Deep Learning for Coral Classification

NON-PRINT ITEMS:

Abstract: This chapter presents a summary of the use of deep learning for underwater image analysis, in particular for coral species classification. Deep learning techniques have achieved the state-of-the-art results in various computer vision tasks such as image classification, object detection and scene understanding. Marine ecosystems are complex scenes and hence difficult to tackle from a computer vision perspective. Automated technology to monitor the health of our oceans can facilitate in detecting and identifying marine species while freeing up experts from the repetitive task of manual annotation. Classification of coral species is a challenging task in itself and deep learning has a potential of solving this problem efficiently.

Keywords: corals, deep learning, convolutional neural networks (CNN), marine images, classification, marine ecosystems.

Chapter starts here:

1. Introduction

Coral reefs are a vital part of marine ecosystems. They provide a nutrient rich habitat and a safe shelter for many marine organisms. They are a rich source of nitrogen and other essential nutrients for benthic species. They also play an essential part in the recycling of nutrients and in protecting coastlines from the devastating effects of waves and sea storms. Coral reefs help in sustaining a growing fishing industry since many fish and other species are found closer to the reefs. Shallow sea coral reefs such as the Great Barrier Reef of Australia also benefit the tourism industry.

Marine scientists have reported a worldwide decreasing trend in the coral population. According to a 2011 research, 19% of the coral reefs were lost and 75% are now being threatened [1]. With the increase in global warming, urbanization, human population, large use of sea for shipping, exploration for minerals, recreational uses such as boating and industrial trade and activities, there has been a huge impact on coral reefs, both positive and negative [2]. Increased water temperatures are responsible for bleaching and death of corals [3], [4]. This has resulted in a rapid decline in our planet's marine biodiversity [5]. In order to minimize the negative impact of these activities on the sea, marine ecosystems need to be monitored regularly. That is where underwater optical imaging comes to the rescue. Long-term monitoring of large areas, remote sensing and tracking of marine species and their associated habitats are now standard requirements in most management strategies. As a result, the automatic annotation of collected marine data is now at the forefront of management applications and thus a research priority [6]. With the development of underwater optical imaging techniques, standard protocols can be developed for analysing and curtailing the negative impacts on seawater environmental sustainability. Additionally, an exponential increase in the use of digital cameras and video implies the need for storage and automated analysis of such data.

Marine scientists have a massive amount of imagery of coral reefs that is yet to be annotated. Monitoring systems like the Integrated Marine Observing System (IMOS) collect millions of images of coral reefs around Australia every year. However, only a small percentage of these images, typically less than 5%, get analysed by marine experts. Moreover, manual annotation is a tediously repetitive

task and demands large human resources and time. Automated technologies to monitor marine ecosystems are crucial for a continuous monitoring without feedback from human experts. With these stated facts in mind, automatic annotation of underwater images can achieve visually comprehensible results. The proposed results could help in curbing a global threat. The extent of the usefulness of this research can be seen from the attention this research field is getting. In this chapter, we aim to address the challenges associated with the automatic analysis of marine data and explore the applications of deep learning for automatic annotation of coral reef images.

The rest of this chapter is organized as follows: Section 2 presents existing methods for the manual annotation of corals used by marine scientists and discusses the challenges involved with coral reef classification. It also highlights previous coral classification research. Section 4 presents a brief introduction of deep learning and the state-of-the-art of deep networks. Section 5 summarises current coral classification studies with deep learning. Finally, Section 6 outlines the future prospects and applications of deep learning for deep sea image analysis. Section 7 concludes this chapter.

2. Annotation of Coral Reefs: Methods and Challenges

Coral reefs are one of the most diverse and economically important ecosystems of our planet. There are three main types of coral reefs: fringing, barrier and atoll. Fringing reefs (e.g., reefs off the coast of Eilat, Israel) are found closer to the shores and they often form borders around islands and coastlines. Barrier reefs (e.g., The Great Barrier Reef, Australia) are separated from coastlines by water. This results in a lagoon of water between the shore and the reef itself. Atoll reefs (e.g., Lighthouse Atoll Reefs, Belize) are found deep below the sea level and are usually circular or oval in shape. Three coral reef examples are shown in Fig. 1.1 (a-c).

The main goal of a long-term monitoring of coral reefs is to investigate how the reefs are changing over time due to the phenomenon of coral bleaching (shown in Fig. 1.1d). This investigation is done on local and global scales. Coral reef data generally consists of the following:

- Site survey (e.g., information about location, depth, water temperatures, and turbidity)
- Coral species survey (e.g., hard corals, soft corals, bleached corals, and dead corals)
- Substrate survey (non-coral species: e.g., macroalgae, sponges, sand, rock, and rubble)

Corals have a large number of sub-species. The Great Barrier Reef in Australia alone have more than 600 sub-species. They are a diverse specie and are found in a variety of size, shapes and colors. The two main categories for corals are: hard corals and soft corals. Hard corals have a limestone skeleton, whereas soft corals are flexible and are often mistaken as plants due to the lack of a skeleton. Hard corals are the best indicator of health of any coral reef. Their percentage cover is the most commonly used parameter to quantify the coral reef population. Adverse climate effects such as pollution and increased temperature of sea floor result in the bleaching of healthy corals and eventually death.

[Figure 1.1 reefs.jpg here. Caption: (a) Fringing reef off the coast of Eilat, Israel. (b) The Great Barrier Reef, Australia. (c) Lighthouse Atoll Reef, Belize. (d) Coral Bleaching: Healthy corals on left and bleached corals on right.]

2.1. Methods for Conventional Annotation

Underwater imaging techniques such as autonomous underwater vehicles (AUVs) have tremendously increased the amount of marine data that is available for analysis. However, the process of manually annotating this data is cumbersome and inefficient [7]. In practice, marine scientists usually adopt random point annotations whereby a pre-defined number of random points are selected on each image, as low as 20 or as high as 200. Afterwards, a marine expert assigns a label to all of the individual points as shown in Fig. 1.2. A single image can take up to 30 minutes to fully annotate it. Repeating the same procedure for millions images is obviously a tedious and a challenging task because the class boundaries are ambiguous and difficult to define in terms of color, shape or texture [7]. This annotation scheme is often facilitated by software such as Coral Point Count [9]. It is free software developed by the National Coral Reef Institute (NCRI) for experts and researchers working in the management and the monitoring of coral reefs. CPCe overlays a given image with a pre-defined set of random pixels. A marine expert then assign a class label to these random pixels.

Furthermore, water turbidity and underwater illumination render the images difficult to analyse [8]. Also, coral reef images consist of an assemblage of corals and non-corals of irregular shapes and sizes. To manually label the full segmentation ground truths for every image is exhausting and time consuming. Bounding box annotations are prone to leave out key details. Assigning one label per image hinders sub-species classification. As a result, the well-known labelling techniques such as bounding boxes, boundary segmentation and whole image labelling are impractical. Random point sampling and labelling are the least cumbersome and most efficient of these techniques.

[Figure 1.2 labels.jpg here. Caption: Sample coral reef image from Benthoz15 dataset illustrating random point sampling annotation method.]

2.2. Challenges

Sea floor exploration and imaging have provided us with a great opportunity to look into the vast and complex marine ecosystem. Data acquisition from the sea bed is vital for the scientific understanding of these intricate ecosystems, but is often hampered by logistical constraints associated with working underwater. Advanced underwater cameras and an increasing interest in exploring underwater environments have initiated the need for improvements in the field of imaging techniques. Seabed observations, archaeology, marine geology, marine biology and biodiversity are mainly conducted by optical imaging [1, 10 and 11]. Digital images of the sea floor are now commonly collected with the help of Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) [12]. To make the marine data useful for analysis, there has to be an accurate automatic annotation system instead of the reliance on manual labelling.

Any system that is used in theory has to face specific challenges in a real underwater environment. In the same manner, when a sufficiently large number of training images are available, the methods derived for real-world object classification can be used to analyse structured textures and objects, but they fall short in a real underwater world. In order to achieve higher classification accuracy, many issues have to be addressed such as blurring, scattering, sun flicker and color attenuation. Therefore, automatic annotation for underwater scene classification is a difficult and challenging topic.

Underwater digital imaging and automatic species classification is an extremely challenging task. Training datasets are created on the basis of underwater classes that are different in terms of shape, color, texture, size, rotation, illumination, view angle, camera distance and light conditions. These challenges mainly include:

- significant intra-class and inter-site diversity of the acquired images
- complex and ambiguous spatial borders between classes
- manual annotation varies from expert to expert
- variations in the spatial and spectral resolution limits, view points, and image quality of the cameras
- partial or complete occlusion of objects of interests
- gradual changes in the structures of the marine seabed over longer periods of time
- lighting artefacts due to refraction from waves and variable depth dependent optical properties
- variable water turbidity, color distortions and inadequate illumination conditions.

Four marine images are shown in Fig. 1.3 to illustrate some of these challenges. These pictures were captured at the same sites but under different illumination conditions. They also portray a significant color distortion. In the next section, we will explore the previous work done for coral classification.

[Figure 1.3 samples.jpg here. Caption: Sample marine images from the Western Australian seabed under different illumination conditions and color distortion]

3. Automatic Coral Classification

3.1. Coral Classification with Hand-crafted Features

Color and texture are the key discriminating factors for classifying corals. Hence, researchers have extensively studied the extraction of color and texture based hand-crafted features for image representation. Features that encode shape information are less suitable because the corals have arbitrary shapes and the class boundaries are unclear. Usually a combination of color and texture based features is preferred. There are no definite combinations of features which are expected to work for any general coral dataset. The features are often selected based on the discriminating characters of corals and non-corals that are present in a given dataset. In this section, we will highlight some of the prominent studies for coral image classification with hand-crafted features.

- Marcos [13] used Normalized Chromaticity Coordinate (NCC) for color and Local Binary Pattern (LBP) for texture. A 3-layer feed-forward back propagation neural network was used to classify five classes: living corals, dead corals, corals with algae, abiotics and algae. It was proposed that NCC color features are invariant to illumination conditions and LBP is robust to brightness changes. However, the NCC and LBP features are not discriminative enough for complex underwater images. This method was then tested on the three coral classes with only 300 images. A higher performance was reported when a combination of LBP and hue information was used compared to the combination of LBP with NCC.
- Stokes and Deane [14] used normalized color histograms as color descriptors and a discrete cosine transform (DCT) based feature vector for texture. Their training set consisted of 3000 images and 18 distinct classes. A novel classification approach titled “probability density

weighted mean distance (PDWMD)" was proposed for classification. This method is easy to implement and fast. However, the weights of the color and texture features are set manually during the feature extraction. Also, DCT descriptors are not very robust and accurate texture descriptors.

- Pizarro [15] used a feature vector based on NCC histogram, bag of words (BoW) for scale-invariant feature Transform (SIFT) and Hue-histograms. A subset of training samples is used to construct a BoW and the test image is then described in terms of this vocabulary of words. They performed classification by voting for the best matches. In their method, each image is classified as one class out of the total eight classes. A total of 453 images were used for training and vocabulary learning. This annotation method does not perform well for pixel annotations and is prone to leaving out key details. The sub-image level classification is not addressed in this work. Defining texture with BoW on SIFT features is not an efficient texture feature in complex underwater conditions.
- Beijbom [8] introduced the Moorea Labelled Coral (MLC) dataset (with four non-coral and five coral classes) and used a Maximum Response (MR) filter bank followed by texton maps for feature extraction at multiple scales. A subset of training images was used alongwith k-means clustering for generating a texture dictionary. They also showed that pre-processing the images in the *L*a*b* color space improves a superior performance compared to RGB. They used an SVM classifier with a Radial Basis Function (RBF) kernel for classification. Coral images from three different years were automatically annotated to yield coral maps across the reef sites.
- In [16], a combination of hand-crafted features and multiple classifiers were analysed to achieve best classification of accuracy for multiple benthic datasets. The descriptors that they used include Completed Local Binary Patterns (CLBP), grey level co-occurrence matrix (GLCM), Gabor feature, and opponent angle and hue channel color histograms. All the feature vectors used in this work were scale invariant and robust to color distortion and low illumination. Support vector machines (SVM), k-nearest neighbours (KNN), neural networks and probability density weighted mean distance (PDWMD) were the selected classifiers. Different combinations of features and classifiers were also employed to get the best performance for the six test datasets. However, issues such as how to choose an optimal scale for patch extraction and identification of overlapping classes were not addressed in this work.

Table 1.1 summarises the feature vectors and the number of classes of the methods explained above.

[Table 1.1 here. Caption: Summary of hand-crafted feature based methods for coral classification.]

3.2. Coral Classification with Learned Features

Deep neural networks are a powerful category of machine learning algorithms implemented by stacking layers of neural networks along the depth and width of smaller architectures. Deep networks have recently demonstrated discriminative and representation learning capabilities over a wide range of applications in the contemporary years. Researchers in ML are expanding the horizons of deep learning by seeking their prospective applications in other diverse domains. One such forthcoming domain is marine scene classification. Deep networks require a large amount of

annotated data for training. With efficient training algorithms, deep neural networks are capable of separating millions of labelled images. Moreover, the trained network can also be used for learning efficient image representations for other similar benthic datasets. Before discussing applications of deep learning in coral classification, we give a brief review on deep learning and its state-of-the-art architectures in the next section.

4. Deep Neural Networks

An excellent performance of any image or video processing task (e.g., classification, object detection, scene understating) relies on the extraction of discriminative features or image representations from the input data. Domain specific hand-crafted image representations have been extensively used in computer vision for decades. Features learned using machine learning algorithms, known as representation learning, have shown better performance in recent years, compared to the traditional hand-crafted representations. Deep learning algorithms employ conventional neural networks with increased complexities and depths.

Neural networks with many hidden layers [17] are capable of extracting high levels of abstractions from raw data. Many state-of-the-art systems in computer vision owe their success to their ability to extract high-level abstractions. Neural networks were popular in the 1990s but support vector machines ascended to the central stage in the 2000s and out-performed NNs. Deep neural networks became very popular in computer vision after the seminal work in [18].

4.1. Convolutional Neural Networks

Convolutional neural networks (CNN) [18] are another important class of neural networks used to learn image representations that can be applied to numerous computer vision problems. Deep CNNs, in particular, consist of multiple layers of linear and non-linear operations that are learned simultaneously, in an end-to-end manner. To solve a particular task, the parameters of these layers are learned over several iterations. CNN based methods have become popular in the recent years for feature extraction from images and video data.

A CNN consists of convolutional layers and pooling layers occurring in an alternating fashion. Sparse connectivity, parameter sharing, sub-sampling and local receptive fields are the key factors that render CNNs invariant to shifting, scaling and distortions of input data. Sparse connectivity is achieved by making the kernel size smaller than the input image which results in a reduced number of connections between the input and the output layer. Inter-channel and intra-channel redundancy can be exploited to maximize sparsity. Moreover, the computation of output requires fewer operations and less memory to store the weights. In a non-convolutional neural network, a weight element is only multiplied once by the input and never used again. However in a convolutional layer, every element of the kernel matrix is convolved with the input image more than once. The convolutional layers consist of stacks of filters of pre-defined sizes that are convolved with the input of the layer. The parameter sharing used by convolutional layers is more efficient (requires fewer computation and memory storage) than a dense matrix multiplication. Parameter sharing can also make the convolutional layers equivariant to linear translations (i.e., any shift in the input will result in a similar shift in the output). However, convolutional layers are not equivariant to distortions in scale or rotation.

The depth of the CNN can be increased by setting the output of the pooling layer to be the input of the next convolutional layer. CNNs with smaller filter size (3x3) and deeper architectures have shown increased performances. One such example is the VGGnet [19] (shown in Fig. 1.4). They have reported a significant improvement on the prior-art configurations by pushing the depth to 16–19 hidden layers. They secured the first and the second places in the ImageNet Challenge 2014 in the localisation and classification tracks, respectively.

[Figure 1.4 vgg.jpg here. Caption: VGGnet Architecture: 16 weighted layers. C1 to C5 are 5 convolutional layers with sub-layers. FC are 3 fully connected layers.]

4.2. Representation Learning

Learning discriminative image representations from data have evolved as a promising research area. A powerful image representation captures the prior distributions of data by learning the image features. These features are usually hierarchical in nature (low and high level features) and hence the image representations learn to define the more abstract concepts in term of the less abstract ones. A good learned representation should be simple (usually linearly dependent), sparse and possess spatial and temporal coherence. The depth of a network is also an important aspect in the representation learning. Representations learned from the higher layers of deep networks encode high level features of data.

Image representations extracted from CNNs, trained on large datasets such as ImageNet [18] and fine-tuned on domain specific datasets, have shown state-of-the-art performance in numerous image classification problems [20]. These learned features can be used as universal image representations and have produced outstanding performances in computer vision tasks e.g., image classification, object detection, fine grained recognition, attribute detection and instance retrieval. The activations of the first fully connected layer of CNNs are the preferred choice of most researchers. However, the activations of intermediate convolutional layers have also shown comparable performances. In [21], subarrays of convolutional layer activations are extracted and used as region descriptors in a 'local feature' setting. The extracted local features from two consecutive convolutional layers are then pooled together and included in the resulting feature vector. This approach, termed "cross-convolutional layer pooling" achieved significant performance improvements in scene classification tasks [21].

Why do these CNN features perform so well across diverse domains? Despite their outstanding performance, the intrinsic behaviour of these deep networks is somewhat of a mystery. A visualization technique was proposed in [22] which investigated the relationship between the output of various layers of the CNN architecture (proposed in [18]) and the input image. The outputs of different convolutional layers were analysed and the following conclusions were drawn: Layer 2 responds to corners and edges, Layer 3 captures complex invariances such as texture and mesh patterns, Layer 4 is more class-specific and Layer 5 captures entire objects irrespective of pose variations.

Visualisation methods that can help us understand computer vision image representations in general and learned deep representation in particular are gaining popularity in computer vision society. Before the development of these methods, CNN based image representations were considered as black boxes for deep feature extraction. A new visualisation method was introduced recently in [23].

This method is based on natural looking pre-images (an image obtained by the inverse transform of the learned representation) which have prominent image representations. Such images are termed "natural pre-images". Three image visualizations were used to investigate the effectiveness of standard hand-crafted representations and CNN representations: inversion, activation maximization and caricaturization. It was demonstrated that representations like HOG can be inverted more precisely compared to CNN features. However, different layers of CNN retain relevant information of the input image along with different pose and illumination variations. Deep layers of a CNN preserved object specific information and global variances. Moreover, fully connected layers captured large variations in the object layouts. Intermediate convolutional layers seemed to preserve the local variances and structures such as lines, edges, curves and parts. These conclusions were a big step towards understanding generic deep features.

To further enhance the invariance of deep features without decreasing their discriminative power, multi-scale order-less pooling (MOP-CNN) was introduced in [24]. CNN features are pooled locally at multiple scales using Vector of Locally Aggregated Descriptors (VLAD) pooling. The final feature vector is then obtained by concatenating these local feature vectors and can be used as a generic descriptor for either supervised or unsupervised recognition tasks, image classification or scene understanding. MOP-CNN features have consistently shown better performance than the global CNN activations. They have also eliminated the need of joint training of prediction layers for a particular domain.

For object detection tasks, regions with CNN features (or R-CNN in short) [25] is a powerful variant of CNNs and has recently been very popular. R-CNN combines two key concepts: (1) region proposals combined with deep CNNs in order to localize and segment objects and (2) supervised pre-training followed by domain-specific fine-tuning for smaller training datasets. This method yielded a significant performance improvement in the case of object detection tasks.

A specific pre-trained conventional deep network requires a fixed input image size (e.g., 224 x 224 for VGGnet). This "artificial" requirement may reduce the recognition accuracy for images of arbitrary size. To overcome this limitation of CNNs, a novel pooling scheme called "spatial pyramid pooling (SPP)" was introduced in [26]. A fixed length feature vector can be obtained using SPP-nets irrespective of input image size. SPP also made the network robust to pose variations and scale deformations. SPP-net achieved state-of-the-art classification results using full-image representations and without any fine-tuning. Feature maps are computed from the entire images only once and hence the repeated computation of the convolutional features can be avoided.

Deep learning methods have achieved state-of-art performances on many computer vision tasks. But these tasks remain challenging and the methods have plenty of room for improvement. The history of deep learning applications to computer vision demonstrates that deeper networks improve performance [18].

4.3. Going Deeper with Neural Nets

A detailed study to explain why deep learning outperforms other shallow networks was presented in [27] and [28]. In [27], the number of distinct linear regions was used as a key parameter to address the complexity of a function encoded by a deep network. It was established that for any given layer of a deep network, the ability to encode pieces of input information was exponential in nature. The

functions computed by the deeper layers were more complex but they still possessed an intrinsic rigidity caused by replicating the hidden layers. This rigidity helps deep networks to generalize unseen input samples better than the shallow models. In [28], a novel method of understanding the expressiveness of a model was presented based on computational geometry for piecewise linear functions. Deep and narrow rectifier MLPs generated more regions of linearity as compared to shallow networks with the same number of computational units.

Increasing the depth and width of a CNN requires a huge computational cost to train the deep neural networks (a larger number of weight parameters to adjust). The staggering success of the CNNs over the last 5 years can be explained by these factors: larger datasets, deeper models, faster hardware, and last but not the least, novel algorithms for optimization and efficient training of deeper networks. Conventional CNNs come with two basic functionalities: partitioning and abstraction. Partitioning can be improved by using very small filters at the start of the network and then increasing the filter size as we go deeper. In a standard CNN, a linear classifier and a non-linear activation function are employed to yield an abstraction from the input patch. These abstractions are not discriminative enough. In [29], a novel structure called “Network in network (NIN)” was proposed to enhance the strength of these abstractions. Micro neural networks are initialized along with multilayer perceptrons. This micro network can be viewed as an additional 1×1 convolutional layer followed typically by the rectified linear activation layer. The 1x1 convolutions (small filters) have twofold advantages: reducing the dimension of the input vector, thereby increasing the width of the network and reducing the computational cost. This combination approximates any given function more effectively than a linear classifier followed by a non-linear activation function. These micro networks are then convolved with the input image within a larger network, hence the name “network in network”. Stacking these structures repeatedly results in deep NINs. Fully connected layers are also replaced with global average pooling layers which are less prone to overfitting. Deep NINs demonstrated state-of-the-art performance on CIFAR-10, CIFAR-100 and SVHN datasets.

Inspired by NIN architecture, Google introduced a deep network codenamed ‘inception’ [30]. This network utilized the concept of depth in two ways: (1) it increased the number of layers and (2) an “inception module” is introduced to add a new level of organization along the width of the network. The performance of any deep network can be enhanced by increasing the depth of the network (adding more layers) and increasing the width (adding more channels at each layer). However, the resulting deeper and wider network is more prone to overfitting and is also computationally expensive. A logical approach to solve these bottlenecks is to make the network connections sparse instead of fully connected. When the dense building blocks of the network are approximated by the optimal sparse structures [31], the resulting network outperformed the shallower networks with a similar computational budget. Small filters (1x1 convolutions) were used to reduce the dimension of the output which preceded the bigger filters. The inception modules were only added to the higher layers of GoogLeNet to keep the computational cost lower. This was a promising start towards creating deeper and sparser networks.

Towards training deeper networks, there is another prominent class that is worth mentioning: ‘highway networks’ [32]. In this novel architecture, optimization is performed using a learned gating mechanism inspired from the concept of the Long Short Term Memory (LSTM) recurrent neural networks [33]. This gating mechanism results in arbitrary paths for information flow between multiple layers. These paths are termed ‘information highways’. The switching information for the

gates is learned using the training set and since, some of the neurons are activated at any given iteration, computational cost is minimized as well. Highway networks as deep as 900 layers can be optimized easily using this approach.

So far we have established that the depth of any given neural network is directly proportional to the computational difficulty involved in training that network. The accuracy of a deep network gets saturated if we keep on stacking layers after layers beyond a certain depth. However, if the training computations are optimized effectively, an increased depth can result in higher performance. One such approach was articulated in [34] and was named residual networks (ResNets). A residual network includes a number of residual blocks each being a small CNN itself. These residual blocks are not only just stacked together; each block also has a shortcut connection to the outputs of the next blocks. An example of a residual block is shown in Fig. 1.5. These shortcut connections decrease the network's complexity. A 34-layer ResNet contains 3.6 billion multiply-add operations whereas a 19-layer VGGnet has 19.6 billion multiply-add operations. Consequently, ResNets are easier to train and the training accuracy does not get saturated. Improved results on CIFAR-10 were reported in a subsequent study [35] using a 1001-layer deep ResNet.

[Figure 1.5 resnet.jpg here. Caption: A residual block of ResNets shown in a red rectangle. The output of one residual block acts as the input of the next block and is also added to the output of the next residual block.]

5. Deep Learning for Coral Classification

Coral reefs exhibit significant within-class variations, complex between-class boundaries and inconsistent image clarity. The accuracy of any classification algorithm depends on the discriminating power of the extracted features from the images. In the light of the challenges outlined in Section 2, hand-crafted feature have a number of limitations. Hand-crafted features usually encode one or two aspects of data such as color, shape or texture. Creating a novel hand-crafted feature representation which addresses all of the challenges involved in marine images is an up-hill task. It is far more feasible to rely on off-the-shelf CNN features extracted from a deep network pre-trained on a large image dataset. CNN features have shown their discriminating power when transferred to a different domain [20]. Combining the CNN features with the domain specific hand-crafted features to improve the classification performance, presents an interesting research problem (as shown below).

5.1. Hybrid and Quantized Features

The idea of combining CNN and hand-crafted features was used in action classification tasks from videos [36]. Most of the entries in THUMOS challenge [36] have combined CNN based features with hand-crafted features for action classification in videos. Hand-crafted features are usually encoded using Fisher vectors [37] and Vector of Locally Aggregated Descriptors (VLAD) [38] before combining them with the CNN features. Wang *et al.,* [39] have cascaded morphology and texture based hand-crafted features with CNN features for mitosis detection. They have trained 3 classifiers; one for CNN features, one for hand-crafted features and a third classifier for the test samples that are misclassified by the first two classifiers. This approach is computationally expensive and impractical for applications with large datasets. Jin *et al.,* [40] have showed that CNN and hand-crafted features complement each other and have shown promising results for RGB-D object detection. They

combined Locality-constrained Linear Coding (LLC) based spatial pyramid matching features with the CNN features.

CNN features cannot be used directly in coral image classification since benthic datasets come with pixel (instead of bounding box) annotations. Deep features have not yet been explored until recently for the coral reef classification problem. An application of generic deep features extracted from VGGnet combined with hand-crafted features for coral reef classification to take advantage of the complementary strengths of these representation types was proposed in [41]. The dataset was not big enough for training a CNN from random initializations. Therefore, pre-trained CNN based features were extracted from patches centred at labelled pixels at multiple scales and a local variant of SPP was implemented to render the image representations scale-invariant. Texture and colour based hand-crafted features extracted from the same patches were used to complement the CNN features. A memory efficient 2-bit feature representation scheme was investigated to reduce the memory requirements by a factor of 16. The proposed method achieved a classification accuracy that is higher than the state-of-art methods on the MLC benchmark dataset for corals. The hybrid (hand-crafted and learned) features performed the best. It is also implied that the CNN features and the hybrid features addressed the problem of class imbalance more efficiently. In the case of corals, the most abundant class overshadows the less frequent classes when the patches are extracted at one scale. Since the patches were extracted at different scales and then max-pooled, the less abundant classes are made more prominent in the resulting feature vectors. This was demonstrated by the experimental results. This helps the classifier to cope with the inherent class imbalance problem effectively. Fig. 1.6 outlines the block diagrams of the different classification pipelines.

[Figure 1.6 Block.jpg here. Caption: Block diagram of coral classification method of [41]: (a) the pipeline for CNN feature extraction (b) the pipeline for the hybrid features. (c) the pipeline for the quantized features.]

Fig. 1.7 shows the confusion matrices (CM) from experiments in [41]. The rows correspond to the ground truth and the columns correspond to the predicted class assignments. An ideal confusion matrix has 1s in its diagonals and 0s elsewhere. A better classification performance was observed when these confusion matrices are compared with the ones in [8]. The presence of high non-zero elements in the first column implies an imbalance towards class 1 which is the most abundant class in our dataset. In practice, high values in the diagonal of the CM represent a good quality classifier. When the first column is compared with the corresponding first columns of [8], we note that the later method copes better with the class imbalance. It was concluded that the local-SPP scheme takes care of the class imbalance problem to an extent.

[Figure 1.7 CM.jpg here. Caption: Confusion matrices for coral classification on MLC dataset: (a-c) Baseline performance [8] for experiments 1, 2 and 3 respectively. (d-f) Our combined features (CF) for three experiments: (1) Trained and tested on 2008 images. (2) Trained on 2008 images and tested on 2009. (3) Trained on 2008 and 2009 images while tested on 2010 images.]

Memory overhead is an important aspect when dealing with large datasets. Feature representations for larger datasets require a lot of storage space. Efficient encoding schemes are necessary to compress these representations without losing the essential information. Therefore, we propose to

quantize the feature vector to a low bit representation to encode the CNN based features. The resulting feature vector takes up to 16 times less storage space. Compact feature representations with lower memory storage are preferred in the case of CNN based features. This should lead to faster training and testing times. In CNNs, the magnitude activation of neurons is not so much important as the spatial location of that particular neuron in the network. To prove this, the individual elements of our combined feature vector were quantized into three values i.e., 0, 1 and -1. The positive elements were replaced by 1 and the negative elements were replaced by -1. Consequently, only two bits are required to store each individual element compared to the commonly used 32-bit single precision floating point format. This quantization effectively reduced the required memory to store the feature vectors by a factor of 16 (32 bits replaced by 2 bits after quantization). This efficient utilization of memory was achieved at the cost of a slight decrease in the classification accuracy. The resulting accuracy for the quantized features (QF) is still comparable with the baseline performance on the MLC dataset.

5.2. Coral Population Analysis

The proposed classification algorithm of [41] was also evaluated on Benthoz15 dataset [42]. This dataset consists of an expert-annotated set of geo-referenced benthic images and associated sensor data, captured by an autonomous underwater vehicle (AUV) across multiple sites from all over Australia. The whole dataset contains 407,968 expert labelled points, on 9,874 distinct images collected at different depths from nine sites around Australia over the past few years. There are almost 40 distinct class labels in this dataset, which make it quite challenging to annotate automatically. A subset of this dataset containing images from Western Australia (WA) was used to train the classifier in [43]. Fig. 1.8 outlines the general approach of their proposed framework. The multi-scale features were extracted using a deep network. The coral population of the Abrolhos Islands (located off the west coast of Western Australia) was also analysed by automatically annotating the unlabelled mosaics using our best classifier. Coral cover maps were then generated and validated by a marine expert as ground-truth labels were not available. This method detected a decreasing trend in the coral population in this region. It was an important step towards investigating the long-term effects of environmental change on the effective sustenance of marine ecosystems automatically.

[Figure 1.8 outline.jpg here. Caption: Block diagram of proposed framework of [41].]

5.3. Cost-sensitive Learning for Corals

Like most real-world computer vision datasets, marine datasets also exhibit class imbalance. Non-coral classes exist in abundance and hence the class balance is skewed towards coral classes. This imbalance in class distribution hinders the classifier to learn distinct class boundaries and a performance drop occurs. A cost sensitive deep network was proposed in [44] to address this issue. This network's architecture was based on VGG-net (16 layer version). Instead of altering the original class distributions (e.g., over-sampling and under-sampling), a cost learning layer was introduced before the soft-max layer of the classifier. An optimization algorithm was proposed to optimize the

network parameters and the cost-sensitive layer parameters. This approach was tested on many datasets (including a coral dataset, MLC) which exhibit class imbalance. Their approach performed better than the baseline performance of MLC reported in [8]. However, this performance is lower than the performance reported in [41].

5.4. CNN with Fluorescent Images

Most common deep networks work with color images and hence the input layer has three distinct channels (R, G and B). However in theory, a CNN can have an arbitrary number of input channels to encode additional information. One such approach was proposed in [45]. RGB images were combined with reflectance images and fluorescent images. A pixel-wise average function was used to obtain the final image. The fluorescent images had rich contrast information for the corals and the reflectance images provided context of the non-fluorescent substrates. After registration, the input image had 5 channels and a CNN was trained with these additional channels. This CNN's architecture was similar to the CIFAR10 architecture defined in Caffe. Patches of 128 x128 were extracted and resized to 32 x 32 before passing them through three consecutive rounds of convolutional layers. The 5-channel CNN performed better than the corresponding traditional CNN for coral classification. The performance of this 5-channel CNN was also compared with the baseline performance of [8]. A 22% reduction of classification error-rate was demonstrated when both reflectance and fluorescent images were used compared to the case when only reflectance images were used.

6. Future Prospects

Deep learning solutions to ecological studies can provide a truly objective measure to detect, discriminate and identify species, and their behaviour and morphology. This will reduce common sources of variations and bias in human observer studies caused by subjective interpretation or the lack of skill or experience. Such automated processing tools will ensure transparency of the study results and standardization of methods for analysts. It will also facilitate comparisons of studies across individuals, populations and species in a systematic and objective manner. It will also enable the processing of datasets at considerably higher speeds compared to human experts. This is particularly relevant for tediously repetitive tasks. Freeing human resources for more complex tasks is becoming increasingly important in budget-limited and data intensive studies. Other transformational ecological outcomes include: (i) rapid quantitative surveying of the massive amount (95%) of the acquired underwater imagery that is yet to be processed. This will enable the construction of large-scale spatially extensive image baselines for marine habitats. Such data could then be used to make a quantitative assessment of the impact of climate change; (ii) monitoring of the growth, mortality, and recruitment rates and competitive abilities of marine species (e.g. coral reef, lobsters, kelps) associated with warming and acidification; and (iii) improved knowledge of marine ecosystems for which very little is known.

Some other future prospects of this research are:

- To develop deep learning methods, not only limited to corals, to classify a huge amount of marine data automatically.
- To compare different deep learning methods to form a solid basis for efficient assessment of marine ecosystems.

- To develop an automatic annotation system that works with diverse datasets while saving human resources that are necessary for manual labelling.
- To investigate the resilience of marine ecosystems to environmental impacts (global warming, marine pollution, resource extraction, coastal development) through economically sustainable monitoring programs.
- To analyse the relationships between marine species and to quantify the trends in the population dynamics.

7. Conclusion

In this chapter, we presented a concise survey on the evolution of deep learning and state-of-the-art deep neural network architectures. We introduced sea floor exploration and the challenges involved in collecting and analysing marine data. Next, we presented a brief literature survey on marine image classification techniques. We further explored the potential applications of deep learning for benthic image classification by discussing the most recent studies which have been conducted by our group and other researchers. We also discussed a few future research directions in the fields of deep learning and underwater scene understanding. We expect that this chapter will encourage researchers from computer vision and marine societies to collaborate on similar long-term joint ventures.

8. Acknowledgments

This research was partially supported by Australian Research Council Grants (DP150104251 and DE120102960) and the Integrated Marine Observing System (IMOS) through the Department of Innovation, Industry, Science and Research (DIISR), National Collaborative Research Infrastructure Scheme. The authors also thank NVIDIA for providing a Titan-X GPU for the experiments involved in this research.

9. References

[1] Hedley JD, Roelfsema CM, Chollett I, Harborne AR, Heron SF, Weeks S, Skirving WJ, Strong AE, Eakin CM, Christensen TR, Ticzon V. Remote sensing of coral reefs for monitoring and management: A review. Remote Sensing. 2016 Feb 6; 8(2):118.

[2] Doney SC, Ruckelshaus M, Duffy JE, Barry JP, Chan F, English CA, Galindo HM, Grebmeier JM, Hollowed AB, Knowlton N, Polovina J. Climate change impacts on marine ecosystems. Marine Science. 2012; 4.

[3] Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E, Harvell CD, Sale PF, Edwards AJ, Caldeira K, Knowlton N. Coral reefs under rapid climate change and ocean acidification. Science. 2007 Dec 14; 318(5857):1737-42.

[4] Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C, Grosberg R, Hoegh-Guldberg O, Jackson JB, Kleypas J, Lough JM. Climate change, human impacts, and the resilience of coral reefs. Science. 2003 Aug 15; 301(5635):929-33.

[5] Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, Jackson JB, Lotze HK, Micheli F, Palumbi SR, Sala E. Impacts of biodiversity loss on ocean ecosystem services. Science. 2006 Nov 3; 314(5800):787-90.

[6] Shafait F, Mian A, Shortis M, Ghanem B, Culverhouse PF, Edgington D, Cline D, Ravanbakhsh M, Seager J, Harvey ES. Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science: Journal du Conseil. 2016 Jul 18:fsw106.

[7] 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 Proc Australas Conf Rob Autom 2012 Dec.

[8] Beijbom O, Edmunds PJ, Kline DI, Mitchell BG, Kriegman D. Automated annotation of coral reef survey images. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on 2012 Jun 16 (pp. 1170-1177).

[9] Kohler KE, Gill SM. Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. Computers & Geosciences. 2006 Nov 30; 32(9):1259-69.

[10] Solan M, Germano JD, Rhoads DC, Smith C, Michaud E, Parry D, Wenzhöfer F, Kennedy B, Henriques C, Battle E, Carey D. Towards a greater understanding of pattern, scale and process in marine benthic systems: a picture is worth a thousand worms. Journal of Experimental Marine Biology and Ecology. 2003 Feb 12; 285:313-38.

[11] Dolan MF, Lucieer VL. A review of marine geomorphometry, the quantitative study of the seafloor. Hydrology and Earth System Sciences. 2016 Aug 1; 20(8):3207.

[12] Patterson MR, Relles NJ. Autonomous underwater vehicles resurvey Bonaire: a new tool for coral reef management. In Proceedings of the 11th International Coral Reef Symposium 2008 Jul (pp. 539-543).

[13] Marcos MS, Soriano M, Saloma C. Classification of coral reef images from underwater video using neural networks. Optics express. 2005 Oct 31; 13(22):8766-71.

[14] Stokes MD, Deane GB. Automated processing of coral reef benthic images. Limnol. Oceanogr. Methods. 2009 Feb 1; 7(157):157-68.

[15] Pizarro O, Rigby P, Johnson-Roberson M, Williams SB, Colquhoun J. Towards image-based marine habitat classification. In OCEANS 2008. 2008 Sep 15 (pp. 1-7). IEEE.

[16] Shihavuddin AS, Gracias N, Garcia R, Gleason AC, Gintert B. Image-based coral reef classification and thematic mapping. Remote Sensing. 2013 Apr 15; 5(4):1809-41.

[17] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006 Jul 28; 313(5786):504-7.

[18] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems 2012 (pp. 1097-1105).

[19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.

[20] Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2014 (pp. 806-813).

[21] Liu L, Shen C, van den Hengel A. The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015 (pp. 4749-4757).

[22] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In European Conference on Computer Vision 2014 Sep 6 (pp. 818-833). Springer International Publishing.

[23] Mahendran A, Vedaldi A. Visualizing deep convolutional neural networks using natural pre-images. International Journal of Computer Vision. 2016 Apr 15:1-23.

[24] Gong Y, Wang L, Guo R, Lazebnik S. Multi-scale orderless pooling of deep convolutional activation features. In European Conference on Computer Vision 2014 Sep 6 (pp. 392-407). Springer International Publishing.

[25] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition 2014 (pp. 580-587).

[26] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In European Conference on Computer Vision 2014 Sep 6 (pp. 346-361). Springer International Publishing.

[27] Montufar GF, Pascanu R, Cho K, Bengio Y. On the number of linear regions of deep neural networks. In Advances in neural information processing systems 2014 (pp. 2924-2932).

[28] Pascanu R, Montufar G, Bengio Y. On the number of inference regions of deep feed forward networks with piece-wise linear activations. In International Conference on Learning Representations 2014 Apr.

[29] Lin M, Chen Q, Yan S. Network in network. arXiv preprint arXiv:1312.4400. 2013 Dec 1.

[30] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015 (pp. 1-9).

[31] Arora S, Bhaskara A, Ge R, Ma T. Provable Bounds for Learning Some Deep Representations. In International Conference on Machine Learning (ICML) 2014 Jun 21 (pp. 584-592).

[32] Srivastava RK, Greff K, Schmidhuber J. Highway networks. In International Conference on Machine Learning (ICML) 2015.

[33] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997 Nov 15; 9(8):1735-80.

[34] 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 2016.

[35] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. arXiv preprint arXiv:1603.05027. 2016 Mar 16.

[36] Xu Z, Zhu L, Yang Y, Hauptmann AG. Uts-cmu at thumos 2015. THUMOS challenge. 2015.

[37] Sánchez J, Perronnin F, Mensink T, Verbeek J. Image classification with the fisher vector: Theory and practice. International journal of computer vision. 2013 Dec 1; 105(3):222-45.

[38] Jégou H, Douze M, Schmid C, Pérez P. Aggregating local descriptors into a compact image representation. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on 2010 Jun 13 (pp. 3304-3311). IEEE.

[39] Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, Tomaszewski J, Gonzalez F, Madabhushi A. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging. 2014 Oct 1; 1(3):034003.

[40] Jin L, Gao S, Li Z, Tang J. Hand-crafted features or machine learnt features? Together they improve rgb-d object recognition. In Multimedia (ISM), 2014 IEEE International Symposium on 2014 Dec 10 (pp. 311-319). IEEE.

[41] Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, Fisher RB. Coral classification with hybrid feature representations. In 2016 IEEE International Conference on Image Processing (ICIP) 2016 Sep 25 (pp. 519-523). IEEE.

[42] Bewley M, Friedman A, Ferrari R, Hill N, Hovey R, Barrett N, Pizarro O, Figueira W, Meyer L, Babcock R, Bellchambers L. Australian sea-floor survey data, with images and expert annotations. Scientific data. 2015; 2.

[43] Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, Fisher RB. Automatic annotation of coral reefs using deep learning. In OCEANS 2016. 2016 Sep 20. IEEE.

[44] Khan SH, Bennamoun M, Sohel F, Togneri R. Cost Sensitive Learning of Deep Feature Representations from Imbalanced Data. arXiv preprint arXiv:1508.03422. 2015 Aug 14.

[45] Beijbom O, Treibitz T, Kline DI, Eyal G, Khen A, Neal B, Loya Y, Mitchell BG, Kriegman D. Improving Automated Annotation of Benthic Survey Images Using Wide-band Fluorescence. Scientific reports. 2016; 6.

Figures and Tables

Fig 1.1

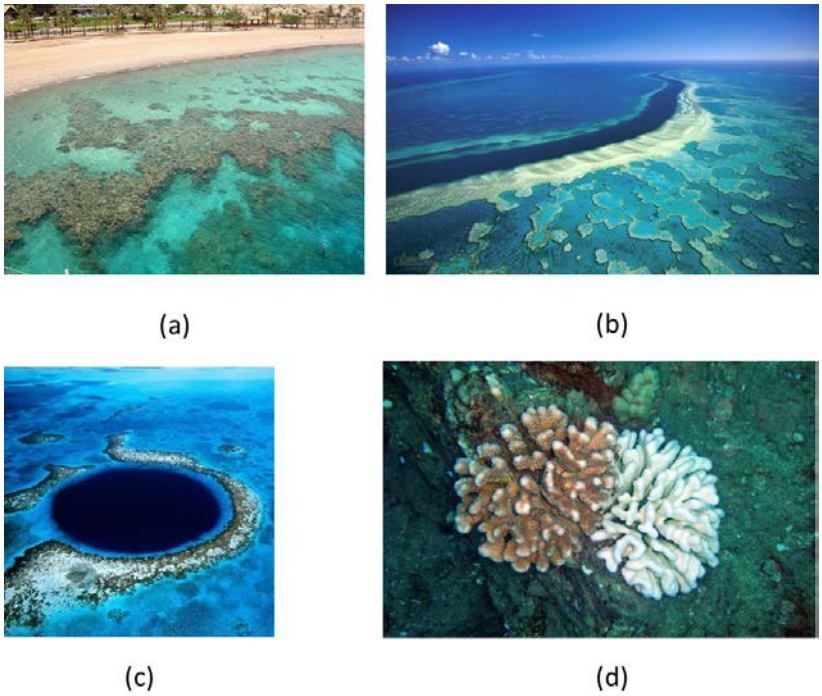

Fig 1.2

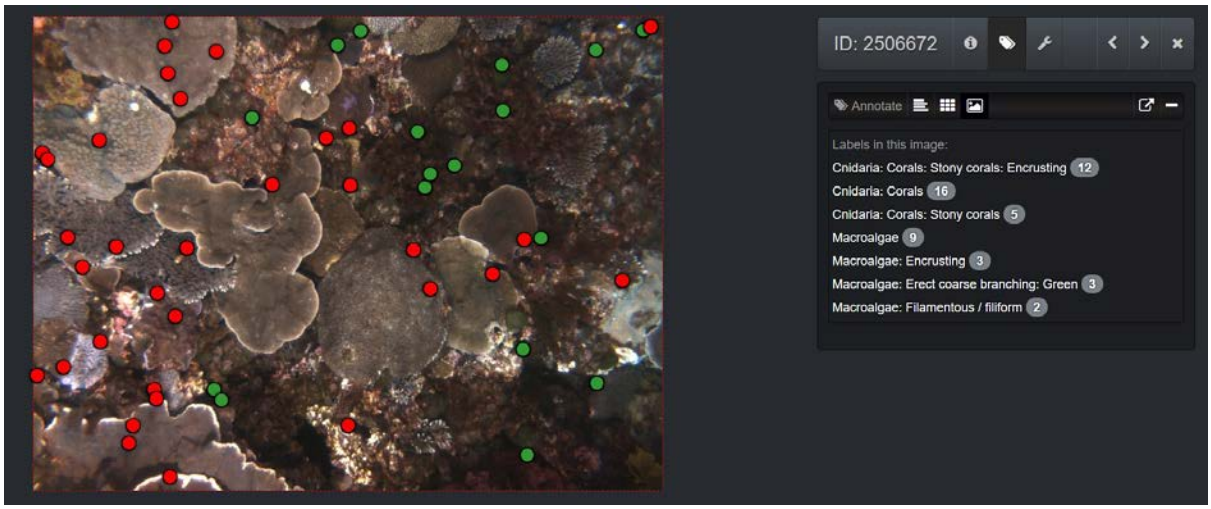

Fig 1.3

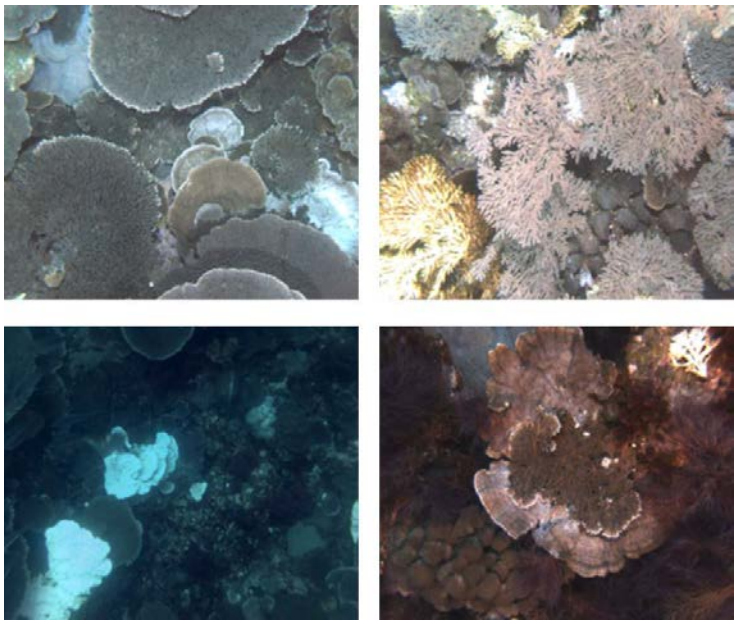

Fig 1.4

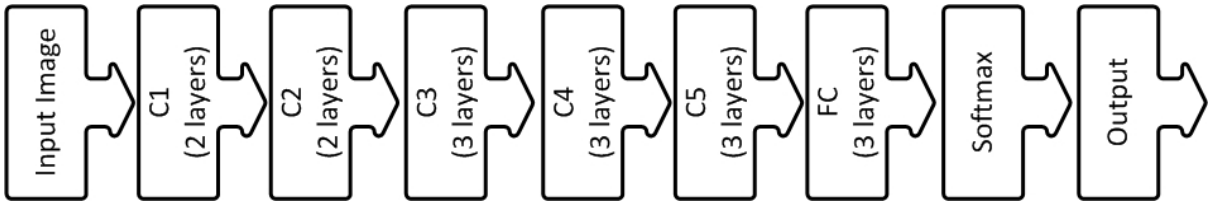

Fig 1.5

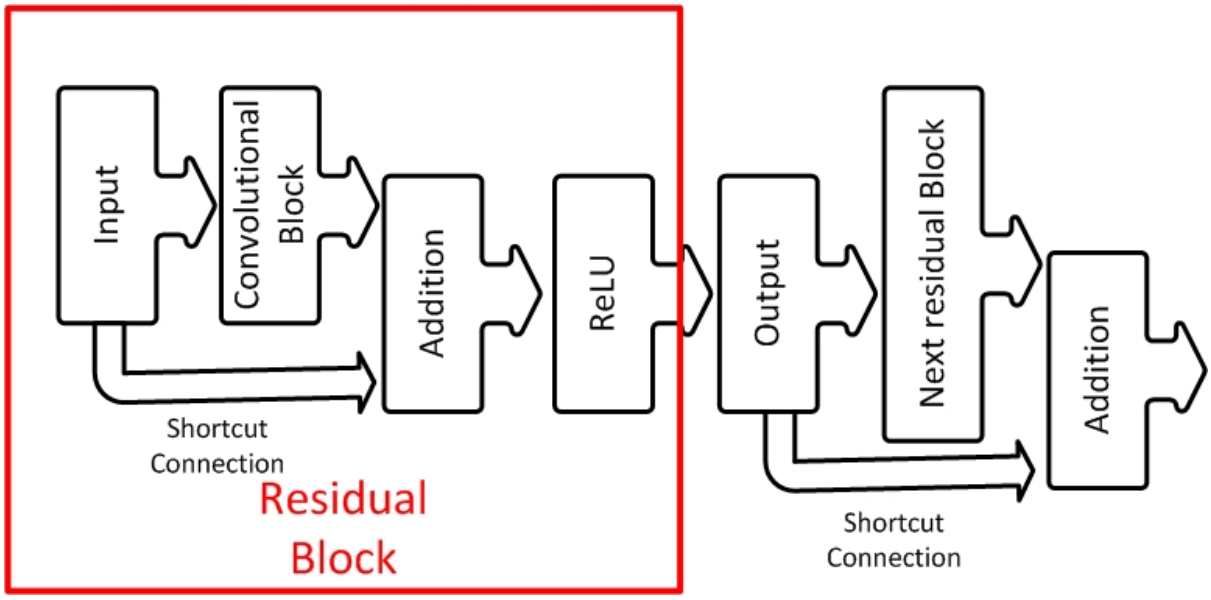

Fig 1.6

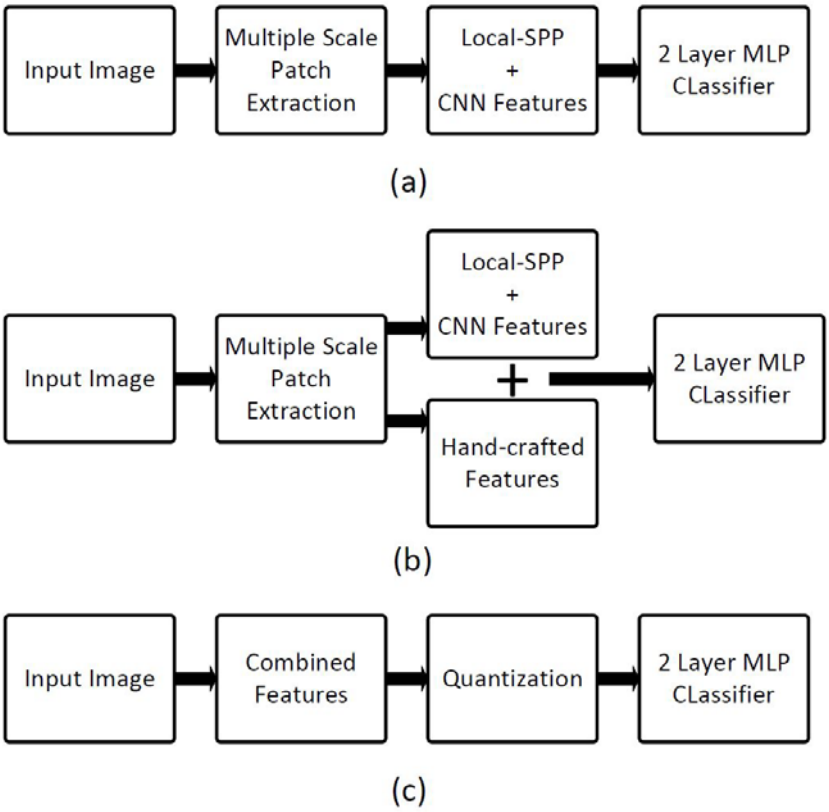
Input Image
Multiple Scale Patch Extraction
Local-SPP + CNN Features
2 Layer MLP CLassifier
(a)
Input Image
Multiple Scale Patch Extraction
Local-SPP + CNN Features
Hand-crafted Features
2 Layer MLP CLassifier
(b)
Input Image
Combined Features
Quantization
2 Layer MLP CLassifier
(c)

Fig 1.7

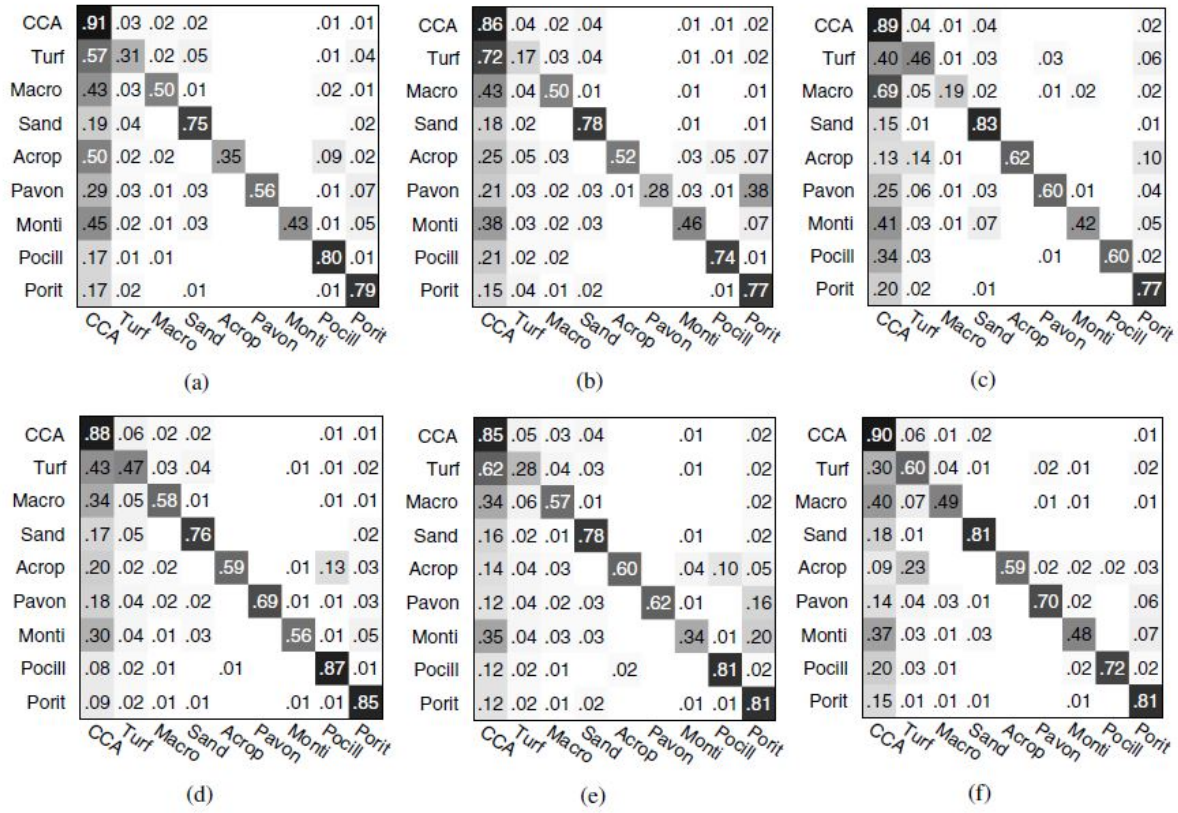
CCA .91 .03 .02 .02 .01 .01
Turf .57 .31 .02 .05 .01 .04
Macro .43 .03 .50 .01 .02 .01
Sand .19 .04 .75 .02
Acrop .50 .02 .02 .35 .09 .02
Pavon .29 .03 .01 .03 .56 .01 .07
Monti .45 .02 .01 .03 .43 .01 .05
Pocill .17 .01 .01 .80 .01
Porit .17 .02 .01 .01 .79
CCA Turf Macro Sand Acrop Pavon Monti Pocill Porit
(a)
CCA .86 .04 .02 .04 .01 .01 .02
Turf .72 .17 .03 .04 .01 .01 .02
Macro .43 .04 .50 .01 .01 .01
Sand .18 .02 .78 .01 .01
Acrop .25 .05 .03 .52 .03 .05 .07
Pavon .21 .03 .02 .03 .01 .28 .03 .01 .38
Monti .38 .03 .02 .03 .46 .07
Pocill .21 .02 .02 .74 .01
Porit .15 .04 .01 .02 .01 .77
CCA Turf Macro Sand Acrop Pavon Monti Pocill Porit
(b)
CCA .89 .04 .01 .04 .02
Turf .40 .46 .01 .03 .03 .06
Macro .69 .05 .19 .02 .01 .02 .02
Sand .15 .01 .83 .01
Acrop .13 .14 .01 .62 .10
Pavon .25 .06 .01 .03 .60 .01 .04
Monti .41 .03 .01 .07 .42 .05
Pocill .34 .03 .01 .60 .02
Porit .20 .02 .01 .77
CCA Turf Macro Sand Acrop Pavon Monti Pocill Porit
(c)
CCA .88 .06 .02 .02 .01 .01
Turf .43 .47 .03 .04 .01 .01 .02
Macro .34 .05 .58 .01 .01 .01
Sand .17 .05 .76 .02
Acrop .20 .02 .02 .59 .01 .13 .03
Pavon .18 .04 .02 .02 .69 .01 .01 .03
Monti .30 .04 .01 .03 .56 .01 .05
Pocill .08 .02 .01 .01 .87 .01
Porit .09 .02 .01 .01 .01 .01 .85
CCA Turf Macro Sand Acrop Pavon Monti Pocill Porit
(d)
CCA .85 .05 .03 .04 .01 .02
Turf .62 .28 .04 .03 .01 .02
Macro .34 .06 .57 .01 .02
Sand .16 .02 .01 .78 .01 .02
Acrop .14 .04 .03 .60 .04 .10 .05
Pavon .12 .04 .02 .03 .62 .01 .16
Monti .35 .04 .03 .03 .34 .01 .20
Pocill .12 .02 .01 .02 .81 .02
Porit .12 .02 .01 .02 .01 .01 .81
CCA Turf Macro Sand Acrop Pavon Monti Pocill Porit
(e)
CCA .90 .06 .01 .02 .01
Turf .30 .60 .04 .01 .02 .01 .02
Macro .40 .07 .49 .01 .01 .01
Sand .18 .01 .81
Acrop .09 .23 .59 .02 .02 .02 .03
Pavon .14 .04 .03 .01 .70 .02 .06
Monti .37 .03 .01 .03 .48 .07
Pocill .20 .03 .01 .02 .72 .02
Porit .15 .01 .01 .01 .01 .81
CCA Turf Macro Sand Acrop Pavon Monti Pocill Porit
(f)

Fig 1.8

Table 1.1

Features	***Number of Classes***	***Ref***
NCC Histogram for color and LBP for texture	*5*	*[27]*
RGB Histogram for color and DCT+LBP for texture	*18*	*[28]*
NCC Histogram for Color and bag of words SIFT	*8*	*[29]*
*L*a*b colorspace and MR filter bank + texton maps*	*9*	*[22]*
CLBP + GLCM + Gabor filter	*Multiple Datasets*	*[30]*