experiments. We used the publicly available model of this network, which was pre-trained on the ImageNet dataset. We implemented our proposed method using MatConvNet [38] and the SVM classifier using LIBLINEAR [39] (Figure 2).

3.5. Experimental Settings and Evaluation Criteria

70% of images from each geographical location were used to form the training set for experiments carried out on the Benthoz15 dataset. However, for Rottnest Island data, the images from years 2010, 2011 and 2012 are included in the training set and the images from year 2013 form the testing set. We performed our experiments with three different classification approaches: flat classification and local binary classification with both inclusive and sibling training policies. The overall classification accuracy is not an effective measure of binary classifier performance for datasets exhibiting a skewed class distribution. Therefore, to evaluate the performance of our classifier, we have used four evaluation criteria: overall classification accuracy, mean f1-score (the average of f1-scores of each class involved in the test data), precision and recall values of kelp.

3.6. Classification Results

In this section, we report the results of three different types of features for the three training methods on the two datasets: (i) Maximum Response (MR) filter and texton maps of [18] as baseline handcrafted features. We used a publicly available implementation of this method; (ii) CNN features extracted from a VGG16 network pretrained on ImageNet dataset [27]; (iii) Our proposed DRFs extracted from a pretrained ResNet-50.

Classification by the DRF method always outperformed the traditional CNN features and MR features in both datasets as it consistently showed higher accuracy, higher f1 scores, higher precision of kelps and higher kelp recall than previously used features. Additionally, hierarchical classification (sibling and inclusive) in comparison to flat classification, also improved f1-score and recall of kelps while providing lower training times. The sibling training method achieved the highest f1-score for both datasets. Because f1-score is an evaluation metric based on both precision and recall, we recommend the sibling training method as the top performing practical method for classification and automated coverage analysis of kelps.

3.6.1. Benthoz15 Dataset

To highlight the superior classification performance of DRF, we have included a comparative study among DRF and the traditionally used CNN features extracted from VGGnet [27] and MR features (Table 4). The DRF method performs better than both the features for all three classification experiments. The lowest overall accuracy was achieved by the flat multi-class classification method (57.6%). Additionally, a very low mean f1-score of 0.05 was observed, since many classes among the total 145 had very few samples for training and testing. Nonetheless, the flat classification method achieved the highest precision (71%) for kelps among all the three methods. Out of every 100 kelp samples, this method correctly identifies 71 samples as kelps. However, this method resulted in the worst recall value of 65% (Table 4).

The best classification accuracy is achieved with the inclusive training method (90%) for which all the non-kelp samples are bundled together in the negative class. This training scheme achieves a mean f1-score of 0.79 which is similar to the highest f1-score of 0.80 obtained using the sibling training method (Table 4).

The sibling training method is more challenging as compared to the inclusive training method because the negative samples only include macroalgae classes and some of these classes are very similar to kelp in appearance. This accounts for a drop in classification accuracy from 90% to 83.4%. However the sibling training method resulted in the highest mean f1-score (0.80) and recall value (78%) for kelp. Moreover, statistical testing supports the hypothesis that all three DRF classifiers are better than their VGG and MR counterparts at significance level of 0.05. For each DRF feature X and competing feature $Y \in (MR, VGG)$, we did a paired t-test over randomly chosen image samples ($N = 50,000$), using the SVM classifier. Statistical results showed that, for each pairing of features (X, Y), feature X gave better classification than feature Y at the 0.05 significance level. The calculated p-value was less than 0.05 which rejected our null hypothesis that both classifiers show similar performance.

Table 4. A comparison of flat, inclusive and sibling classification methods for kelp classification on Benthoz15 dataset for MR, VGG and DRF methods. The flat classification focuses on all the classes present in the dataset whereas the inclusive and sibling classification only includes kelps and non-kelps. Mean f1-score corresponds to the average of the individual f1-score of each class involved in the experiment. Best scores are shown in bold font.

Method	Accuracy (%)	Mean f1-score	Precision of Kelps (%)	Recall of Kelps (%)
MR: Flat	51.6 ± 0.3	0.03 ± 0.00	64 ± 0.5	59 ± 0.5
MR: Inclusive	82.8 ± 0.4	0.70 ± 0.03	43 ± 0.0	69 ± 0.0
MR: Sibling	79.6 ± 0.3	0.72 ± 0.02	55 ± 0.0	73 ± 0.0
VGG: Flat	54.4 ± 0.6	0.03 ± 0.01	67 ± 0.5	63 ± 0.5
VGG: Inclusive	89.0 ± 0.5	0.75 ± 0.02	47 ± 0.0	73 ± 0.0
VGG: Sibling	82.1 ± 0.4	0.76 ± 0.01	57 ± 0.0	75 ± 0.0
DRF: Flat	57.6 ± 0.5	0.05 ± 0.02	**71 ± 1.0**	65 ± 1.0
DRF: Inclusive	**90.0 ± 0.07**	0.79 ± 0.02	58 ± 0.0	73 ± 0.0
DRF: Sibling	83.4 ± 0.2	**0.80 ± 0.01**	65 ± 0.0	**78 ± 0.0**

3.6.2. Rottnest Island Dataset

The DRF was then applied to the Rottnest Island data and once again confirmed that the DRF outperformed the VGG and MR features for all the classification experiments (Table 5). The hierarchical methods performed better than the flat classification method for all evaluation criteria except for precision. However, the recall value achieved by this method is the worst. This is consistent with the results obtained on Benthoz15 dataset. The mean f1-score for flat classifier (0.03) is again very low given the fact that all 78 classes are classified at the same time. The sibling training method comes out as the best method with respect to accuracy (77.2%), mean f1-score (0.76) and recall value (79%) of kelps. Moreover, the sibling training method is also the fastest method because it has less negative examples than the inclusive method.

Fine-tuning a deep network is also a popular approach for transfer learning [40]. We also compared our proposed method with fine-tuning. Fine-tuning a ResNet-50 on Rottnest Island data achieved an overall classification accuracy of 58.8% as compared to the 59.0% achieved by our proposed method. For Benthoz15 dataset, fine-tuning a ResNet-50 resulted in an overall classification accuracy of 57.1% which is 0.5% lower than our proposed method. The performance change was marginal for both datasets. Hence, we concluded that the classification accuracy achieved by both methods on benthic marine datasets is comparable. One important aspect to compare is the computational time required by these two approaches. The time needed to extract off-the-shelf features from a ResNet and classify them using an SVM classifier is far less than the time required to fine-tune a 50 layer

ResNet on a dataset as large as 297,800 input images. Our proposed method requires a few hours to run. However, fine-tuning a ResNet-50 with Rottnest Island dataset takes at least 2 days on an Nvidia Titan-X GPU. Given these considerations, we selected our proposed method over fine-tuning a ResNet with a marine dataset approach.

Table 5. A comparison of flat, inclusive and sibling classification methods for kelp classification on Rottnest Island dataset for MR, VGG, and DRF methods. The flat classification focuses on all the classes present in the dataset whereas the inclusive and sibling classification only includes kelps and non-kelps. Mean f1-score corresponds to the average of the individual f1-score of each class involved in the experiment. Best scores are shown in bold font.

| Method	Accuracy (%)	Mean f1-score	Precision of Kelps (%)	Recall of Kelps (%)
MR: Flat	52.9 ± 0.4	0.02 ± 0.00	90 ± 2.0	62 ± 1.0
MR: Inclusive	73.2 ± 0.6	0.70 ± 0.01	77 ± 0.0	74 ± 0.0
MR: Sibling	71.7 ± 0.4	0.71 ± 0.01	80 ± 0.0	73 ± 0.0
VGG: Flat	58.6 ± 0.6	0.02 ± 0.01	95 ± 1.5	65 ± 1.0
VGG: Inclusive	74.7 ± 0.4	0.74 ± 0.02	81 ± 0.0	75 ± 0.0
VGG: Sibling	74.5 ± 0.3	0.73 ± 0.02	84 ± 0.0	75 ± 0.0
DRF: Flat	59.0 ± 0.7	0.03 ± 0.01	**95 ± 1.0**	66 ± 1.0
DRF: Inclusive	75.0 ± 0.5	0.75 ± 0.01	82 ± 0.0	75 ± 0.0
DRF: Sibling	**77.2 ± 0.4**	**0.76 ± 0.02**	86 ± 0.0	**79 ± 0.0**

One of many challenges in benthic cover estimations through image analysis is the large amount of time required to manually classify the imagery. The average time for manual annotation with 50 sample points per image is 8 minutes. A trained marine expert can annotate up to 8 images per hour. The proposed method is significantly less time consuming as it results in an annotation rate of 1800 images per hour using a Nvidia Titan-X GPU. This is approximately 225 times faster than manual annotation by experts. Nonetheless, note that the proposed machine learning algorithm is only classifying 'kelp' vs. 'non kelp'. Although it is faster, it is not yet trained to classify 145 potential benthic classes. This paper evaluates the technique for a single class and presents a way forward to develop the methodology for other classes and faster processing times, which will allow scientists to promptly analyze changes in benthic community composition.

3.7. Kelp Coverage Analysis

We extended our method to estimate kelp cover for the Rottnest Island dataset. The expert identified coverage was calculated by aggregating the pixel level ground truth labels in every image. We calculated the estimated kelp coverage by aggregating the predicted labels for the same locations for which the expert labels were available. Kelp cover estimated by the annotations generated by our proposed method was compared to the cover based on expert classification (Figure 5; Table 6). Scatter plots were generated for each of five sites and all the data included in the 2013 test set. An important application of our proposed method is to estimate the population trends of kelp across spatial and time scales. To accomplish this task, we split the Rottnest Island data into sites and trained a classifier on this basis instead of years. The three sites from the north constitute the training set and the two southern sites form the test set.

The first sub-plot in Figure 5 shows kelp coverage for all of the data included in the test set. The slope of the line generated by linear regression is very close to the ideal case. This highlights the robustness of our proposed algorithm. The remaining sub-plots show kelp coverage for each of the five sites. These sub-plots show a good agreement between the annotations generated by our proposed method and the annotations provided by the human experts (Table 6). Moreover, we also calculated the R-squared (R^2) value for each plot to show correlation between the actual and predicted cover. Our proposed method achieved a high R^2 value for each individual site and then all sites combined.

It is important to note that the DRF classification seems to over-fit kelp cover at high percentages of cover and to under-fit kelp cover at lower ones.
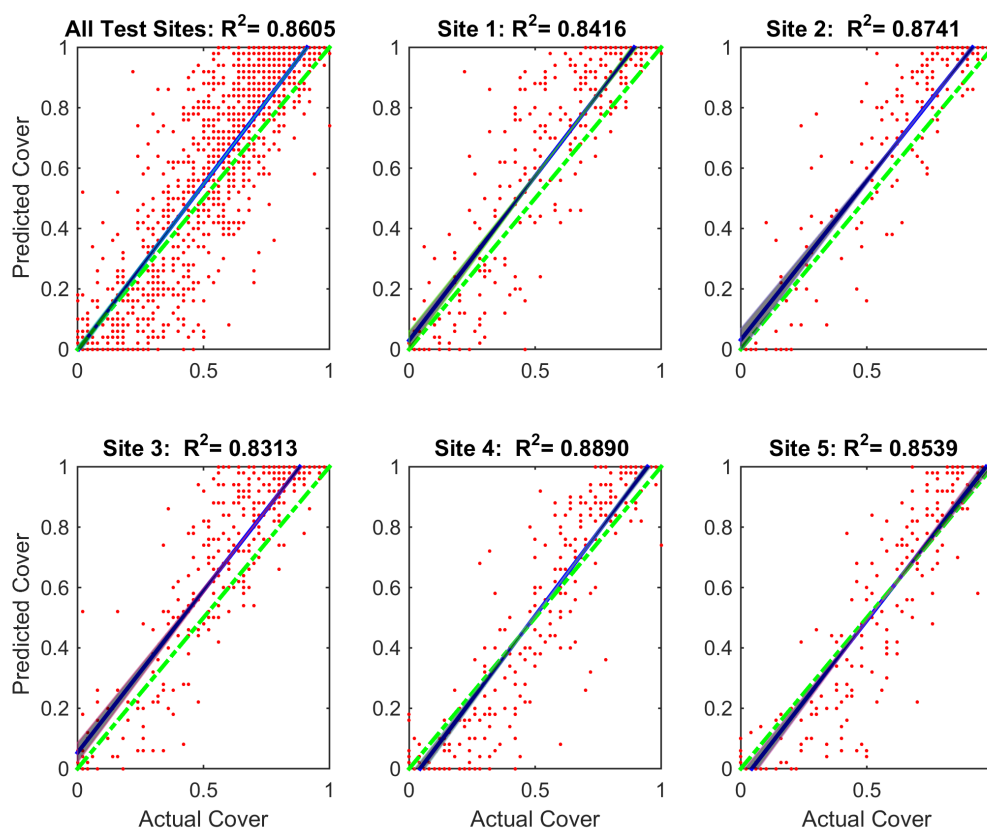
Figure 5. Coverage estimation scatter plots for Rottnest Island Data for the DRF: Sibling Training experiment. Each dot indicates the estimated cover and the actual cover per image. The dashed green line represents the perfect estimation. The blue line on each plot is the linear regression model and the shaded area represent the 95% confidence intervals. The first plot is the aggregated plot of the remaining plots of the five sites included in the 2013 test data. R^2 value for each sub-plot is shown in the respective title.
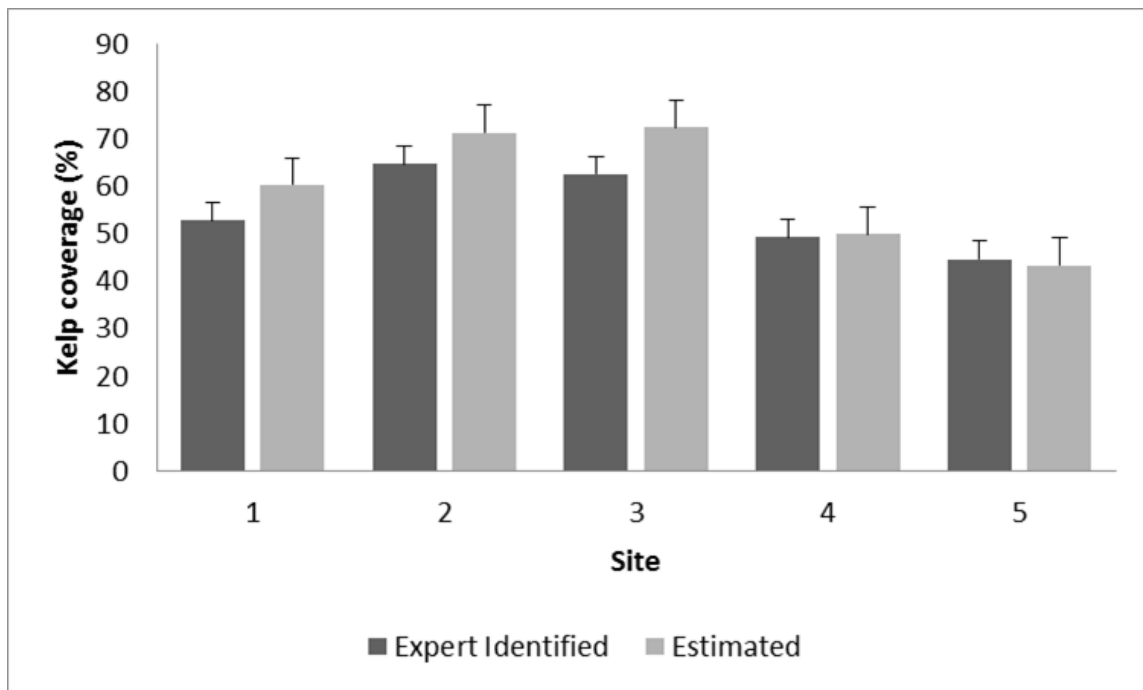
Table 6. Expert identified and estimated kelp coverage for all five sites of Rottnest Island data for the year 2013 along with the R^2 values.

| Site	Depth and Location	Expert Identified (%)	Estimated (%)	R^2
1	15 m North	52.65	60.19	0.84
2	15 m South	64.64	71.23	0.87
3	25 m North	62.44	72.32	0.83
4	25 m South	49.24	49.78	0.89
5	40 m North	44.60	43.28	0.85

The estimated kelp coverage is not significantly different from the coverage calculated by the experts from the ground truth labels (Figure 6). This indicates the robustness of our proposed method for estimating kelp coverage. These results are beneficial to marine scientists since many surveys focus on estimating kelp coverage, which is an important metric to indicate the health of kelp forests.

Figure 7 shows the expert identified and estimated percent cover of kelp across years of sites 2 and 4. For site 2, a slight over estimation of kelp cover by the DRF classification is visible, however no distinct trend of change across years is observable in either manual or automatic classification.

On the other hand, the estimation of kelp cover for site 4 shows no overestimation and similarly to site 2, no trend change in kelp cover over the years.

Figure 6. Expert identified and estimated kelp coverage for all five sites of Rottnest Island data for the year 2013.
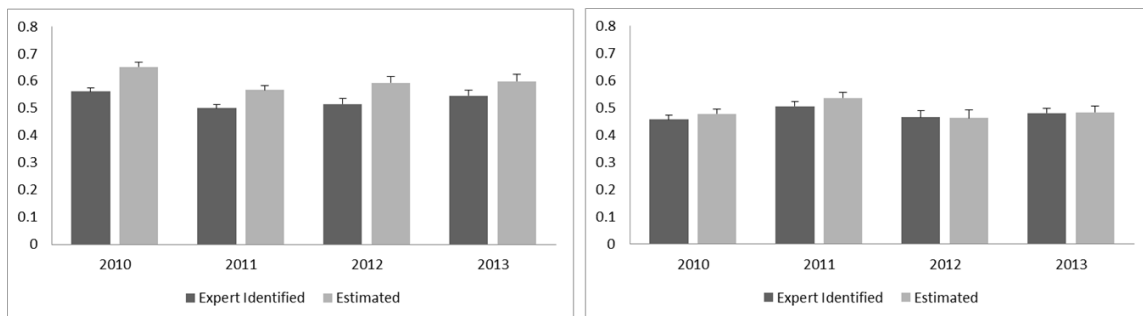
Figure 7. Expert identified and estimated kelp coverage for the two southern sites of the Rottnest Island data. Left: Site 2, Right: Site 4.

4. Discussion

The use of AUVs to survey benthic marine habitats has allowed scientists to investigate remote locations such as off-shore and deep sites, which are beyond the limits of traditional SCUBA diving. Nonetheless, the efficiency of image collection does not match the availability of data for ecological analysis, as image classification is time consuming and costly given that it is performed manually by marine experts. Additionally, manual classification has other disadvantages such as observer discrepancies and biases. Automated analysis of imagery is thus essential to fully benefit from the advantages of remote surveying technologies such as AUV's. In this study, we have addressed this problem by evaluating a machine learning automated image classification method using Deep Residual Features (DRF) for a key marine benthic species: the kelp *Ecklonia radiata*.

We have demonstrated that the image representations extracted from pre-trained deep residual networks can be effectively used for benthic marine image classification in general and kelps in particular. These powerful and generic features outperform traditional off-the-shelf CNN features, which have already shown superior performance over conventional hand-crafted features [19,20].

The sibling and inclusive hierarchical training methods further enhance performance when compared to flat multi-class classification methods. The sibling and inclusive training methods show comparatively similar performance. However, the sibling method is superior because it has lower training time than the inclusive method. Furthermore, estimations of kelp cover by automated DRF classification closely resemble those of manual expert classifications with the added advantage of faster processing times. This work provides evidence that automatic annotations may save resources and time while providing effective estimates of benthic cover.

This method was also applied on a dataset to compare kelp coverage for multiple sites, across three depths and for a consecutive time series of four years (2010–2013) at Rottnest Island. The patterns observed showed differences in percent cover of the kelp *Ecklonia radiata* between sites (with higher percentage cover of kelp in shallower sites compared to deeper sites) and no considerable change of kelp cover across years. These trends were similar to those observed by manually classified data once more confirming the usefulness of automated image classifying methods and the ability to use them for ongoing monitoring of kelp beds with AUV technology.

In this study, we found no evidence of catastrophic loss of kelp over the years at any of the sites surveyed at Rottnest Island. These results are comparable to previous estimates of change in *E. radiata* cover across depth in Australia, performed with manually classified images [14]. They are in contrast with trends of significant and continuous kelp decline reported in the region after an extreme marine heatwave which resulted in widespread mortality of benthic species including corals, seagrasses, invertebrates and kelp [6]. The loss of kelp in Western Australia resulted in a range contraction of 100 km [6] and in crab and scallop fishery closures of benthic species associated with kelp habitat. Importantly, the kelp loss was reported in habitats shallower than 15 m, with little attention to the response of deeper habitats to the heatwave [9]. This may be why our results contrast with studies reporting catastrophic loss of kelp, since our shallowest locations were at 15 m of depth, and most in situ studies take place even shallower (about 12 m). Additionally, all our sites were located off-shore (even the shallow ones), which may indicate that off-shore sites are less impacted by environmental pressures. This may be due to the lack of other environmental disturbances that coastal habitats are exposed to, due to their distance to shore and human populations. The interaction of several disturbances was shown to cause ecological responses such as wide spread mortality of marine benthic species [41]. Kelps growing offshore and in deeper locations (>15 m of depth) appear to be less impacted by extreme warming in contrast to coastal shallow reefs [42]. As a result of the catastrophic consequences that extreme climatic events may have on key habitat building species, such as kelp, deeper marine regions were identified as potential refugia for shallow marine species [43–45]. This emphasizes the importance of AUV surveys to provide information on offshore and deep locations which may be influenced by different factors to their inshore counterparts [9]. The use of automated image analysis for processing AUV images will streamline the processing of these images to efficiently identify patterns observed in deep and remote locations and compare them with patterns observed in shallow and inshore sites.

The rapid characterization of ecological changes is crucial in light of the catastrophic threats to marine biodiversity posed by the rise of extreme climatic events driven by climate change and other anthropogenic stressors. Technology has enabled the rapid collection of images even in remote locations through autonomous underwater vehicles, remotely operated vehicles, automated cameras and even satellite imagery. The subsequent annotation of such imagery is typically time consuming and consequently, the automation of marine species classification from digital images has become a priority. This study focuses on the kelp species *E. radiata*, which is the dominant habitat builder of temperate reefs in Australia, though automated classification of marine species was applied to other important marine species. For example, progress in automated tropical coral identification has resulted in accurate classification the level of genera [46] . Other successful automated classification techniques for coral reefs include the collection of multifaceted data, minimum manual classification effort (around 2% of pixels) and machine learning techniques which result in cm-scale benthic habitat

maps of high taxonomic resolution and accuracy of up to 97% [47]. Similarly, in pelagic species such as fish automated classification has advanced rapidly, with automated fish detection and identification algorithms also measuring basic fish morphological features such as total length [48,49]. In contrast, automated methods for identification of marine macroalgae from benthic images still result in low agreement [46], highlighting the need for more research into unequivocal definitions of algal groups for image classification.

Although the proposed DRF classification method allowed us to compare kelp cover in different sites and across different years providing marginal differences with the estimations from manual annotations, there were some errors associated with the proposed technique. We observed an over-prediction of kelp at high percentage cover and under-prediction at low cover. Nonetheless, the over prediction was smaller when data was divided per site and in some sites was negligible (4 and 5). Overall, the estimated kelp cover closely resembles manual classification and taking into consideration the cost effectiveness of automated DRF classification methods, the benefits of the automated classification method out-weight the drawbacks. As such, automated classification of kelp from AUV-derived images constitute a cost-effective method for estimations of kelp abundance across space and time.

A comparison of the best overall accuracies of hierarchical classification across the two used datasets shows that both the sibling and inclusive DRF classifiers has shown better classification accuracy on Benthoz15 dataset as compared with Rottnest Island dataset. For example, the inclusive DRF classifier for Benthoz15 dataset (Table 4) has an absolute gain of 15% over the respective classifier for the Rottnest dataset (Table 5). This substantial difference is possibly due to the high presence of the brown algae *Scytothalia dorycarpa* in the Rottnest Island data. *Scytothalia dorycarpa* is very similar to kelp in appearance and usually occurs in areas of the sea floor with high cover of kelp. Therefore, marine scientists may mis-classify it as kelp in poor quality images. This misclassification is possible if the point falls on the edge of *Scytothalia dorycarpa*, where the boundary between the two species is not clear. The expert misclassification of *Scytothalia dorycarpa* as kelp may also explain the over-prediction of kelp by the DRF classification method at high percentage cover. The over-prediction of the automated classification is actually an overestimation of the kelp cover by the manual annotation method. The subjectivity in the classification is removed by the automated analysis, which uses several features to classify kelp. Figure 8 illustrates the similarity of appearance of these two species.
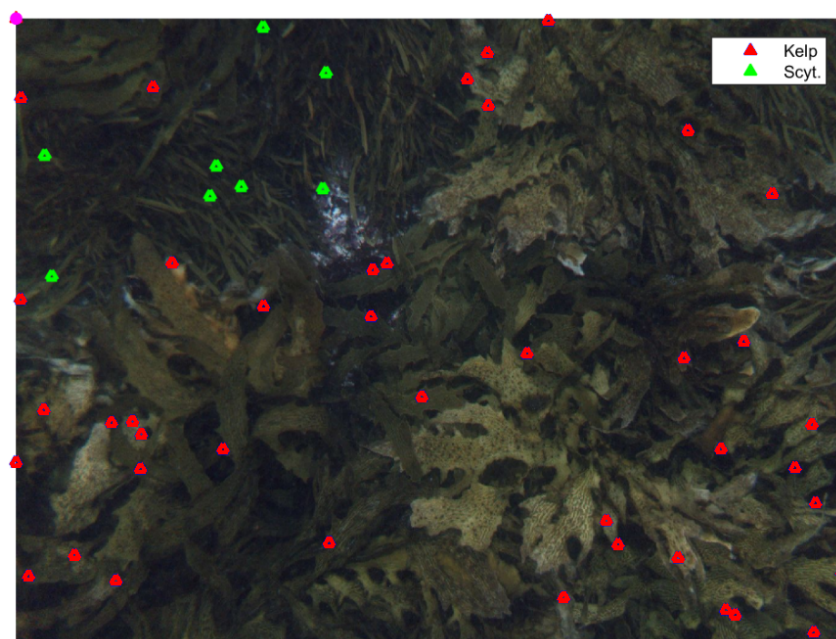
Figure 8. An example image from Rottnest Island Dataset with manual annotations showing similarity in appearance between *Scytothalia dorycarpa* (green) and the kelp *Ecklonia radiata* (blue).

Poor quality images (low light and resolution) will also affect the manual classification of other classes of algae such as 'turf matrix', 'fine branching red algae' or other canopy forming brown algae. These and other algae classes are not as common as kelp at the sites surveyed at Rottnest Island. Thus, misclassification associated with manual annotations may also explain the over prediction of kelp at low percentage covers. At low cover of kelp, a turf and foliose matrix of red algae occurs on the rocks. In areas of low kelp cover it is easy for an expert to distinguish kelp from other classes, but perhaps due to the imbalance of data for training the classifier sometimes other classes are classified as kelp resulting in over-prediction by the DRF classification method. These issues highlight the need for larger training datasets for deep learning-based automatic annotation. Extensive and comprehensive training sets will allow for better classifier training and give the opportunity to increase the amount of biota classified automatically (e.g., other algae species, corals, sponges, invertebrates such as sea urchins, and lobsters). Future work will explore multi-class classification of benthic marine species across diverse benthic habitats so methods based on deep learning algorithms can be applied to numerous ecological problems that include other benthic marine species. Scientists who use data extracted from image classification should keep these considerations in mind when manually annotating images since these datasets are extremely valuable for deep learning-based automatic classification.

5. Conclusions

The aim of this study was to investigate deep learning techniques for automatic annotation of kelp species in a complex underwater scenery. Towards this end, we evaluated a Deep Residual Features (DRF)-based method to carry out this task and showed that it outperformed the widely adopted off-the-shelf CNN based classification. We also established that hierarchical classification with the sibling method gave superior results compared to the flat multi-class approach with the added advantage of faster training times. Our results suggest that the proposed automatic kelp annotation method can significantly reduce the number of human-hours spent in manual annotations. Additionally, our proposed method can enhance the effectiveness of AUV monitoring campaigns by facilitating the early detection of changes in the population of key species though rapid image processing times, as demonstrated with examples from the Rottnest Island dataset. To conclude, the proposed DRF based automatic annotation of benthic images is to this date the most accurate machine learning technique for estimation of kelp cover.

Author Contributions: A.M. conceived the idea, designed the experimental protocols and led the writing of the manuscript. A.G.O. and R.H. collected the data and provided manual annotations. M.B. and F.B. provided critical feedback for the overall manuscript. S.A. and F.S. helped to develop the methodology from a machine learning perspective. A.G.O. and G.A.K. developed the discussion section and helped in interpreting the results from a marine scientist's perspective. R.B.F. assisted with the statistical analysis of the results and provided important revisions. All authors contributed critically to the drafts and gave final approval for publication. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Class Distribution of Rottnest Island Data.

Label	Training Samples	Test Samples	CATAMI Class ID
1	1	0	AUC
2	0	1	AUS
3	2	0	BMC

Table A1. *Cont.*

Label	Training Samples	Test Samples	CATAMI Class ID
4	483	294	BRYH
5	20	13	BRYS
6	20	0	CB
7	1	0	CBBF
8	2	0	CBBH
9	7	0	CBOT
10	0	3	CNHYC
11	3	0	CNHYD
12	7	1	CSBL
13	44	19	CSBR
14	1	1	CSBRBL
15	15	3	CSCOLBL
16	2	0	CSCOR
17	2	2	CSCORBL
18	7	3	CSDBL
19	265	38	CSE
20	24	1	CSEBL
21	887	355	CSF
22	46	2	CSFBL
23	7	3	CSM
24	50	8	CSSO
25	1	0	CSSOBL
26	0	2	CSST
27	1	0	CSSUBL
28	1	1	CST
29	1	0	CSTBL
30	10	7	EF
31	47	2	ESC
32	15	1	ESS
33	102	31	FELR
34	0	3	MAAG
35	2644	2561	MAAR
36	37	0	MACAU
37	66	113	MAECB
38	1	1	MAECG
39	112762	43014	MAECK (Kelp)
40	2419	1124	MAECR
41	1733	173	MAEFB
42	1	1	MAEFG
43	2839	586	MAEFR
44	6744	1300	MAENB
45	29948	11686	MAENR
46	1252	2073	MAFR
47	2	0	MAGB
48	9	0	MAGG
49	1	0	MAGR
50	4	0	MALAB
51	2	0	MALAR
52	285	87	MALCB
53	3	1	MAPAD
54	1177	2391	MASAR
55	52	6	MASB
56	16571	3366	MASCY
57	137	0	MASR
58	24637	4846	MATM
59	2	0	RH
60	1505	163	SC

Table A1. *Cont.*

Label	Training Samples	Test Samples	CATAMI Class ID
61	14	13	SCC
62	2	0	SEAGSAA
63	18	3	SEAGSAG
64	0	3	SEAGSPA
65	1	3	SEAGSPC
66	2	0	SEAGSPS
67	1	0	SEAGSZ
68	106	15	SHAD
69	2013	1201	SPC
70	400	214	SPCL
71	110	125	SPEB
72	123	36	SPEL
73	289	347	SPES
74	69	0	SPM
75	23	6	SUPBC
76	164	4	SUPBR
77	9340	1893	SUS
78	68	1	UNK

References

1. Doney, S.C.; Ruckelshaus, M.; Duffy, J.E.; Barry, J.P.; Chan, F.; English, C.A.; Galindo, H.M.; Grebmeier, J.M.; Hollowed, A.B.; Knowlton, N.; et al. Climate change impacts on marine ecosystems. *Mar. Sci.* **2012**, *4*, 11–37.
2. Moy, F.E.; Christie, H. Large-scale shift from sugar kelp (Saccharina latissima) to ephemeral algae along the south and west coast of Norway. *Mar. Biol. Res.* **2012**, *8*, 309–321.
3. Fernández, C. The retreat of large brown seaweeds on the north coast of Spain: The case of Saccorhiza polyschides. *Eur. J. Phycol.* **2011**, *46*, 352–360.
4. Johnson, C.R.; Banks, S.C.; Barrett, N.S.; Cazassus, F.; Dunstan, P.K.; Edgar, G.J.; Frusher, S.D.; Gardner, C.; Haddon, M.; Helidoniotis, F.; et al. Climate change cascades: Shifts in oceanography, species' ranges and subtidal marine community dynamics in eastern Tasmania. *J. Exp. Mar. Biol. Ecol.* **2011**, *400*, 17–32.
5. Ridgway, K. Long-term trend and decadal variability of the southward penetration of the East Australian Current. *Geophys. Res. Lett.* **2007**, *34*, doi:10.1029/2007GL030393.
6. Wernberg, T.; Bennett, S.; Babcock, R.C.; de Bettignies, T.; Cure, K.; Depczynski, M.; Dufois, F.; Fromont, J.; Fulton, C.J.; Hovey, R.K.; et al. Climate-driven regime shift of a temperate marine ecosystem. *Science* **2016**, *353*, 169–172.
7. Bennett, S.; Wernberg, T.; Connell, S.D.; Hobday, A.J.; Johnson, C.R.; Poloczanska, E.S. The 'Great Southern Reef': Social, ecological and economic value of Australia's neglected kelp forests. *Mar. Freshw. Res.* **2016**, *67*, 47–56.
8. Williams, S.B.; Pizarro, O.R.; Jakuba, M.V.; Johnson, C.R.; Barrett, N.S.; Babcock, R.C.; Kendrick, G.A.; Steinberg, P.D.; Heyward, A.J.; Doherty, P.J.; et al. Monitoring of benthic reference sites: Using an autonomous underwater vehicle. *IEEE Robot. Autom. Mag.* **2012**, *19*, 73–84.
9. Smale, D.A.; Kendrick, G.A.; Harvey, E.S.; Langlois, T.J.; Hovey, R.K.; Van Niel, K.P.; Waddington, K.I.; Bellchambers, L.M.; Pember, M.B.; Babcock, R.C.; et al. Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). *ICES J. Mar. Sci.* **2012**, *69*, 1108–1118.
10. Barrett, N.; Seiler, J.; Anderson, T.; Williams, S.; Nichol, S.; Hill, S.N. Autonomous Underwater Vehicle (AUV) for mapping marine biodiversity in coastal and shelf waters: Implications for marine management. In Proceedings of the OCEANS 2010 IEEE-Sydney, Sydney, Australia, 24–27 May 2010; pp. 1–6.
11. Bridge, T.C.; Ferrari, R.; Bryson, M.; Hovey, R.; Figueira, W.F.; Williams, S.B.; Pizarro, O.; Harborne, A.R.; Byrne, M. Variable responses of benthic communities to anomalously warm sea temperatures on a high-latitude coral reef. *PLoS ONE* **2014**, *9*, e113079.
12. Sherman, A.D.; Smith, K. Deep-sea benthic boundary layer communities and food supply: A long-term monitoring strategy. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2009**, *56*, 1754–1762.

13. Camilli, R.; Reddy, C.M.; Yoerger, D.R.; Van Mooy, B.A.; Jakuba, M.V.; Kinsey, J.C.; McIntyre, C.P.; Sylva, S.P.; Maloney, J.V. Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon. *Science* **2010**, *330*, 201–204.

14. Marzinelli, E.M.; Williams, S.B.; Babcock, R.C.; Barrett, N.S.; Johnson, C.R.; Jordan, A.; Kendrick, G.A.; Pizarro, O.R.; Smale, D.A.; Steinberg, P.D. Large-scale geographic variation in distribution and abundance of Australian deep-water kelp forests. *PLoS ONE* **2015**, *10*, e0118390.

15. Marcos, M.S.A.; Soriano, M.; Saloma, C. Classification of coral reef images from underwater video using neural networks. *Opt. Express* **2005**, *13*, 8766–8771.

16. Denuelle, A.; Dunbabin, M. Kelp detection in highly dynamic environments using texture recognition. In Proceedings of the Australasian Conference on Robotics & Automation (ACRA), Brisbane, Australia, 1–3 December 2010.

17. Bewley, M.; Douillard, B.; Nourani-Vatani, N.; Friedman, A.; Pizarro, O.; Williams, S. Automated species detection: An experimental approach to kelp detection from sea-floor AUV images. In Proceedings of the Australasian Conference on Robotics and Automation, Wellington, New Zealand, 3–5 December 2012.

18. Beijbom, O.; Edmunds, P.J.; Kline, D.; Mitchell, B.G.; Kriegman, D. Automated annotation of coral reef survey images. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1170–1177.

19. Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F.; Hovey, R.; Kendrick, G.; Fisher, R. Coral classification with hybrid feature representations. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 519–523.

20. Razavian, A.S.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Columbus, OH, USA, 23–28 June 2014; pp. 512–519.

21. Stokes, M.D.; Deane, G.B. Automated processing of coral reef benthic images. *Limnol. Oceanogr. Methods* **2009**, *7*, 157–168.

22. Pizarro, O.; Rigby, P.; Johnson-Roberson, M.; Williams, S.B.; Colquhoun, J. Towards image-based marine habitat classification. In Proceedings of the OCEANS 2008, Quebec, QC, Canada, 15–18 September 2008; pp. 1–7.

23. Beijbom, O.; Treibitz, T.; Kline, D.I.; Eyal, G.; Khen, A.; Neal, B.; Loya, Y.; Mitchell, B.G.; Kriegman, D. Improving Automated Annotation of Benthic Survey Images Using Wide-band Fluorescence. *Sci. Rep.* **2016**, *6*, 23166.

24. Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F.; Hovey, R.; Kendrick, G.; Fisher, R. Automatic annotation of coral reefs using deep learning. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016; pp. 1–5.

25. Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F.A.; Boussaid, F.; Hovey, R.; Kendrick, G.A.; Fisher, R.B. Deep image representations for coral image classification. *IEEE J. Ocean. Eng.* **2018**, *44*, 121–131.

26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

27. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

28. Bewley, M.; Nourani-Vatani, N.; Rao, D.; Douillard, B.; Pizarro, O.; Williams, S.B. Hierarchical classification in AUV imagery. In *Field and Service Robotics*; Springer: New York, NY, USA, 2015; pp. 3–16.

29. Cortes, C.; Vapnik, V. Support vector machine. *Mach. Learn.* **1995**, *20*, 273–297.

30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. 2012. Available online: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ (accessed on 12 January 2020)

31. Khan, S.H.; Hayat, M.; Bennamoun, M.; Sohel, F.; Togneri, R. Cost Sensitive Learning of Deep Feature Representations from Imbalanced Data. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 3573–3587.

32. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 630–645.

33. Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F. Resfeats: Residual network based features for image classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1597–1601.

34. Bewley, M.; Friedman, A.; Ferrari, R.; Hill, N.; Hovey, R.; Barrett, N.; Pizarro, O.; Figueira, W.; Meyer, L.; Babcock, R.; et al. Australian sea-floor survey data, with images and expert annotations. *Sci. Data* **2015**, *2*, 150057.

35. Kohler, K.E.; Gill, S.M. Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. *Comput. Geosci.* **2006**, *32*, 1259–1269.

36. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.

37. Silla Jr, C.N.; Freitas, A.A. A survey of hierarchical classification across different application domains. *Data Min. Knowl. Discov.* **2011**, *22*, 31–72.

38. Vedaldi, A.; Lenc, K. Matconvnet: Convolutional neural networks for matlab. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 689–692.

39. Fan, R.E.; Chang, K.W.; Hsieh, C.J.; Wang, X.R.; Lin, C.J. LIBLINEAR: A library for large linear classification. *J. Mach. Learn. Res.* **2008**, *9*, 1871–1874.

40. Azizpour, H.; Sharif Razavian, A.; Sullivan, J.; Maki, A.; Carlsson, S. From generic to specific deep representations for visual recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 36–45.

41. Fraser, M.W.; Kendrick, G.A.; Statton, J.; Hovey, R.K.; Zavala-Perez, A.; Walker, D.I. Extreme climate events lower resilience of foundation seagrass at edge of biogeographical range. *J. Ecol.* **2014**, *102*, 1528–1536.

42. Giraldo Ospina, A.; Kendrick, G.A.; Renae, H. Depth moderates loss of marine foundation species after and extreme marine heatwave: Will deep temperate reefs act as climate refugia? 2019, under review.

43. Graham, M.H.; Kinlan, B.P.; Druehl, L.D.; Garske, L.E.; Banks, S. Deep-water kelp refugia as potential hotspots of tropical marine diversity and productivity. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16576–16580.

44. Lesser, M.P.; Slattery, M.; Leichter, J.J. Ecology of mesophotic coral reefs. *J. Exp. Mar. Biol. Ecol.* **2009**, *1*, 1–8.

45. Kahng, S.; Garcia-Sais, J.; Spalding, H.; Brokovich, E.; Wagner, D.; Weil, E.; Hinderstein, L.; Toonen, R. Community ecology of mesophotic coral reef ecosystems. *Coral Reefs* **2010**, *29*, 255–275.

46. Beijbom, O.; Edmunds, P.J.; Roelfsema, C.; Smith, J.; Kline, D.I.; Neal, B.P.; Dunlap, M.J.; Moriarty, V.; Fan, T.Y.; Tan, C.J.; et al. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. *PLoS ONE* **2015**, *10*, e0130312.

47. Chennu, A.; Färber, P.; De'ath, G.; de Beer, D.; Fabricius, K.E. A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Sci. Rep.* **2017**, *7*, 7122.

48. Williams, K.; Lauffenburger, N.; Chuang, M.C.; Hwang, J.N.; Towler, R. Automated measurements of fish within a trawl using stereo images from a Camera-Trawl device (CamTrawl). *Methods Oceanogr.* **2016**, *17*, 138–152.

49. Shortis, M.R.; Ravanbakhsh, M.; Shafait, F.; Mian, A. Progress in the automated identification, measurement, and counting of fish in underwater image sequences. *Mar. Technol. Soc. J.* **2016**, *50*, 4–16.