

Computer Vision and Deep Learning for Fish Classification in Underwater Habitats: A Survey

Alzayat Saleh¹ | Marcus Sheaves¹ |

Mostafa Rahimi Azghadi^{1,2}

¹College of Science and Engineering, James Cook University, Townsville, QLD, Australia

²ARC Research Hub for Supercharging Tropical Aquaculture through Genetic Solutions, James Cook University, Townsville, QLD, Australia

Correspondence

Mostafa Rahimi Azghadi, PhD, College of Science and Engineering, James Cook University, Townsville, QLD, Australia
Email: mostafa.rahimiazghadi@jcu.edu.au

Present address

College of Science and Engineering, James Cook University, Townsville, QLD, Australia

Funding information

This research is supported by an Australian Research Training Program (RTP) Scholarship, and the Australian Research Council funding through their Industrial Transformation Research Program.

Marine scientists use remote underwater image and video recording to survey fish species in their natural habitats. This helps them get a step closer toward understanding and predicting how fish respond to climate change, habitat degradation, and fishing pressure. This information is essential for developing sustainable fisheries for human consumption, and for preserving the environment. However, the enormous volume of collected videos makes extracting useful information a daunting and time-consuming task for a human. A promising method to address this problem is the cutting-edge Deep Learning (DL) technology. DL can help marine scientists parse large volumes of video promptly and efficiently, unlocking niche information that cannot be obtained using conventional manual monitoring methods. In this paper, we first provide a survey of Computer Vision (CV) and DL studies conducted between 2003-2021 on fish classification in underwater habitats. We then give an overview of the key concepts of DL, while analyzing and synthesizing DL studies. We also discuss the main challenges faced when developing DL for underwater image processing and propose approaches to address them. Finally, we provide insights into the marine habitat monitoring research domain and shed light on what the future of DL for underwater image processing may hold. This paper aims to inform marine scientists who would like to gain a high-level understanding of essential DL concepts and survey state-of-the-art DL-based fish classification in their underwater habitat.

KEYWORDS

Fish Habitat, Monitoring, Computer Vision, Deep Learning.

NOMENCLATURE

AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUV	Autonomous Underwater Vehicle
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
DNN	Deep Neural Networks
FCN	Fully Convolutional Network
LSTM	Long short-term memory
ML	Machine Learning
OCR	Optical character recognition
RNN	Recurrent Neural Network
ROV	Remotely Operated Vehicles
RUV	Remote Underwater Video

1 | INTRODUCTION

Understanding and modelling how fish respond to climate change, habitat degradation, and fishing pressure are critical for environmental protection, and are crucial steps toward ensuring sustainable natural fisheries, to support ever-growing human consumption (Zarco-Perello and Enríquez, 2019). Effective monitoring is a vital first step underpinning decision support mechanisms for identifying problems and planning actions to preserve and restore the habitats. However, there is still a gap between the complexity of marine ecosystems and the available monitoring mechanisms.

Marine scientists use underwater cameras to record, model, and understand fish habitats and fish behaviour. Remote Underwater Video (RUV) recording in marine applications (Zarco-Perello and Enríquez, 2019) has shown great potential for fisheries, ecosystem management, and conservation programs (Piggott et al., 2020). With the introduction of consumer-grade high-definition cameras, it is now feasible to deploy a large number of RUVs or Autonomous Underwater Vehicles (AUVs) to collect substantial volumes of data and to perform more effective monitoring (Pope et al., 2010; Rasmussen and Morrissey, 2008; Thorstad et al., 2013). However, underwater habitats introduce diverse video monitoring challenges such as adverse water conditions, high similarity

between fish species, cluttered backgrounds, and occlusions among fish. In addition, the volume of data generated by deployed RUVs and AUVs rapidly surpasses the capacity of human video viewers, making video analysis prohibitively expensive (Konovalov et al., 2019a). Moreover, humans are more prone to error than a well-designed machine-centred monitoring algorithm. Therefore, an automated, comprehensive monitoring system could significantly reduce labour expenses while improving throughput and accuracy, increasing the precision in estimates of fish stocks, fish distribution and biodiversity in general (Hilborn and Walters, 1992). Implementing such systems necessitates effective Computer Vision (CV) processes. As a result, significant research has been conducted on implementing monitoring tools and techniques that build upon CV algorithms for determining how fish exploit various maritime environments and differentiating between fish species (Zion, 2012).

In image analysis and CV domains, Deep Learning (DL) approaches have consistently produced state-of-the-art results in a variety of applications from agriculture (Olsen et al., 2019) to medicine (Saleh et al., 2021; Azghadi et al., 2020) using Deep Neural Networks (DNNs) (Zheng et al., 2020; Miikkulainen et al., 2019; Montavon et al., 2018). Notably, a video is inherently composed of images or frames, which are processed using image analysis techniques. Therefore, image- and video-based monitoring tasks can be done using DL models such as Convolutional Neural Networks (CNNs) that receive an image (frame) as their input. Therefore, the methods mentioned for image-based tasks are useful for both images and videos.

Many of DNN-based approaches outperform conventional methods in marine applications, including ecological and habitat monitoring, using video trap data (Willi et al., 2019; Tabak et al., 2019). DL is a technique that mimics how people acquire knowledge by continuous analysis of input data. The main drivers of DNN success over the past decade have been architectural progress by a large community of computer scientists, more powerful computers and processors, and access to massive amounts of data, which is critical for developing successful generalizable DL applications.

DNNs have been successfully employed in many CV applications such as object classification, identification, and segmentation as a result of the invention of CNN. CNN is



FIGURE 1 Illustration of four typical types of CV tasks From left: Image Classification (*i.e.* is there a fish in the image, or what type (class) of fish is in the image?), Object Detection/Localisation, Semantic Segmentation, Instance Segmentation.

a class of DNN, most commonly applied to visual analyses. For instance, CNNs have been successfully used for analysis of fish habitats (Xu et al., 2019; Kononov et al., 2019a; Pope et al., 2010). In comparison to other image recognition algorithms, CNNs have the significant benefit that they require limited pre-processing. CNNs are not hand-engineered but uncover and learn hidden features in the data on their own. They learn level-by-level with various levels of abstraction. For instance, they learn simple shapes (edges, lines, etc.) in the first few layers, understand more sophisticated patterns in their next layers, and learn classes of objects in their final layers.

A putative challenge with CNNs is that they require a large number of images to be fully trained and generalise their learning to unseen scenarios. On the other hand, CNNs have an interesting and powerful feature that enables transfer of their learning and knowledge across different domains. This means that they can be fine-tuned to work on new datasets (e.g. fish datasets) other than the one that they have been trained on (e.g. general objects). However, fine-tuning with annotated datasets specific for a given domain implies cost/effort/time needed to generate the annotations, and also requires a larger set of data which may not always be available.

Equipping CV algorithms with the powerful learning and inference capabilities of CNNs can provide marine scientists and ecologists with powerful tools to help them better understand and manage marine environments. However, although DL, and its variants such as CNNs, have been applied to various applications across a multitude of domains (Deng and Yu, 2013; Pathak et al., 2018; Min et al., 2017), their use in conjunction with computer vision for marine science and fish habitat monitoring is not broadly appreciated, meaning they remain under utilised. To address this, in this paper, we introduce key concepts and typical architectures of DL, and provide a comprehensive survey of key CV techniques for

underwater fish habitat monitoring. In addition, we provide insights into challenges and opportunities in the underwater fish habitat monitoring domain. It is worth noting that our article is written to provide a general and high-level, as opposed to detailed, introduction of deep learning and its relevant contexts for marine scientists. This is useful in understanding the follow-up discussions on the use of deep learning in the marine task of underwater fish classification.

Although a recent survey reviews deep learning techniques for marine ecology (Goodwin et al., 2022) and briefly discusses DL-based fish image analysis, to the best of our knowledge, no comprehensive survey and overview of deep learning with a specific focus on fish classification in underwater habitats currently exists. Our paper tries to address this gap and to facilitate the application of modern deep learning approaches into the challenging underwater fish images analysis and monitoring domains. We do this by comprehensively reviewing and analysing the literature providing information about the DL model the previous works have used, their training dataset, their annotation techniques, their performance and a comparison to other similar works. This detailed analysis is not provided in (Goodwin et al., 2022).

In addition, another survey (Li and Du, 2021) exists that focuses on five different tasks of classification, detection, counting, behaviour recognition, and biomass estimation. Compared to (Li and Du, 2021), we provide a different analysis and review of the literature because we mainly focus on the classification of fish in underwater images. Li and Du's work (Li and Du, 2021) fits mostly in the domain of aquaculture, while our paper is mostly a review of "fish classification techniques in underwater habitats" and the challenges they bring. Li and Du introduce a background to many different DL architectures, one of which is CNN, which is the focus of our paper. Also, the challenges and opportunities that Li and Du introduce are different to our paper, which is mainly about

underwater fish classification in their natural habitat.

Furthermore, we provide a historical review of the CV and DL research using underwater cameras for fish classification, and analyse how their accuracy has evolved over years. This is not covered by previous works including (Goodwin et al., 2022; Li and Du, 2021).

2 | BACKGROUND TO COMPUTER VISION AND MACHINE LEARNING

Humans, have a natural ability to comprehend the three-dimensional structure of the world around us. Vision scientists (Oomes, 2001) have spent decades attempting to understand how the human visual system functions (Wang and Weiland, 2017). Inspired by their findings, CV researchers (Ballard and Brown, 1982; Huang, 1996; Sonka et al., 2008) have also been working on ways to recover the 3D shape and appearance of objects from photos. The automatic retrieval, interpretation, and comprehension of useful information from a single image or collection of images can be referred to as CV. In another definition, CV is a field of Artificial Intelligence (AI) that focuses on training computers to detect, recognise, and understand images similarly to processes used by humans. This necessitates the development of logical and algorithmic foundations for automated visual understanding (Mader et al., 2018). This understanding can include image classification, object localisation, object recognition, semantic segmentation, and instance segmentation, as shown in Figure 1. Today, computers with CV powers can extract, analyse, and interpret significant information from a single image or a sequence of images.

Despite this progress, the goal of making a computer to understand a picture at the same level as a two-year-old child remains unattainable. This is due, in part, to the fact that CV is an inverse problem in which we attempt to recover specific unknowns despite having inadequate knowledge to completely describe the solution. In CV applications, the cause is usually an exploration process, while the effects are the observed data. The corresponding forward problems then consist of predicting empirical data given complete knowledge of the exploration process. In some sense, solving inverse

problems means “computing backwards”, which is usually more difficult than forward problem solving (Hohage et al., 2020).

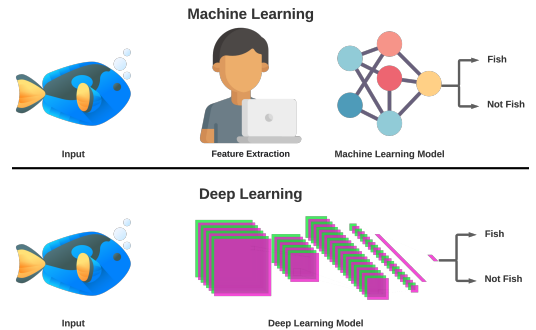


FIGURE 2 Comparison between Machine Learning (ML) and DL. In ML techniques, the features need to be extracted by domain expert while DL relies on layers of artificial neural networks to extract these features.

The problem of backward computation was eased by the introduction of ML techniques more than 6 decades ago. However, in conventional ML approaches, the majority of complex features of the learning subject must be identified by a domain expert in order to decrease the complexity of the data and make patterns more evident for successful learning (see Figure 2-top). However, DL offered a fundamentally new method to ML. Most DL algorithms possess the groundbreaking ability of automatically learning high-level features from data with minimal or no human intervention (see Figure 2-bottom).

DL is based on neural networks, which are general-purpose functions that can learn almost any data type that can be represented by many instances. When you feed a neural network a large number of labelled instances of a certain type of data, it will be able to uncover common patterns between those examples and turn them into a mathematical equation that will assist in categorising future data. Empowered by this fundamental feature, DL and DNN have progressed from theory to practice as a result of advancements in hardware and cloud computing resources (Azghadi et al., 2020). In recent years, DL approaches have outperformed previous state-of-the-art ML techniques in a variety of areas, with CV being one of the most notable examples.

Before the introduction of DL, the capabilities of CV were severely limited, necessitating a great deal of manual coding and effort. However, owing to improved research in DL and neural networks, CV is now able to outperform humans in several tasks related to object recognition and classification (Sarigül and Avci, 2017; Salman et al., 2016; Qin et al., 2016; Sun et al., 2017). CV equipped with DL, is being used today in a wide variety of real-world applications, that include, but are not limited to:

- *Optical character recognition (OCR)* (Permaloff and Grafton, 1992): automatic number plate recognition and reading handwritten postal codes on letters;
- *Machine inspection* (Park et al., 2016): fast quality assurance inspection of components using stereo vision with advanced lighting to assess tolerance levels on aircraft wings or car body parts, or to spot flaws in steel castings using X-ray technology;
- *Retail* (Trinh et al., 2012): object detection for automatic checkout lanes;
- *Medical imaging* (Erickson et al., 2017): registration of preoperative and intra-operative imaging or long-term analyses of human brain anatomy as they age;
- *Automotive safety* (Falcini et al., 2017): detection of unforeseen objects such as pedestrians on the street (e.g. fully autonomously driving vehicles);
- *Surveillance* (Brunetti et al., 2018): Monitoring of trespassers, studies of highway traffic, and monitoring pools for drowning victims;
- *Fingerprint recognition and bio-metrics* (Kim et al., 2016): For both automatic entry authentication and forensic software.

This demonstrates the significant impact of DL on CV and demonstrates its potential for marine visual analysis applications.

3 | THE EVOLUTION OF COMPUTER VISION APPROACHES TO FISH CLASSIFICATION

The last two decades have witnessed the emergence of novel computer vision approaches for fish classification including the design and evaluation of complex algorithms that could not be applied before and became possible with the availability of sufficiently large data and the use of powerful Graphical Processing Units (GPUs). Here, we perform a systematic literature review of the evolution of computer vision applications and their different approaches over the past two decades.

3.1 | Search and Selection Criteria

We systematically reviewed the literature for underwater fish classification using computer vision from 2003 to 2021. The search terms used included "underwater fish classification", "Deep Learning", "Computer Vision", "Machine vision". The databases searched included Wiley Online Library, IEEE Xplore, Elsevier/ScienceDirect, and ACM Digital Library. We believe that combining these four databases accurately represents global research on this topic.

We divided the search into two stages. First, we queried the databases for articles with the above-mentioned keywords in their titles and contents. Secondly, we independently reviewed the titles and abstracts of each article in order to check its relevance to our research topic. After the individual title and abstract reviews, we considered 64 articles for full-text reading. In the full-reading phase, we extracted information relevant to our research topic. In this phase, it became clear that 21 papers were not relevant to our work and therefore were excluded. This left us with 43 papers for fish classification, 26 of which were classical Computer Vision methods, and 17 Deep Learning papers. Figure 4 presents an overview of the methods used in the identified studies and classifies them into several groups, based on their classification algorithms that can be categorized into two general category of conventional CV, and modern DL models.

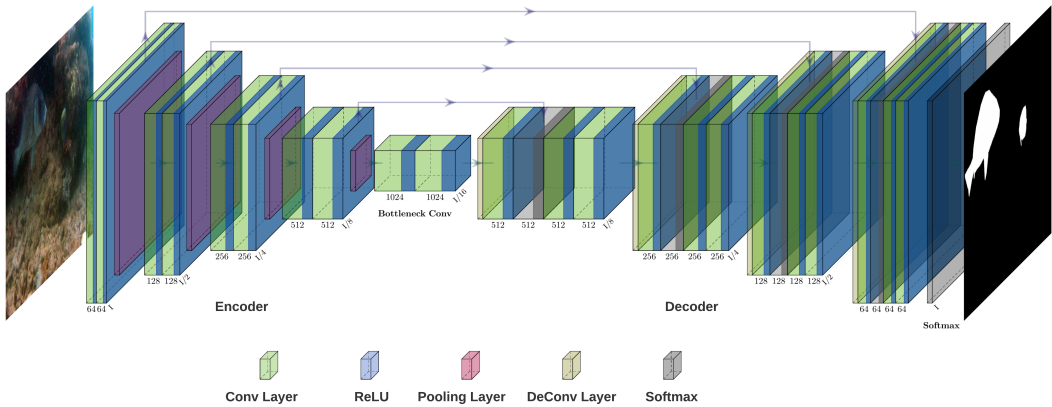


FIGURE 3 A popular CNN architecture, named UNET (Ronneberger et al., 2015) is demonstrated. The first component of UNET is the encoder, which is used to extract features from the input image. The second component is the decoder that outputs per-pixel scores. The network is composed of five different layers including convolutional (Conv Layer), Rectified Linear Unit (ReLU), Pooling, Deconvolutional (DeConv), and Softmax. Here, the task of the DNN layers has been to give a high score to only the pixels in the input image that belong to the fish body, resulting in the demonstrated white blobs output, showing where the fish are.

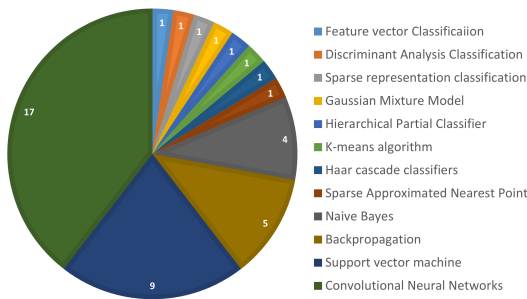


FIGURE 4 An overview of the methods used for fish classification using different Computer Vision techniques from 2003 to 2021. It is evident from the graph that DL and its CNNs have attracted more attention than classical ML methods.

3.2 | The Evolution of Fish Classification Algorithms over Two Decades

The **publication trend** for fish classification studies is summarized in Fig. 5. The figure shows the cumulative number of publications and how the studies evolved over the past two decades. It is evident that the number of publications has been gradually increasing, but in 2016, when the first few studies using deep learning were combined with CV meth-

ods, the study numbers have seen the highest increase and a fast upward trajectory for a few years (2015–2019) after DL burgeoned in fish classification, and before slowing down.

Fig. 5 also shows the highest classification accuracy achieved in each year, as a **quality assessment metric**. It is evident that since 2016, when DL techniques were first proposed for fish classification, the accuracy has seen its highest value. At the same time, it can be seen that there are large differences in the accuracies achieved over years. The main reasons for this difference include (i) using different classification and CV methods, and (ii) using different fish image sources that were captured differently and in different environments. These bring huge variations among studies, such as different image resolutions and inconsistent resolutions and image qualities across time. For example, some fish image datasets are in grayscale (Chuang et al., 2014, 2016; Kartika and Herumurti, 2017), while others are in colour (Zion et al., 2008, 2007; Shafait et al., 2016). Some datasets contain only images (Islam et al., 2019; Kartika and Herumurti, 2017), while others include videos (Lopez-Villa et al., 2015; Cutter et al., 2015; Hossain et al., 2016). Also, some datasets (Huang et al., 2014) used low-quality images from the internet, which negatively affects the accuracy, due to their wide

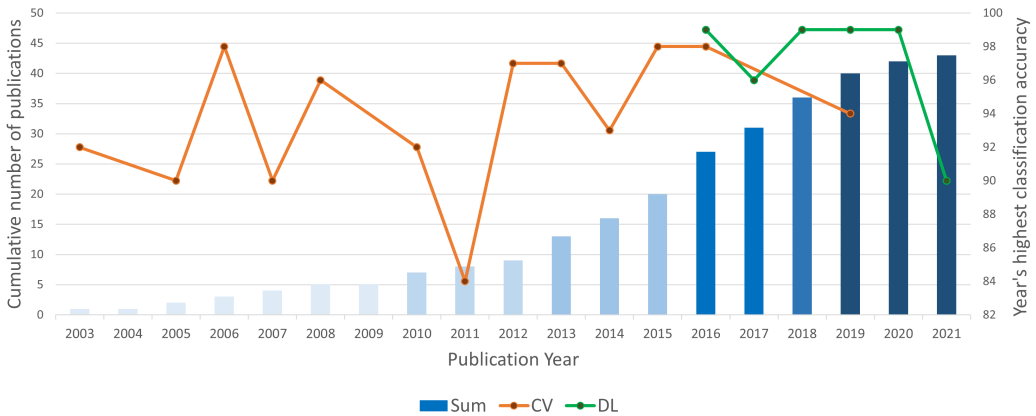


FIGURE 5 An overview of the publication trend and performance of an extensive range of fish classification Computer Vision (CV) and Deep Learning (DL) models from 2003 to 2021. Here the bars show the cumulative number of publications over years and the growth thereof, while the line graphs demonstrate the highest classification accuracy in each year in literature on the right-hand-side vertical axis.

range of resolutions, colours, and angles. They are also taken at random locations. Due to these factors in various studies, direct comparison of accuracy values is unfeasible, though the accuracy trend can be still observed in Fig. 5.

Computer vision for fish classification in the early 2000s and up to 2016, when first DL works started, has been mainly to manually extract fish features and then build classifiers that recognize these features. These conventional studies are listed, in a chronological order, in Table 1. Although there are many existing models, most of the classical non-DL models are based on local and engineered features. These include works using Haar features (Mutneja and Singh, 2021), Scale-Invariant Feature Transform (SIFT) (Lindeberg, 2012), and Histogram of Oriented Gradient (HOG) (Dalal and Triggs, 2005), which need hand-engineered algorithms. Because these algorithms are not suitable for recognizing images of untrained animals and cannot capture fish features from complex backgrounds, they usually use a large number of manually extracted samples to build classifiers.

As shown in Table 1, support vector machines (Rova et al., 2007; Hu et al., 2012; Fouad et al., 2014; Huang et al., 2014; Chuang et al., 2016; Ogunlana et al., 2015; Hossain et al., 2016; Wang et al., 2017a; Islam et al., 2019) were one of the most commonly used classifiers for fish recognition, but they are prone to overfitting when trained with too many

samples. This problem limits the scale of application. Another popular classification technique used in early works was backpropagation to train a simple feed-forward shallow neural network (Alsmadi et al., 2010, 2011; Pornpanomchai et al., 2013; Badawi and Alsmadi, 2014; Boudhane et al., 2016). Although this technique can handle simple samples, it is difficult to scale because of the neural network shallow layers, which will be explained in the next Section. Naive Bayes (Nery et al., 2005; Zion et al., 2007, 2008; Kartika and Herumurti, 2017) have also been used to classify fish since the early 2000s and up to 2017. The technique does not require much training data, and as shown in Table 1 can reach good accuracy levels. Table 1 also shows some other CV classification techniques, which while not as popular as the above-mentioned methods, could demonstrate good performance. However, it should be noted that, most of the CV techniques in Table 1, were carefully engineered for their target datasets and are not capable of showing a similar performance level if used for another similar dataset. They will perhaps require an overhaul in their design, starting from manual feature engineering, to designing the detailed classification models.

In contrast, deep learning can extract features and perform classification tasks automatically. The features are invariant to data scaling, translation, rotation, and distortion. Because

these features are better for classification, the classification performance can be better than that conventional CV tasks using manually designed features. Also, DL classification models, compared to traditional CV one, usually require a simpler redesign procedure to work on a new similar dataset, due to the ability to extract features on their own.

Although DL emerged in 2012 (Krizhevsky et al., 2012), its first use for underwater fish classification was in 2016 (Salman et al., 2016). After that, 16 other works also used DL and its CNNs, as shown in Fig. 4, to develop models that learn features from large amounts of data without manual interference. These studies have shown that, by using deep learning, some of the usual fish image classification challenges such as image noise reduction, classification of difficult or rare-seen fish, and classifying small fish, can be solved.

In the following parts of this paper, we mainly focus on deep learning, how it works, and how it can be applied to develop efficient and high-performance underwater fish classifiers. We will also critically analyse the 17 DL studies found as part of our systematic literature review described earlier.

4 | BACKGROUND TO DEEP LEARNING

Deep Learning (DL) (Goodfellow et al., 2016; LeCun et al., 2015) is a subset of ML algorithms that employs a neural network with several layers to very loosely replicate the function of the human brain by enabling it to "learn" from huge quantities of data. The learning happens when the neural network extracts higher-level features from input training data. The term "deep" refers to the usage of several layers in the neural network. Lower layers, for example in image processing, could detect edges, whereas higher layers might identify parts of the object.

4.1 | How Deep Learning differs from Machine Learning

Machine Learning (ML) is usually referred to as a class of algorithms that can recognise patterns in data and create prediction models automatically. Deep Learning (DL) is a sub-

class of standard ML because it uses the same type of data and learning methods that ML applies. However, when dealing with unstructured data, e.g. text and images, ML usually goes through some pre-processing to convert it to a structured format for learning. DL, on the other hand, does not usually require the data pre-processing needed by ML. It is capable of recognising and analysing unstructured data, as well as automating feature extraction, significantly reducing the need for human knowledge (see Figure 2-bottom).

For example, to recognise fish in an image, ML requires that specific fish features (such as shape, colour, size, and patterns) be explicitly defined in terms of pixel patterns. This may be a challenge for non-ML specialists because it typically requires a deep grasp of the domain knowledge and good programming skills. DL techniques, on the other hand, skip this step entirely. Using general learning techniques, DL systems can automatically recognise and extract features from data. This means that we just need to tell a DL algorithm whether a fish is present in an image, and it will be able to figure out what a fish looks like given enough examples. Decomposing the data into layers with varying levels of abstraction enables the algorithm to learn complex traits defining the data, allowing for an automatic learning approach. DL algorithms may be able to determine which features (such as fishtail) are most important in differentiating one animal from another. Prior to DL, this feature hierarchy needed to be determined and created by hand by an ML expert.

4.2 | How Deep Learning works

Deep Neural Network (DNN), also known as artificial neural network, is the basis of deep learning. DNNs use a mix of data inputs, weights, and biases to learn the data, by properly detecting, categorising, and characterising objects in a given dataset of interest. DNNs are made up of several layers of linked nodes, each of which improves and refines the network prediction or categorisation capabilities. For instance, Fig. 3 shows a popular DNN architecture for image processing, called UNET (Ronneberger et al., 2015). UNET, which is a fairly complex deep learning architecture, is composed of a few different components and layers, to achieve a specific learning goal, i.e. to segment fish body in an input image.

Any DNN is composed of three types of layers, namely

TABLE 1 A list of computer vision studies for underwater fish classification between 2003-2021 using conventional classifiers and based on engineered features. The last column presents the work's achieved accuracy.

Article	Year	Classification Method	AC
An Automated Fish Species Classification and Migration Monitoring System (Lee et al., 2003)	2003	Feature vector Classification	92
Determining the appropriate feature set for fish classification tasks (Nery et al., 2005)	2005	Naive Bayes	90
Real-time underwater sorting of edible fish species (Zion et al., 2007)	2006	Naive Bayes	98
One Fish, Two Fish, Butterfish, Trumpeter: Recognizing Fish in Underwater Video (Rova et al., 2007)	2007	Support vector machine	90
Classification of guppies' (<i>Poecilia reticulata</i>) gender by computer vision (Zion et al., 2008)	2008	Naive Bayes	96
Automatic Fish Classification for Underwater Species Behavior Understanding (Spampinato et al., 2010)	2010	Discriminant Analysis Classification	92
Fish Recognition Based on Robust Features Extraction from Size and Shape Measurements Using Neural Network (Alsmadi et al., 2010)	2010	Backpropagation	86
Fish Classification Based on Robust Features Extraction From Color Signature Using Back-Propagation Classifier (Alsmadi et al., 2011)	2011	Backpropagation	84
Fish species classification by color, texture and multi-class support vector machine using computer vision (Hu et al., 2012)	2012	Support vector machine	97
Real-world underwater fish recognition and identification, using sparse representation (Hsiao et al., 2014a)	2013	Sparse representation classification	81
A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage (Boom et al., 2014)	2013	Gaussian Mixture Model	97
Automatic Nile Tilapia Fish Classification Approach using Machine Learning Techniques (Fouad et al., 2014)	2013	Support vector machine	94
Shape- and Texture-Based Fish Image Recognition System (Pornpanomchai et al., 2013)	2013	Backpropagation	90
A General Fish Classification Methodology Using Meta-heuristic Algorithm With Back Propagation Classifier (Badawi and Alsmadi, 2014)	2014	Backpropagation	80
GMM improves the reject option in hierarchical classification for fish recognition (Huang et al., 2014)	2014	Support vector machine	74
Supervised and Unsupervised Feature Extraction Methods for Underwater Fish Species Recognition (Chuang et al., 2014)	2014	Hierarchical Partial Classifier	93
A Feature Learning and Object Recognition Framework for Underwater Fish Images (Chuang et al., 2016)	2015	Support vector machine	98
A novel tool for ground truth data generation for video-based object classification (Lopez-Villa et al., 2015)	2015	K-means algorithm	93
Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: labeled fishes in the wild (Cutter et al., 2015)	2015	Haar cascade classifiers	89
Fish Classification Using Support Vector Machine (Ogunlana et al., 2015)	2015	Support vector machine	79
Fish identification from videos captured in uncontrolled underwater environments (Shafait et al., 2016)	2016	Sparse Approximated Nearest Point	94
Fish Activity Tracking and Species Identification in Underwater Video (Hossain et al., 2016)	2016	Support vector machine	91
Koi Fish Classification based on HSV Color Space (Kartika and Herumurti, 2017)	2016	Naive Bayes	97
Optical Fish Classification Using Statistics of Parts (Boudhane et al., 2016)	2016	Backpropagation	95
Shrinking Encoding with Two-Level Codebook Learning for Fine-Grained Fish Recognition (Wang et al., 2017a)	2017	Support vector machine	98
Indigenous Fish Classification of Bangladesh using Hybrid Features with SVM Classifier (Islam et al., 2019)	2019	Support vector machine	94

input, output, and hidden layers. The visible layers are the input and output layers (see Figure 6). The DL model gets the data for processing in the input layer, and the final prediction or classification is generated in the output layer. In a typical neural network, including a DNN, the learning happens through two general processes, *i.e.* forward and backward propagations. Forward propagation refers to the propagation of input data through the network layers to generate a prediction or classification result. Backward propagation or, backpropagation in short, is where the learning happens in the network. Backpropagation uses a training model that determines prediction errors and then changes the weights and biases of the neural network by going backwards through its layers. Forward propagation and backpropagation work together to allow a neural network to generate predictions and reduce the network errors. Through many iterations of backward and forward propagation, the neural network prediction or classification accuracy improves.

Almost all DNNs work on and through the same principles described above. However, different DL networks and architectures are used to solve different tasks. For instance, CNNs, which are commonly used in computer vision and image classification applications, can recognise characteristics and patterns within an image, allowing tasks such as object detection and recognition to be accomplished. However, in tasks with a different nature, such as natural language processing, speech recognition, or timeseries forecasting (Jahanbakht et al., 2022), Recurrent Neural Networks (RNNs) are commonly employed. Despite the differences in their architectures, many DL techniques, use the concept of supervised learning to process their input data and accomplish different tasks.

4.3 | Supervised Learning

Supervised learning is a method used to enable finding and optimising a function that maps an input to its corresponding output in an input-output object pair, also known as training example (Kotsiantis, 2007). Supervised learning uses a set of training examples based on manually-labelled training data prepared by human observers or 'supervisors', hence the name for the learning method.

The aim of supervised learning is to generate an inferred

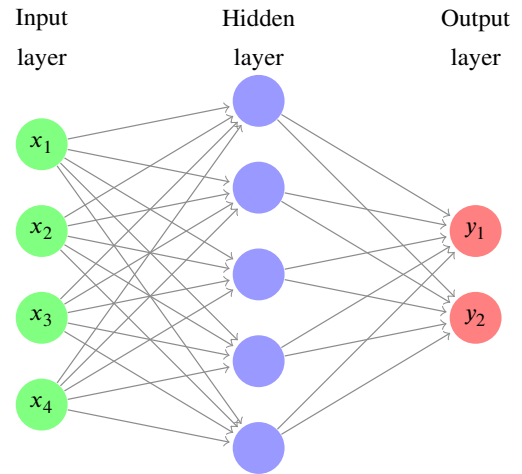


FIGURE 6 A diagram of a single-layer neural network, composed of input, hidden, and output layers.

function, f , that maps to the training examples, and can then be used to map to new examples outside of the training examples. In order to accomplish any general task, a computer can be programmed to find function f to map X to Y , *i.e.* ($f : X \mapsto Y$), where X is an input domain and Y is an output domain. For example, in an image classification task, X is the dataset of images and Y is a set of corresponding classification labels, which determine whether an object is present in the respective image in the dataset or not.

To determine the function f that can recognise, for instance, a fish in an image using DL, one solution is to do feature engineering. However, it is usually very difficult to perform this, *i.e.* hand-pick features of the fish, based on the domain knowledge that comes from the training dataset. In addition, most of the time, the hand-picked features need to be pruned to reduce their pixel dimensionality. Comparatively, it is often more feasible to collect a large dataset of $(x, y) \in X \times Y$ to find the mapping function f , and this affords supervised learning advantage as an alternative mapping technique compared with direct feature engineering. Specifically, in the fish classification task, a large dataset of fish images is collected, where each image x is labelled with y that shows the presence or absence of a fish, without the need to hand-pick its features.

One of the main supervised learning approaches is training

a neural network, which is the foundation of deep learning, especially for computer vision applications such as fish image processing. We, therefore, dedicate the next subsection to neural networks and their underlying working principles.

4.4 | Neural Networks

A 'neural network' (Cook, 2020) is a computer program originally conceived by mimicking actual cerebral neural networks that make up the brain's grey matter. A computer's neural network, a.k.a. an artificial neural network, "learns" to do a specific task by using a large amount of data, usually through supervised network training that does not involve any task-specific rules. As briefly mentioned, a neural network is constructed from three types of layers: an input layer, hidden or latent layers, and an output layer (see Figure 6). These layers include processing neurons within them (coloured circles in Figure 6), and connecting synapses (weights) between them (edges in the figure).

The input layer is the gate to the network. It provides information to the network from outside data, and no calculation is made in this layer. Instead, input nodes pass the information on to the hidden layer. This layer is not visible to the outside world and serves as an abstraction of the inputs, independent of the neural network structure. The hidden layer (layers) processes the data received from the input layer and transfers the results to the output layer. Finally, the output layer brings the information that the network has learned into the outside world.

Learning in a neural network happens through minimising a loss function. Generally, a loss function is a function that returns a scalar value to represent how well the network performs a specific task. For example, in image classification, the network is expected to correctly classify all the images containing a fish as fish, and all those not including a fish, as no fish, returning a loss value of zero. During learning, the network receives a large amount of input data, e.g. thousands of fish images, and eventually learns to minimise the loss between its predicted output and the true target value. In the case of supervised learning, these true target values are provided to the network, to find function f described in the previous section, to minimise the loss function. This minimisation happens through optimising f using an algorithm

such as Stochastic Gradient Descent (SGD) (Loshchilov and Hutter, 2017) that helps find network weights/parameters that minimise the loss.

4.5 | Convolutional Neural Network

CNNs are probably the most commonly used artificial neural networks. They have been the dominant deep learning tool in computer vision and have been widely used in underwater marine habitat monitoring (Saleh et al., 2020). CNNs are broadly designed after the neuronal architecture of the human cortex but on much smaller scales (Schmidhuber, 2015). A CNN (LeCun et al., 1998) is specifically designed for dealing with datasets that have some spatial or topological features (e.g. images, videos), where each of the neurons are placed in such a manner that they overlap and thus react to multiple spots in the visual field. A CNN neuron is a simple mathematical design of the human brain's neuron that is utilised to transform nonlinear relationships between inputs and outputs in parallel. There are two primary layer types in a CNN, i.e. convolutional layers and pooling layers, which generate feature maps, as explained in the following subsections.

4.5.1 | Convolutional Layer

In this layer, the convolutional processes (*i.e.*, the multiplication of a small matrix of the input neurons by a small array of weights called filter) are used on limited fields (which depend on the size of the filter) to avoid the need to learn billions of weights (parameters), which would be required if all the neurons in one layer are connected to all the neurons in the next layer. This excessive computation is avoided through the weight-sharing of convolutional layers combined with filters for their corresponding feature maps. In a convolution operation, a small matrix of the input neurons is multiplied in its same-sized matrix, called a filter. In a convolutional layer, this convolution operation happens by sliding the filter on the entire input neurons, generating a feature map. Filters work on a reduced area of the input (convolutional kernel). Convolutional layers can either use the same kernel size or they can use different kernel sizes, which makes it possible to extract complex features from the input using fewer parameters. In addition, weight-sharing is useful in avoiding model overfit-

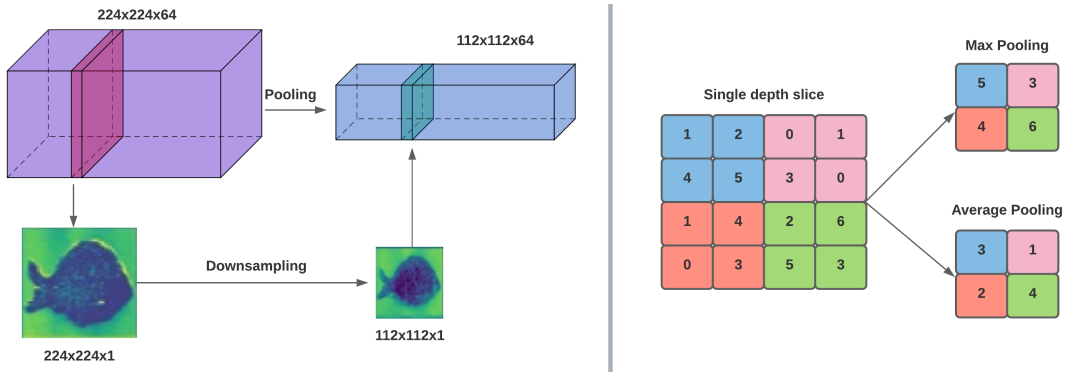


FIGURE 7 Schematic diagram of pooling layer: **(Left)** single feature map spatially downsampled from a representation block with shape $224 \times 224 \times 1$ to a new representation of shape $112 \times 112 \times 1$. **(Right)** types of pooling layer (max-pooling and average-pooling).

ting, i.e. memorising the training data, (Abdel-Hamid et al., 2013), while also reducing computing memory requirements and enhancing learning performance (Korekado et al., 2003).

4.5.2 | Pooling Layer

This layer is used to reduce the spatial dimension (not depth) of the input features and add control for avoiding overfitting by reducing the number of representations with a specified spatial size. Pooling operations can be done in two different ways, i.e. Max and Average pooling. In both methods (see Figure 7), an input image is down-scaled in size, by taking the maximum of 4 pixels and down-sampling them to one pixel. Pooling layers are systematically implemented between convolutional layers in conventional CNN architectures. The pooling layers work on each channel (activation map) individually and downsample them spatially. By having fewer spatial information, pooling layers make a CNN more computationally efficient.

4.5.3 | Feature Maps

Feature Maps, also called Activation Maps, are the result of applying convolutional filters or feature detectors to the preceding layer image. The filters are moved on the preceding layer by a specified number of pixels. For instance, in Figure 8, there are 37 filters of the size 3×3 that move across the

input image with a stride of 1 and result in 37 feature maps.

The majority of CNN layers are convolutional layers. These layers are used to apply the same convolutional filtering operation to different parts of the image, creating “neurons” that can then be used to detect features, like the edges and corners. A collection of weights connects each neuron in a convolutional layer to the preceding layer’s feature maps, or to the input layer image. The feature maps help visualise the features that the CNN is learning to give an understanding of the network learning process, as shown in Figure 8.

5 | APPLICATIONS OF DEEP LEARNING IN FISH-HABITAT MONITORING

In a recent special issue titled “Applications of machine learning and artificial intelligence in marine science” published in the International Council for the Exploration of the Sea (ICES) journal of marine science (Proud et al., 2020), many uses of deep learning and CNNs have been shown. These include identifying the species of harvested fish (Lu et al., 2020), analysis of fisheries surveillance videos (French et al., 2020), and natural mortality estimation (Liu et al., 2020). Other published works have used CNN for other marine applications such as automatic vessel detection (Chen et al., 2019), and analysis of deep-sea mineral exploration (Juliani and Juliani, 2021). However, in this paper we focus on using

CNNs for CV tasks.

These tasks are mainly designed to extract knowledge from underwater videos and images. Despite the recent use of CNNs for various visual analysis tasks such as segmentation (Garcia et al.; Alshdaifat et al., 2020; Islam et al., 2020; Zhang et al., 2022), localisation (Su et al., 2020; Jalal et al., 2020; Knausgård et al., 2021), and counting (Tarling et al., 2021; Schneider and Zhuang, 2020; Ditría et al., 2021), the most common and the widest studied CV task in underwater fish habitat monitoring has been classification. Therefore, in this paper, we focus mainly on classification of underwater fish images. We survey some of the latest works on fish classification and provide a high-level technical discussion of these works.

The task of classification is defined as classifying the input samples into different categories, usually based on the presence or absence of a certain object/class, in binary classification; or the presence of several different objects belonging to different classes, in multi-class classification (Ismail Fawaz et al., 2019). Similarly, image classification is concerned with assigning a label to a whole image based on the objects in that image. Conceivably, an image can be labelled as fish, when there is a fish present in it, or negative when no fish is present. Similarly, images of different species should be automatically assigned to their respective classes or given a label representing their class.

Classification is a difficult process if done manually, because an image may need to be categorised into more than one class. In addition, there may be thousands of images to be classified, which makes the task very time-consuming and prone to human error. Consequently, automation can help perform classification quicker and more efficiently.

In the context of fish and marine habitat monitoring, CV offers a low-cost, long-term, and non-destructive observation opportunity. One of the initial tasks performed using deep learning on CV-collected marine habitat images is fish classification, which is a key component of any intelligent fish monitoring systems, because it may activate further processing on the fish image. However, underwater monitoring based on image and video processing pose numerous challenges related to the hostile condition under which the fish images are collected. These include poor underwater image quality due to low light and water turbidity, which result in low resolution

and contrast. Additionally, fish movements in an uncontrolled environment can create distortion, deformations, occlusion, and overlapping. Many previous works (Boom et al., 2012; Takada et al., 2014; Martinez-de Dios et al., 2003) have tried to address these challenges. Some of these works focused on devising new methods to properly extract traditional low-level features such as colours and textures using mean shift algorithm (Boudhane and Nsiri, 2016), in the presence of the challenges. However, these works have not been very successful compared to DL approaches.

With the inception of CNNs, many researchers utilised them to extract both high-level and low-level features of input images. These features, which can be automatically detected by the CNN, carry extensive semantic information that can be applied to recognise objects in an image. In addition, CNNs have the ability to address the challenges outlined above. Therefore, they are currently the main underwater image processing tool in literature for fish classification, as shown in Tables 2 and 3. These tables list some of the latest classification works, while providing details about the DL models used and the framework within which the model was implemented. It also provides information about the data source, as well as the pre-processing of the data and its labels, while reporting the Classification Accuracy (CA) and a short comparison with other methods if the reviewed work has provided it. One of the main metrics when comparing different methods for classification is their CA, which is defined as the percentage of correct predictions by the network.

$$CA = (TP + TN)/(TP + TN + FP + FN), \quad (1)$$

where TP (True Positive) and TN (True Negative) represent the number of correctly classified instances, while FP (False Positive) and FN (False Negative) represent the number of incorrectly classified instances. For multi-class classification, CA is averaged among all the classes.

DL algorithms are gaining momentum in their growing accuracy in different applications. However, they have inherent limitations, which should be considered before choosing a DL algorithm for a given application. This is because accuracy, for example in a fish classification task, may significantly differ from true accuracy due to the distribution of samples in the training and testing populations. To address

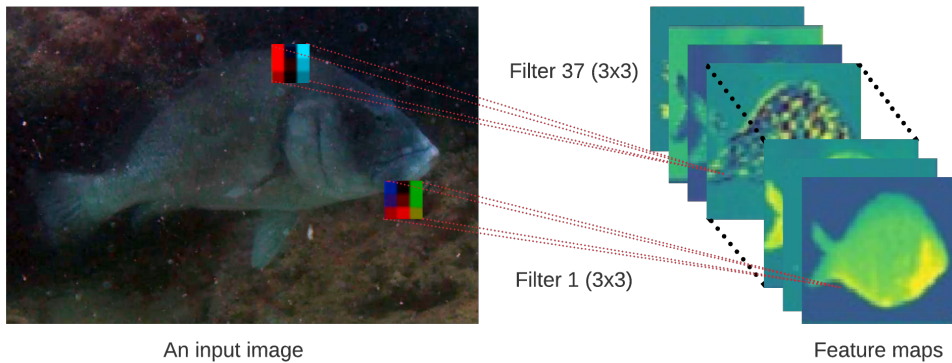


FIGURE 8 Schematic diagram of feature maps of the CNN used in the classification task. The feature map is a two-dimensional representation of an input image. Here (3×3) is the size of the filter slid over the entire image to generate feature maps.

this limitation of classification accuracy, the Receiver Operating Characteristics (ROC) (Krupinski, 2017) and Area Under The Curve (AUC) (Janssens and Martens, 2020) are widely used as a standard measure for determining the performance of a model in a binary classification setting. Their definition is very similar to accuracy but they help one understand the probability that the classifier produces correct outputs with desired levels of true positives and false negatives, using a certain classification threshold.

The works in Tables 2 and 3 can be divided into two general categories. The first category deals with designing effective CNNs that address the challenge of unconstrained, complex, and noisy underwater scenes, while the second category also tries to address the usual problem of limited fish training datasets.

As mentioned, when processing unconstrained underwater scenes specific attention should be paid to implementing a classification approach that is capable of handling variations in light intensity, fish orientation, and background environments, and similarity in shape and patterns among fish of various species. In order to overcome these challenge and to improve classification accuracy, various works have devised different methodologies. In (Varalakshmi and Julanta Leela Rachel, 2019), the authors used different activation functions to examine the most suitable for fish classification, while in (Sarigül and Avci, 2017) different number of convolutional layers and different filter sizes were examined. In

(Salman et al., 2016), the authors used a CNN model in a hierarchical feature combination setup to learn species-dependent visual features for better accuracy. In another work (Qin et al., 2016), principal-component analysis was used in two convolutional layers, followed by binary hashing in the non-linear layer and block-wise histograms in the feature pooling layer. Furthermore, a single-image super-resolution method was used in (Sun et al., 2017) to resolve the problem of limited discriminative information of low-resolution images. Moreover, (Chen et al., 2018) used two independent classification branches, with the first branch aiming to handle the variation of pose and scale of fish and extract discriminative features, and the second branch making use of context information to accurately infer the type of fish. The reviewed works show that depending on the type of environment and fish species similarities in the dataset under consideration, various techniques should be considered and investigated to find the best classification accuracy.

As already mentioned, data gathering in the wild is sometimes very difficult and challenging, thus to maximize the success rate of training, it is essential to consider gathering field data from the beginning of the project. This ensures that the collected training dataset has good sample diversity including samples collected at different environmental conditions such as water turbidity and salinity, and it captures fish species similarities. Diversity and comprehensiveness in the dataset is one of the key factors in reaching high classifica-

tion accuracies when the model is deployed in the real world. Data augmentation is another important method that can help improve the classification accuracy, through increasing the dataset size and diversity. An alternative to data augmentation is transfer learning, but the model should be always fine-tuned to the new dataset to maximize accuracy. Image pre-processing is another important technique that can help improve classification accuracy, and should be considered when working with new fish datasets.

Dataset limitation, i.e. having limited number of fish images from different species, and/or having few numbers of different fish etc, is another challenge in underwater fish habitat monitoring in general and in fish classification, in specific. This challenge has been addressed in (Saleh et al., 2020; Jin and Liang, 2017; Rathi et al., 2017; Tamou et al., 2018) using transfer learning.

Transfer learning is a ML method that works by transferring information obtained while learning one problem or domain to a different but related problem or domain. Comparing a randomly initialised classifier with another one pre-trained on ImageNet (Russakovsky et al., 2015), Saleh *et al.* (Saleh et al., 2020) achieved a fish classification accuracy of 99%, outperforming the randomly-initialised classifier, significantly. This finding shows that transfer learning can bring learned information from the ImageNet learning domain to fish classification domain and can be a useful and crucial method for evaluating fish environments. Transfer learning was also used in (Kononov et al., 2019b) where general-domain above-water fish image learning was transferred and used for underwater fish classification. In the same way, to train large-scale models that are able to generate reasonable results, (Zhuang et al., 2020) collected 1000 fish categories with 54,459 unconstrained images from various professional fish websites and Google engine.

In addition to transfer learning, some works have developed specific machine learning techniques suiting their applications. For instance, in a previous study (Siddiqui et al., 2018), a pre-trained CNN was used as a generalised feature extractor to avoid the need for a large amount of training data. The authors showed that by feeding the CNN-extracted features to a Support Vector Machine (SVM) classifier (Pisner and Schnyer, 2019), a CA of 94.3% for fish species classification can be achieved, which significantly outperforms

a stand-alone CNN achieving an accuracy of 53.5%. Also, (Deep and Dash, 2019) used the same techniques in (Siddiqui et al., 2018) to achieve a CA of 98.79%. In addition, (Iqbal et al., 2021) developed a new technique for fish classification by modifying AlexNet (Krizhevsky et al., 2012) model with fewer number of layers. Moreover, (Kononov et al., 2019a) presented a labelling efficient method of training a CNN-based fish-detector on a small dataset by adding 27,000 above-water and underwater fish images.

CNNs are sometimes capable of surpassing human performance in identifying fish in underwater images. By training a CNN on 900,000 images, Villon *et al.* (Villon et al., 2018) could achieve a CA of 94.9% while human CA was only 89.3%. This result was achieved mainly because the CNN was able to successfully distinguish fish that were partially occluded by corals or other fish, while human could not. Furthermore, the best CNN model developed in (Villon et al., 2018) takes 0.06 seconds on average to identify each fish using typical hardware (Titan X GPU). This demonstrates that DL techniques can conduct accurate fish classification on underwater images cost-effectively and efficiently. This facilitates monitoring underwater fish and can advance marine studies concerned with fish ecology.

If DL methods are going to be deployed widely for different marine applications such as fish classification, there is a need to implement them efficiently, so that they can run on low-power embedded systems, which can run in real-time on mobile devices such as underwater drones. To that end, Meng *et al.* (Meng et al., 2018) have developed an underwater drone with a panoramic camera for recognising fish species in a natural lake to help protect the environment. They have trained an efficient CNN for fish recognition and achieved 87% accuracy while requiring only 6 seconds to identify 115 images. This promising result shows that, DL can be used to classify underwater fish while also satisfying the real-time conditions of mobile monitoring devices. In addition, other efficient hardware design approaches that have proven useful in reducing power consumption and increasing speed in classification task in other domains such as agriculture (Lammie et al., 2019) can be adopted on edge underwater processors.

In DL applications, video storage is currently a bottleneck that may be bypassed with real-time algorithms, because they only need to store some and not all the video frames in

memory and process them in-situ, as they become available. This eliminates the time it takes for all the frames to be stored and retrieved from memory. This is helpful in situations where large amounts of data have to be processed quickly, for example, in an underwater fish observation camera, where frames are collected continuously and should either be stored locally or transferred to surface, which are both costly and mostly impossible. Using real-time processing algorithms, the frames are processed and only the information obtained, *i.e.* the number of fish in a frame are sent or stored, which is much lighter than the entire frame.

6 | CHALLENGES AND APPROACHES TO ADDRESS THEM

Despite the rapid improvement of DL for marine habitat monitoring through visual analysis, four main challenges still exist. The first challenge is to develop models that can generalise their learning and perform well on new unseen data samples. The second challenge is limited datasets available for general DL tasks, and in particular for marine visual processing tasks. The third challenge is lower image quality in underwater scenarios. The fourth challenge is the gap between DL and ecology.

To address these challenges, various computer algorithms and techniques have been developed. In the following subsections, we explain the challenges in detail and briefly review various approaches to address them. However, we do not intend to include details of these approaches as they are out of the scope of this paper. The interested reader is invited to refer to relevant DL materials and the cited papers.

6.1 | Model Generalisation

One of the most difficult challenges in DL is to improve deep convolutional networks generalisation abilities. This refers to the gap between a model's performance on previously observed data (*i.e.* training data) and data it has never seen before (*i.e.* testing data). A wide gap between the training and validation accuracy is usually a sign of overfitting. Overfitting occurs when the model accurately predicts the training data, mostly because it has memorised the training data instead of

learning their features.

One way to monitor overfitting is by plotting the training and validation accuracy at each epoch during training. That way, we will see that if the gap between the validation and training accuracy/error is widening (over- or under-fitting) or narrowing (learning). A well-known and effective method for improving the generalisability of a DL model is to use regularisation (Kukačka et al., 2017). Some of the regularisation methods applied to fish and marine habitat monitoring domains include transfer learning (Zurowietz and Nattkemper, 2020), batch normalisation (Islam et al., 2020), dropout (Iqbal et al., 2021), and using a regularisation term (Tarling et al., 2021).

6.2 | Dataset Limitation

Another challenge of training DL models is the limited dataset. DL models require enormous datasets for training. Unfortunately, most datasets are large, expensive, and time-consuming to build. For this reason, model training is usually conducted by collecting samples from a small number of datasets, rather than from a large number of datasets.

A dataset can be categorised into two parts: labelled data and unlabeled data. The labelled data is the set of data that needs the labelling of classes, *e.g.* fish species in an image, or absence or presence of fish in an image. The unlabeled data is the set of data that has not been processed. The labelled data forms the training set whose size is closely related to the accuracy of the trained model. The larger the training set, the more accurate the trained model. Large training set, however, are expensive to build. They require a large number of resources, such as people-hours, space, and money, making it very difficult for many researchers to achieve them, and in turn hinders their research.

Since it is difficult to obtain a large labelled dataset, various techniques have been proposed to address this challenge. Some of the techniques applied to the fish and marine habitat monitoring domains include transfer learning (Qiu et al., 2018), data augmentation (Saleh et al., 2020; Sarigül and Avci, 2017), using hybrid features (Mahmood et al., 2016; Cao et al., 2016; Blanchet et al., 2016), weakly supervised learning (Laradji et al., 2021), and active learning (Nilssen et al., 2017).

TABLE 2 Classification

Article	DL Model	Framework	Data	Annotation/Pre-processing/Augmentation	Classes and Labels	Perf. Metric	Metric Value	Comparisons with other methods
Recognition of Fish Categories Using Deep Learning Technique (Varalakshmi and Julanta Leela Rachel, 2019)	CNN	Keras, TensorFlow	Authors-created dataset containing 560 fish images, 400 training and 160 test images.	Each image is assigned the fish species name as a label	10 classes of 10 different fish species	CA	95%	NA
Comparison of Different DL Structures for Fish Classification (Sarigül and Avci, 2017)	CNN	Torch	The public QUT fish dataset contains 3960 images of 468 fish species in different environments.	Each image is assigned the fish species name as a label	468 classes of 468 different fish species	CA	46.02%	NA
Fish Species Classification in Unconstrained Underwater Environments Based on DL (Salman et al., 2016)	CNN	NA	The images are from the public Fish4Knowledge dataset (LifeCLEF 2014, LifeCLEF 2015)	Each image is assigned the fish species name as a label	25 classes of 25 different fish species	CA	96.75%	Comparison with the conventional SVM machine learning tool that achieved 83.94%
Deep-Fish: Accurate Underwater Live Fish Recognition with a DL Architecture (Qin et al., 2016)	CNN	Matlab	The images are from the public Fish4Knowledge dataset	Each image is assigned the fish species name as a label	23 classes of 23 Different fish species	CA	98.64%	Comparison with conventional machine learning tools as baseline methods achieving 93.58%
Fish Recognition from Low-resolution Underwater Images (Sun et al., 2017)	CNN	NA	93 videos from LifeCLEF 2015 fish dataset	Each image was annotated by drawing a bounding box and labelling by species name	15 classes of 15 different fish species	CA	76.57%	Authors used the traditional gabor features and dense sift features that generated CA of 38.28% and 28.63%, respectively.
Automatic Fish Classification System Using DL (Chen et al., 2018)	CNN	NA	Eight target categories: Albacore tuna, Bigeye tuna, Yellowfin tuna, Mahi Mahi, Opah, Sharks, Other.	Each image is assigned the fish species name as a label	8 classes of 8 different fish species	CE	0.578, 1.387	Ranked 17th on Kaggle leaderboard on test set at stage 1 and 16th at stage 2.
A Realistic Fish-habitat Dataset to Evaluate DL Algorithms For Underwater Visual Analysis (Saleh et al., 2020)	ResNet-50 CNN	PyTorch	Authors-created database containing 39,766 images from 20 habitats in remote coastal marine environments of tropical Australia	point-level and semantic segmentation labels	20 classes of 20 different fish species	CA	0.99	NA
Deep Learning for Underwater Image Recognition in Small Sample Size Situations (Jin and Liang, 2017)	CNN	Caffe	The images are from the public Fish4Knowledge dataset	Each image is assigned the fish species name as a label	10 classes of 10 different fish species	CA	85.08%	NA
Underwater Fish Species Classification using CNN and DL (Rathi et al., 2017)	CNN	NA	27000 images from the public Fish4Knowledge dataset	Each image is assigned the fish species name as a label	23 fish classes	CA	96.29%	NA

TABLE 3 Classification

Article	DL Model	Frame-work	Data	Annotation/Pre-processing/Augmentation	Classes and Labels	Perf. Metric	Matrix Value	Comparisons with other methods
Underwater Live Fish Recognition by Deep Learning (Tamou et al., 2018)	AlexNet CNN	Matlab	27000 images from the public Fish4Knowledge dataset	Each image is assigned the fish species name as a label	23 classes of 23 different fish species	CA	99.45%	NA
WildFish++: A Comprehensive Fish Benchmark for Multimedia Research (Zhuang et al., 2020)	CNN	NA	Authors-created dataset of 54,459 labelled images from various professional websites and Google engine	Each image is assigned the fish species name as a label	100 classes of 1000 different fish species	CA	74.7%	Comparison with other state-of-the-art approaches
Automatic Fish Species Classification in Underwater Videos: Exploiting Pre-trained DNN Models to Compensate for Limited Labelled Data (Siddiqui et al., 2018)	CNN	MAT-LAB	The dataset contains 50 to 120 10-second video clips of 16 species from Western Australia during 2011 to 2013.	Each image is assigned the fish species name as a label	16 classes of 16 Different fish species	CA	89.0%	Comparison of their proposed method of CNN+SVM achieving a CA of 89.0% with two previous works: SRC (Hsiao et al., 2014b)) 49.1% and CNN (Salman et al., (Salman et al., 2016)) 53.5%
Underwater Fish Species Recognition using Deep Learning Techniques (Deep and Dash, 2019)	CNN	Keras, TensorFlow	35000 images from the public Fish4Knowledge dataset	Each image is assigned the fish species name as a label	23 classes of 23 different fish species	CA	98.79%	NA
Automatic Fish Species Classification Using Deep Convolutional Neural Networks (Iqbal et al., 2021)	modified AlexNet CNN	Tensorflow	The images are from two public datasets: QUT fish dataset and LifeClef-15	Each image is assigned the fish species name as a label	6 classes of 6 different fish species	CA	90.48%	Comparing their proposed modified AlexNet achieving a CA of 90.48% with original AlexNet CA of 86.65%
Underwater Fish Detection with Weak Multi-Domain Supervision (Kononov et al., 2019a)	CNN	Keras, TensorFlow	Authors-created dataset of 40000 labelled fish images from video sequences	Each image is labelled as Fish or no fish	2 classes	CA	99.94%	NA
A Deep Learning Method for Accurate and Fast Identification of Coral Reef Fishes in Underwater Images (Villon et al., 2018)	GoogLeNet CNN	Caffe	Authors-created dataset containing 450,000 images from over 50 reef sites around the Mayotte island	Annotation included drawing a rectangle around the fish and associating the species name as label.	20 classes of 20 different fish species	CA	94.9%	Comparing accuracy to human experts. The rate of correct identification was 94.9%, greater than the rate of correct identification by humans (89.3%).
Underwater-Drone Panoramic Camera for Automatic Fish Recognition Based on Deep Learning (Meng et al., 2018)	CNN	NA	Authors-created dataset of 100 labelled images from Google search engine	Each image is assigned the fish species name as a label	4 classes of 4 different fish species.	CA	87%	NA

6.3 | Image Quality

Underwater image recognition's average accuracy lags significantly behind that of terrestrial image recognition. This is mostly owing to the low quality of underwater photos, which frequently exhibit blurring, and colour deterioration, caused by the physical characteristics of the water and the hostile underwater environment.

Most CV applications perform some initial preprocessing of images before feeding them to their image processor. In underwater scenarios, these preprocessing techniques are typically used to enhance the image quality. Preprocessing can also help with the red channel information loss problem, which is required for obtaining relevant colour data. The red channel information loss problem is about losing the actual intensity of the red colour in the scene, for instance, compared to the blue and green colour channels. This is more pronounced in the underwater environment and as the depth increases, which attenuates red channel values more strongly than the other colour channels. We should, therefore, consider that the red channel value depends not only on the distance from the subject but also on the intensity of the light reflected by the subject, as the reflection of intense light is typically much stronger than that of a light of a very low intensity. Another issue that arises in the detection of a specific target in an underwater image is the fact that multiple pixels can potentially be activated in the image in the form of an object. For example, sunlight shining through a periscope lens can cause spurious activation of a given pixel. There is a need for a reliable method and system for determining whether a given pixel in a remote underwater image is activated by some cause other than the presence of a target in the area of the image.

Preprocessing of underwater photos has been extensively researched, and several solutions have been devised for correcting typical underwater image artefacts (Carlevaris-Bianco et al., 2010; Kumar and Prabhaka, 2011). However, the image quality produced by these approaches is subjective to the observer, and because acquisition settings vary so widely, these methods may not be applicable to all datasets. According to empirical results (Beijbom et al., 2012; Shihavuddin et al., 2013), the current tendency appears to be to perform picture repair and enhancement processes based on the dataset, i.e.

determining the most appropriate preprocessing strategy for a specific dataset. This strategy also depends on the purpose (e.g. labelling, classification or both) of the images in the dataset.

In addition, basic image enhancement techniques have been shown to be effective in improving image quality. For instance, in (Cao et al., 2016) increasing the uniformity of the background was used to boost picture contrast in underwater images for marine animal classification. This is a strong indicator that simple enhancing approaches might result in increased performance. Furthermore, some recent studies have employed DL algorithms to enhance image quality using low-quality images. In (He and Li, 2019), for example, end-to-end mapping is performed between low-resolution and high-resolution images.

When compared to state-of-the-art handcrafted and traditional image enhancement methods, DL-based algorithms typically perform better in addressing picture quality in terrestrial photos. However, significant new research is required to customise these DL-based techniques for underwater images and maritime datasets. This poses as a future research opportunity for image quality enhancement in fish monitoring applications. Below, we discuss some more opportunities.

6.4 | Deep Learning Gap

DL is an emerging field that has a lot to offer in terms of ecology. The first and most obvious ecological applications are fish classification or fish count. However, there is still a gap between the DL-predicted fish counts and, for example, absolute abundance (fish per area or volume unit). The existing DL literature discusses mainly the use of CNNs for the ecological problems of species classification or fish counting. However, the absolute abundance of fish is important for ecological research and species conservation.

Another important problem in ecological research is fish population dynamics. A step in addressing this problem is to analyze long-term data on fish movements and fish densities. However, such long-term datasets are relatively rare and expensive to obtain. Hence, there is a need to obtain as much information as possible from the small amount of data given. This requires novel methods to give an accurate long-term estimate of fish densities or, even better, an estimate of the

absolute abundance of fish.

Other exemplar ecological questions that can be addressed using DL include species habitat selection, or the relationship between the physical environment and the life history of species (Van Allen et al., 2012; Shryock et al., 2014; Vincenzi et al., 2019). DL methods can help us with this because they can take advantage of all the available information. The current state of DL research can be improved by considering alternative network architectures, more complex training algorithms, and more detailed knowledge of the problem domain. The existing DL literature suggests that we may see many new methods in the future. Most of them still do not have sufficient data to prove that they can outperform existing methods. There are, however, examples of successful applications, such as fish classification. For many ecological problems, a DL method can give very accurate predictions of fish densities or absolute abundance. However, it remains unclear whether this accuracy can be obtained only with the appropriate method or whether this is a property of the particular dataset on which the method was trained. From this perspective, the development of a general method for predicting fish densities and absolute abundance from very little data is a major problem in ecology.

One potential approach to solving this problem is to take advantage of DL models trained on other datasets, as long as they are related to the fish density/abundance problem. The ecological literature suggests that the relationship between the physical environment and the life history of species (e.g., fish density) is likely to be complex because the physical environment differs from species to species. Therefore, we may be able to find many similar datasets on other related problems (e.g., environmental science or engineering). In addition to developing and testing general methods to estimate the absolute abundance of fish from very little data, there is a need to develop general methods that can take advantage of the ecological knowledge and domain-specific data from a particular problem.

7 | OPPORTUNITIES IN APPLICATION OF DL TO FISH HABITAT MONITORING

New methods and techniques will need to be devised to improve the accuracy of deep learning models for various marine habitat monitoring applications and to bring them closer to their terrestrial counterparts.

7.1 | Spatio-temporal and Image Data Fusion

Most of the current marine habitat monitoring and visual processing tools only use image-based data to train their model to understand the habitats and monitor the environment. In such tools, each frame or image is separately processed and spatiotemporal correlations across neighbouring frames are simply overlooked. Exploiting this extra information and fusing it with the image-processing model can be beneficial (Yang et al., 2020). For instance, fusing a master-slave camera setup with LSTM (Wang et al., 2017b) can help to learn the kinematic model of fish in a 3D fish tracking system. Future works should consider including spatiotemporal information in training their model and understanding the scene. In particular, approaches similar to Long short-term memory (LSTM) networks or other RNN models can be used in conjunction with CNNs, to obtain improved classification or prediction outcomes by taking advantage of the time-domain information. For example, An RNN and a CNN model are combined in (Måløy et al., 2019) to achieve better performance for salmon feeding action recognition from underwater videos. In (Peng et al., 2019), the authors propose a spatio-temporal recurrent network to classify behavioural patterns. Similar schemes have been proposed in (Xu et al., 2021). However, their performance and complexity heavily rely on the ability of the RNN to track the temporal relations of the frames and on the effectiveness of the CNN.

For instance, estimating and monitoring fish development based on previous continuous observations, and analysing fish behaviour are some of the applications where time domain information will be not only useful but also critical. Such models can also be used to build novel video-based protocols for the surveillance of critically endangered reef

fish biodiversity.

7.2 | Underwater Embedded and Edge Processing

DNNs have proven to be successful in both industry and research in recent years, particularly for CV tasks. Specifically, large-scale DL models have had a lot of success in real-world scenarios with large-scale data. This is mainly due to their capacity to encode vast amounts of data and handle millions of model parameters that enhance generalisation performance when new data is evaluated. However, this high computational complexity and substantial storage requirement makes them difficult to use in real-time applications, especially on devices with restricted resources (e.g. embedded devices and underwater edge processors for online monitoring). One approach to address this is to use compressed networks such as binarised neural networks, which have shown promise toward reaching low-power and high-speed edge inference engines (Lammie et al., 2019), for near-underwater-sensor processing. This can significantly improve underwater image analysis capabilities, because the collected large-volume images do not need to be transferred to surface for processing, and only the low-volume results can be communicated to shore. This also solves another problem, which is the challenging underwater communication (Jahanbakht et al., 2021).

7.3 | Combining Data from Multiple Platforms

The use of different data collection platforms such as autonomous underwater vehicles (AUVs) or occupied submarines, can provide different image data from different perspectives of the same or different underwater habitats, to train more effective DNNs. In addition, using simultaneous data from multiple platforms can give more monitoring information, for instance, of fish distribution patterns, especially in situations where the number of platforms is limited. However, combining data from multiple platforms introduces some challenges such as the lack of ground truth (e.g., the number of fish in the sampled area for all the platforms), and the need to develop techniques that can integrate these data in a robust manner. Future research can work toward addressing

these challenges to exploit the significant benefits of multiple platform data combination.

7.4 | Automated Fish Measurement and Monitoring

DL can be used to achieve automated fish measurements, which may be useful in underwater fish monitoring, for instance to survey fish growth (Yang et al., 2020) through monitoring of fish length (Palmer et al., 2022) and abundance (Ditria et al., 2019). Here, abundance means the number of fish in an image or video frame, and not the fish count per area or volume unit. In addition, automated measurements can realise remote fish assessments, for example when the monitoring locations are remote, or the environmental conditions and or potential hazards do not allow frequent underwater scouting by human.

DL can also be used for automation of monitoring of other fish biological variables such as their movement dynamics, present species, and their abundance and biomass. On top of these, DL can be used to automate understanding of environmental and habitat features. To achieve these, new datasets should be collected, and new or existing DL techniques should be devised or customised in future research.

8 | CONCLUSION

Deep Learning (DL) sits at the forefront of the machine learning technologies providing the processing power needed to enable underwater video to fulfill its promise as a critical tool for visual sampling of fish. It offers efficient and accurate solutions to the challenges of adverse water conditions, high similarity between fish species, cluttered backgrounds, occlusions among fish, that have limited the spatio-temporal consistency of underwater video quality. As a result, DL, complemented by many other advances in monitoring hardware and underwater communication technologies, opens the way for underwater video to provide comprehensive fish sampling. This can span from shallow fresh and marine waters to the deep ocean, opening the way for the development of the truly comparative understanding of marine and aquatic fish fauna and ecosystems that has hitherto been impossible.

At least as importantly, DL solves the problem of handling the vast quantities of data produced by underwater video in a consistent and cost-effective way, converting a prohibitively expensive activity into a simple issue of computer processing. By enabling the processing of vast quantities of data, DL allows underwater fish video surveys to be conducted with unprecedented levels of spatial and temporal replication enabling the massive knowledge advances that flow from the ability of underwater videos to be deployed contemporaneously across many habitats, and at many spatial scales, or to provide continuous data over time.

DL, and associated techniques, have the potential for widespread use in marine habitat monitoring for (1) data classification and feature extraction to improve the quality of automatic monitoring tools; or (2) to provide a reliable means of surveying fish habitats and understanding their movement dynamics. While this will allow marine ecosystem researchers and practitioners to increase the efficiency of their monitoring efforts, effective development of DL will require concentrated and coordinated data collection, model development, and model deployment efforts, as well as transparent and reproducible research data and tools, which help us reach our target sooner.

ACKNOWLEDGEMENT

This research is supported by an Australian Research Training Program (RTP) Scholarship. We acknowledge the Australian Research Council for funding awarded under their Industrial Transformation Research Program.

CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no data sets were generated or analysed during the current study.

ORCID

Alzayat Saleh [0000-0001-6973-019X](https://orcid.org/0000-0001-6973-019X)

Marcus Sheaves [0000-0003-0662-3439](https://orcid.org/0000-0003-0662-3439)

Mostafa Rahimi Azghadi [0000-0001-7975-3985](https://orcid.org/0000-0001-7975-3985)

REFERENCES

- Abdel-Hamid, O., Deng, L. and Yu, D. (2013) Exploring convolutional neural network structures and optimization techniques for speech recognition. In *Interspeech*, vol. 11, 73–75.
- Alshdaifat, N. F. F., Talib, A. Z. and Osman, M. A. (2020) Improved deep learning framework for fish segmentation in underwater videos. *Ecological Informatics*, **59**, 101121.
- Alsmadi, M. K., Omar, K. B. and Mohd Noah, S. A. (2011) Fish classification based on robust features extraction from color signature using back-propagation classifier. *Journal of Computer Science*, **7**, 52–58.
- Alsmadi, M. K., Omar, K. B., Noah, S. A. and Almarashdeh, I. (2010) Fish recognition based on robust features extraction from size and shape measurements using neural network. *Journal of Computer Science*, **6**, 1088–1094.
- Azghadi, M. R., Lammie, C., Eshraghian, J. K., Payvand, M., Donati, E., Linares-Barranco, B. and Indiveri, G. (2020) Hardware Implementation of Deep Network Accelerators towards Healthcare and Biomedical Applications. *IEEE Transactions on Biomedical Circuits and Systems*, **14**, 1138–1159.
- Badawi, U. A. and Alsmadi, M. K. (2014) A general fish classification methodology using meta-heuristic algorithm with back propagation classifier. *Journal of Theoretical and Applied Information Technology*, **66**, 803–812.
- Ballard, D. H. and Brown, C. M. (1982) *Computer Vision*. Prentice Hall. URL: <https://archive.org/details/computervision0000ball>.
- Beijbom, O., Edmunds, P. J., Kline, D. I., Mitchell, B. G. and Kriegman, D. (2012) Automated annotation of coral reef survey images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1170–1177.
- Blanchet, J.-N., Déry, S., Landry, J.-A. and Osborne, K. (2016) Automated annotation of corals in natural scene images using multiple texture representations. *PeerJ*.

- Boom, B. J., He, J., Palazzo, S., Huang, P. X., Beyan, C., Chou, H.-M., Lin, F.-P., Spampinato, C. and Fisher, R. B. (2014) A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. *Ecological Informatics*, **23**, 83–97.
- Boom, B. J., Huang, P. X., Beyan, C., Spampinato, C., Palazzo, S., He, J., Beauxis-Aussalet, E., Lin, S. I., Chou, H. M., Nadarajan, G., Chen-Burger, Y. H., van Ossenbruggen, J., Giordano, D., Hardman, L., Lin, F. P. and Fisher, R. B. (2012) Long-term underwater camera surveillance for monitoring and analysis of fish populations. *Workshop on Visual observation and Analysis of Animal and Insect Behavior (VAIB), in conjunction with ICPR 2012*.
- Boudhane, M. and Nsiri, B. (2016) Underwater image processing method for fish localization and detection in submarine environment. *Journal of Visual Communication and Image Representation*, **39**, 226–238. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1047320316300840>.
- Boudhane, M., Nsiri, B. and Toulni, H. (2016) Optical fish classification using statistics of parts. *International Journal of Mathematics and Computers in Simulation*, **10**, 18–22.
- Brunetti, A., Buongiorno, D., Trotta, G. F. and Bevilacqua, V. (2018) Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing*.
- Cao, Z., Principe, J. C., Ouyang, B., Dagleish, F. and Vuorenkoski, A. (2016) Marine animal classification using combined CNN and hand-designed image features. In *OCEANS 2015 - MTS/IEEE Washington*.
- Carlevaris-Bianco, N., Mohan, A. and Eustice, R. M. (2010) Initial results in underwater single image dehazing. In *OCEANS 2010 MTS/IEEE SEATTLE*, 1–8. IEEE. URL: <http://ieeexplore.ieee.org/document/5664428/>.
- Chen, G., Sun, P. and Shang, Y. (2018) Automatic fish classification system using deep learning. In *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*.
- Chen, L., Xia, Y., Pan, D. and Wang, C. (2019) Deep learning based active monitoring for anti-collision between vessels and bridges. In *IABSE Symposium, Guimaraes 2019: Towards a Resilient Built Environment Risk and Asset Management - Report*.
- Chuang, M. C., Hwang, J. N., Kuo, F. F., Shan, M. K. and Williams, K. (2014) Recognizing live fish species by hierarchical partial classification based on the exponential benefit. In *Proc. Int. Conf. Image Process.*, 5232–5236. Paris, France: IEEE.
- Chuang, M. C., Hwang, J. N. and Williams, K. (2016) A feature learning and object recognition framework for underwater fish images. *IEEE Trans. Image Process.*, **25**, 1862–1872.
- Cook, T. R. (2020) Neural Networks. In *Advanced Studies in Theoretical and Applied Econometrics*, 161–189. URL: http://link.springer.com/10.1007/978-3-030-31150-6_6.
- Cutter, G., Stierhoff, K. and Zeng, J. (2015) Automated detection of rockfish in unconstrained underwater videos using haar cascades and a new image dataset: Labeled fishes in the wild. In *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2015*, 57–62.
- Dalal, N. and Triggs, B. (2005) Histograms of oriented gradients for human detection. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. 1, 886–893.
- Deep, B. V. and Dash, R. (2019) Underwater Fish Species Recognition Using Deep Learning Techniques. In *2019 6th International Conference on Signal Processing and Integrated Networks, SPIN 2019*.
- Deng, L. and Yu, D. (2013) Deep learning: Methods and applications.
- Martinez-de Dios, J. R., Serna, C. and Ollero, A. (2003) Computer vision and robotics techniques in fish farms. *Robotica*, **21**, 233–243. URL: https://www.cambridge.org/core/product/identifier/S0263574702004733/type/journal_article.
- Ditria, E., Lopez-Marcano, S., Sievers, M., Jinks, E., Brown, C. and Connolly, R. (2019) Automating the analysis of fish abundance using object detection: optimising animal ecology with deep learning. *Frontiers in Marine Science*.
- Ditria, E. M., Connolly, R. M., Jinks, E. L. and Lopez-Marcano, S. (2021) Annotated Video Footage for Automated Identification and Counting of Fish in Unconstrained Seagrass Habitats. *Frontiers in Marine Science*, **8**. URL: <https://www.frontiersin.org/articles/10.3389/fmars.2021.629485/full>.
- Erickson, B. J., Korfiatis, P., Akkus, Z. and Kline, T. L. (2017) Machine Learning for Medical Imaging. *RadioGraphics*, **37**, 505–515. URL: <http://pubs.rsna.org/doi/10.1148/rg.2017160130>.

- Falcini, F., Lami, G. and Costanza, A. M. (2017) Deep Learning in Automotive Software. *IEEE Software*, **34**, 56–63. URL: <https://ieeexplore.ieee.org/document/7927925/>.
- Fouad, M. M. M., Zawbaa, H. M., El-Bendary, N. and Hassanien, A. E. (2014) Automatic Nile Tilapia fish classification approach using machine learning techniques. In *13th International Conference on Hybrid Intelligent Systems, HIS 2013*, 173–178.
- French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N. and Needle, C. (2020) Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. *ICES Journal of Marine Science*, **77**, 1340–1353. URL: <https://academic.oup.com/icesjms/article/77/4/1340/5542623>.
- Garcia, R., Prados, R., Quintana, J., Tempelaar, A., Gracias, N., Rosen, S., Vågstøl, H., Løvall, K., Vagstol, H. and Lovall, K. (2022) Automatic segmentation of fish using deep learning with application to fish size measurement. *ICES Journal of Marine Science*, **77**, 1354–1366. URL: <https://doi.org/10.1093/icesjms/fsz186>.
- Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y. (2016) *Deep learning*, vol. 1. MIT Press.
- Goodwin, M., Halvorsen, K. T., Jiao, L., Knausgård, K. M., Martin, A. H., Moyano, M., Oomen, R. A., Rasmussen, J. H., Sjørdalen, T. K. and Thorbjørnsen, S. H. (2022) Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. *ICES Journal of Marine Science*, **79**, 319–336. URL: <https://arxiv.org/abs/2109.14737v1>.
- He, T. and Li, X. (2019) Image quality recognition technology based on deep learning. *Journal of Visual Communication and Image Representation*, **65**.
- Hilborn, R. and Walters, C. J. (1992) Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, **2**, 177–178. URL: <http://link.springer.com/10.1007/BF00042883>.
- Hohage, T., Sprung, B. and Weidling, F. (2020) Inverse Problems. In *Topics in Applied Physics*, 145–164. URL: http://link.springer.com/10.1007/978-3-030-34413-9_5.
- Hossain, E., Alam, S. M., Ali, A. A. and Amin, M. A. (2016) Fish activity tracking and species identification in underwater video. In *2016 5th International Conference on Informatics, Electronics and Vision, ICIEV 2016*, 62–66.
- Hsiao, Y. H., Chen, C. C., Lin, S. I. and Lin, F. P. (2014a) Real-world underwater fish recognition and identification, using sparse representation. *Ecological Informatics*, **23**, 13–21.
- Hsiao, Y.-H., Chen, C.-C., Lin, S.-I. and Lin, F.-P. (2014b) Real-world underwater fish recognition and identification, using sparse representation. *Ecological Informatics*, **23**, 13–21.
- Hu, Y., Mian, A. S. and Owens, R. (2012) Face Recognition Using Sparse Approximated Nearest Points between Image Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 1992–2004.
- Huang, P. X., Boom, B. J. and Fisher, R. B. (2014) GMM improves the reject option in hierarchical classification for fish recognition. In *IEEE Winter Conference on Applications of Computer Vision*, 371–376. IEEE. URL: <https://ieeexplore.ieee.org/document/6836076>.
- Huang, T. (1996) *Computer Vision : Evolution And Promise*. Geneva: CERN. URL: <http://cds.cern.ch/record/400313/files/p21.pdf>.
- Iqbal, M. A., Wang, Z., Ali, Z. A. and Riaz, S. (2021) Automatic Fish Species Classification Using Deep Convolutional Neural Networks. *Wireless Personal Communications*.
- Islam, M. A., Howlader, M. R., Habiba, U., Faisal, R. H. and Rahman, M. M. (2019) Indigenous Fish Classification of Bangladesh using Hybrid Features with SVM Classifier. In *5th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, IC4ME2 2019*.
- Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S. S. and Sattar, J. (2020) Semantic Segmentation of Underwater Imagery: Dataset and Benchmark. <http://arxiv.org/abs/2004.01241>. URL: <http://arxiv.org/abs/2004.01241>.
- Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.-A. (2019) Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, **33**, 917–963. URL: <https://link.springer.com/10.1007/s10618-019-00619-1>.
- Jahanbakht, M., Xiang, W. and Azghadi, M. R. (2022) Sea Surface Temperature Forecasting With Ensemble of Stacked Deep Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1–5. URL: <https://ieeexplore.ieee.org/document/9507522/>.
- Jahanbakht, M., Xiang, W., Hanzo, L. and Azghadi, M. R. (2021) Internet of Underwater Things and Big Marine Data Analytics - A Comprehensive Survey. *IEEE Communications Surveys and Tutorials*, **23**, 904–956. URL: <https://ieeexplore.ieee.org/document/9328873/>.

- Jalal, A., Salman, A., Mian, A., Shortis, M. and Shafait, F. (2020) Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, **57**, 101088.
- Janssens, A. C. J. and Martens, F. K. (2020) Reflection on modern methods: Revisiting the area under the ROC Curve. *International Journal of Epidemiology*, **49**, 1397–1403.
- Jin, L. and Liang, H. (2017) Deep learning for underwater image recognition in small sample size situations. In *OCEANS 2017 - Aberdeen*.
- Juliani, C. and Juliani, E. (2021) Deep learning of terrain morphology and pattern discovery via network-based representational similarity analysis for deep-sea mineral exploration. *Ore Geology Reviews*.
- Kartika, D. S. Y. and Herumurti, D. (2017) Koi fish classification based on HSV color space. In *Proceedings of 2016 International Conference on Information and Communication Technology and Systems, ICTS 2016*, 96–100.
- Kim, S., Park, B., Song, B. S. and Yang, S. (2016) Deep belief network based statistical feature learning for fingerprint liveness detection. *Pattern Recognition Letters*, **77**, 58–65. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0167865516300198>.
- Knausgård, K. M., Wiklund, A., Sjørdalen, T. K., Halvorsen, K. T., Kleiven, A. R., Jiao, L. and Goodwin, M. (2021) Temperature fish detection and classification: a deep learning based approach. *Applied Intelligence*.
- Konovalov, D. A., Saleh, A., Bradley, M., Sankupellay, M., Marini, S. and Sheaves, M. (2019a) Underwater Fish Detection with Weak Multi-Domain Supervision. In *2019 International Joint Conference on Neural Networks (IJCNN)*, vol. 2019-July, 1–8. IEEE. URL: <https://ieeexplore.ieee.org/document/8851907/>.
- (2019b) Underwater Fish Detection with Weak Multi-Domain Supervision. In *2019 International Joint Conference on Neural Networks (IJCNN)*, vol. 2019-July, 1–8. IEEE. URL: <https://ieeexplore.ieee.org/document/8851907/>.
- Korekado, K., Morie, T., Nomura, O., Ando, H., Nakano, T., Matsugu, M. and Iwata, A. (2003) A convolutional neural network VLSI for image recognition using merged/mixed analog-digital architecture. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 169–176. Springer.
- Kotsiantis, S. B. (2007) Supervised machine learning: A review of classification techniques.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25* (eds. F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger), 1097–1105. Curran Associates, Inc.
- Krupinski, E. A. (2017) Receiver operating characteristic (ROC) analysis. *Frontline Learning Research*, **5**, 31–42.
- Kukačka, J., Golkov, V. and Cremers, D. (2017) Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*.
- Kumar, C. and Prabhaka, P. (2011) An Image Based Technique for Enhancement of Underwater Images. *International Journal of Machine Intelligence*, **3**, 217–224. URL: <http://arxiv.org/ftp/arxiv/papers/1212/1212.0291.pdf>.
- Lammie, C., Olsen, A., Carrick, T. and Rahimi Azghadi, M. (2019) Low-power and high-speed deep FPGA inference engines for weed classification at the edge. *IEEE Access*.
- Laradji, I. H., Saleh, A., Rodriguez, P., Nowrouzezahrai, D., Azghadi, M. R. and Vazquez, D. (2021) Weakly supervised underwater fish segmentation using affinity LCFCN. *Scientific reports*, **11**, 17379. URL: <https://www.nature.com/articles/s41598-021-96610-2><http://www.ncbi.nlm.nih.gov/pubmed/34462458><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC8405733>.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature*, **521**, 436–444. URL: <http://www.nature.com/articles/nature14539>.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. and others (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**, 2278–2324.
- Lee, D. J., Redd, S., Schoenberger, R., Xu, X. and Zhan, P. (2003) An Automated Fish Species Classification and Migration Monitoring System. In *IECON Proceedings (Industrial Electronics Conference)*, vol. 2, 1080–1085.
- Li, D. and Du, L. (2021) Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish. *Artificial Intelligence Review*, 1–40. URL: <https://link.springer.com/article/10.1007/s10462-021-10102-3><https://link.springer.com/10.1007/s10462-021-10102-3>.
- Lindeberg, T. (2012) Scale Invariant Feature Transform. *Scholarpedia*, **7**, 10491. URL: http://www.scholarpedia.org/article/Scale_Invariant_Feature_Transform.

- Liu, C., Zhou, S., Wang, Y. G. and Hu, Z. (2020) Natural mortality estimation using tree-based ensemble learning models. *ICES Journal of Marine Science*, **77**, 1414–1426. URL: <https://academic.oup.com/icesjms/article/77/4/1414/5854079>.
- Lopez-Villa, J. S., Insuasti-Ceballos, H. D., Molina-Giraldo, S., Alvarez-Meza, A. and Castellanos-Dominguez, G. (2015) A novel tool for ground truth data generation for video-based object classification. In *2015 20th Symposium on Signal Processing, Images and Computer Vision, STSIVA 2015 - Conference Proceedings*.
- Loshchilov, I. and Hutter, F. (2017) SGDR: Stochastic Gradient Descent with Warm Restarts. In *International Conference on Learning Representations*.
- Lu, Y. C., Tung, C. and Kuo, Y. F. (2020) Identifying the species of harvested tuna and billfish using deep convolutional neural networks. *ICES Journal of Marine Science*, **77**, 1318–1329. URL: <https://academic.oup.com/icesjms/article/77/4/1318/5509966>.
- Mader, A. O., Lorenz, C., Bergtholdt, M., von Berg, J., Schramm, H., Modersitzki, J. and Meyer, C. (2018) Detection and localization of spatially correlated point landmarks in medical images using an automatically learned conditional random field. *Computer Vision and Image Understanding*.
- Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G. and Fisher, R. B. (2016) Coral classification with hybrid feature representations. In *Proceedings - International Conference on Image Processing, ICIP*.
- Måløy, H., Aamodt, A. and Misimi, E. (2019) A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture. *Computers and Electronics in Agriculture*, **167**, 105087.
- Meng, L., Hirayama, T. and Oyanagi, S. (2018) Underwater-Drone with Panoramic Camera for Automatic Fish Recognition Based on Deep Learning. *IEEE Access*.
- Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A., Duffy, N. and Hodjat, B. (2019) Evolving Deep Neural Networks. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, 293–312. Elsevier. URL: <https://linkinghub.elsevier.com/retrieve/pii/B9780128154809000153>.
- Min, S., Lee, B. and Yoon, S. (2017) Deep learning in bioinformatics.
- Montavon, G., Samek, W. and Müller, K. R. (2018) Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, **73**, 1–15.
- Mutneja, V. and Singh, S. (2021) Haar-features training parameters analysis in boosting based machine learning for improved face detection. *International Journal of Advanced Technology and Engineering Exploration*, **8**, 919–931.
- Nery, M. S., Machado, A. M., Campos, M. F., Pádua, F. L., Carceroni, R. and Queiroz-Neto, J. P. (2005) Determining the appropriate feature set for fish classification tasks. In *Brazilian Symposium of Computer Graphic and Image Processing*, vol. 2005, 173–180.
- Nilssen, I., Moller, T. and Nattkemper, T. W. (2017) Active Learning for the Classification of Species in Underwater Images from a Fixed Observatory. In *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, vol. 2018-Janua, 2891–2897.
- Ogunlana, S. O., Olabode, O. and Oluwadare, S. A. A. (2015) Fish Classification Using Support Vector Machine. *African Journal of Computing & ICT*, **8**, 75–82.
- Olsen, A., Konovalov, D. A., Philippa, B., Ridd, P., Wood, J. C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J., Calvert, B., Azghadi, M. R. and White, R. D. (2019) DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Scientific Reports*, **9**, 2058. URL: <https://www.nature.com/articles/s41598-018-38343-3><http://www.nature.com/articles/s41598-018-38343-3>.
- Oomes, A. H. J. (2001) Perception: Theory, Development and Organisation. *Optometry and Vision Science*, **78**, 477. URL: <http://journals.lww.com/00006324-200107000-00005>.
- Palmer, M., Álvarez-Ellacuría, A., Moltó, V. and Catalán, I. A. (2022) Automatic, operational, high-resolution monitoring of fish length and catch numbers from landings using deep learning. *Fisheries Research*, **246**, 106166. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0165783621002940>.
- Park, J. K., Kwon, B. K., Park, J. H. and Kang, D. J. (2016) Machine learning-based imaging system for surface defect inspection. *International Journal of Precision Engineering and Manufacturing - Green Technology*.
- Pathak, A. R., Pandey, M. and Rautaray, S. (2018) Application of Deep Learning for Object Detection. *Procedia Computer Science*, **132**, 1706–1717. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1877050918308767>.

- Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Wulandari, Yoshioaka, H. and Itoyama, E. (2019) Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. *Computers and Electronics in Agriculture*, **157**, 247–253.
- Permaloff, A. and Grafton, C. (1992) Optical Character Recognition. *PS: Political Science and Politics*, **25**, 523. URL: <https://www.jstor.org/stable/419444?origin=crossref>.
- Piggott, C. V., Depczynski, M., Gagliano, M. and Langlois, T. J. (2020) Remote video methods for studying juvenile fish populations in challenging environments. *Journal of Experimental Marine Biology and Ecology*, **532**, 151454. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022098119304319>.
- Pisner, D. A. and Schnyer, D. M. (2019) Support vector machine. In *Machine Learning: Methods and Applications to Brain Disorders*.
- Pope, K. L., Lochmann, S. E. and Young, M. K. (2010) Methods for Assessing Fish Populations. *Inland fisheries management in North America*.
- Pornpanomchai, C., Lurstwut, B., Leerasakultham, P. and Kitiyanan, W. (2013) Shape- and texture-based fish image recognition system. *Kasetsart Journal - Natural Science*, **47**, 624–634.
- Proud, R., Mangeni-Sande, R., Kayanda, R. J., Cox, M. J., Nyamweya, C., Ongore, C., Natugonza, V., Everson, I., Elison, M., Hobbs, L., Kashindy, B. B., Mlaponi, E. W., Taabu-Munyaho, A., Mwainge, V. M., Kagoya, E., Pegado, A., Nduwayesu, E. and Brierley, A. S. (2020) Applications of machine learning and artificial intelligence in marine science: all articles. *ICES Journal of Marine Science*, **77**, 1267–1455. URL: <https://academic.oup.com/icesjms/article/77/4/1267/5873749>.
- Qin, H., Li, X., Liang, J., Peng, Y. and Zhang, C. (2016) DeepFish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, **187**, 49–58. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0925231215017312>.
- Qiu, C., Zhang, S., Wang, C., Yu, Z., Zheng, H. and Zheng, B. (2018) Improving transfer learning and squeeze- and-excitation networks for small-scale fine-grained fish image classification. *IEEE Access*, **6**, 78503–78512.
- Rasmussen, R. S. and Morrissey, M. T. (2008) Methods for the Commercial Fish and Seafood Species. *Comprehensive reviews in food science and food safety*.
- Rathi, D., Jain, S. and Indu, S. (2017) Underwater Fish Species Classification using Convolutional Neural Network and Deep Learning. *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*. URL: <http://dx.doi.org/10.1109/icapr.2017.8593044>.
- Ronneberger, O., Fischer, P. and Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241.
- Rova, A., Mori, G. and Dill, L. M. (2007) One fish, two fish, butterflyfish, trumpeter: Recognizing fish in underwater video. In *Proceedings of IAPR Conference on Machine Vision Applications, MVA 2007*, 404–407.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, **115**.
- Saleh, A., Laradji, I. H., Kononov, D. A., Bradley, M., Vazquez, D. and Sheaves, M. (2020) A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports*, **10**, 14671. URL: <http://www.ncbi.nlm.nih.gov/pubmed/32887922http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7473859https://www.nature.com/articles/s41598-020-71639-x>.
- Saleh, A., Laradji, I. H., Lammie, C., Vazquez, D., Flavell, C. A. and Azghadi, M. R. (2021) A Deep Learning Localization Method for Measuring Abdominal Muscle Dimensions in Ultrasound Images. *IEEE Journal of Biomedical and Health Informatics*, **25**, 3865–3873. URL: <https://ieeexplore.ieee.org/document/9444630/>.
- Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J. and Harvey, E. (2016) Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, **14**, 570–585.
- Sarigül, M. and Avci, M. (2017) Comparison of Different Deep Structures for Fish Classification. *International Journal of Computer Theory and Engineering*.
- Schmidhuber, J. (2015) Deep learning in neural networks: An overview. *Neural Networks*, **61**, 85–117. URL: <http://dx.doi.org/10.1016/j.neunet.2014.09.003>.
- Schneider, S. and Zhuang, A. (2020) Counting Fish and Dolphins in Sonar Images Using Deep Learning. *arXiv preprint arXiv:2007.12808*. URL: <http://arxiv.org/abs/2007.12808>.

- Shafait, F., Mian, A., Shortis, M., Ghanem, B., Culverhouse, P. F., Edgington, D., Cline, D., Ravanbakhsh, M., Seager, J. and Harvey, E. S. (2016) Fish identification from videos captured in uncontrolled underwater environments. *ICES Journal of Marine Science*, **73**, 2737–2746.
- Shihavuddin, A., Gracias, N., Garcia, R., Gleason, A. and Gintert, B. (2013) Image-Based Coral Reef Classification and Thematic Mapping. *Remote Sensing*, **5**, 1809–1841. URL: <http://www.mdpi.com/2072-4292/5/4/1809>.
- Shryock, D. F., Defalco, L. A. and Esque, T. C. (2014) Life-history traits predict perennial species response to fire in a desert ecosystem. *Ecology and Evolution*, **4**, 3046–3059.
- Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R. and Harvey, E. S. (2018) Automatic fish species classification in underwater videos: Exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES Journal of Marine Science*.
- Sonka, M., Hlavac, V. and Boyle, R. (2008) *Image Processing, Analysis, and Machine Vision*. Thomson.
- Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.-H. J., Fisher, R. B. and Nadarajan, G. (2010) Automatic fish classification for underwater species behavior understanding. In *Proc. 1st Int. Worksh. Anal. Retrieval Tracked Events Motion Imagery Streams*, 45–50. Firenze, Italy: ACM.
- Su, Y., Guo, L., Jin, Z. and Fu, X. (2020) A mobile-beacon based iterative localization mechanism in large-scale underwater acoustic sensor networks. *IEEE Internet Things J.*
- Sun, C., Shrivastava, A., Singh, S. and Gupta, A. (2017) Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, 843–852.
- Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., Lukacs, P. M., Moeller, A. K., Mandeville, E. G., Clune, J. and Miller, R. S. (2019) Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*.
- Takada, Y., Koyama, K. and Usami, T. (2014) Position Estimation of Small Robotic Fish Based on Camera Information and Gyro Sensors. *Robotics*, **3**, 149–162. URL: <http://www.mdpi.com/2218-6581/3/2/149>.
- Tamou, A. B., Benzinou, A., Nasreddine, K. and Ballihi, L. (2018) Underwater live fish recognition by deep learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Tarling, P., Cantor, M., Clapés, A. and Escalera, S. (2021) DEEP LEARNING WITH SELF-SUPERVISION AND UNCERTAINTY REGULARIZATION TO COUNT FISH IN UNDERWATER IMAGES. *Tech. rep.*, Other. URL: <http://www.echoview.com>.
- Thorstad, E. B., Rikardsen, A. H., Alp, A. and Okland, F. (2013) The Use of Electronic Tags in Fish Research - An Overview of Fish Telemetry Methods. *Turkish Journal of Fisheries and Aquatic Sciences*.
- Trinh, H., Fan, Q., Gabbur, P. and Pankanti, S. (2012) Hand tracking by binary quadratic programming and its application to retail activity recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Van Allen, B. G., Dunham, A. E., Asquith, C. M. and Rudolf, V. H. (2012) Life history predicts risk of species decline in a stochastic world. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 2691–2697.
- Varalakshmi, P. and Julanta Leela Rachel, J. (2019) Recognition of Fish Categories Using Deep Learning Technique. In *2019 Proceedings of the 3rd International Conference on Computing and Communications Technologies, ICCCT 2019*.
- Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T. and Villéger, S. (2018) A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological Informatics*.
- Vincenzi, S., Crivelli, A. J., JeseŃsek, D., Campbell, E. and Garza, J. C. (2019) Effects of species invasion on population dynamics, vital rates and life histories of the native species. *Population Ecology*, **61**, 25–34.
- Wang, B. and Weiland, J. D. (2017) Visual system. In *Neuroprosthetics: Theory and Practice: Second Edition*.
- Wang, G., Hwang, J. N., Williams, K., Wallace, F. and Rose, C. S. (2017a) Shrinking encoding with two-level codebook learning for fine-grained fish recognition. In *Proceedings - 2nd Workshop on Computer Vision for Analysis of Underwater Imagery, CVAUI 2016 - In Conjunction with International Conference on Pattern Recognition, ICPR 2016*, 31–36.
- Wang, S. H., Zhao, J., Liu, X., Qian, Z.-M., Liu, Y. and Chen, Y. Q. (2017b) 3D tracking swimming fish

- school with learned kinematic model using LSTM network. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1068–1072. IEEE. URL: <http://ieeexplore.ieee.org/document/7952320/>.
- Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldhuis, M. and Fortson, L. (2019) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, **10**, 80–91. URL: <https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13099>.
- Xu, J.-L., Hugelier, S., Zhu, H. and Gowen, A. A. (2021) Deep learning for classification of time series spectral images using combined multi-temporal and spectral features. *Analytica Chimica Acta*, **1143**, 9–20. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0003267020311429>.
- Xu, L., Bennamoun, M., An, S., Sohel, F. and Boussaid, F. (2019) Deep learning for marine species recognition. In *Smart Innovation, Systems and Technologies*, vol. 136, 129–145. Springer Science and Business Media Deutschland GmbH. URL: http://link.springer.com/10.1007/978-3-030-11479-4_7.
- Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S. and Zhou, C. (2020) Deep learning for smart fish farming: applications, opportunities and challenges.
- Zarco-Perello, S. and Enríquez, S. (2019) Remote underwater video reveals higher fish diversity and abundance in seagrass meadows, and habitat differences in trophic interactions. *Scientific reports*, **9**, 6596. URL: <http://www.ncbi.nlm.nih.gov/pubmed/31036932><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6488625>.
- Zhang, W., Wu, C. and Bao, Z. (2022) DPANet: Dual Pooling and Aggregated Attention Network for fish segmentation. *IET Computer Vision*, **16**, 67–82. URL: <https://onlinelibrary.wiley.com/doi/10.1049/cvi2.12065>.
- Zheng, H., Wang, R., Ji, W., Zong, M., Wong, W. K., Lai, Z. and Lv, H. (2020) Discriminative deep multi-task learning for facial expression recognition. *Information Sciences*, **533**, 60–71. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0020025520303601>.
- Zhuang, P., Wang, Y. and Qiao, Y. (2020) WildFish++: A Comprehensive Fish Benchmark for Multimedia Research. *IEEE Transactions on Multimedia*.
- Zion, B. (2012) The use of computer vision technologies in aquaculture – A review. *Computers and Electronics in Agriculture*, **88**, 125–132. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168169912001950>.
- Zion, B., Alchanatis, V., Ostrovsky, V., Barki, A. and Karplus, I. (2007) Real-time underwater sorting of edible fish species. *Computers and Electronics in Agriculture*, **56**, 34–45. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168169906001244>.
- (2008) Classification of guppies’ (*Poecilia reticulata*) gender by computer vision. *Aquacultural Engineering*, **38**, 97–104.
- Zurowietz, M. and Nattkemper, T. W. (2020) Unsupervised Knowledge Transfer for Object Detection in Marine Environmental Monitoring and Exploration. *IEEE Access*, **8**, 143558–143568.