Deep Learning on Underwater Marine Object Detection: A Survey

Md. Moniruzzaman[1], Syed Mohammed Shamsul Islam[1,2(✉)],
Mohammed Bennamoun[2], and Paul Lavery[1]

[1] School of Science, Edith Cowan University, Joondalup, WA 6027, Australia
mmoniruz@our.ecu.edu.au,
{syed.islam,p.lavery}@ecu.edu.au
[2] School of Computer Science and Software Engineering,
University of Western Australia, Crawley, WA 6009, Australia
mohammed.bennamoun@uwa.edu.au

Abstract. Deep learning, also known as deep machine learning or deep structured learning based techniques, have recently achieved tremendous success in digital image processing for object detection and classification. As a result, they are rapidly gaining popularity and attention from the computer vision research community. There has been a massive increase in the collection of digital imagery for the monitoring of underwater ecosystems, including seagrass meadows. This growth in image data has driven the need for automatic detection and classification using deep neural network based classifiers. This paper systematically describes the use of deep learning for underwater imagery analysis within the recent past. The analysis approaches are categorized according to the object of detection, and the features and deep learning architectures used are highlighted. It is concluded that there is a great scope for automation in the analysis of digital seabed imagery using deep neural networks, especially for the detection and monitoring of seagrass.

Keywords: Deep learning · Underwater · Marine · Object detection · Seagrass · Neural network · Convolutional architecture

1 Introduction

Oceans are like the lifeblood of Mother Nature, holding 97% of the earth's water. They produce more than half of the oxygen and absorb most of the carbon from our environment. Maintaining these and other oceanic ecosystem services requires maintenance of critical marine habitats. Important among these are seagrass meadows and coral reefs, which are critical to marine foodwebs, habitat provision and nutrient cycling [29]. For example, dredging, physically remove benthic marine species, like seagrasses, can lead to their burial and can reduce the light necessary for photosynthesis [3]. Tourism, shipping, urbanization and human intervention are damaging coral colonies, with 19% of the world's coral reefs having been destroyed by 2011 and 75% threatened [4]. Monitoring is an important aspect of any robust effort to manage these destructive

impacts but can be an arduous task. Marine optical imaging technology offers enormous potential to make monitoring more efficient, in terms of both cost and time.

Many marine management strategies incorporate remote sensing and tracking of marine habitats and species. In recent years, the use of digital cameras, autonomous underwater vehicles (AUV) and unmanned underwater vehicles (UUV) has led to an exponential increase of availability of underwater imagery [9]. The Integrated Marine Observing System (IMOS) collects millions of images of coral reefs around Australia, but less than 5% go through expert marine analysis. For the National Oceanic and Atmospheric Administration, the rate is even lower, only 1–2% [1]. For this reason, it is now a research priority to analyse marine digital data automatically. To solve this issue, deep learning, the state-of-art machine learning technology, provides potentially unprecedented opportunities for many underwater objects [12].

Low-level manually designed features have been used in traditional classification solution so far. Face and texture classification is done by Gabor and Local Binary Patterns (LBP) while features and object recognition is regularly done by Scale Invariant Feature Transform (SIFT) and Histogram of oriented gradients (HOG) hand-crafted features. In the case of specific task and data, careful execution of hand-crafted features have achieved good performances. But many of them cannot be reused for a new situation without core changing. Moreover, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA) and other machine learning conventional tools are quickly saturated when the training data volume increases. Hinton et al. [5] proposed learning features using deep neural networks (DNNs) to address these short comings. To make sense of texts, images, sounds etc., deep learning transforms input data through more layers than shallow learning algorithms [19]. At each layer, the signal is transformed by a processing unit, like an artificial neuron, whose parameters are 'learned' through training [20]. Deep learning is replacing handcrafted features, with efficient algorithms for feature learning and hierarchical feature extraction [21]. Deep learning attempts to make better representations of an observation (e.g. an image) and create models to learn these representations from large-scale data.

By the use of large amounts of training data, large and deep networks demonstrated excellent success. For example, convolutional neural network which is trained through ImageNet has achieved unprecedented accuracy in image classification [6]. They have been applied in the field of object detection [7], image classification [6], face verification [22], digits and traffic signs recognition [23] etc. and demonstrated high performance. However, deep learning has not been widely applied in marine object detection and classification.

A survey on the current deep learning approaches for various marine object detection and classification would help researchers understanding the challenges and explore more efficient possibilities. ***To the best of our knowledge, this paper is the first survey on such approaches.***

The rest of the paper is organized as follows. The existing approaches for automated marine object detection on digital data are discussed in Sect. 2. Associated challenges, especially for seagrass identification have been outlined in Sect. 3 and finally, conclusions are drawn in Sect. 4.

2   Approaches for Underwater Marine Object Detection

All the known machine learning approaches especially those using deep neural network in digital marine data analysis, image annotation, object detection and classification are discussed in this section. The approaches are categorized according to the object of detection. Features and classifiers used in each of the approaches are also highlighted and summarized in Table 1 and discussed in the follow sections.

Table 1. Summary of deep learning approaches for marine object detection

Target group	Author (publication year)	Type of image or dataset	Feature used	Classifier
Fish	Li et al. (2015) [8]	RBG photos and videos from LifeCLIEF Fish Task of ImageCLIEF	RGB color space	Fast R-CNN
	Villon et al. (2016) [25]	Marine Biodiversity Exploitation and Conservation Dataset	Motion from previous sliding window	Soft max Classifier with Deep Network
Planktons	Dieleman (2015) [30]	Grey scale images provided by National Data Science Bowl	Shapes and rotational symmetry	ConvNNet inspired by OxfordNet
	Py et al. (2016) [26]	Grey scale images provided by National Data Science Bowl	Inception module for multi scale architecture	Deep CNN inspired by GoogleNet
	Lee et al. (2016) [27]	Woods Hole Oceanographic Institution (WHOI-Plankton) dataset	Transfer Learning to reduce Class imbalance	CIFAR 10 Convolutional Neural Network
	Dai et al. (2016) [28]	ZooScane System Dataset	Data Augmentation to increase the dataset	ZooPlanktoNet inspired by AlexNet and VGGNet
Corals	Shiela et al. (2008) [14]	Video stills of coral reef transects from the Great Barrier Reef	Local Binary Pattern (LBP) for texture & Normalized Chromaticity Coordinates histogram for color	Linear Discriminant Analysis followed by a three layer back propagation neural network.
	Elawady (2014) [24]	Moorea Labeled Corals and Heriot-Watt University Atlantic Deep Sea Digital dataset	Color Shape Texture feature Descriptors	Supervised Convolutional Neural Networks (CNNs)
	Mahmood et al. (2016) [10]	Moorea Labelled Coral (MLC) dataset	Texton and color based hand-crafted features Spatial Pyramid Pooling (SPP)	VGGNet

2.1 Deep Learning in Fish Detection and Classification

Before 2015, very few attempts were taken to integrate deep learning on fish recognition. Haar classifiers were used by Ravanbakhsh et al. [13] to classify shape features. Principal Component Analysis (PCA) modelled the features. To get a balance of accuracy and processing time for underwater fish detection, Spampinato et al. [15] used moving average algorithm. Both of these methods have limited ability to process large amount of underwater imagery. Li et al. [8] first introduced deep convolution network for fish detection and recognition. They used Fast Region-based Convolutional Neural Network (Fast R-CNN) to detect fish efficiently and accurately. They also constructed a clean fish dataset of 24272 images over 12 classes, a subset of ImageCLIEF training and test dataset. As illustrated in Fig. 1, they pre-trained an AlexNet on a large auxiliary dataset (ILSVRC2012) with five convolutional layers and fully connected three layers by caffe CNN library which is an open source one. They modified AlexNet so that the Fast R-CNN can be adopted to train the Fast R-CNN parameters; they used stochastic gradient descent (SGD). Their experimental outcome showed better performance with a higher maximum a posteriori estimation (mAP). They got an average 9.4% higher precision than Deformable Parts Model (DPM). Table 2 shows the performance of their approach in fish detection compared to different other approaches using non-deep learning techniques.
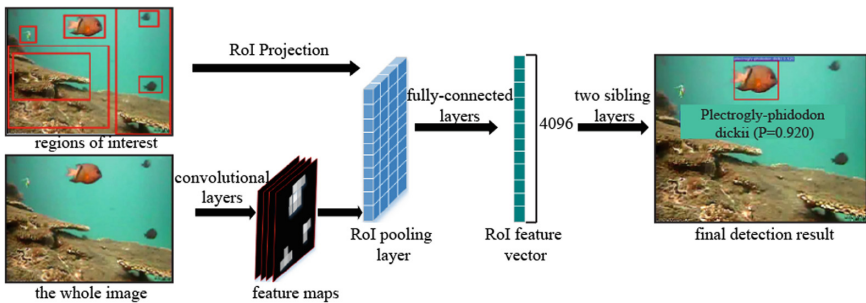

Fig. 1. Architecture of fish detection and recognition using fast R-CNN (adapted from [8]).

Table 2. Fish recognition accuracy comparison (adopted from [12])

Method	Accuracy (%)
LDA + SVM	80.14
Raw-pixel SVM	82.92
Raw-pixel Softmax	87.56
Raw-pixel Nearest Neighbor	89.79
VLFeat Dense-SIFT	93.58
Deep-CNN	**98.57**

Villon et al. [25] evaluated the effectiveness of the deep learning against Ground-Truth dataset made by the Fish4Knowledge project. They also compared the performance of deep learning for fish detection with a traditional system combined with Support Vector Machines (SVM) classification and HOG feature extraction. The architecture of their deep network had nine inception layers, 27 layers with a soft max classifier and was inspired by the GoogleNet [32].

2.2   Deep Learning in Plankton Classification

Planktons are frequently the foundation for aquatic food webs and therefore are frequently monitored as indicators of ecosystem condition. Conventional plankton monitoring and measurement systems are not adequate to meet the scope of large scale studies. In 2015, The National Data Science Bowl [30], a data science competition, was held to classify the images of plankton with the support of Hatfield Marine Science Centre of Oregon State University. The winning team was a group of researchers lead by Prof. Joni Dambre from Ghent University in Belgium using convolutional neural network. While it generally thought that enormous datasets are required for the deep learning approaches, the classification accuracy in this case was 81.52% where there were about 30000 examples for 121 classes and some of the classes had less than 20 examples in total. The winning team's output feature maps were the same as the input maps and the pooling and overlapping had window size 3 and stride 2. By starting with a fairly shallow model of six layers and, gradually increasing the number of layers, the final structure had 16 layers. To give network the ability to use the same feature extraction pipeline to look at the input from different angles, a cyclic pooling technique was used where the same stack of convolutional layers were applied and fed into a stack of dense layers and at the top the feature maps were pooled together. Finally, the stacks of cyclic pooling output feature maps from different orientations were combined into one large stack and then the next layer was learned on this combined input which adds four times more filters than it actually had. The operation that combines feature maps from different orientations was named a 'roll' (Fig. 2).
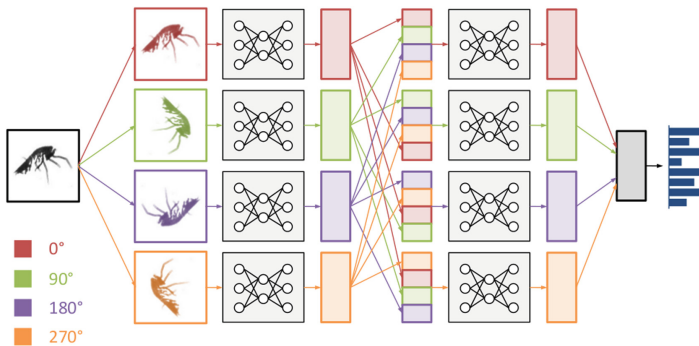

Fig. 2. Roll operation with cyclic pooling (adopted from [31]).

Using the same dataset of National Dataset Bowl of 2015 and inspired by GoogleNet, another published approach of plankton classification was done by Py et al. [26]. They proposed and developed an inception module with a convolutional layer for distortion minimization and maximization of image information extraction. Inside the network, improved utilization of computing resources was the hallmark of their network architecture. Data augmentation was done to co-op with rotational and translational invariant and rotational affine was applied to data augmentation. They divided a deep convolutional Neural Network into classifier part and feature part. But they found, this kind of design of classifier part is prone to overfitting if the dataset is not large enough and, replacing the last two fully connected layers with small kernels was better for such dataset. Performance of their model was better than the state of the art models for particular size of images [26].

Deep network approach for classification of plankton using a much larger dataset was done by Lee et al. [27]. They worked with the WHOI-Plankton dataset (developed by Woods Hole Oceanographic Institution) which had 3.4 million expert-labeled images of 103 classes. In their approach, they mainly focused on solving the class imbalance problem of a large dataset. For the reduction of bias from class imbalance, they chose the CIFAR 10 CNN model as a classifier. Their proposed architecture had three convolutional layers followed by two fully connected layers. At first their classifier was pre-trained on class normalized data and then re-trained on the original data which helped reducing the class imbalance biasness [27].

Introduction of deep convolutional network solely for the classification of Zooplankton was done by Dai et al. [28]. Their dataset was consisting of 9460 microscopic and grey scale zooplankton images of 13 different classes captured by ZooScan system. They proposed a new deep learning architecture called ZooplanktoNet for zooplankton classification which is strongly inspired by AlexNet and VGGNet. After experimenting with different sizes of convolution, they concluded that ZooplanktoNet with 11 layers can provide the best performance so far. To support their claim, they did a comparative experiment with other deep learning architectures like AlexNet, CaffeNet, VGGNet and GoogleNet and found that ZooplanktoNet performs better with an accuracy of 93.7% [28].

2.3 Deep Learning in Coral Classification

The color, size, shape and texture of corals may vary according to the class difference. Moreover, the boundary differences are ambiguous and organic. Furthermore, currents, algal blooms, density of planktons can change the turbidity of water and light availability, affecting the image color. These kinds of challenges make conventional annotation techniques like, bounding boxes, situ analysis in line or point transects, image labels or full segmentation inappropriate [1, 16].

Local Binary Pattern (LBP) for texture and Normalized Chromaticity Coordinate (NCC) for color were used by Shiela et al. [14]. They used a three layer back propagation neural network for classification purposes. However, Beijbom et al. [1] first addressed automated annotation on a large scale for coral reef survey image by introducing the Moorea Labelled Corals (MLC) dataset. They proposed a method based on color and texture descriptors over multiple scales and it out performed traditional

methods for texture classification. Elawady et al. [24] used supervised Convolutional Neural Networks (CNNs) for coral classification. They worked on Moorea Labeled Corals and Heriot-Watt University's Atlantic Deep Sea Dataset and computed Phase Congruency (PC), Zero Component Analysis (ZCA) and Weber Local Descriptor (WLD). With spatial color channels they also considered shape and texture features for input images [24].

For making the conventional point-annotated marine data compatible to the input constraints of CNNs, Mahmood et al. [10] proposed a feature extraction scheme based on Spatial Pyramid Pooling (SPP) (as shown in Fig. 3). They used deep features extracted from the VGGNet [10] for coral classification. They also combined texton and color based hand-crafted features to improve capability of classification. The block diagram of the combined approach is illustrated in Fig. 4.
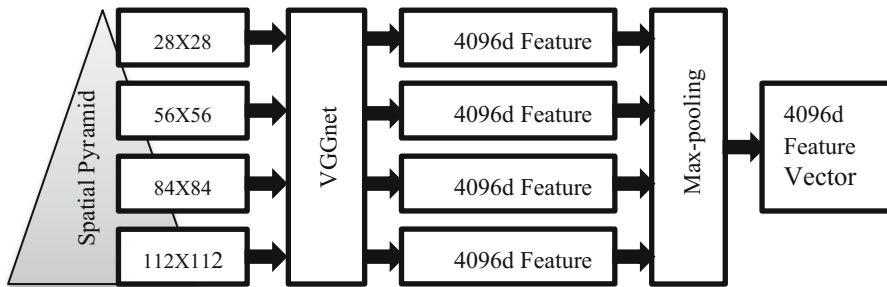
Fig. 3. Local-SPP based feature extraction scheme from the VGGNet for coral classification (adopted from [10]).
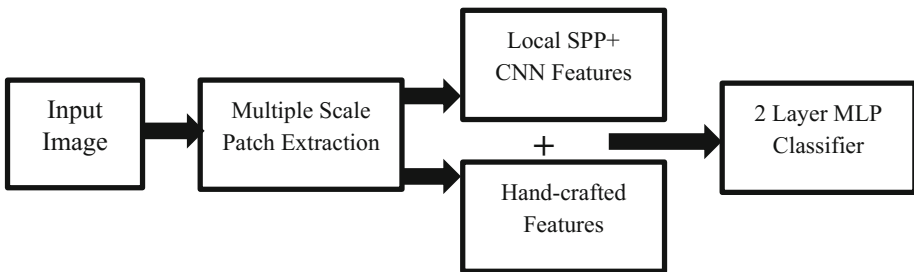
Fig. 4. CNNs architecture combined with Texton and color based hand-crafted features to improve capability for coral identification and classification (adopted from [10]).

2.4   Deep Learning Opportunities for Seagrass Detection and Classification

For the stabilization of sediment, sequestration of carbon and provision of food and habitat for enormous oceanic animals, sea grasses are very vital [7]. To improve the understanding of the temporal and spatial patterns in species composition, reproductive

phenology and abundance of seagrass and the influence of commercialization and human interaction, it is very important to monitor seagrass in more and more areas.

In 2013, Teng et al. [17] performed the binary classification of seagrass using hyperspectral images from seagrass habitats to separate tube worms from rest of the seagrass surface. More specific work to quantify the presence of the seagrass *Posidonia oceanica* in Palma Bay was performed by Campos et al. [2]. They used analogic RGB data. They chose Logistic Model Tree (LMT) classifier and Law's energy measurements. Grey level co-occurrence matrix was used to identify the differences in texture. Oguslu et al. [11] used sparse coding and morphological filter to detect seafloor propeller seagrass scars in shallow water using panchromatic images captured using WorldView2 orbiting satellite. This approach was only effective in the shallow coastline and for detecting the scars in the shore line.

Presently, as a conventional digital imagery approach approved by Commonwealth Scientific and Industrial Research Organization (CSIRO) and Health Safety and Environment Policies (HSE), Australia, images approximately 60×80 cm are taken from a digital camera every three seconds. The camera is normally kept attach to a frame towed behind a boat travelling at 1.5–3 knots which ensures the images are spaced approximately 2–3 m apart. These images are then analyzed using photo Grid or TranscetMeasure (®SeaGIS) software. A regular grid of 20 dots are superimposed (Fig. 5) and a human operator identifies the presence and species of seagrasses [18]. It typically takes a technician several hours to process image data for a single transect of 50 m and with 25–50 images. As most surveys require several hundreds of meters of seabed to be covered, it can require several days to perform the analysis. Furthermore, different technicians may vary in their ability to detect seagrass within images. Deep learning approaches may increase efficiency and simultaneously remove observer bias for the analyses. However, to best of our knowledge, ***there is no approach existing that applies deep learning to digital images for seagrass detection***. Therefore, there is
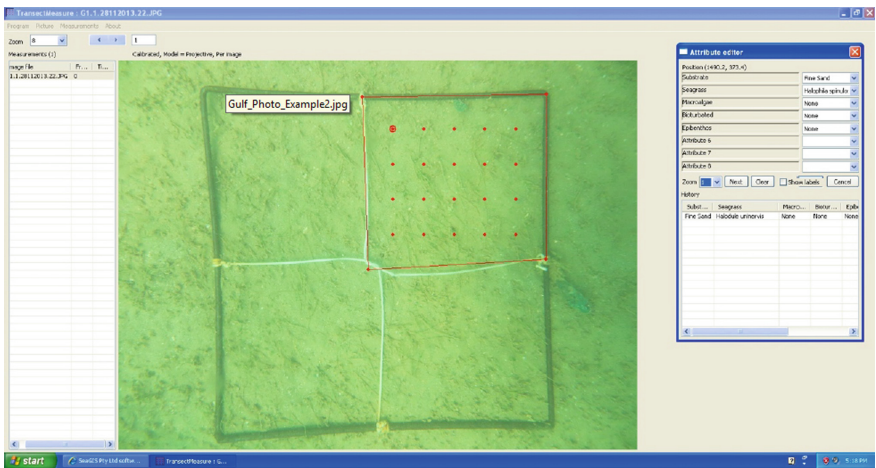

Fig. 5. A screenshot of the TransectMeasure software, used to analyze seagrass [18].

a great opportunity to use deep neural network to analyse the deep sea bed, detect and classify the species of seagrasses. We are going to focus in this matter in our future work.

3   Challenges

Visual content recognition is the most important problem and a quite challenging task for underwater imagery analysis. Intra-class variability produces the variation of visual content through views, scales, illumination and non-rigid deformation. Especially, for the detection and classification of seagrasses, the boundary differences in different classes are much more ambiguous than for fish or corals. Also in the digital images, visual content becomes more ambiguous as the depth of the water increases.

4   Conclusion

In this paper, recent approaches for detecting and classifying various underwater marine objects using deep learning are discussed. Approaches are categorized according to the targets of detection. Features and deep learning architectures used are summarized. It was necessary to highlight all the approaches of marine data analysis in a single paper so that it becomes easy to focus on the possibilities of future work based on deep neural network method. It has been found that more works have be done for coral detection and classification using deep learning but no work has been done for the case of seagrass which is equally vital for oceanic ecosystem. The effectiveness, accuracy and robustness of any detection and classification algorithm can be increased significantly if both color and texture based features are combined. Accumulation of hand-crafted features and neural network may bring better results for seagrass detection and classification. Therefore, the opportunity exists to develop an efficient and effective deep learning approach for underwater seagrass imagery, which will be the focus of our future work.

References

1. Beijbom, O., Edmunds, P.J., Kline, D.I., Mitchell, B.G., Kriegman, D.: Automated annotation of coral reef survey images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1170–1177. IEEE Press, Rhode Island (2012)
2. Campos, M.M., Codina, G.O., Amengual, L.R., Julia, M.M.: Texture analysis of seabed images: quantifying the presence of Posidonia Oceanica at palma bay. In: OCEANS 2013 - MTS/IEEE Bergen, pp. 1–6. IEEE Press, Bergen (2013)
3. Erftemeijer, P.L.A., Lewis III, R.R.R.: Environmental impacts of dredging on seagrasses: a review. Marine Pollut. Bull. **52**, 1553–1572 (2006)

4. Hedley, J.D., Roelfsema, C.M., Chollett, I., Harborne, A.R., Heron, S.F., Weeks, S.J., Mumby, P.J.: Remote sensing of coral reefs for monitoring and management : a review. Remote Sens. **8**(118), 1–40 (2016)
5. Hinton, G.E., Osindero, S.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25 (NIPS) (2012)
7. Lavery, P.S., McMahon, K., Weyers, J., Boyce, M.C., Oldham, C.E.: Release of dissolved organic carbon from seagrass wrack and its implications for trophic connectivity. Mar. Ecol. Prog. Ser. **494**, 121–133 (2013)
8. Li, X., Shang, M., Qin, H., Chen, L.: Fast accurate fish detection and recognition of underwater images with fast R-CNN. In: OCEANS 2015 - MTS/IEEE Washington, pp. 1–5. IEEE Press, Washington, DC (2015)
9. Lu, H., Li, Y., Zhang, Y., Chen, M., Serikawa, S., Kim, H.: Underwater optical image processing: a comprehensive review. Mob. Netw. Appl. **22**, 1–8 (2017)
10. Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., Fisher, R.B.: Coral classification with hybrid feature representations. In: IEEE International Conference on Image Processing, pp. 519–523. IEEE Press, Arizona (2016)
11. Oguslu, E., Erkanli, S., Victoria, J., Bissett, W.P., Zimmerman, R.C., Li, J.: Detection of seagrass scars using sparse coding and morphological filter. In: Proceedings of SPIE 9240, Remote Sensing of the ocean, Sea Ice, Coastal Waters, and Large Water Regions, pp. 92400G, Amsterdam (2014)
12. Qin, H., Li, X., Yang, Z., Shang, M.: When underwater imagery analysis meets deep learning : a solution at the age of big visual data. In: OCEANS 2015 - MTS/IEEE Washington, pp. 1–5. IEEE Press, Washington, DC (2015)
13. Ravanbakhsh, M., Shortis, M., Shafait, F., Mian, A., Harvey, E., Seager, J.: Automated fish detection in underwater images using shape-based level sets. Photogram. Rec. **30**(149), 46–62 (2015)
14. Shiela, M.M.A., Soriano, M., Saloma, C.: Classification of coral reef images from underwater video using neural networks. Opt. Express **13**(22), 8766–8771 (2005)
15. Spampinato, C., Chen-Burger, Y.H., Nadarajan, G., Fisher, R.B.: Detecting, tracking and counting fish in low quality unconstrained underwater videos. In: 3rd International Conference on Computer Vision Theory and Applications (VISAPP), pp. 514–519. Funchal, Madeira (2008)
16. Stokes, M.D., Deane, G.B.: Automated processing of coral reef benthic images. Limnol. Oceanogr. Methods. **7**, 157–168 (2009)
17. Teng, M.Y., Mehrubeoglu, M., King, S.A., Cammarata, K., Simons, J.: Investigation of Epifauna coverage on seagrass blades using spatial and spectral analysis of hyperspectral images. In: 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, pp. 25–28, Gainsville, Florida (2013)
18. Vanderklift, M., Bearham, D., Haywood, M., Lozano-Montes, H., Mccallum, R., Mclaughlin, J., Lavery, P.: Natural dynamics: understanding natural dynamics of seagrasses in North-Western Australia. Theme 5 Final Report of the Western Australian Marine Science Institution (WAMSI) Dredging Science Node. 39 (2016)
19. Deng, L., Yu, D.: Methods and applications. Found. Trends. Sig. Process. **7**, 197–387 (2013)
20. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)

21. Song, H.A., Lee, S.-Y.: Hierarchical representation using NMF. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013. LNCS, vol. 8226, pp. 466–473. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-42054-2_58

22. Liu, M.: AU-aware deep networks for facial expression recognition. In: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1–6. IEEE Press, Shanghai (2013)

23. Ciresan, D., Meier, U., Schmidhuber, J: Multi-column deep neural networks for image classification. In: Computer Vision and Pattern Recognition, pp. 3642–3649 (2012)

24. Elawady, M.: Sparsem: coral classification using deep convolutional neural networks. M.sc. thesis. Hariot-Watt University (2014)

25. Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T., Mouillot, D.: Coral reef fish detection and recognition in underwater videos by supervised machine learning: comparison between deep learning and HOG+SVM methods. In: Blanc-Talon, J., Distante, C., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2016. LNCS, vol. 10016, pp. 160–171. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48680-2_15

26. Py, Q., Hong, H., Zhongzhi, S.: Plankton classification with deep convolutional neural network. In: IEEE Information Technology, Networking, Electronic and Automation Control Conference, pp. 132–136, Chongqin (2016)

27. Lee, H., Park, M., Kim, J.: Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In: IEEE International Conference on Image Processing (ICIP), pp. 3713–3717, Phoenix (2016)

28. Dai, J., Wang, R., Zheng, H., Ji, G., Qiao, X.: ZooplanktoNet: deep convolutional network for Zooplankton classification. In: OCEANS, pp. 1–6, Shanghai (2016)

29. Hemminga, M.A., Duarte, C.M.: Seagrass Ecology. Cambridge University Press, Cambridge (2004)

30. National Data Science Bowl. https://www.kaggle.com/c/datasciencebowl

31. Dieleman, S.: Classifying Planktons with Deep Neural Networks. http://benanne.github.io/2015/03/17/plankton.html

32. Szegedy, C., Liu, W., Jia, Y.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)