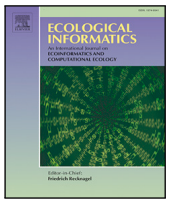
Ecological Informatics 86 (2025) 102989

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Surveying the deep: A review of computer vision in the benthos

Cameron Trotter *, Huw J. Griffiths, Rowan J. Whittle

British Antarctic Survey, Cambridge, UK

ARTICLE INFO

Dataset link: https://github.com/antarctica/surveying-the-deep

Keywords:
Benthos
Biodiversity monitoring
Computer vision
Deep learning
Instance segmentation
Image classification
Machine learning
Object detection
Semantic segmentation

ABSTRACT

The analysis of image data for benthic biodiversity monitoring is now commonplace within the domain of marine ecology. Whilst advances in imaging technologies have allowed for the collection of vast quantities of data, the curation of this has traditionally been performed manually, resulting in a bottleneck whereby data is collected faster than it can be processed. Recent years have seen marine ecologists turn to the domain of computer vision to help automate this curation process. However, as the knowledge required to build such systems spans both domains, there is a high barrier to entry. To help reduce this barrier, this paper aims to provide an introduction to computer vision-based benthic biodiversity monitoring via a comprehensive literature review. To aid ecologists, key computer vision concepts are described and example use-cases highlighted. The major challenges inherent to benthic imagery for computer vision systems are explored, alongside a discussion of how current systems attempt to mitigate against these. To aid computer scientists wishing to enter the domain, an exploration of currently available open-source benthic datasets is also provided. Recommendations for future research are explored, including a move towards human-centric techniques, committing to ablation studies, reaching community agreement on open-source benchmarking datasets, and an increased use of innovative methods to allow for improved answering of key benthic ecology questions.

Contents

1. Introduction 2
2. Computer vision for benthic biodiversity monitoring 2
   2.1. Image classification 3
      2.1.1. Coverage estimation via image patching 4
   2.2. Object detection 5
   2.3. Semantic segmentation 5
   2.4. Instance segmentation 6
3. Challenges of working with benthic imagery from a computer vision perspective 7
   3.1. Required labelling effort 7
   3.2. Concept drift 8
   3.3. Other notable issues 8
4. Available benthic imagery datasets 9
5. Future research 10
   5.1. Human-centric techniques 10
   5.2. Ablation studies 10
   5.3. Benchmark datasets 10
   5.4. Vision transformers 11
   5.5. Foundation models 11
   5.6. Semi-supervised learning 11
   5.7. Open-set recognition 11
   5.8. Multi-modal models 12
   5.9. Other innovative methods 12
6. Conclusion 12

* Corresponding author.
E-mail address: cater@bas.ac.uk (C. Trotter).

https://doi.org/10.1016/j.ecoinf.2024.102989
Received 20 September 2024; Received in revised form 31 December 2024; Accepted 31 December 2024
Available online 22 January 2025

CRediT authorship contribution statement 14
Declaration of competing interest 14
Acknowledgements 14
Appendix A. Supplementary data 14
Appendix. Data availability 14
References 14

1. Introduction

Benthos, meaning 'the depths of the sea' in ancient Greek, is the name given to the community of organisms that live on, in, or near the bottom of a body of water. The number of benthic animal species is thought to exceed one million (Sokolova, 2000), located in freshwater and intertidal environments down to the deepest ocean trenches, with benthic algae restricted to the reach of daylight (Lalli and Parsons, 1997). Marine macrobenthic fauna play important roles in global ecosystem function and regulating the fluxes of energy, nutrients, and matter within global cycles. However, the consequences of anthropogenic change and direct human impacts are altering the structure and function of benthic communities (Lam-Gordillo et al., 2020). In order to understand the nature, scale, and intensity of these ecological changes and the potential impacts on the ecosystem services provided by the ocean, quantifying and monitoring benthic ecosystems has become increasingly important.

For most of human history, our knowledge of what lives in the benthos has relied upon destructive nets or devices (e.g. grabs and cores) to bring organisms to the surface (Rees, 2009). The development of technology that enabled us to visit the seafloor (e.g. submersibles and SCUBA) and later remotely operated or autonomous devices have enabled humans to observe and record these organisms in situ, enabling us to better understand community distributions, structure and function (Du Preez et al., 2016; Roelfsema et al., 2021; Medelytė et al., 2022). The rapid evolution of photography and video as a tool to study the benthos has led to a rise in use of these non-invasive study methods, with increasing volumes of image-based data now collected from the world's oceans (Gomes-Pereira et al., 2016).

However, identification of organisms in images is a very slow process, taking hours per individual photograph (Williams et al., 2019; Alicia et al., 2023). It also requires a high level of specialist taxonomic expertise over a wide range of organisms. The mis-match between the volume of data that can now be collected and the speed such data can be curated has caused a bottleneck in analyses of benthic communities (Williams et al., 2019).

Recent years have seen the introduction of a range of automated solutions to benthic biodiversity monitoring, mostly thanks to advances in *computer vision* (CV, definitions of italicised words can be found in the Glossary) techniques. Such advances were first observed thanks to progress in the domain of *image processing* (IP), which allows for the extraction of representative visual features from imagery through pre-determined algorithmic steps. Later, advances in the field of *machine learning* (ML) turned attention away from extraction via hand-engineered algorithms to more robust statistical models. The rise of *deep learning* (DL) expedited this further thanks to the development of complex multi-layered *Convolutional Neural Networks* (CNNs) capable of generalised feature extraction learnt via large-scale training datasets.

Whilst these techniques have been widely embraced by other fields of ecology to answer key questions, their use in the benthos is not as widespread. The use of existing CV-based methods, and the development of new approaches, often requires an understanding of their underlying technologies, posing a high barrier to entry. On the other hand, computer scientists wishing to use their skills to aid in the domain of benthic ecology are often unaware of potential use-cases, publicly available sources of high quality data, and may lack the domain (i.e. benthic ecology) knowledge to ensure any system they develop is both useful and correct. Furthermore, the inherent and unique properties of benthic imagery pose an exciting computational challenge, most notably overcoming issues surrounding adverse environmental or sensor conditions, occlusion and species aggregations, as well as new-to-science or invasive organisms.

Whilst a host of reviews into CV-based automation of ecology data exist, these primarily focus on terrestrial data, such as those from camera traps (Norouzzadeh et al., 2018). Kumar et al. (2023) highlight some marine use cases in their review of species localisation and identification, though this focuses only on DL-based approaches rather than a complete view including IP and ML systems also. Of the reviews focusing on the marine environment, most often concentrate on *midwater* rather than benthic environments (Moniruzzaman, Md. et al., 2017; Saleh et al., 2022; Goodwin et al., 2022) or, as before, only discuss DL-based approaches (Wang et al., 2023b). Other reviews examining the benthos often provide an overview of coral-focused systems only (Raphael et al., 2020).

As such, this review aims to further bridge the gap between benthic ecology and computer science by providing a comprehensive overview of the work so far into automated benthic biodiversity monitoring CV systems, an area where to the best of our knowledge no survey currently exists. Along the way, key terms are introduced and explained. Next, the inherent properties of benthic imagery which may pose a challenge to the current state of the art in CV are explored, aiming to guide researchers entering the area. An overview of the currently available open-source benthic imagery datasets is then presented. Finally, potential avenues for future research are provided.

2. Computer vision for benthic biodiversity monitoring

This section explores works that apply CV techniques to benthic biodiversity monitoring and is structured as follows. Sub-sections are categorised by system output type, from classification to segmentation, beginning with those that make use of IP, moving onto ML, and finally DL. When discussing ML and DL, particular attention is paid to *supervised learning* techniques, given that it is likely researchers wishing to make use of CV are in possession of an existing catalogue of labelled data.

Fig. 1 illustrates the progression of research into CV-based benthic biodiversity monitoring over time. This includes works that have both developed their own methods and made use of existing ones. The number of publications for each CV technique, IP, ML, and DL, represented as stacked bars, highlights the shifts in popularity and the relative prominence of each approach over time. Note that if a publication contained more than a single technique, it is represented multiple times.

Early research focused predominately on IP methods. This continued until 2006 when the use of ML was also explored, marking the beginning of data-driven techniques. The use of IP and ML continued, with researchers often utilising methods from both domains in tandem to achieve improved results.

2016 onwards has seen a dominance of DL-based methods in the space of automated benthic biodiversity monitoring. Indeed, whilst only 47% of papers surveyed use DL overall, 69% of papers since 2016 use this technique.[1] This has likely been led by rapid advancements

[1] Percentage of papers surveyed overall using IP: 33%; ML: 20%; DL: 47%. Since 2016, IP: 18%; ML: 13%; DL: 69%.

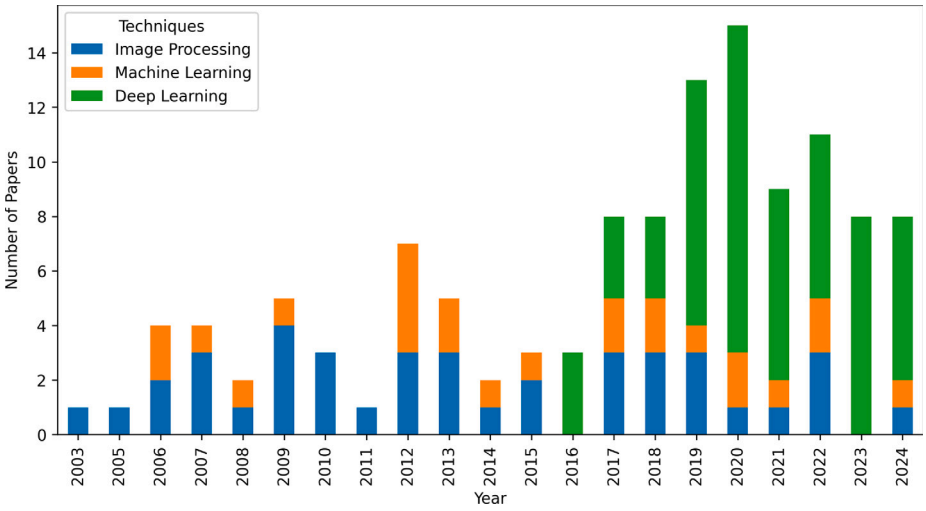


Fig. 1. The progression of computer vision-based benthic biodiversity monitoring literature over time, subdivided by techniques utilised. If a publication contained more than a single technique, it is represented multiple times. For a detailed overview of each paper included, see Table S1.

Table 1
A comparison of key selection criteria across different computer vision techniques.

Criteria	Computer vision technique		
	Image processing	Machine learning	Deep learning
Computational Cost	Low	Medium	High
Data Requirements	Low	Medium	High
Feature Engineering	High	Medium	Low
Feature Extraction Automation	Low	Medium	High
Generalisability	Low	Medium	High
Handle Unstructured Data	Medium	Low	High
Interpretability	High	Medium	Low
Pattern Recognition	Low	Medium	High
Training Time	None	Short	Long

in DL model architectures providing the ability to perform generalised, domain-agnostic feature extraction (Ahmed et al., 2023), in comparison to IP and ML-based techniques which often require extensive domain knowledge and hand-engineering in order to produce efficient, though task-specific, feature extractors. The rise of DL in benthic ecology has also been helped by the ease of access to such models provided by open-source code repositories, model zoos, and their inclusion in data labelling tools (see Section 3.1). Despite this, DL does have its disadvantages, notably a typical need for larger data volumes and more expensive training processes than IP or ML methods. For a comparison of IP, ML, and DL, see Table 1.

As a result of these trade-offs, there is space in the benthos to answer ecological questions using all three CV techniques — though understanding when one is more appropriate over the others is non-trivial. To afford benthic biodiversity researchers the ability to make an informed decision regarding their use of CV to answer their desired ecological questions, a comprehensive literature review into previous CV use in the benthos was undertaken. The Scopus and Google Scholar databases were queried for relevant publications using the specific terms 'benthos/benthic', 'computer vision', and 'biodiversity monitoring', alongside 'classification', 'detection', and 'segmentation'. To locate literature which may have been missed by the initial search, the bibliographies of publications deemed within scope were examined. Additional relevant literature were also highlighted during peer review.

This review defines biodiversity monitoring to encompass a wide variety of ecological questions, including abundance estimation, coverage estimation, heath assessment, impact assessment, population dynamics, taxonomic identification, and the detection of new-to-science organisms. Publications were considered in scope if they outlined work making use of automated CV methodologies, processing benthic imagery to extract information which could be utilised to perform biodiversity monitoring as previously defined. The use of CV in the benthos for other tasks, such as resource abundance assessment (e.g. Schoening et al., 2016) or debris detection (e.g. Huang et al., 2023) are considered out of scope for this review, as are works which make use of supervised learning for benthic biodiversity monitoring on non-image based data sources (e.g. Kwon et al., 2024). Note that due to the context-specific nature of these systems, it is currently not possible to chart a sequence of improvements over time — this is discussed in greater detail in Section 5.

To help provide an overview of the literature examined in this study, Table 2 outlines the CV task-technique pairs utilised by the reviewed literature in this study, alongside example studies and the ecological question they aimed to answer. A full breakdown of all papers included in this study can be found in Table S1. For each, the CV techniques and system output types used were extracted alongside geographic location and data availability information.

2.1. Image classification

Works discussed in the following section all perform *image classification*. This technique aims to produce a single *class* label for a given image, which typically contains a lone dominant object. This can be useful to benthic ecologists wishing to classify imagery that has been previously cropped to contain a single taxa or *region of interest* (RoI), such as diseased coral (Ani Brown Mary and Dharma, 2019) or tube worms (Lüdtke et al., 2012). An example of benthic image classification can be seen in Fig. 2.

Of the articles reviewed, only Šaškov et al. (2015) perform classification through IP techniques alone. However, a large volume of work first makes use of IP techniques to extract relevant features before passing these to ML models for classification. Extracted features may

Table 2
An overview of the computer vision task-technique pairs utilised by the literature reviewed in this study. For each, an example study is highlighted, alongside the ecological question the work aimed to answer. Table ordered based on task-technique complexity.

Computer vision			Example	Ecological question
Task	Output	Technique	study	answered
Image Classification	Single class label	Image Processing	Šaškov et al. (2015)	Coverage estimation
		Machine Learning	Beijbom et al. (2015)	Coverage estimation
			Ani Brown Mary and Dharma (2019)	Health assessment
		Deep Learning	Zhou et al. (2023)	Taxonomic ID
Object Detection	Area-level localisation and class label	Image Processing	Clement et al. (2005)	Presence-absence survey
		Machine Learning	Dawkins et al. (2013)	Abundance estimation
		Deep Learning	Zhang et al. (2024a)	Abundance estimation
			Cuvelier et al. (2024)	Impact assessment
			Naseer et al. (2020)	Presence-absence survey
Semantic Segmentation	Pixel-level localisation and lass label (single mask per class)	Image Processing	Smith and Dunbabin (2007)	Abundance estimation
				Morphological analysis
		Machine Learning	Mohamed et al. (2022)	Habitat mapping
			Tan et al. (2018)	Abundance estimation
			Manderson et al. (2017)	Health assessment
		Deep Learning	Pavoni et al. (2021)	Habitat mapping
			Harrison et al. (2021)	Behaviour analysis
Instance Segmentation	Pixel-level localisation and class label (multiple masks per class)	Deep Learning	Lütjens and Sternberg (2021)	Abundance estimation

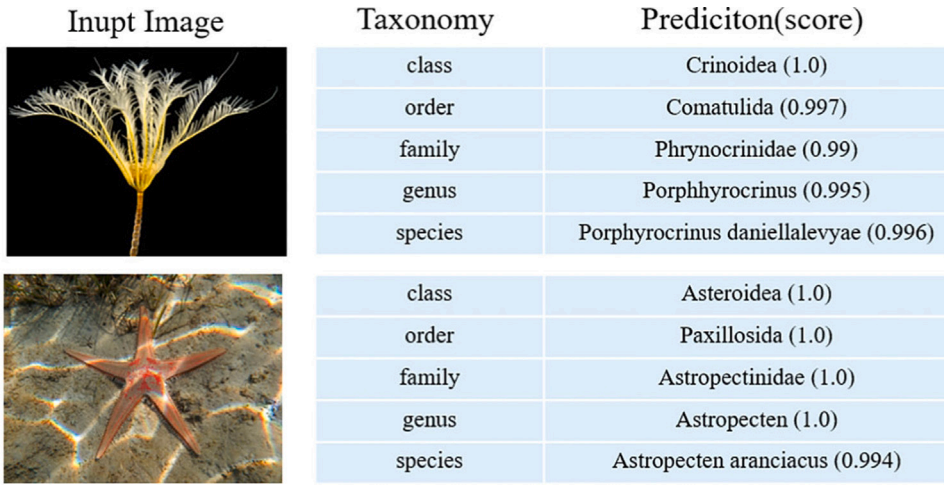


Fig. 2. An example of benthic image classification. All pixels in each input image are provided a single predicted class per taxonomic level. A new classification network is trained to output labels at different levels of the taxonomic tree.
Source: Figure adapted from Zhou et al. (2023, CC BY 4.0).

be solely texture-based (Lüdtke et al., 2012; Gonzalez-Cid et al., 2017; Ani Brown Mary and Dharma, 2019), or based on both colour and texture (Beijbom et al., 2012; Shihavuddin et al., 2013; Beijbom et al., 2015).

More recent works in benthic image classification forego the need for hand-engineered feature extraction pipelines, instead relying on features derived by training DL models such as CNNs (Boulais et al., 2020; Langenkämper et al., 2020; Zhou et al., 2023). CNNs demonstrate superior ability to directly extract important classification features from benthic image data compared to pre-determined IP or ML techniques, as evidenced by multiple studies (Gonzalez-Cid et al., 2017; Rimavicius and Gelzinis, 2017). Further, Zhou et al. (2023) show that CNNs are capable of classifying benthic taxa at multiple taxonomic levels, though this does require creating a new network for each given taxonomy.

2.1.1. *Coverage estimation via image patching*

Benthic imagery is often captured at high resolution. When working with such data, it may be computationally infeasible to process the whole image at once. To combat this, images may be split into smaller sub-images, or patches, processed independently, and recombined. An example of image patching can be seen in Fig. 3. The use of patching can be seen as advantageous over simply resizing the whole image, as doing so can result in a loss of image quality which may hinder the detection of smaller RoIs due to a subsequent loss of representative features.

Patching large-scale imagery allows for species coverage estimates to be performed using image classification, rather than more computationally expensive techniques such as *semantic segmentation* (see Section 2.3). Multiple works in the literature generate coverage estimations via patched image classification. This is common for the purposes of coral coverage estimation (Marcos et al., 2008; Shihavuddin et al., 2013; Mahmood et al., 2016; Gómez-Ríos et al., 2019; González-Rivero et al., 2020; Chen et al., 2021), though has also been used for estimating coverage of bacterial mats (Lüdtke et al., 2012), tapeweed (Gonzalez-Cid et al., 2017), and substrate (Jackett et al., 2023).

Whilst patches are typically generated via a sliding window over the original image (Shihavuddin et al., 2013; Gómez-Ríos et al., 2019; Chen et al., 2021), Rimavicius and Gelzinis (2017) make use of IP algorithms to generate non-uniform patches, whilst Marburg and Bigham

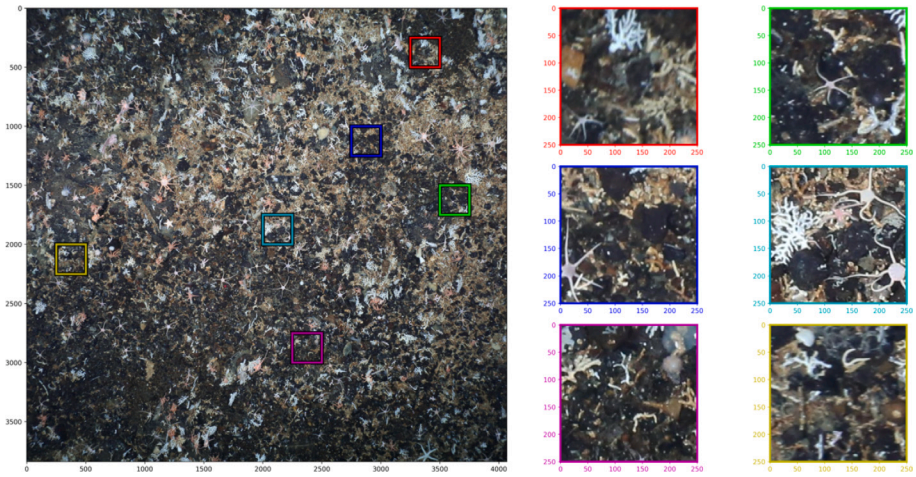


Fig. 3. An example of image patching. A large-scale, high-resolution image (left) may be computationally infeasible to process in a single instance. As such, it can be split into multiple smaller patches of the same resolution which are processed independently then recombined. Six example patches are shown (right). Images resized for clarity, actual sizes shown on axes in pixels.
Source: Figure generated using data from Purser et al. (2021, CC BY 4.0).

(2016), Piechaud et al. (2019) and Durden et al. (2021) generate patches by extracting pixels surrounding labelled RoIs. As with whole image classification, patch classification may be performed using IP and ML (Shihavuddin et al., 2013; Lüdtke et al., 2012; Gonzalez-Cid et al., 2017), or directly via a CNN (Chen et al., 2021; Wyatt et al., 2022; Lumini et al., 2023).

2.2. Object detection

The works outlined in Section 2.1 classify input images as a single class. If the image contains multiple RoIs these must first be extracted before processing, reducing the time and effort savings relative to a fully manual approach. This limitation can be overcome through the use of *object detection* algorithms that perform both RoI localisation and classification. This allows multiple labels to be associated with a single image, removing the need for prior RoI extraction. As such, object detection techniques are well-suited to ecological questions related to community structure or abundance (Lopez-Vazquez et al., 2020; Cuvelier et al., 2024).

Earlier works in the area of benthic object detection make use of IP techniques to extract hand-engineered features for the purposes of single class RoI localisation. Works utilising IP alone often focus on visually distinct organisms such as starfish (Di Gesu et al., 2003; Clement et al., 2005; Smith and Dunbabin, 2007), though these techniques can be combined with ML models to detect less visually distinct organisms such as crabs (Gobi, 2010), kelp (Bewley et al., 2012), or scallops (Enomoto et al., 2009; Kannappan and Tanner, 2013; Dawkins et al., 2013). Aguzzi et al. (2009) is the only reviewed work to make use of IP and ML to detect multiple benthic classes.

More recently, the focus of benthic object detection works has shifted to the use of DL techniques. Typically these methods will output detected RoIs in the form of a *bounding box*, in contrast to ML methods where an RoI may take a range of forms. An example of bounding box outputs for a given input image can be seen in Fig. 4.

A large volume of benthic DL object detection work makes use of data provided by the Underwater Robot Picking Challenge (URPC, Liu et al., 2021), aiming to detect commercially valuable taxa such as scallops, sea cucumbers, and sea urchins (Lv et al., 2019; Chen et al., 2020; Wang et al., 2021; Fu et al., 2022; Wang et al., 2023c). Huang et al. (2019) utilise a Faster R-CNN (Ren et al., 2015) model to detect sea cucumbers, sea urchins, and scallops in underwater imagery, though based on the literature it is not clear if the authors make use of a URPC dataset or data from elsewhere. At the time of writing, no URPC datasets are publicly available, though Liu et al. (2022) develop a similar open-source dataset called The Underwater Open-sea Farm Object Detection Dataset (UDD, see Section 4). The authors make use of this dataset to train their novel AquaNet CNN, taking advantage of data augmentation strategies (see Section 3.1) such as multi-scale blur-sampling and feature fusion to increase model robustness and aid in the detection of small RoIs.

There are however also multiple works which do not make use of the URPC dataset. Whilst Naseer et al. (2020) use DL to detect lobster and their burrows, other works such as Yeh et al. (2022) and Zhang et al. (2023) aim instead to detect multiple distinct objects. Boulais et al. (2020) makes use of the FathomNet dataset (Katija et al., 2022, see Section 4) to train their benthic object detector. Due to the long-tail problem inherent in ecology datasets, where a small number classes make up the majority of data samples leading to degraded performance (Van Horn and Perona, 2017; Van Horn et al., 2018; Miao et al., 2021), the work found that training for fine-grained classification at the genus or family level did not lead to speed ups compared to hand labelling. Instead, they recommend labelling at a coarse-grained level (e.g. 'fish' or 'crustacean').

Liu et al. (2024) developed DeepSeaNet, a pipeline which employs DL, ML, and IP techniques to detect species in benthic imagery. First, they make use of a modified YOLOv7 Tiny (Wang et al., 2023a) architecture, adapted with specialised 'deep-sea modules' to aid the detection of organisms with scale variability, evasive behaviour, and camouflage. Detections from the DL model are then passed through IP-based feature extractors to generate organism descriptors. These are then clustered using ML to generate a feature latent space, generating groupings corresponding to organism species. Previously unseen species or model mislabels can be detected by comparing an input's location in the latent space against all other existing clusters. Such methodologies may be thus be useful for detecting new-to-science or invasive organisms.

2.3. Semantic segmentation

One disadvantage of bounding box-based object detection algorithms is the need to include background features within RoIs in order to fully encompass an object. This can cause generalisation issues, particularly if training data is limited to a small spatio-temporal scale. Under these circumstances, the network may learn to classify based on background features rather than those present on the object (Tian et al., 2018; Curry et al., 2021; Shepley et al., 2021; Trotter et al., 2022).

The problem of background inclusion can be solved through the use of *semantic segmentation*. Here, RoIs are defined as *masks*, allowing

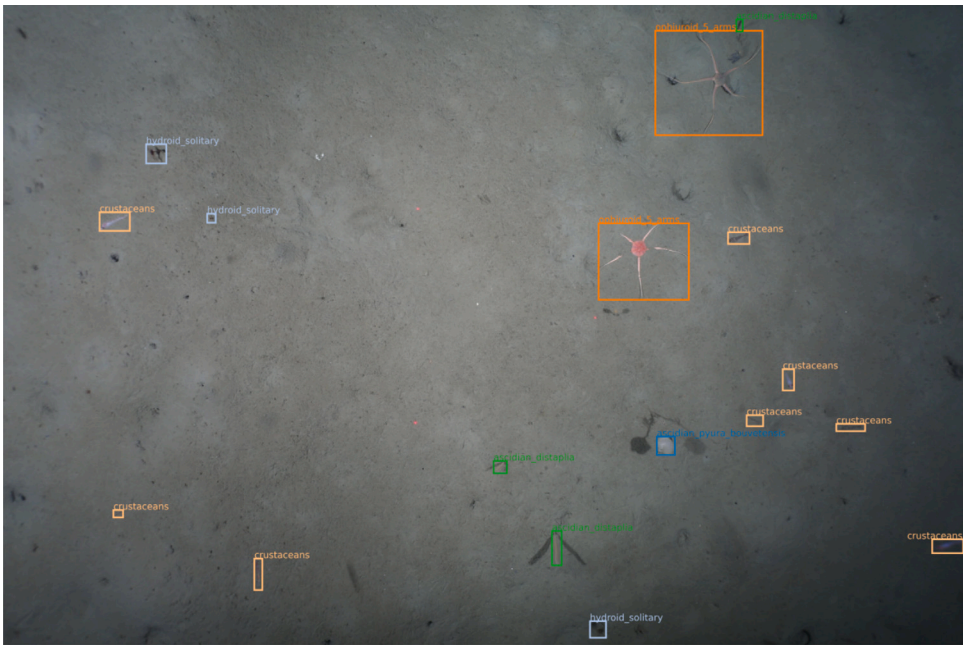


Fig. 4. An example of benthic object detection. Detected RoIs are encapsulated by bounding boxes. Boxes and labels are colour coded based on classification, with labels displayed above each box.
Source: Figure generated using data from Purser et al. (2021, CC BY 4.0).

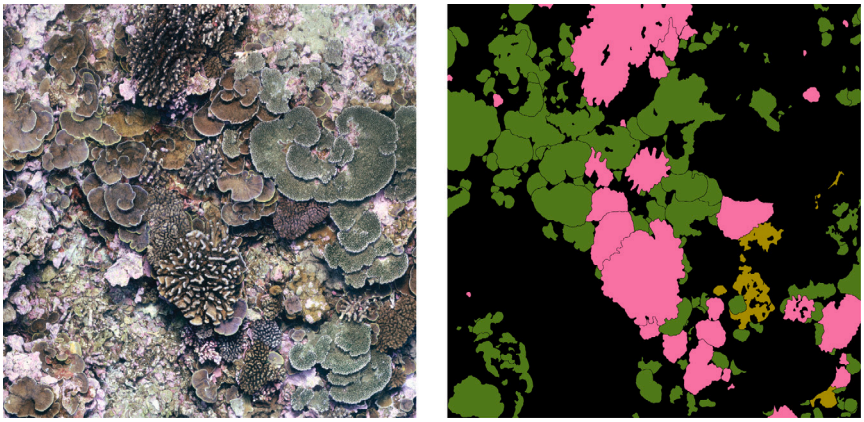


Fig. 5. An example benthic image (Left) and corresponding ground truth mask (Right) produced for the task of coral semantic segmentation. Pixel colour denotes class, with black denoting background.
Source: Figure adapted from Pavoni et al. (2022, CC BY 4.0).

for finer-grained localisation (i.e. pixel level categorisation) when compared to object detection tasks. Benthic ecologists wishing to perform coral coverage estimates (Alonso et al., 2019; Pavoni et al., 2021) or classify organism behaviour (Harrison et al., 2021) may find semantic segmentation methods useful. Furthermore, semantic segmentation also has the advantage of capturing temporal dynamics of organisms when they change their geometry over spatio-temporal scales (Harrison et al., 2021). An example input image and a visualisation of its corresponding semantically segmented pixel masks can be seen in Fig. 5.

The creation of segmentation masks for benthic imagery can be performed solely using classical IP techniques (Gleason et al., 2007; Beuchel et al., 2010; Aguzzi et al., 2011; Tan et al., 2014) or through a combination of IP and ML (Johnson-Roberson et al., 2006; Johnson-Roberson et al., 2006; Fearn et al., 2007; Schoening et al., 2012, 2014; Manderson et al., 2017; Ani Brown Mary and Dejey, 2018). However, feature extraction pipelines to perform semantic segmentation using these methods can often be complex and cumbersome for multi-class tasks. For example, Schoening et al. (2012) makes use of a pipeline of *Support Vector Machines* organised in a tree structure, with each trained to classify a single class.

The use of DL for the task of benthic semantic segmentation can help reduce the complexity of these pipelines, training directly using images and their corresponding hand-annotated masks. However, this often comes at the cost of increased computation requirements. A range of newer works make use of DL to semantically segment benthic environments, most notably for the tasks of coral coverage estimation (Alonso et al., 2017; King et al., 2018; Mizuno et al., 2020; Pavoni et al., 2020; Song et al., 2021), though use cases also exist for other fauna (Osterloff et al., 2019; Buškus et al., 2021).

2.4. Instance segmentation

In research areas such as coral reef size estimation, extracting estimates of class coverage using semantic segmentation may be sufficient for drawing conclusions. However, for tasks such as abundance estimation this may not hold true. When 90% of pixels in an image

Fig. 6. An example of instance segmentation. Unlike object detection, RoIs are provided as masks rather than bounding boxes, outputted on an RoI rather than class basis as in semantic segmentation. Each mask is visualised as a coloured overlay.
Source: Figure reproduced from Lütjens and Sternberg (2021, CC BY-NC-ND 4.0).

belong to a single class, has one dominant instance or numerous small instances been captured? This example highlights the unsuitability of semantic segmentation for tasks where distinguishing between different instances of the same class is required.

In these cases, techniques such as *instance segmentation* may be more beneficial. As with object detection each individual RoI is localised, however rather than providing this at a coarse-grained level such as a bounding box, the RoI is represented as a fine-grained pixel mask. This is in contrast to semantic segmentation, where masks are provided at a class level. Each RoI can thus be processed in isolation. An example of instance segmentation can be seen in Fig. 6.

Despite the usefulness of instance segmentation within benthic biodiversity monitoring, few examples of its use exist within the literature. Both Zurowietz et al. (2018) and Shashidhara et al. (2020) utilise a two-step pipeline to propose and filter RoIs, with the former generating coarse-grained ellipses with the latter generating fine-grained pixel-level masks. In contrast, Lütjens and Sternberg (2021) adopt a more efficient approach, training a CNN capable of producing instance segmentation masks directly for a small number of benthic classes.

3. Challenges of working with benthic imagery from a computer vision perspective

When applying CV-based automation techniques to benthic imagery for the purposes of biodiversity monitoring, several challenges must be overcome. This section discusses these challenges in detail, and highlights how existing literature aims to overcome them.

3.1. Required labelling effort

Multiple works available in the literature show that, for benthic imagery automation, ML and DL techniques provide a more generalisable and accurate solution when compared to more traditional IP (Gonzalez-Cid et al., 2017; Lopez-Vazquez et al., 2020; González-Rivero et al., 2020). However, the need for extensive manually labelled data can often be a challenge in this domain.

As the level of required annotation granularity increases, from image classification to instance segmentation, so does the cost and effort of labelling. This is exacerbated if objects are aggregated, a situation commonly observed within benthic imagery. Whilst the use of citizen scientists could expedite this process, this is only viable when classifying at high taxonomic levels, such as in Zhang et al. (2023). Labelling at low levels often requires the use of marine ecologists, resulting in prolonged and expensive processes (Katija et al., 2022).

The use of semi-automated labelling tools may help reduce the data curation workload. An overview of semi-automated labelling tools utilised by the reviewed literature is provided in Table 3. Other semi-automated labelling tools exist outside of those listed, focussing on both marine imagery specifically such as VIAME (Dawkins et al., 2017) as well as more general labelling platforms such as AIDE (Kellenberger et al., 2020).

Further to this, the use of *data augmentation* during model training may help increase the amount of varied data available. Data augmentation is prevalent in the explored works, with some making use of simple perturbations such as rotating, cropping, or re-scaling (e.g. Rimavicius and Gelzinis, 2017; Durden et al., 2021; Pavoni et al., 2021), whilst others make use of techniques such as copy-pate (Doig et al., 2024) or domain-specific strategies (e.g. Alonso et al., 2017; Huang et al., 2019; Lütjens and Sternberg, 2021; Yeh et al., 2022).

The use of *transfer learning* may also aid model generalisability when data volumes are low. Typically when training a DL model from scratch the initial model parameters are randomly assigned and updated during the training process to reflect the target dataset's distribution. In transfer learning, the learnt parameters from a model trained on a large source dataset are used as a starting point when training on the target dataset. This allows for a more generalised model with higher convergent rate when the target dataset is small, as is often the case in benthic image analysis, in essence performing a knowledge transfer (Pan and Yang, 2010). Multiple reviewed works make use of transfer learning to aid model generalisation (see Table S2).

In spite of this, once a labelled benthic dataset has been obtained there is no guarantee that it is perfectly annotated. Whilst this issue is not present solely in the benthic domain (e.g. Van Horn et al., 2015), the risk of mislabelling is inherently high due to the challenging environmental conditions images are often captured in and the morphological similarity between some taxa. As such the performance of any automated biodiversity data curation system can only be interpreted as how well the system agrees with the labeller (Boulent et al., 2023). Even still, errors may be present due to the labeller becoming fatigued or distracted, resulting in a missed or misclassified organism. The use of multiple labellers alongside some merging process can reduce this risk, though this does increase time and monetary expenses.

Table 3
A comparison of semi-automated labelling tools utilised by the reviewed literature.

Tool name	Label style	Local install	Export Format	Features			Quality control	Multi-User support	Code availability	Example study
				Active learning	Pre-trained models	Custom models				
BIIGLE (Langenkämper et al., 2017)	Point, Line, Circle, Bounding Box, Polygon	✓	CSV, JSON	✗	✓	✓[a]	✓	✓	Open	Cuvelier et al. (2024)
CoralNet (Beijbom et al., 2015)	Point	✗	CSV	✗	✓	✓	✗	✓	Open	Miller et al. (2023)
EISeg (Hao et al., 2022)	Polygon	✓	JSON	✗	✓	✓	✗	✗	Open	Gu et al. (2023)
Roboflow[b]	Image, Point, Bounding Box, Polygon	✗	CSV, JSON, TXT, XML	✓	✓	✓	✓	✓	Closed (Freemium)	Monari et al. (2023)
RootPainter (Smith et al., 2022)	Polygon	✓	CSV	✗	✗	✓	✗	✗	Open	Clark et al. (2024)
SQUIDLE+[c]	Image, Point, Bounding Box, Polygon	✗	CSV, HTML, JSON	✓	✓	✓	✓	✓	Closed (Free)	Deo et al. (2024)
TagLab (Pavoni et al., 2022)	Polygon	✓	CSV, GeoTIFF, JSON	✗	✓	✓	✓	✗	Open	Amir et al. (2023)

[a] Model training and inference in BIIGLE, with the exception of MAIA (Zurowietz et al., 2018), is currently performed outside of the user interface via the API.
[b] Roboflow: roboflow.com/annotate.
[c] SQUIDLE+: squidle.org.

3.2. Concept drift

When creating automated CV systems, it is often the case that a consistent data distribution is assumed. This is unlikely to hold true in the domain of benthic ecology. Data is often collected over multiple surveys and spread over a wide spatio-temporal scale (data collected in multiple locations or over multiple years). This may result in changes to collection methodology (e.g. diver, autonomous underwater vehicle), as well as varying environmental (e.g. weather, lighting, occlusions, turbidity) or sensor (e.g. noise, blur, lens distortion, colour aberration) conditions (Drenkow et al., 2022). These factors can lead to *concept drift*.

Multiple works in the literature examine the effect of concept drift on automated benthic CV systems caused by changes in camera type, altitude, or time (Langenkämper et al., 2020; Zurowietz and Nattkemper, 2020; Wyatt et al., 2022). Further, a large volume of the work examined mention issues with, or the need to mitigate against, adverse environmental conditions (see Table S3). Despite this, no existing literature extensively explores how these conditions affect model performance, for example through an ablation study (see Section 5.2).

The problem of concept drift for ML and DL systems can be mitigated through the use of *active learning*. This technique allows a system to prioritise unlabelled data based on their value to the training process. By strategically selecting which data points should be labelled by humans and retraining the model, active learning enables the system to adapt to changes in data distribution over time whilst significantly reducing the required manual labelling effort (Ren et al., 2022). As a result, the use of active learning approaches, or other *human in the loop* techniques, are present in a range of the literature examined (Mahmood et al., 2016; Chen et al., 2021; Pavoni et al., 2022; Zhang et al., 2023).

Such techniques may help mitigate the effects of imagery containing adverse environmental or sensor conditions. If the system is intended to operate over large spatio-temporal scales, it may be beneficial to ensure as wide a range of operating conditions are observed during training as possible to help aid system generalisation. Depending on the active learning selection criteria, it may be practical to rank such images as having a higher value to the training process. However, if a system is intended to operate in a constrained environment where such adverse conditions are unlikely, then such measures may not be necessary — researchers may find it less expensive to train their model only to handle nominal conditions and perform labelling of organisms in adverse conditions manually.

3.3. Other notable issues

Further to the previous challenges, there are other notable issues with benthic imagery mentioned in the literature which can prove challenging for biodiversity monitoring systems. Uneven or insufficient illumination is the main issue discussed in the literature (e.g. Clement et al., 2005; Tan et al., 2014; Gonzalez-Cid et al., 2017; Huang et al., 2019; Chen et al., 2020). The performance of these systems can be significantly influenced by illumination, impacting colour perception and potentially causing the misclassification or oversight of RoIs.

Alongside illumination, issues may also arise due to high turbidity causing a reduction in water clarity (e.g. Beijbom et al., 2015; Rimavicius and Gelzinis, 2017; Lopez-Vazquez et al., 2020). This can dampen an RoI's colour or cause occlusion, degrading model performance. A handful of works also note further occlusion of objects thanks to sediment and marine snow present in the water column (e.g. Aguzzi et al., 2009; Gobi, 2010; Schoening et al., 2012; González-Rivero et al., 2020).

Whilst illumination and occlusion are common problems in automated benthic imagery analysis, these issues are not distinct to the domain and could be considered issues fundamental to the realm of CV. One unique problem however is the challenge of new-to-science organisms. As many of the animals in the ocean are currently thought to be undiscovered (Appeltans et al., 2012), this presents a unique challenge for supervised benthic CV systems – one which even most terrestrial systems do not have to account for. However, the current literature primarily treats benthic organisms as closed-set, where the number of classes is static, with very few works accounting for the potential detection of new-to-science or invasive organisms. If such an animal was observed, current systems may misidentify it as a previously learnt class. This has the potential to bias human observers, which may result in such new-to-science organisms remaining unknown.

More recent advancements, such as the work proposed by Liu et al. (2024) aim to flag new-to-science and invasive species through the use of feature extractors and *unsupervised learning* clustering methods. This

Table 4
A comparison of publicly available benthic imagery datasets capable of training CV systems.

Dataset	Collection method	Annotation type	Geographic location	Depth (m)	Taxonomic levels	Class types	Number of Images	Classes	Annotations	Example studies
Beijbom et al. (2015)	Photoquadrant	Point	Australia, French Polynesia, Kiribati, Taiwan, USA	1–17	Phylum–Genus	Anthropogenic Material, Biota, Substrate[a]	5,090	17	218,418	Beijbom et al. (2015)
BENTHOZ-2015 (Bewley et al., 2015)	AUV	Point	Australia	15–30	Phylum–Species	Biota, Substrate[b]	9,874	148	407,968	Bewley et al. (2015), Mahmood et al. (2016)
Brackish Dataset (Pedersen et al., 2019)	Stationary Camera	Bounding Box	Denmark	9	Phylum–Order	Biota	14,518	6	25,613	Pedersen et al. (2019), Fu et al. (2022)
EILAT (Beijbom et al., 2016)	Photoquadrant (Reflectance & Flurescence)	Point	Israel	3–15	Genus–Species	Biota, Substrate	212	10	42,400	Beijbom et al. (2016), (Gómez-Ríos et al., 2019)
FathomNet (Katija et al., 2022)[c]	AUV, ROV, Drop Camera, Stationary Camera	Bounding Box	Canada, USA, Taiwan	28–10,641	Class–Species	Anthropogenic Material, Biota, Substrate	84,454	2,244	175,873	Boulais et al. (2020), Belcher et al. (2023)
Marini et al. (2022b)	Stationary Camera	Bounding Box	Antarctica	20	Family–Species	Biota	775	13	23,881	(Marini et al., 2022a)
Moorea Labeled Corals (Moorea Coral Reef LTER and Edmunds, 2019)	Photoquadrant	Point	French Polynesia	Unknown	Order–Genus	Biota	2055	9	400,000	Beijbom et al. (2012)
Š.iaulys et al. (2021)	ROV, Drop Camera	Semantic Mask	Norway	3–65	Subphylum–Species	Biota	47	12	2,242	Buškus et al. (2021)
Underwater Open-Sea Farm Object Detection Dataset (UDD) (Liu et al., 2022)	AUV, Diver	Bounding Box	China	Unknown	Class–Family	Biota	2,227	3	15,022	Liu et al. (2022), Zhang et al. (2024a)

[a] Contains multiple dataset classes for coral and algae genera, though other biota such as sponges are grouped into a single class regardless of genus. Also contains classes for objects such as sand, bare space, transect hardware. An 'all other labels' class is present for objects not encapsulated by another defined dataset class.
[b] Dataset is classified in a hierarchical manner according to the Collaborative and Automated Tools for Analysis of Marine Imagery class hierarchy (Althaus et al., 2015).
[c] Also contains midwater images. Dataset size and taxa can increase as community-provided images are added. Information correct as of July 2022.

work is a promising step towards mitigating an inherent and unique problem in the space of automated benthic biodiversity monitoring, though more research is needed to improve the accuracy of such methods – Liu et al. (2024) currently achieves a top-1 accuracy of 43.43% when identifying unfamiliar species.

4. Available benthic imagery datasets

Whilst most of the examined literature makes use of closed-source data in their work, a handful provide their data open-source. To guide researchers who wish to train or benchmark automated benthic biodiversity monitoring systems, but do not have access to private data sources, this section details the currently available open-source datasets – an overview of which is provided in Table 4. Note that publications which have processed existing open-source data for their work but have subsequently not released this, or require a formal application to access the used data, have not been included.

Publicly available datasets cover a wide range of collection methods and geographies. The types of classes labelled within these datasets is also varied, with some focussing solely on biota whilst others also include anthropogenic material or substrate. These properties are often dependant on the use case of the dataset. For example, the Brackish dataset (Pedersen et al., 2019) focuses on coastal and estuarine areas, whilst the UDD dataset (Liu et al., 2022) focuses on farmed epibenthic taxa. Further, as previously explored in Section 3.1, due to inter-observer variability there is no certainty that the discussed datasets are perfectly labelled.

Alongside this, the taxonomic level biota are labelled to varies greatly between, and within, datasets. Whilst this is also influenced by use case, this range also highlights the difficulty in identifying organisms in the benthos. For some biota such as brittle star, it may be impossible to identify at a species level from imagery alone, instead requiring detailed morphological or molecular analysis (Stöhr et al., 2020). As a result, many of the datasets examined either group certain taxa into a single dataset class (e.g. Beijbom et al., 2015; Pedersen et al., 2019) or make use of a hierarchical labelling structure (e.g. Bewley et al., 2015; Katija et al., 2022).

As discussed in Section 3.1, annotating benthic imagery for use in developing automated systems is often prohibitively expensive, particularly when labelling RoIs in a granular manner. Publicly available benthic datasets reflect this, with most making use of either point-based (Beijbom et al., 2015; Bewley et al., 2015; Beijbom et al., 2016; Moorea Coral Reef LTER and Edmunds, 2019) or bounding box (Pedersen et al., 2019; Katija et al., 2022; Marini et al., 2022b; Liu et al., 2022) annotations. Only Š.iaulys et al. (2021) provides RoIs directly usable for semantic segmentation, though the number of images and annotations present here is comparatively lower than the other datasets analysed. None of the datasets analysed annotate to a level of granularity which allows for instance segmentation.

Of the datasets analysed, the FathomNet dataset (Katija et al., 2022) is unique in its aim to continuously grow through the inclusion of global, community-provided, data. As a result, the number of classes, and example images per class, within FathomNet may be updated at any time. This could potentially address the long-tailed distribution

challenge by offering data for rarer species, assuming distribution similarity and effective mitigation of challenges tied to combining marine data from diverse sources (Schoening et al., 2022). Whilst the ultimate aim of FathomNet is to be global in scope, it should be noted that the data in the initial dataset (defined as the seed data by Katija et al., 2022) is only from the USA, Canada, and Taiwan. Further, whilst the majority of datasets are made up of imagery from shallow benthic environments collected at depths between 1–30 m, only FathomNet contains data from deeper waters.

5. Future research

From the literature examined, it is clear that the application of CV to benthic imagery has shown great promise in its ability to reduce the data processing workload of marine ecologists and aid in the answering of key ecological questions. However, the field is also still in its infancy. As other areas of underwater imagery research advance, notably improvements in camera hardware and vehicles (see Whitt et al. (2020) and Wibisono et al. (2023) for further discussion), the volume and resolution of image data collected in the benthos will continue to increase. Such data will be collected over wider spatio-temporal scales, exhibit a wider variety of environmental and sensor conditions, and include a wider variety of organisms. As a result, the software used for curating this data will also need to advance at pace, else the data bottleneck will never clear. To provide direction to this advancement, the following section outlines potential avenues for future research into automated benthic biodiversity monitoring.

5.1. Human-centric techniques

The use of human-centric techniques such as human in the loop and active learning should be fully embraced. The use of these methods can help mitigate the risk of concept drift, a phenomenon which may be inevitable for biodiversity monitoring projects undertaken over large spatio-temporal scales. Furthermore, the use of such techniques can help mitigate the substantial expense and effort associated with labelling benthic data, which currently sees research groups curating only small subsets of the total data collected (Gutt et al., 2019) or focussing their analysis on particular taxonomic groups (Segelken-Voigt et al., 2016).

As highlighted in Section 3.2, only a small subset of existing work in this domain which is intended to operate over large spatio-temporal scales make use of human-centric techniques. Given the advantages provided by such methods however, future automated monitoring works should aim to utilise these approaches wherever possible, provided the CV system is intended to operate over large spatio-temporal scales or over imagery collected by multiple camera set-ups. In such cases, researchers should ensure that their training datasets include as wide a range of conditions as possible to help aid system generalisability. If the system is not intended to operate in such conditions however, then a standard training procedure may suffice, with out-of-distribution imagery manually labelled.

5.2. Ablation studies

Several of the works examined discuss or aim to mitigate environmental conditions which they believe may cause the performance of their developed systems to degrade (see Table S3). Despite this, these works often fall short of fully evaluating the effect these changes have on system performance, for example through ablation study. As a result, it is not yet clear how, and by how much, the performance of such systems would degrade should these mitigations not be present. Future works in this domain should aim to fully quantify the effect of such mitigations on their systems, as without this it is impossible to know whether built-in mitigations are necessary or working as intended.

5.3. Benchmark datasets

Recent years have seen the development of novel DL architectures designed specifically for use on benthic biodiversity data (Chen et al., 2020; Song et al., 2021; Liu et al., 2022; Yeh et al., 2022; Zhang et al., 2024b; Liu et al., 2024). Whilst these architectures all report high accuracies on the data used in the studies, it is currently extremely difficult to compare their relative performance without datasets agreed upon by the community with which to perform benchmarking studies. This is in contrast to other areas of CV research, where novel architectures will often be benchmarked against community-agreed open-source datasets to allow for fair comparison.

As highlighted by Fig. 7, whilst data from relatively few areas of the global benthos have been utilised to train the automated systems described by this review (largely due to the difficulty and cost associated with collecting or curating such data (Bell et al., 2022; Crosby et al., 2023)), enough labelled data currently exists to allow for the development of a sufficiently varied benchmark dataset. Once generated, such a dataset would allow researchers to gauge the adaptability of current and future automated benthic analysis approaches when exposed to varied data. For example, it is plausible that system performance may decline when exposed to data from different geographies due to distinctive environmental conditions or organism structures, particularly at lower taxonomic levels, although quantification currently remains challenging.

The introduction of benchmark benthic image datasets would vastly lower the barrier to entry for new researchers into a domain where advancements are limited to multi-disciplinary research groups made up of both marine ecologists and computing scientists, reducing the speed of progress in this area. These benchmarks may be task-specific or more general, though should cover diverse conditions to sufficiently evaluate system robustness.

One key consideration when generating any benchmark dataset is ensuring that it is not biased towards organisms with economic value. A significant proportion of work into automated benthic biodiversity monitoring focusses on the detection of taxa like lobsters (Aguzzi et al., 2011; Tan et al., 2015, 2018; Naseer et al., 2020), scallops (Fearn et al., 2007; Enomoto et al., 2009, 2010; Gobi, 2010; Dawkins et al., 2013; Kannappan and Tanner, 2013), mussels (Gu et al., 2023), or sea-urchins (Lv et al., 2019; Wang et al., 2021; Fu et al., 2022; Liu et al., 2021, 2022; Wang et al., 2023c). Such organisms have intrinsic value to humans as a food-source, and thus advancements in their detection may be motivated both by answering ecological questions and improvements in aquaculture efficiency. Indeed, one of the most used datasets by reviewed DL object detection works is the URPC dataset (see Section 2.2), designed to aid research into the automated capture of economically valuable organisms. Whilst data for such organisms is abundant, any community-agreed benchmark dataset must ensure as wide a variety of organisms is represented as possible, not just those favoured by humans.

Community driven datasets such as FathomNet should ensure that static releases are available for benchmarking purposes, alongside focussing collation efforts towards the closed-source studies highlighted in this review. If these datasets are not check-pointed in this way, it will be impossible to objectively evaluate systems using them. Inspiration could be taken from platforms such as iNaturalist,[2] where biodiversity data collection is community driven, but static check-pointed versions are released to facilitate the development and evaluation of CV models (Van Horn et al., 2018).

[2] iNaturalist: inaturalist.org

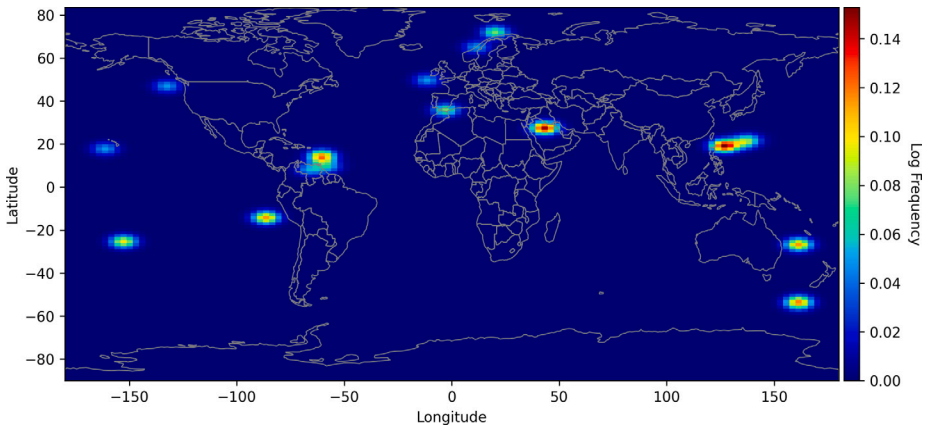


Fig. 7. Heatmap showing the geographic origin of image data used to train the automated benthic analysis systems reviewed in this study. Both closed and open-source works are included.

5.4. Vision transformers

One major issue typical in benthic imagery is the high levels of noise present, such as marine snow or other adverse environmental conditions, as well as large potential for organism aggregations. Such conditions pose major hurdles for current feature extraction methods, and may lead to highly degraded CV system performance. One avenue to explore in helping to reduce this performance drop when using DL is the use of *Vision Transformer* architectures (ViT, Dosovitskiy et al., 2021) over the current standard of CNNs. Such architectures have been shown to be more robust than CNNs when processing noisy images, leading to more generalised models (Maurício et al., 2023). As a result, ViTs have seen widespread adoption in other domains, such as medicine (Shamshad et al., 2023) and robotics (Gupta et al., 2024), with works in terrestrial ecology now also adopting them for use (e.g. Gong et al., 2023; Thakur et al., 2023). Given the challenges the benthos often poses to the feature extraction mechanisms of CNNs, the use of ViTs should be explored to help alleviate this. However, given the considerably larger datasets ViTs require to train compared to CNNs (Maurício et al., 2023), it may not be feasible to train such architectures from scratch for use in the benthos.

5.5. Foundation models

As highlighted in Table 4, the quantity of open-source benthic imagery available to train automated benthic biodiversity monitoring systems is, compared to other areas of CV, low. Given the high cost of data collection and required labelling effort (see Section 3.1), as well as the continued need for automated curation systems, it is reasonable to expect that only small proportions of closed-source datasets held by benthic research labs are currently labelled. If it is possible to achieve accurate curation results without the need for large quantities of pre-labelled data, this could see a considerable speed-up in the adoption of CV for benthic biodiversity monitoring.

The advent of *foundation models* may make this a reality. Such ViT-based models are trained on vast quantities of varied data sources, resulting in extremely generalised models capable of being applied to a wide variety of tasks, including those not represented during training (known as *zero-shot learning*). Examples of CV-focused foundation models include CLIP (Radford et al., 2021), DINOv2 (Oquab et al., 2023), and Segment Anything (SAM, Kirillov et al., 2023). Such models are seeing widespread use in the CV community as a first step in downstream task processing, including in ecology (Vyskočil and Picek, 2024; Gong et al., 2024; Zhang et al., 2024b). Recent research has also seen the development of foundation models with ecology focuses like BioCLIP (Stevens et al., 2024), and marine-focused foundation models such as MarineInst (Zheng et al., 2024).

Thanks to their ability to perform well on zero-shot tasks, foundation models should be explored as a component in future automated benthic biodiversity monitoring systems. Indeed, labelling tools are now beginning to include foundation models as built-in components, such as BIIGLE and SQUIDLE+ (see Table 3) which provide SAM. Use of such models could greatly reduce data curation times, for example by automatically segmenting objects in a benthic scene, requiring the researcher to only label the resultant masks, an example of this is provided by Doig et al. (2024). Future research could extend this by passing masks from SAM downstream to an automatic mask classification model. Further, the ability of foundational models to perform zero-shot detection may prove useful for the detection of new-to-science or invasive organisms.

5.6. Semi-supervised learning

So far, this review has focused primarily on the domain of supervised learning, or utilising solely the available labelled data to train automated benthic biodiversity monitoring systems. As previously mentioned however, it is often the case that benthic research labs have far more unlabelled data than labelled. In supervised learning scenarios, this large unlabelled set must be ignored during initial development, processed only once a well performing model has been created using the available labelled data.

Techniques such as *semi-supervised* learning allow researchers to make use of all available data, whether labelled or not, during initial model development. Here, the smaller labelled data is utilised to provide concrete ground truth examples to the model, as in supervised learning, whilst the larger unlabelled data is used to provide further understanding of the overall data distribution and structure (Van Engelen and Hoos, 2020). This can be achieved through methods such as pseudo-labelling, whereby a model is trained using the labelled data to generate predicted labels for the unlabelled data with some confidence. Data with high confidence labels are then added to the labelled data, and the model is re-trained in an iterative process. Whilst this may lead to labelling errors, pseudo-labelling often improves overall model performance relative to utilising only the original labelled data (Sohn et al., 2020), and has recently been utilised with success in biodiversity monitoring efforts in benthic and midwater environments (Sharma et al., 2024). Future research should thus aim to greatly expand the use of semi-supervised approaches in this domain.

5.7. Open-set recognition

Current research into automated benthic biodiversity monitoring systems focuses heavily on closed-set recognition, where the range of potential classes is considered static. However this does not reflect real

world conditions. Organisms initially thought to be rare may actually be abundant after further exploration of an area, and thus warrant inclusion in the training dataset. Further, in closed-set recognition the model may not be able to handle any new-to-science or recently invasive organisms, both key to biodiversity monitoring.

These issues can be mitigated by re-framing benthic biodiversity monitoring as an open-set problem, one where the total number of classes is not known at training time – a shift which has begun to occur in terrestrial studies (Lang et al., 2024; Lu et al., 2024). Advances here could be aided by work from outside of the ecology domain, such as those into anomaly detection (Hojjati et al., 2024) or out of distribution detection (Yang et al., 2024b), though classes in this context may be too difficult for such methods to accurately differentiate.

One promising avenue is the use of *self-supervised learning*, a subset of unsupervised learning that generates a latent feature space by creating predictive tasks within the data itself, allowing the model to learn representations that capture meaningful patterns and similarities between inputs (Rani et al., 2023). This latent space is typically generated by performing a proxy task, such as distinguishing the similarity between two images (Jaiswal et al., 2020). Where distinguishing labels are known prior, as would be the case with previously identified benthic taxa, these could be utilised to generate similar or dissimilar image pairs during training, without providing the labels explicitly to the model. When labels are not present, *contrastive learning* could be utilised. This technique makes use of aggressive data augmentations to create two different images from the same source, and using these to generate a latent space (Jaiswal et al., 2020). Self-supervised learning has been utilised with success to answer a variety of ecological questions using terrestrial data sources such as species identification (Pantazis et al., 2021; Yang et al., 2024a) and individual animal re-identification (Schneider et al., 2020). As such, it follows that the use of such methods in the benthos could help in key tasks such as the identification of new-to-science or invasive organisms.

The use of unsupervised clustering methods for the detection of invasive or new-to-science organisms as proposed by Liu et al. (2024) is an innovative first step in solving this problem. Future research should focus on the refinement of such techniques, both through improved feature extraction and clustering methods. Taking a self-supervised approach here may allow for the generation of more fine-grained species clusters, improving accuracy when detecting previously unseen species.

5.8. Multi-modal models

When performing taxonomic identification manually, ecologists often utilise contextual data alongside an organism's features. In the case of benthic ecology, this data may include the geographic location an image was collected at, as well as other auxiliary data obtained by the camera system such as conductivity, temperature, or depth (Sheehan et al., 2010). As benthic CV systems traditionally only ingest image data to make a prediction, this may result in incorrect classification of organisms which share a high degree of morphological similarity but are found in dissimilar habitats.

One possible avenue to help mitigate such confusion is the inclusion of contextual data through the use of multi-modal models. Capable of processing multiple different modalities of data – imagery, audio, text, etc. – such models have been shown to perform more accurately in fine-grained domains such as medicine (Stahlschmidt et al., 2022) and other ecological areas (Blair et al., 2022; Gu et al., 2024). Indeed, the use of multi-modal models has been shown to improve the classification accuracy of morphologically similar but habitat dissimilar organisms in terrestrial studies (Terry et al., 2020); it is reasonable to assume the same would be observed within the benthos.

5.9. Other innovative methods

Outside of the previously described future research directions, the expanded use of the following innovative methods within the benthos may also improve researchers' efficiency and ability to answer key ecological questions. First, researchers should explore the use of synthetic data to improve rare organism generalisation. Unlike augmented data which is generated using existing samples (see Section 3.1), synthetic data is generated using 3D graphics engines. The use of synthetic data has been shown to improve the accuracy of terrestrial models when detecting rare classes (Bondi et al., 2018; Beery et al., 2020), though its use is still untested in the benthos.

The problem of concept drift is well described in the literature, with multiple examined works making use of active learning techniques in an attempt to mitigate against this (see Section 3.2). However, Doig et al. (2023) have recently shown that such shifts can also be mitigated in benthic environments through the use of unsupervised domain adaption. Future work should aim to understand the effectiveness of both techniques, when one should be favoured over the other, and their integration into biodiversity monitoring tools.

A large volume of work examined in this review makes use of image patching techniques, classifying patches using image classification (see Section 2.1.1). Whilst the discussed methods may be useful for answering ecological questions such as coverage estimation, they may not be appropriate for other use-cases such as abundance estimation via object detection. Here, methods such as Slicing Aided Hyper Inference (SAHI, Akyon et al., 2022) may be best. This technique has seen wide ranging use in other applied-CV domains (Chaurasia and Patro, 2023; Muzammul et al., 2024; Gia et al., 2024), and future research should aim to explore the use of SAHI in the benthos. Such techniques may be useful in improving the detection accuracy of smaller taxa, as well as the efficiency of high-resolution benthic imagery patching compared to previously described task-specific methodologies.

Finally, benthic imagery collected in areas with high levels of biomass often contain dense organism aggregations, where animals overlap and occlude each other. When answering ecological questions relating to abundance, an accurate count of organisms, regardless of their condition, must be obtained. Typical methods for object counting like object detection often struggle with overlapping or occluded objects and may undercount (Chattopadhyay et al., 2017). Further, such methods often make use of techniques like non-max suppression to refine model predictions. When dense organism aggregations are observed, such methods may inadvertently merge correct localisations, reducing the final count. This issue may be mitigated through the use of DL techniques such as density estimation (Lempitsky and Zisserman, 2010). Rather than learning to localise each individual organism, the model instead predicts a density map over the whole image, with counts obtained via integration of said map. Density estimation has seen success in other domains where aggregations and occlusions are common, such as counting humans in crowds (Ma et al., 2019; Lin et al., 2021) or animal colonies in satellite imagery (Hoekendijk et al., 2021; Qian et al., 2023), and as such may prove useful for benthic abundance estimation.

6. Conclusion

The use of image data to monitor biodiversity in-situ for the purpose of answering key ecological questions is now commonplace within the domain of marine ecology. However, the curation of this data has traditionally been performed manually, a costly process which has led to a bottleneck whereby data is collected faster than it can be curated. As a result, there is a dire need to automate some or all of this process.

Thanks to advances in the domain of computer vision, recent years have seen an increase in the development of automated benthic biodiversity monitoring systems, aiming to reduce curation efforts and relieve the aforementioned data bottleneck. To help navigate and provide

context to these advances, this review surveys computer vision-based systems for analysing benthic biodiversity imagery. The key challenges presented by such data are discussed, alongside analysis of how the existing literature overcomes these. The current state of available open-source benthic biodiversity datasets is analysed, and potential avenues for future research in the area are outlined.

Whilst the current issues and future research directions discussed here are by no means exhaustive, it is hoped that this review will spur advancement in the domain of automated benthic biodiversity monitoring and ultimately help reduce the barrier to entry, both to marine ecologists wishing to make use of these systems in their own work and to computing scientists aiming to bring the latest advancements into the domain. If this can be achieved the data bottleneck can be cleared, and benthic biodiversity monitoring studies can make use of far greater volumes of data, curated in shorter time-frames than is currently achievable. This will allow researchers to better identify at risk benthic communities and implement mitigation strategies, regardless of where, when, or how their imagery is collected thanks to verifiable automated analysis systems.

Glossary

Active Learning A technique allowing for the labelling prioritisation of newly collected, unlabelled data based on estimated value to the training of a machine or deep learning model (Settles, 2009). Labelling priority is typically evaluated against some defined metric, with the data not utilised for learning until it is labelled.

Bounding Box A rectangle defining the location of an object within an image, typically provided alongside a class label. All pixels within the bounding box are denoted as forming part of the class.

Class A qualitative category which is to be predicted by an algorithm. Typically the predicted variable is made up of a fixed number of possible categories. For example, a benthic computer vision algorithm may be designed to detect the classes ‘starfish’ and ‘coral’.

Computer Vision (CV) A field of research which aims to automate the processing of visual data such as images. Can employ traditional IP methods, as well as more advanced machine learning or deep learning processes.

Concept Drift A phenomenon whereby the relationship between a system’s input and output features changes due to shifts in the input data distribution.

Contrastive Learning A learning approach where a model learns to differentiate between similar and dissimilar data points by bringing representations of similar pairs closer together and pushing apart representations of dissimilar pairs (Jaiswal et al., 2020).

Convolutional Neural Network (CNN) A type of deep neural network designed primarily for processing structured grid-like data, such as images, using convolutional layers. These layers apply filters to detect patterns, such as edges and textures, through a series of transformations. After each convolution, a non-linear activation function is applied to introduce non-linearity, allowing CNNs to learn complex representations.

Data Augmentation A technique whereby existing data is perturbed to generate new synthetic samples.

Deep Learning (DL) A subset of machine learning that makes use of complex neural networks consisting of multiple layers (deep neural networks).

Foundation Model A large-scale, pre-trained model that serves as a base for a wide range of cross-domain downstream tasks. Typically trained on extensive datasets, foundation models can be fine-tuned or adapted to specific applications with minimal task-specific data.

Human in the Loop The name given to a broad range of techniques where humans are involved in the decision-making loop, often in conjunction with automated algorithms.

Image Classification A computer vision task aiming to categorise whole images into classes.

Image Processing (IP) The task of manipulating or extracting features from visual data through pre-determined algorithmic steps.

Instance Segmentation A computer vision task which aims to provide pixel-level labelling over an image. A segmentation mask is provided on a per-object basis.

Machine Learning (ML) The development of statistical models capable of analysing data through some learned function via training, rather than a pre-determined algorithm. In the case of image-based machine learning, models are trained to extract relevant features from some input image.

Mask A binary matrix that indicates a pixel-level region of interest for an image. Masks can be provided on a per-class basis for tasks such as semantic segmentation, or a per-object basis for instance segmentation.

Midwater The section of a water body located vertically between the surface and the benthos.

Object Detection A computer vision task aiming to localise and classify multiple regions of interest within a single image.

Region of Interest (RoI) A specific area within an image identified as requiring attention.

Self-Supervised Learning A subset of *unsupervised learning* where the model generates its own data labels, allowing it to learn meaningful representations. Typically, a proxy task, like identifying similarities between image pairs, is used to capture relationships in a latent feature space (Rani et al., 2023).

Semantic Segmentation A computer vision task which aims to provide pixel-level labelling over an image. A segmentation mask is provided on a per-class basis.

Semi-Supervised Learning A learning approach that combines a small amount of labelled data with a larger set of unlabelled data to improve model performance. The labelled data provides ground truth, whilst the unlabelled data teaches broader data distribution and structure, enhancing generalisation with minimal labelled examples (Van Engelen and Hoos, 2020).

Supervised Learning The name given to machine or deep learning techniques which train using labelled data, such as image data alongside a textual label in the case or image classification or a list of bounding box locations for object detection.

Support Vector Machine A machine learning algorithm used for classification and regression tasks. Aims to find the optimal hyperplane that maximally separates different classes when plotted into a latent space.

Transfer Learning A technique allowing a machine or deep learning model to learn better data representations by "transferring information from a representation built using a data rich or clean modality to a data scarce or noisy modality" (Baltrusaitis et al., 2019).

Unsupervised Learning A learning approach that identifies patterns and structure within unlabelled data.

Vision Transformer (ViT) A deep learning model that applies transformer architectures, originally developed for Natural Language Processing, to image data. ViTs process images by dividing them into patches, treating each patch as a token, and using self-attention to capture relationships between patches, enabling effective computer vision tasks (Dosovitskiy et al., 2021).

Zero-Shot Learning A technique enabling a model to perform tasks not previously encountered during training by leveraging semantic information to generalise across tasks without specific labelled examples.

CRediT authorship contribution statement

Cameron Trotter: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Huw J. Griffiths:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology. **Rowan J. Whittle:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

CT, HJG and RJW are funded by the UKRI Future Leaders Fellowship MR/W01002X/1 'The past, present and future of unique cold-water benthic (sea floor) ecosystems in the Southern Ocean' awarded to RJW.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102989.

Data availability

CSV files used to generate Figs 1 and 7 are provided as supplementary material. Code used to generate these can be found at https://github.com/antarctica/surveying-the-deep.

References

Aguzzi, J., Costa, C., Fujiwara, Y., Iwase, R., Ramirez-Llorda, E., Menesatti, P., 2009. A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna. Sensors 9 (11), 8438–8455. http://dx.doi.org/10.3390/s91108438.

Aguzzi, J., Costa, C., Robert, K., Matabos, M., Antonucci, F., Juniper, S.K., Menesatti, P., 2011. Automated image analysis for the detection of benthic crustaceans and bacterial mat coverage using the VENUS undersea cabled network. Sensors 11 (11), 10534–10556. http://dx.doi.org/10.3390/s111110534.

Ahmed, S.F., Alam, M.S.B., Hassan, M., Rozbu, M.R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A.B.M., Gandomi, A.H., 2023. Deep learning modelling techniques: Current progress, applications, advantages, and challenges. Artif. Intell. Rev. 56 (11), 13521–13617. http://dx.doi.org/10.1007/s10462-023-10466-8.

Akyon, F.C., Onur Altinuc, S., Temizel, A., 2022. Slicing aided hyper inference and fine-tuning for small object detection. In: 2022 IEEE International Conference on Image Processing (ICIP). IEEE, Bordeaux, France, pp. 966–970. http://dx.doi.org/10.1109/ICIP46576.2022.9897990.

Alicia, R.-R., Laura, M.L.H., Piotr, K., Maciej, C., Piotr, B., 2023. Image analysis and benthic ecology: proceedings to analyze in situ long-term image series. Limnol. Oceanography: Methods 21 (4), 169–177. http://dx.doi.org/10.1002/lom3.10537.

Alonso, I., Cambra, A., Munoz, A., Treibitz, T., Murillo, A.C., 2017. Coral-segmentation: training dense labeling models with sparse ground truth. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). IEEE, Venice, Italy, pp. 2874–2882. http://dx.doi.org/10.1109/ICCVW.2017.339.

Alonso, I., Yuval, M., Eyal, G., Treibitz, T., Murillo, A.C., 2019. CoralSeg: learning coral segmentation from sparse annotations. J. Field Robotics 36 (8), 1456–1477. http://dx.doi.org/10.1002/rob.21915.

Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schönberg, C.H.L., Stuart-Smith, R., Barrett, N., Edgar, G., Colquhoun, J., Tran, M., Jordan, A., Rees, T., Gowlett-Holmes, K., 2015. A standardised vocabulary for identifying benthic Biota and substrata from underwater imagery: the CATAMI classification scheme. In: Hewitt, J. (Ed.), PLOS ONE 10 (10), e0141039. http://dx.doi.org/10.1371/journal.pone.0141039.

Amir, C., Oliver, T., Lamirand, M., Couch, C., 2023. Measuring coral vital rates using TagLab semi-automatic coral annotation and temporal linking across fixed sites: standard operating procedures and time savings estimate. NOAA Tech. Memo. NMFS PIFSC 39, 32. http://dx.doi.org/10.25923/S2YM-TN10.

Ani Brown Mary, N., Dejey, D., 2018. Classification of coral reef submarine images and videos using a novel z with tilted z local binary pattern (Z⊕TZLBP). Wirel. Pers. Commun. 98 (3), 2427–2459. http://dx.doi.org/10.1007/s11277-017-4981-x.

Ani Brown Mary, N., Dharma, D., 2019. A novel framework for real-time diseased coral reef image classification. Multimedia Tools Appl. 78 (9), 11387–11425. http://dx.doi.org/10.1007/s11042-018-6673-2.

Appeltans, W., Ahyong, S.T., Anderson, G., Angel, M.V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C.B., Brandão, S.N., Bray, R.A., Bruce, N.L., Cairns, S.D., Chan, T.-Y., Cheng, L., Collins, A.G., Cribb, T., Curini-Galletti, M., Dahdouh-Guebas, F., Davie, P.J., Dawson, M.N., De Clerck, O., Decock, W., De Grave, S., de Voogd, N.J., Domning, D.P., Emig, C.C., Erséus, C., Eschmeyer, W., Fauchald, K., Fautin, D.G., Feist, S.W., Fransen, C.H., Furuya, H., Garcia-Alvarez, O., Gerken, S., Gibson, D., Gittenberger, A., Gofas, S., Gómez-Daglio, L., Gordon, D.P., Guiry, M.D., Hernandez, F., Hoeksema, B.W., Hopcroft, R.R., Jaume, D., Kirk, P., Koedam, N., Koenemann, S., Kolb, J.B., Kristensen, R.M., Kroh, A., Lambert, G., Lazarus, D.B., Lemaitre, R., Longshaw, M., Lowry, J., Macpherson, E., Madin, L.P., Mah, C., Mapstone, G., McLaughlin, P.A., Mees, J., Meland, K., Messing, C.G., Mills, C.E., Molodtsova, T.N., Mooi, R., Neuhaus, B., Ng, P.K., Nielsen, C., Norenburg, J., Opresko, D.M., Osawa, M., Paulay, G., Perrin, W., Pilger, J.F., Poore, G.C., Pugh, P., Read, G.B., Reimer, J.D., Rius, M., Rocha, R.M., Saiz-Salinas, J.I., Scarabino, V., Schierwater, B., Schmidt-Rhaesa, A., Schnabel, K.E., Schotte, M., Schuchert, P., Schwabe, E., Segers, H., Self-Sullivan, C., Shenkar, N., Siegel, V., Sterrer, W., Stöhr, S., Swalla, B., Tasker, M.L., Thuesen, E.V., Timm, T., Todaro, M.A., Turon, X., Tyler, S., Uetz, P., van der Land, J., Vanhoorne, B., van Ofwegen, L.P., van Soest, R.W., Vanaverbeke, J., Walker-Smith, G., Walter, T.C., Warren, A., Williams, G.C., Wilson, S.P., Costello, M.J., 2012. The magnitude of global marine species diversity. Curr. Biol. 22 (23), 2189–2202. http://dx.doi.org/10.1016/j.cub.2012.09.036.

Baltrusaitis, T., Ahuja, C., Morency, L.-P., 2019. Multimodal machine learning: a survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. 41 (2), 423–443. http://dx.doi.org/10.1109/TPAMI.2018.2798607.

Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Meister, M., Joshi, N., Perona, P., 2020. Synthetic examples improve generalization for rare classes. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, Snowmass Village, CO, USA, pp. 852–862. http://dx.doi.org/10.1109/WACV45572.2020.9093570.

Beijbom, O., Edmunds, P.J., Kline, D.I., Mitchell, B.G., Kriegman, D., 2012. Automated annotation of coral reef survey images. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Providence, RI, pp. 1170–1177. http://dx.doi.org/10.1109/CVPR.2012.6247798.

Beijbom, O., Edmunds, P.J., Roelfsema, C., Smith, J., Kline, D.I., Neal, B.P., Dunlap, M.J., Moriarty, V., Fan, T.-Y., Tan, C.-J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B.G., Kriegman, D., 2015. Towards automated annotation of benthic survey images: variability of human experts and operational modes of automation. In: Chen, C.A. (Ed.), PLOS ONE 10 (7), e0130312. http://dx.doi.org/10.1371/journal.pone.0130312.

Beijbom, O., Treibitz, T., Kline, D.I., Eyal, G., Khen, A., Neal, B., Loya, Y., Mitchell, B.G., Kriegman, D., 2016. Improving automated annotation of benthic survey images using wide-band fluorescence. Sci. Rep. 6 (1), 23166. http://dx.doi.org/10.1038/srep23166.

Belcher, B.T., Bower, E.H., Burford, B., Celis, M.R., Fahimipour, A.K., Guevara, I.L., Katija, K., Khokhar, Z., Manjunath, A., Nelson, S., Olivetti, S., Orenstein, E., Saleh, M.H., Vaca, B., Valladares, S., Hein, S.A., Hein, A.M., 2023. Demystifying image-based machine learning: A practical guide to automated analysis of field imagery using modern machine learning tools. Front. Mar. Sci. 10, 1157370. http://dx.doi.org/10.3389/fmars.2023.1157370.

Bell, K.L.C., Chow, J.S., Hope, A., Quinzin, M.C., Cantner, K.A., Amon, D.J., Cramp, J.E., Rotjan, R.D., Kamalu, L., De Vos, A., Talma, S., Buglass, S., Wade, V., Filander, Z., Noyes, K., Lynch, M., Knight, A., Lourenço, N., Girguis, P.R., De Sousa, J.B., Blake, C., Kennedy, B.R.C., Noyes, T.J., McClain, C.R., 2022. Low-cost, deep-sea imaging and analysis tools for deep-sea exploration: a collaborative design study. Front. Mar. Sci. 9, 873700. http://dx.doi.org/10.3389/fmars.2022.873700.

Beuchel, F., Primicerio, R., Lønne, O.J., Gulliksen, B., Birkely, S.-R., 2010. Counting and measuring epibenthic organisms from digital photographs: A semiautomated approach. Limnol. Oceanography: Methods 8 (5), 229–240. http://dx.doi.org/10.4319/lom.2010.8.229.

Bewley, M.S., Douillard, B., Nourani-Vatani, N., Friedman, A., Pizarro, O., Williams, S.B., 2012. Automated species detection: an experimental approach to kelp detection from sea-floor AUV images. In: Proceedings of Australasian Conference on Robotics and Automation. 1, Australian Robotics and Automation Association (ARAA), Wellington, New Zealand.

Bewley, M., Friedman, A., Ferrari, R., Hill, N., Hovey, R., Barrett, N., Marzinelli, E.M., Pizarro, O., Figueira, W., Meyer, L., Babcock, R., Bellchambers, L., Byrne, M., Williams, S.B., 2015. Australian sea-floor survey data, with images and expert annotations. Sci. Data 2 (1), 150057. http://dx.doi.org/10.1038/sdata.2015.57.

Blair, J., Weiser, M.D., De Beurs, K., Kaspari, M., Siler, C., Marshall, K.E., 2022. Embracing imperfection: machine-assisted invertebrate classification in real-world datasets. Ecol. Inform. 72, 101896. http://dx.doi.org/10.1016/j.ecoinf.2022.101896.

Bondi, E., Dey, D., Kapoor, A., Piavis, J., Shah, S., Fang, F., Dilkina, B., Hannaford, R., Iyer, A., Joppa, L., Tambe, M., 2018. AirSim-w: a simulation environment for wildlife conservation with UAVs. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. ACM, Menlo Park and San Jose CA USA, pp. 1–12. http://dx.doi.org/10.1145/3209811.3209880.

Boulais, O.E., Woodward, B., Schlining, B., Lundsten, L., Barnard, K., Bell, C., Katija, K., 2020. FathomNet: an underwater image training database for ocean exploration and discovery. arXiv:2007.00114 [Cs.CV].

Boulent, J., Charry, B., Kennedy, M.M., Tissier, E., Fan, R., Marcoux, M., Watt, C.A., Gagné-Turcotte, A., 2023. Scaling whale monitoring using deep learning: A human-in-the-loop solution for analyzing aerial datasets. Front. Mar. Sci. 10, 1099479. http://dx.doi.org/10.3389/fmars.2023.1099479.

Buškus, K., Vaičiukynas, E., Verikas, A., Medelytė, S., Šiaulys, A., Šaškov, A., 2021. Automated quantification of brittle stars in seabed imagery using computer vision techniques. Sensors 21 (22), 7598. http://dx.doi.org/10.3390/s21227598.

Chattopadhyay, P., Vedantam, R., Selvaraju, R.R., Batra, D., Parikh, D., 2017. Counting everyday objects in everyday scenes. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, pp. 4428–4437. http://dx.doi.org/10.1109/CVPR.2017.471.

Chaurasia, D., Patro, B., 2023. Real-time detection of birds for farm surveillance using YOLOv7 and SAHI. In: 2023 3rd International Conference on Computing and Information Technology (ICCIT). IEEE, Tabuk, Saudi Arabia, pp. 442–450. http://dx.doi.org/10.1109/ICCIT58132.2023.10273929.

Chen, Q., Beijbom, O., Chan, S., Bouwmeester, J., Kriegman, D., 2021. A new deep learning engine for CoralNet. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). IEEE, Montreal, Canada, pp. 3686–3695. http://dx.doi.org/10.1109/ICCVW54120.2021.00412.

Chen, L., Liu, Z., Tong, L., Jiang, Z., Wang, S., Dong, J., Zhou, H., 2020. Underwater object detection using invert multi-class adaboost with deep learning. In: 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, Glasgow, United Kingdom, pp. 1–8. http://dx.doi.org/10.1109/IJCNN48605.2020.9207506.

Clark, H.P., Smith, A.G., McKay Fletcher, D., Larsson, A.I., Jaspars, M., De Clippele, L.H., 2024. New interactive machine learning tool for marine image analysis. R. Soc. Open Sci. 11 (5), 231678. http://dx.doi.org/10.1098/rsos.231678.

Clement, R., Dunbabin, M., Wyeth, G., 2005. Toward robust image detection of crown-of-thorns starfish for autonomous population monitoring. Proc. the 2005 Australas. Conf. Robot. Autom. 1–8.

Crosby, A., Orenstein, E.C., Poulton, S.E., Bell, K.L., Woodward, B., Ruhl, H., Katija, K., Forbes, A.G., 2023. Designing ocean vision AI: an investigation of community needs for imaging-based ocean conservation. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, pp. 1–16. http://dx.doi.org/10.1145/3544548.3580886.

Curry, R., Trotter, C., McGough, A.S., 2021. Application of deep learning to camera trap data for ecologists in planning / engineering – can captivity imagery train a model which generalises to the wild? In: 2021 IEEE International Conference on Big Data (Big Data). IEEE, Orlando, FL, USA, pp. 4011–4020. http://dx.doi.org/10.1109/BigData52589.2021.9671661.

Cuvelier, D., Zurowietz, M., Nattkemper, T.W., 2024. Deep learning–assisted biodiversity assessment in deep-sea benthic megafauna communities: A case study in the context of polymetallic nodule mining. Front. Mar. Sci. 11, 1366078. http://dx.doi.org/10.3389/fmars.2024.1366078.

Dawkins, M., Sherrill, L., Fieldhouse, K., Hoogs, A., Richards, B., Zhang, D., Prasad, L., Williams, K., Lauffenburger, N., Wang, G., 2017. An open-source platform for underwater image and video analytics. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, Santa Rosa, CA, USA, pp. 898–906. http://dx.doi.org/10.1109/WACV.2017.105.

Dawkins, M., Stewart, C., Gallager, S., York, A., 2013. Automatic scallop detection in benthic environments. In: 2013 IEEE Workshop on Applications of Computer Vision (WACV). IEEE, Clearwater Beach, FL, USA, pp. 160–167. http://dx.doi.org/10.1109/WACV.2013.6475014.

Deo, R., John, C.M., Zhang, C., Whitton, K., Salles, T., Webster, J.M., Chandra, R., 2024. Deepdive: leveraging pre-trained deep learning for deep-sea rov Biota identification in the great barrier reef. Sci. Data 11 (1), 957. http://dx.doi.org/10.1038/s41597-024-03766-3.

Di Gesu, V., Isgro, F., Tegolo, D., Trucco, E., 2003. Finding essential features for tracking starfish in a video sequence. In: 12th International Conference on Image Analysis and Processing, 2003. Proceedings.. pp. 504–509. http://dx.doi.org/10.1109/ICIAP.2003.1234100.

Doig, H., Pizarro, O., Monk, J., Williams, S., 2024. Detecting endangered marine species in autonomous underwater vehicle imagery using point annotations and few-shot learning. arXiv arXiv:2406.01932.

Doig, H., Pizarro, O., Williams, S.B., 2023. Improved benthic classification using resolution scaling and SymmNet unsupervised domain adaptation. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, London, United Kingdom, pp. 3124–3130. http://dx.doi.org/10.1109/ICRA48891.2023.10160255.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: transformers for image recognition at scale. arXiv arXiv:2010.11929.

Drenkow, N., Sani, N., Shpitser, I., Unberath, M., 2022. A systematic review of robustness in deep learning for computer vision: mind the gap?. arXiv:2112.00639 [Cs.CV].

Du Preez, C., Curtis, J.M.R., Clarke, M.E., 2016. The structure and distribution of benthic communities on a shallow seamount (Cobb Seamount, Northeast Pacific Ocean). In: Patterson, H.M. (Ed.), PLOS ONE 11 (10), e0165513. http://dx.doi.org/10.1371/journal.pone.0165513.

Durden, J.M., Hosking, B., Bett, B.J., Cline, D., Ruhl, H.A., 2021. Automated classification of fauna in seabed photographs: the impact of training and validation dataset size, with considerations for the class imbalance. Prog. Oceanogr. 196, 102612. http://dx.doi.org/10.1016/j.pocean.2021.102612.

Enomoto, K., Toda, M., Kuwahara, Y., 2009. Scallop detection from sand-seabed images for Fishery investigation. In: 2009 2nd International Congress on Image and Signal Processing. IEEE, Tianjin, China, pp. 1–5. http://dx.doi.org/10.1109/CISP.2009.5305438.

Enomoto, K., Toda, M., Kuwahara, Y., 2010. Extraction method of scallop area in gravel seabed images for Fishery investigation. IEICE Trans. Inf. Syst. E93-D (7), 1754–1760. http://dx.doi.org/10.1587/transinf.E93.D.1754.

Fearn, R., Williams, R., Cameron-Jones, M., Harrington, J., Semmens, J., 2007. Automated intelligent abundance analysis of scallop survey video footage. In: Orgun, M.A., Thornton, J. (Eds.), AI 2007: Advances in Artificial Intelligence. vol. 4830, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 549–558. http://dx.doi.org/10.1007/978-3-540-76928-6_56.

Fu, X., Liu, Y., Liu, Y., 2022. A case study of utilizing YOLOT based quantitative detection algorithm for marine benthos. Ecol. Inform. 70, 101603. http://dx.doi.org/10.1016/j.ecoinf.2022.101603.

Gia, B.T., Bui Cong Khanh, T., Trong, H.H., Tran Doan, T., Do, T., Le, D.-D., Ngo, T.D., 2024. Enhancing road object detection in fisheye cameras: an effective framework integrating SAHI and hybrid inference. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Seattle, WA, USA, pp. 7227–7235. http://dx.doi.org/10.1109/CVPRW63382.2024.00718.

Gleason, A.C.R., Reid, R.P., Voss, K.J., 2007. Automated classification of underwater multispectral imagery for coral reef monitoring. In: OCEANS 2007. pp. 1–8. http://dx.doi.org/10.1109/OCEANS.2007.4449394.

Gobi, A.F., 2010. Towards generalized benthic species recognition and quantification using computer vision. In: OCEANS'10 IEEE SYDNEY. pp. 1–6. http://dx.doi.org/10.1109/OCEANSSYD.2010.5603995.

Gomes-Pereira, J.N., Auger, V., Beisiegel, K., Benjamin, R., Bergmann, M., Bowden, D., Buhl-Mortensen, P., De Leo, F.C., Dionísio, G., Durden, J.M., Edwards, L., Friedman, A., Greinert, J., Jacobsen-Stout, N., Lerner, S., Leslie, M., Nattkemper, T.W., Sameoto, J.A., Schoening, T., Schouten, R., Seager, J., Singh, H., Soubigou, O., Tojeira, I., Van Den Beld, I., Dias, F., Tempera, F., Santos, R.S., 2016. Current and future trends in marine image annotation software. Prog. Oceanogr. 149, 106–120. http://dx.doi.org/10.1016/j.pocean.2016.07.005.

Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A., Krawczyk, B., Herrera, F., 2019. Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. Expert Syst. Appl. 118, 315–328. http://dx.doi.org/10.1016/j.eswa.2018.10.010.

Gong, H., Luo, T., Ni, L., Li, J., Guo, J., Liu, T., Feng, R., Mu, Y., Hu, T., Sun, Y., Guo, Y., Li, S., 2023. Research on facial recognition of sika deer based on vision transformer. Ecol. Inform. 78, 102334. http://dx.doi.org/10.1016/j.ecoinf.2023.102334.

Gong, Z., Shan, K., Zhao, Y., Peng, D., Liang, H., Zou, B., 2024. Real-time algal bloom detection technique: a zero-shot approach using DINOv2 and HSV filtering. In: Proceedings of the 2024 12th International Conference on Communications and Broadband Networking. ACM, Nyingchi China, pp. 102–107. http://dx.doi.org/10.1145/3688636.3688644.

Gonzalez-Cid, Y., Burguera, A., Bonin-Font, F., Matamoros, A., 2017. Machine learning and deep learning strategies to identify posidonia meadows in underwater images. In: OCEANS 2017 - Aberdeen. pp. 1–5. http://dx.doi.org/10.1109/OCEANSE.2017.8084991.

González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, D.E.P., Ganase, A., Gonzalez-Marrero, Y., Herrera-Reveles, A., Kennedy, E.V., Kim, C.J.S., Lopez-Marcano, S., Markey, K., Neal, B.P., Osborne, K., Reyes-Nivia, C., Sampayo, E.M., Stolberg, K., Taylor, A., Vercelloni, J., Wyatt, M., Hoegh-Guldberg, O., 2020. Monitoring of coral reefs using artificial intelligence: a feasible and cost-effective approach. Remote Sens. 12 (3), 489. http://dx.doi.org/10.3390/rs12030489.

Goodwin, M., Halvorsen, K.T., Jiao, L., Knausgård, K.M., Martin, A.H., Moyano, M., Oomen, R.A., Rasmussen, J.H., Sørdalen, T.K., Thorbjørnsen, S.H., 2022. Unlocking the potential of deep learning for marine ecology: Overview, applications, and outlook. In: Demer, D. (Ed.), ICES J. Mar. Sci. 79 (2), 319–336. http://dx.doi.org/10.1093/icesjms/fsab255.

Gu, Y.-F., Hu, J., Williams, G.A., 2023. A comparative study on CNN-based semantic segmentation of intertidal mussel beds. Ecol. Inform. 75, 102116. http://dx.doi.org/10.1016/j.ecoinf.2023.102116.

Gu, Y., Hu, Z., Zhao, Y., Liao, J., Zhang, W., 2024. MFGTN: A multi-modal fast gated transformer for identifying single trawl marine fishing vessel. Ocean Eng. 303, 117711. http://dx.doi.org/10.1016/j.oceaneng.2024.117711.

Gupta, A., Aggarwal, M., Khetan, D., Mishra, K., 2024. Vision-based transformers survey for robotics. http://dx.doi.org/10.36227/techrxiv.172470864.43128162/v1.

Gutt, J., Arndt, J., Kraan, C., Dorschel, B., Schröder, M., Bracher, A., Piepenburg, D., 2019. Benthic communities and their drivers: A spatial analysis off the antarctic peninsula. Limnol. Oceanogr. 64 (6), 2341–2357. http://dx.doi.org/10.1002/lno.11187.

Hao, Y., Liu, Y., Chen, Y., Han, L., Peng, J., Tang, S., Chen, G., Wu, Z., Chen, Z., Lai, B., 2022. Eiseg: an efficient interactive segmentation tool based on PaddlePaddle. arXiv arXiv:2210.08788.

Harrison, D., De Leo, F.C., Gallin, W.J., Mir, F., Marini, S., Leys, S.P., 2021. Machine learning applications of convolutional neural networks and unet architecture to predict and classify demosponge behavior. Water 13 (18), 2512. http://dx.doi.org/10.3390/w13182512.

Hoekendijk, J.P.A., Kellenberger, B., Aarts, G., Brasseur, S., Poiesz, S.S.H., Tuia, D., 2021. Counting using deep learning regression gives value to ecological surveys. Sci. Rep. 11 (1), 23209. http://dx.doi.org/10.1038/s41598-021-02387-9.

Hojjati, H., Ho, T.K.K., Armanfard, N., 2024. Self-supervised anomaly detection in computer vision and beyond: A survey and outlook. Neural Netw. 172, 106106. http://dx.doi.org/10.1016/j.neunet.2024.106106.

Huang, B., Chen, G., Zhang, H., Hou, G., Radenkovic, M., 2023. Instant deep sea debris detection for maneuverable underwater machines to build sustainable ocean using deep neural network. Sci. Total Environ. 878, 162826. http://dx.doi.org/10.1016/j.scitotenv.2023.162826.

Huang, H., Zhou, H., Yang, X., Zhang, L., Qi, L., Zang, A.-Y., 2019. Faster R-CNN for marine organisms detection and recognition using data augmentation. Neurocomputing 337, 372–384. http://dx.doi.org/10.1016/j.neucom.2019.01.084.

Jackett, C., Althaus, F., Maguire, K., Farazi, M., Scoulding, B., Untiedt, C., Ryan, T., Shanks, P., Brodie, P., Williams, A., 2023. A benthic substrate classification method for seabed images using deep learning: application to management of deep-sea coral reefs. J. Appl. Ecol. 60 (7), 1254–1273. http://dx.doi.org/10.1111/1365-2664.14408.

Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F., 2020. A survey on contrastive self-supervised learning. Technologies 9 (1), 2. http://dx.doi.org/10.3390/technologies9010002.

Johnson-Roberson, M., Kumar, S., Pizarro, O., Willams, S., 2006. Stereoscopic imaging for coral segmentation and classification. In: OCEANS 2006. pp. 1–6. http://dx.doi.org/10.1109/OCEANS.2006.306876.

Johnson-Roberson, M., Kumar, S., Willams, S., 2006. Segmentation and classification of coral for oceanographic surveys: a semi-supervised machine learning approach. In: OCEANS 2006 - Asia Pacific. pp. 1–6. http://dx.doi.org/10.1109/OCEANSAP.2006.4393835.

Kannappan, P., Tanner, H.G., 2013. Automated detection of scallops in their natural environment. In: 21st Mediterranean Conference on Control and Automation. IEEE, Platanias, Chania - Crete, Greece, pp. 1350–1355. http://dx.doi.org/10.1109/MED.2013.6608895.

Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., Boulais, O., Cromwell, M., Butler, E., Woodward, B., Bell, K.L.C., 2022. FathomNet: A global image database for enabling artificial intelligence in the ocean. Sci. Rep. 12 (1), 15914. http://dx.doi.org/10.1038/s41598-022-19939-2.

Kellenberger, B., Tuia, D., Morris, D., 2020. AIDE: accelerating image-based ecological surveys with interactive machine learning. In: Graham, L. (Ed.), Methods Ecol. Evol. 11 (12), 1716–1727. http://dx.doi.org/10.1111/2041-210X.13489.

King, A., Bhandarkar, S.M., Hopkinson, B.M., 2018. A comparison of deep learning methods for semantic segmentation of coral reef survey images. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Salt Lake City, UT, pp. 1475–14758. http://dx.doi.org/10.1109/CVPRW.2018.00188.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything. arXiv:2304.02643 [Cs.CV].

Kumar, P., Luo, S., Shaukat, K., 2023. A comprehensive review of deep learning approaches for animal detection on video data. Int. J. Adv. Comput. Sci. Appl. 14 (11), http://dx.doi.org/10.14569/IJACSA.2023.01411144.

Kwon, Y.S., Kang, H., Pyo, J., 2024. Estimation of aquatic ecosystem health using deep neural network with nonlinear data mapping. Ecol. Inform. 81, 102588. http://dx.doi.org/10.1016/j.ecoinf.2024.102588.

Lalli, C., Parsons, T.R., 1997. Biological Oceanography: An Introduction. Elsevier.

Lam-Gordillo, O., Baring, R., Dittmann, S., 2020. Ecosystem functioning and functional approaches on marine macrobenthic fauna: A research synthesis towards a global consensus. Ecol. Indic. 115, 106379. http://dx.doi.org/10.1016/j.ecolind.2020.106379.

Lang, N., Snaebjarnarson, V., Cole, E., Aodha, O.M., Igel, C., Belongie, S., 2024. From coarse to fine-grained open-set recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17804–17814.

Langenkämper, D., Van Kevelaer, R., Purser, A., Nattkemper, T.W., 2020. Gear-induced concept drift in marine images and its effect on deep learning classification. Front. Mar. Sci. 7, 506. http://dx.doi.org/10.3389/fmars.2020.00506.

Langenkämper, D., Zurowietz, M., Schoening, T., Nattkemper, T.W., 2017. BIIGLE 2.0 - browsing and annotating large marine image collections. Front. Mar. Sci. 4, 83. http://dx.doi.org/10.3389/fmars.2017.00083.

Lempitsky, V., Zisserman, A., 2010. Learning to count objects in images. In: Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., Culotta, A. (Eds.), Advances in Neural Information Processing Systems. 23, Curran Associates, Inc..

Lin, H., Hong, X., Ma, Z., Wei, X., Qiu, Y., Wang, Y., Gong, Y., 2021. Direct measure matching for crowd counting. http://dx.doi.org/10.48550/ARXIV.2107.01558.

Liu, C., Li, H., Wang, S., Zhu, M., Wang, D., Fan, X., Wang, Z., 2021. A dataset and benchmark of underwater object detection for robot picking. In: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 1–6. http://dx.doi.org/10.1109/ICMEW53276.2021.9455997, arXiv:2106.05681.

Liu, A., Liu, Y., Xu, K., Zhao, F., Zhou, Y., Li, X., 2024. DeepSeaNet: a bio-detection network enabling species identification in the deep sea imagery. IEEE Trans. Geosci. Remote Sens. 62, 1–13. http://dx.doi.org/10.1109/TGRS.2024.3359350.

Liu, C., Wang, Z., Wang, S., Tang, T., Tao, Y., Yang, C., Li, H., Liu, X., Fan, X., 2022. A new dataset, Poisson GAN and AquaNet for underwater object grabbing. IEEE Trans. Circuits Syst. Video Technol. 32 (5), 2831–2844. http://dx.doi.org/10.1109/TCSVT.2021.3100059.

Lopez-Vazquez, V., Lopez-Guede, J., Marini, S., Fanelli, E., Johnsen, E., Aguzzi, J., 2020. Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. Sensors 20 (3), 726. http://dx.doi.org/10.3390/s20030726.

Lu, Y., Li, J., Zhao, Z., Zhang, Y., Tong, Y., Teng, B., Liu, N., Shaw, J.J., Bai, M., 2024. Accelerating the discovery of biodiversity by detecting “new” species based on machine learning method. http://dx.doi.org/10.21203/rs.3.rs-3832815/v1.

Lüdtke, A., Jerosch, K., Herzog, O., Schlüter, M., 2012. Development of a machine learning technique for automatic analysis of seafloor image data: case example, pogonophora coverage at mud volcanoes. Comput. Geosci. 39, 120–128. http://dx.doi.org/10.1016/j.cageo.2011.06.020.

Lumini, A., Nanni, L., Maguolo, G., 2023. Deep learning for plankton and coral classification. Appl. Comput. Informatics 19 (3/4), 265–283. http://dx.doi.org/10.1016/j.aci.2019.11.004.

Lütjens, M., Sternberg, H., 2021. Deep learning based detection, segmentation and counting of benthic megafauna in unconstrained underwater environments. IFAC-Pap. 54 (16), 76–82. http://dx.doi.org/10.1016/j.ifacol.2021.10.076.

Lv, X., Wang, A., Liu, Q., Sun, J., Zhang, S., 2019. Proposal-refined weakly supervised object detection in underwater images. In: Zhao, Y., Barnes, N., Chen, B., Westermann, R., Kong, X., Lin, C. (Eds.), Image and Graphics. vol. 11901, Springer International Publishing, Cham, pp. 418–428. http://dx.doi.org/10.1007/978-3-030-34120-6_34.

Ma, Z., Wei, X., Hong, X., Gong, Y., 2019. Bayesian loss for crowd count estimation with point supervision. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Seoul, Korea (South), pp. 6141–6150. http://dx.doi.org/10.1109/ICCV.2019.00624.

Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., Fisher, R., 2016. Automatic annotation of coral reefs using deep learning. In: OCEANS 2016 MTS/IEEE Monterey. IEEE, Monterey, CA, USA, pp. 1–5. http://dx.doi.org/10.1109/OCEANS.2016.7761105.

Manderson, T., Li, J., Dudek, N., Meger, D., Dudek, G., 2017. Robotic coral reef health assessment using automated image analysis: robotic coral reef health assessment using automated image analysis. J. Field Robotics 34 (1), 170–187. http://dx.doi.org/10.1002/rob.21698.

Marburg, A., Bigham, K., 2016. Deep learning for benthic fauna identification. In: OCEANS 2016 MTS/IEEE Monterey. pp. 1–5. http://dx.doi.org/10.1109/OCEANS.2016.7761146.

Marcos, M.S.A., David, L., Peñaflor, E., Ticzon, V., Soriano, M., 2008. Automated benthic counting of living and non-living components in ngedarrak reef, palau via subsurface underwater video. Environ. Monit. Assess. 145 (1–3), 177–184. http://dx.doi.org/10.1007/s10661-007-0027-2.

Marini, S., Bonofiglio, F., Corgnati, L.P., Bordone, A., Schiaparelli, S., Peirano, A., 2022a. Long-term automated visual monitoring of antarctic benthic fauna. Methods Ecol. Evol. 13 (8), 1746–1764. http://dx.doi.org/10.1111/2041-210X.13898.

Marini, S., Bonofiglio, F., Corgnati, L.P., Bordone, A., Schiaparelli, S., Peirano, A., 2022b. Long-term high resolution image dataset of Antarctic Coastal benthic fauna. Sci. Data 9 (1), 750. http://dx.doi.org/10.1038/s41597-022-01865-7.

Maurício, J., Domingues, I., Bernardino, J., 2023. Comparing vision transformers and convolutional neural networks for image classification: a literature review. Appl. Sci. 13 (9), 5521. http://dx.doi.org/10.3390/app13095521.

Medelytė, S., Šiaulys, A., Daunys, D., Włodarska-Kowalczuk, M., Węsławski, J.M., Olenin, S., 2022. Application of underwater imagery for the description of upper sublittoral benthic communities in glaciated and ice-free arctic fjords. Polar Biol. 45 (12), 1655–1671. http://dx.doi.org/10.1007/s00300-022-03096-3.

Miao, Z., Liu, Z., Gaynor, K.M., Palmer, M.S., Yu, S.X., Getz, W.M., 2021. Iterative human and automated identification of wildlife images. Nat. Mach. Intell. 3 (10), 885–895. http://dx.doi.org/10.1038/s42256-021-00393-0.

Miller, S.D., Dubel, A.K., Adam, T.C., Cook, D.T., Holbrook, S.J., Schmitt, R.J., Rassweiler, A., 2023. Using machine learning to achieve simultaneous, georeferenced surveys of fish and benthic communities on shallow coral reefs. Limnol. Oceanography: Methods lom3.10557. http://dx.doi.org/10.1002/lom3.10557.

Mizuno, K., Terayama, K., Hagino, S., Tabeta, S., Sakamoto, S., Ogawa, T., Sugimoto, K., Fukami, H., 2020. An efficient coral survey method based on a large-scale 3-D structure model obtained by speedy sea scanner and u-net segmentation. Sci. Rep. 10 (1), 12416. http://dx.doi.org/10.1038/s41598-020-69400-5.

Mohamed, H., Nadaoka, K., Nakamura, T., 2022. Automatic semantic segmentation of benthic habitats using images from towed underwater camera in a complex shallow water environment. Remote Sens. 14 (8), 1818. http://dx.doi.org/10.3390/rs14081818.

Monari, D., Larkin, J., Machado, P., Bird, J.J., Ihianle, I.K., Yahaya, S.W., Tash, F.F., Hasan, M.M., Lotfi, A., 2023. UDEEP: edge-based computer vision for in-situ underwater crayfish and plastic detection. arXiv arXiv:2401.06157.

Moniruzzaman, Md., Islam, S.M.S., Bennamoun, M., Lavery, P., 2017. Deep learning on underwater marine object detection: a survey. In: Blanc-Talon, J., Penne, R., Philips, W., Popescu, D., Scheunders, P. (Eds.), Advanced Concepts for Intelligent Vision Systems. vol. 10617, Springer International Publishing, Cham, pp. 150–160. http://dx.doi.org/10.1007/978-3-319-70353-4_13.

Moorea Coral Reef LTER, Edmunds, P., 2019. MCR LTER: Coral Reef: Computer Vision: Moorea Labeled Corals. Environmental Data Initiative, http://dx.doi.org/10.6073/PASTA/88DDE0E68AB5232A470389F4BEDD1892.

Muzammul, M., Algarni, A., Ghadi, Y.Y., Assam, M., 2024. Enhancing UAV aerial image analysis: integrating advanced sahi techniques with real-time detection models on the VisDrone dataset. IEEE Access 12, 21621–21633. http://dx.doi.org/10.1109/ACCESS.2024.3363413.

Naseer, A., Baro, E.N., Khan, S.D., Gordillo, Y.V., 2020. Automatic detection of nephrops norvegicus burrows in underwater images using deep learning. In: 2020 Global Conference on Wireless and Optical Technologies (GCWOT). IEEE, Malaga, Spain, pp. 1–6. http://dx.doi.org/10.1109/GCWOT49901.2020.9391590.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. 115 (25), http://dx.doi.org/10.1073/pnas.1719367115.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2023. DINOv2: learning robust visual features without supervision. http://dx.doi.org/10.48550/ARXIV.2304.07193.

Osterloff, J., Nilssen, I., Järnegren, J., Van Engeland, T., Buhl-Mortensen, P., Nattkemper, T.W., 2019. Computer vision enables short- and long-term analysis of lophelia pertusa polyp behaviour and colour from an underwater observatory. Sci. Rep. 9 (1), 6578. http://dx.doi.org/10.1038/s41598-019-41275-1.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. http://dx.doi.org/10.1109/TKDE.2009.191.

Pantazis, O., Brostow, G.J., Jones, K.E., Aodha, O.M., 2021. Focus on the positives: self-supervised learning for biodiversity monitoring. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, pp. 10563–10572. http://dx.doi.org/10.1109/ICCV48922.2021.01041.

Pavoni, G., Corsini, M., Callieri, M., Fiameni, G., Edwards, C., Cignoni, P., 2020. On improving the training of models for the semantic segmentation of benthic communities from orthographic imagery. Remote Sens. 12 (18), 3106. http://dx.doi.org/10.3390/rs12183106.

Pavoni, G., Corsini, M., Pedersen, N., Petrovic, V., Cignoni, P., 2021. Challenges in the deep learning-based semantic segmentation of benthic communities from ortho-images. Appl. Geomatics 13 (1), 131–146. http://dx.doi.org/10.1007/s12518-020-00331-6.

Pavoni, G., Corsini, M., Ponchio, F., Muntoni, A., Edwards, C., Pedersen, N., Sandin, S., Cignoni, P., 2022. TagLab: AI-assisted annotation for the fast and accurate semantic segmentation of coral reef orthoimages. J. Field Robotics 39 (3), 246–262. http://dx.doi.org/10.1002/rob.22049.

Pedersen, M., Haurum, J.B., Gade, R., Moeslund, T.B., 2019. Detection of marine animals in a new underwater dataset with varying visibility. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, pp. 18–26.

Piechaud, N., Hunt, C., Culverhouse, P., Foster, N., Howell, K., 2019. Automated identification of benthic epifauna with computer vision. Mar. Ecol. Prog. Ser. 615, 15–30. http://dx.doi.org/10.3354/meps12925.

Purser, A., Dreutter, S., Griffiths, H., Hehemann, L., Jerosch, K., Nordhausen, A., Piepenburg, D., Richter, C., Schröder, H., Dorschel, B., 2021. Seabed video and still images from the northern weddell sea and the western flanks of the Powell basin. Earth Syst. Sci. Data 13 (2), 609–615. http://dx.doi.org/10.5194/essd-13-609-2021.

Qian, Y., Humphries, G.R.W., Trathan, P.N., Lowther, A., Donovan, C.R., 2023. Counting animals in aerial images with a density map estimation model. Ecol. Evol. 13 (4), e9903. http://dx.doi.org/10.1002/ece3.9903.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning transferable visual models from natural language supervision. http://dx.doi.org/10.48550/ARXIV.2103.00020.

Rani, V., Nabi, S.T., Kumar, M., Mittal, A., Kumar, K., 2023. Self-supervised learning: a succinct review. Arch. Comput. Methods Eng. 30 (4), 2761–2775. http://dx.doi.org/10.1007/s11831-023-09884-2.

Raphael, A., Dubinsky, Z., Iluz, D., Netanyahu, N.S., 2020. Neural network recognition of marine benthos and corals. Diversity 12 (1), 29. http://dx.doi.org/10.3390/d12010029.

Rees, H.L. (Ed.), 2009. Guidelines for the study of the epibenthos of subtidal environments. ICES Techniques in Marine Environmental Sciences, (42), International Council for the Exploration of the Sea, Copenhagen.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-CNN: towards real-time object detection with region proposal networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. 28, Curran Associates, Inc..

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B.B., Chen, X., Wang, X., 2022. A survey of deep active learning. ACM Comput. Surv. 54 (9), 1–40. http://dx.doi.org/10.1145/3472291.

Rimavicius, T., Gelzinis, A., 2017. A comparison of the deep learning methods for solving seafloor image classification task. In: Damaˇsevičius, R., Mikaˇsytė, V. (Eds.), Information and Software Technologies. vol. 756, Springer International Publishing, Cham, pp. 442–453. http://dx.doi.org/10.1007/978-3-319-67642-5_37.

Roelfsema, C., Kovacs, E.M., Markey, K., Vercelloni, J., Rodriguez-Ramirez, A., Lopez-Marcano, S., Gonzalez-Rivero, M., Hoegh-Guldberg, O., Phinn, S.R., 2021. Benthic and coral reef community field data for heron reef, southern great barrier reef, Australia, 2002–2018. Sci. Data 8 (1), 84. http://dx.doi.org/10.1038/s41597-021-00871-5.

Saleh, A., Sheaves, M., Rahimi Azghadi, M., 2022. Computer vision and deep learning for fish classification in underwater habitats: A survey. Fish Fish. 23 (4), 977–999. http://dx.doi.org/10.1111/faf.12666.

Schneider, S., Taylor, G.W., Kremer, S.C., 2020. Similarity learning networks for animal individual re-identification - beyond the capabilities of a human observer. In: 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW). IEEE, Snowmass, CO, USA, pp. 44–52. http://dx.doi.org/10.1109/WACVW50321.2020.9096925.

Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., Nattkemper, T.W., 2012. Semi-automated image analysis for the assessment of megafaunal densities at the arctic deep-sea observatory HAUSGARTEN. In: Slomp, C.P. (Ed.), PLoS ONE 7 (6), e38179. http://dx.doi.org/10.1371/journal.pone.0038179.

Schoening, T., Durden, J.M., Faber, C., Felden, J., Heger, K., Hoving, H.-J.T., Kiko, R., Köser, K., Krämmer, C., Kwasnitschka, T., Möller, K.O., Nakath, D., Naß, A., Nattkemper, T.W., Purser, A., Zurowietz, M., 2022. Making marine image data FAIR. Sci. Data 9 (1), 414. http://dx.doi.org/10.1038/s41597-022-01491-3.

Schoening, T., Kuhn, T., Jones, D.O., Simon-Lledo, E., Nattkemper, T.W., 2016. Fully automated image segmentation for benthic resource assessment of poly-metallic nodules. Methods Ocean. 15–16, 78–89. http://dx.doi.org/10.1016/j.mio.2016.04.002.

Schoening, T., Kuhn, T., Nattkemper, T.W., 2014. Seabed classification using a bag-of-prototypes feature representation. In: 2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery. pp. 17–24. http://dx.doi.org/10.1109/CVAUI.2014.9.

Segelken-Voigt, A., Bracher, A., Dorschel, B., Gutt, J., Huneke, W., Link, H., Piepenburg, D., 2016. Spatial distribution patterns of ascidians (ascidiacea: tunicata) on the continental shelves off the northern antarctic peninsula. Polar Biology 39 (5), 863–879. http://dx.doi.org/10.1007/s00300-016-1909-y.

Settles, B., 2009. Active learning literature survey. In: Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H., 2023. Transformers in medical imaging: A survey. Med. Image Anal. 88, 102802. http://dx.doi.org/10.1016/j.media.2023.102802.

Sharma, T., Cline, D.E., Edgington, D., 2024. Making use of unlabeled data: comparing strategies for marine animal detection in long-tailed datasets using self-supervised and semi-supervised pre-training. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Seattle, WA, USA, pp. 1224–1233. http://dx.doi.org/10.1109/CVPRW63382.2024.00129.

Shashidhara, B.M., Scott, M., Marburg, A., 2020. Instance segmentation of benthic scale worms at a hydrothermal site. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, Snowmass Village, CO, USA, pp. 1303–1312. http://dx.doi.org/10.1109/WACV45572.2020.9093574.

Sheehan, E.V., Stevens, T.F., Attrill, M.J., 2010. A quantitative, non-destructive methodology for habitat characterisation and benthic monitoring at offshore renewable energy developments. In: Ropert-Coudert, Y. (Ed.), PLoS ONE 5 (12), e14461. http://dx.doi.org/10.1371/journal.pone.0014461.

Shepley, A., Falzon, G., Meek, P., Kwan, P., 2021. Automated location invariant animal detection in camera trap images using publicly available data sources. Ecol. Evol. 11 (9), 4494–4506. http://dx.doi.org/10.1002/ece3.7344.

Shihavuddin, A., Gracias, N., Garcia, R., Gleason, A., Gintert, B., 2013. Image-based coral reef classification and thematic mapping. Remote Sens. 5 (4), 1809–1841. http://dx.doi.org/10.3390/rs5041809.

Smith, D., Dunbabin, M., 2007. Automated counting of the Northern Pacific Sea Star in the Derwent using shape recognition. In: 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007). pp. 500–507. http://dx.doi.org/10.1109/DICTA.2007.4426838.

Smith, A.G., Han, E., Petersen, J., Olsen, N.A.F., Giese, C., Athmann, M., Dresbøll, D.B., Thorup-Kristensen, K., 2022. Rootpainter: Deep learning segmentation of biological images with corrective annotation. New Phytol. 236 (2), 774–791. http://dx.doi.org/10.1111/nph.18387.

Sohn, K., Berthelot, D., Li, C.-L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C., 2020. FixMatch: simplifying semi-supervised learning with consistency and confidence.

Sokolova, M.N., 2000. Feeding and Trophic Structure of the Deep-Sea Macrobenthos. Science, Enfield, N.H.

Song, H., Mehdi, S.R., Zhang, Y., Shentu, Y., Wan, Q., Wang, W., Raza, K., Huang, H., 2021. Development of coral investigation system based on semantic segmentation of single-channel images. Sensors 21 (5), 1848. http://dx.doi.org/10.3390/s21051848.

Stahlschmidt, S.R., Ulfenborg, B., Synnergren, J., 2022. Multimodal deep learning for biomedical data fusion: A review. Brief. Bioinform. 23 (2), bbab569. http://dx.doi.org/10.1093/bib/bbab569.

Stevens, S., Wu, J., Thompson, M.J., Campolongo, E.G., Song, C.H., Carlyn, D.E., Dong, L., Dahdul, W.M., Stewart, C., Berger-Wolf, T., Chao, W.-L., Su, Y., 2024. Bioclip: a vision foundation model for the tree of life. Proc. the IEEE/ CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) 19412–1942.

Stöhr, S., Weber, A.A.-T., Boissin, E., Chenuil, A., 2020. Resolving the ophioderma longicauda (echinodermata: ophiuroidea) cryptic species complex: Five sisters, three of them new. Eur. J. Taxon. (600), http://dx.doi.org/10.5852/ejt.2020.600.

Tan, C.S., Lau, P.Y., Correia, P.L., Campos, A., 2018. Automatic analysis of deep-water remotely operated vehicle footage for estimation of Norway lobster abundance. Front. Inf. Technol. Electron. Eng. 19 (8), 1042–1055. http://dx.doi.org/10.1631/FITEE.1700720.

Tan, C.S., Lau, P.Y., Correia, P., Fonseca, P., Campos, A., 2015. A tracking scheme for Norway lobster and burrow abundance estimation in underwater video sequences. Proc. Int. Work. Adv. Image Technol..

Tan, C.S., Lau, P.Y., Low, T.J., Fonseca, P., Campos, A., 2014. Detection of marine species on underwater video images. Proc. Int. Work. Adv. Image Technol. 6–8.

Terry, J.C.D., Roy, H.E., August, T.A., 2020. Thinking like a naturalist: enhancing computer vision of citizen science images by harnessing contextual data. In: Altwegg, R. (Ed.), Methods Ecol. Evol. 11 (2), 303–315. http://dx.doi.org/10.1111/2041-210X.13335.

Thakur, P.S., Chaturvedi, S., Khanna, P., Sheorey, T., Ojha, A., 2023. Vision transformer meets convolutional neural network for plant disease classification. Ecol. Inform. 77, 102245. http://dx.doi.org/10.1016/j.ecoinf.2023.102245.

Tian, M., Yi, S., Li, H., Li, S., Zhang, X., Shi, J., Yan, J., Wang, X., 2018. Eliminating background-bias for robust person re-identification. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, UT, pp. 5794–5803. http://dx.doi.org/10.1109/CVPR.2018.00607.

Trotter, C., Wright, N., Stephen McGough, A., Sharpe, M., Cheney, B., Civil, M.A., Tyson Moore, R., Allen, J., Berggren, P., 2022. Towards automatic cetacean photo-identification: a framework for fine-grain, few-shot learning in marine ecology. In: 2022 IEEE International Conference on Big Data (Big Data). IEEE, Osaka, Japan, pp. 1942–1949. http://dx.doi.org/10.1109/BigData55660.2022.10020942.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109 (2), 373–440. http://dx.doi.org/10.1007/s10994-019-05855-6.

Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeirotis, P., Perona, P., Belongie, S., 2015. Building a bird recognition app and large scale dataset with citizen scientists: the fine print in fine-grained dataset collection. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston, MA, USA, pp. 595–604. http://dx.doi.org/10.1109/CVPR.2015.7298658.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S., 2018. The inaturalist species classification and detection dataset. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, UT, pp. 8769–8778. http://dx.doi.org/10.1109/CVPR.2018.00914.

Van Horn, G., Perona, P., 2017. The devil is in the tails: fine-grained classification in the wild. http://dx.doi.org/10.48550/ARXIV.1709.01450.

Šaškov, A., Dahlgren, T.G., Rzhanov, Y., Schläppy, M.-L., 2015. Comparison of manual and semi-automatic underwater imagery analyses for monitoring of benthic hard-bottom organisms at offshore renewable energy installations. Hydrobiologia 756 (1), 139–153. http://dx.doi.org/10.1007/s10750-014-2072-5.

Š.iaulys, A., Vaičiukynas, E., Medelytė, S., Olenin, S., Šaškov, A., Buškus, K., Verikas, A., 2021. A fully-annotated imagery dataset of sublittoral benthic species in svalbard, arctic. Data Brief 35, 106823. http://dx.doi.org/10.1016/j.dib.2021.106823.

Vyskočil, J., Picek, L., 2024. Towards zero-shot camera trap image categorization. http://dx.doi.org/10.48550/ARXIV.2410.12769.

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M., 2023a. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Vancouver, BC, Canada, pp. 7464–7475. http://dx.doi.org/10.1109/CVPR52729.2023.00721.

Wang, N., Chen, T., Liu, S., Wang, R., Karimi, H.R., Lin, Y., 2023b. Deep learning-based visual detection of marine organisms: A survey. Neurocomputing 532, 1–32. http://dx.doi.org/10.1016/j.neucom.2023.02.018.

Wang, H., Sun, S., Bai, X., Wang, J., Ren, P., 2023c. A reinforcement learning paradigm of configuring visual enhancement for object detection in underwater scenes. IEEE J. Ocean. Eng. 48 (2), 443–461. http://dx.doi.org/10.1109/JOE.2022.3226202.

Wang, H., Sun, S., Wu, X., Li, L., Zhang, H., Li, M., Ren, P., 2021. A YOLOv5 baseline for underwater object detection. In: OCEANS 2021: San Diego – Porto. pp. 1–4. http://dx.doi.org/10.23919/OCEANS44145.2021.9705896.

Whitt, C., Pearlman, J., Polagye, B., Caimi, F., Muller-Karger, F., Copping, A., Spence, H., Madhusudhana, S., Kirkwood, W., Grosjean, L., Fiaz, B.M., Singh, S., Singh, S., Manalang, D., Gupta, A.S., Maguer, A., Buck, J.J.H., Marouchos, A., Atmanand, M.A., Venkatesan, R., Narayanaswamy, V., Testor, P., Douglas, E., De Halleux, S., Khalsa, S.J., 2020. Future vision for autonomous ocean observations. Front. Mar. Sci. 7, 697. http://dx.doi.org/10.3389/fmars.2020.00697.

Wibisono, A., Piran, M.J., Song, H.-K., Lee, B.M., 2023. A survey on unmanned underwater vehicles: challenges, enabling technologies, and future research directions. Sensors 23 (17), 7321. http://dx.doi.org/10.3390/s23177321.

Williams, I.D., Couch, C.S., Beijbom, O., Oliver, T.A., Vargas-Angel, B., Schumacher, B.D., Brainard, R.E., 2019. Leveraging automated image analysis tools to transform our capacity to assess status and trends of coral reefs. Front. Mar. Sci. 6, 222. http://dx.doi.org/10.3389/fmars.2019.00222.

Wyatt, M., Radford, B., Callow, N., Bennamoun, M., Hickey, S., 2022. Using ensemble methods to improve the robustness of deep learning for image classification in marine environments. Methods Ecol. Evol. 13 (6), 1317–1328. http://dx.doi.org/10.1111/2041-210X.13841.

Yang, F., Shen, N., Xu, F., 2024a. Automatic bird species recognition from images with feature enhancement and contrastive learning. Appl. Sci. 14 (10), 4278. http://dx.doi.org/10.3390/app14104278.

Yang, J., Zhou, K., Li, Y., Liu, Z., 2024b. Generalized out-of-distribution detection: a survey. Int. J. Comput. Vis. http://dx.doi.org/10.1007/s11263-024-02117-4.

Yeh, C.-H., Lin, C.-H., Kang, L.-W., Huang, C.-H., Lin, M.-H., Chang, C.-Y., Wang, C.-C., 2022. Lightweight deep neural network for joint learning of underwater object detection and color conversion. IEEE Trans. Neural Netw. Learn. Syst. 33 (11), 6129–6143. http://dx.doi.org/10.1109/TNNLS.2021.3072414.

Zhang, L., Fan, J., Qiu, Y., Jiang, Z., Hu, Q., Xing, B., Xu, J., 2024a. Marine zoobenthos recognition algorithm based on improved lightweight YOLOv5. Ecol. Inform. 80, 102467. http://dx.doi.org/10.1016/j.ecoinf.2024.102467.

Zhang, Z., Kaveti, P., Singh, H., Powell, A., Fruh, E., Clarke, M.E., 2023. An iterative labeling method for annotating marine life imagery. Front. Mar. Sci. 10, 1094190. http://dx.doi.org/10.3389/fmars.2023.1094190.

Zhang, P., Yan, T., Liu, Y., Lu, H., 2024b. Fantastic animals and where to find them: segment any marine animal with dual SAM. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2578–2587.

Zheng, Z., Chen, Y., Zeng, H., Vu, T.-A., Hua, B.-S., Yeung, S.-K., 2024. MarineInst: a foundation model for marine image analysis with instance visual description. In: Leonardis, A.s., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (Eds.), Computer Vision – ECCV 2024. vol. 15060, Springer Nature Switzerland, Cham, pp. 239–257. http://dx.doi.org/10.1007/978-3-031-72627-9_14.

Zhou, Z., Fu, G.-Y., Fang, Y., Yuan, Y., Shen, H.-B., Wang, C.-S., Xu, X.-W., Zhou, P., Pan, X., 2023. Echoai: A deep-learning based model for classification of echinoderms in global oceans. Front. Mar. Sci. 10, 1147690. http://dx.doi.org/10.3389/fmars.2023.1147690.

Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H.A., Nattkemper, T.W., 2018. MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration. In: Sarder, P. (Ed.), PLOS ONE 13 (11), e0207498. http://dx.doi.org/10.1371/journal.pone.0207498.

Zurowietz, M., Nattkemper, T.W., 2020. Unsupervised knowledge transfer for object detection in marine environmental monitoring and exploration. IEEE Access 8, 143558–143568. http://dx.doi.org/10.1109/ACCESS.2020.3014441.