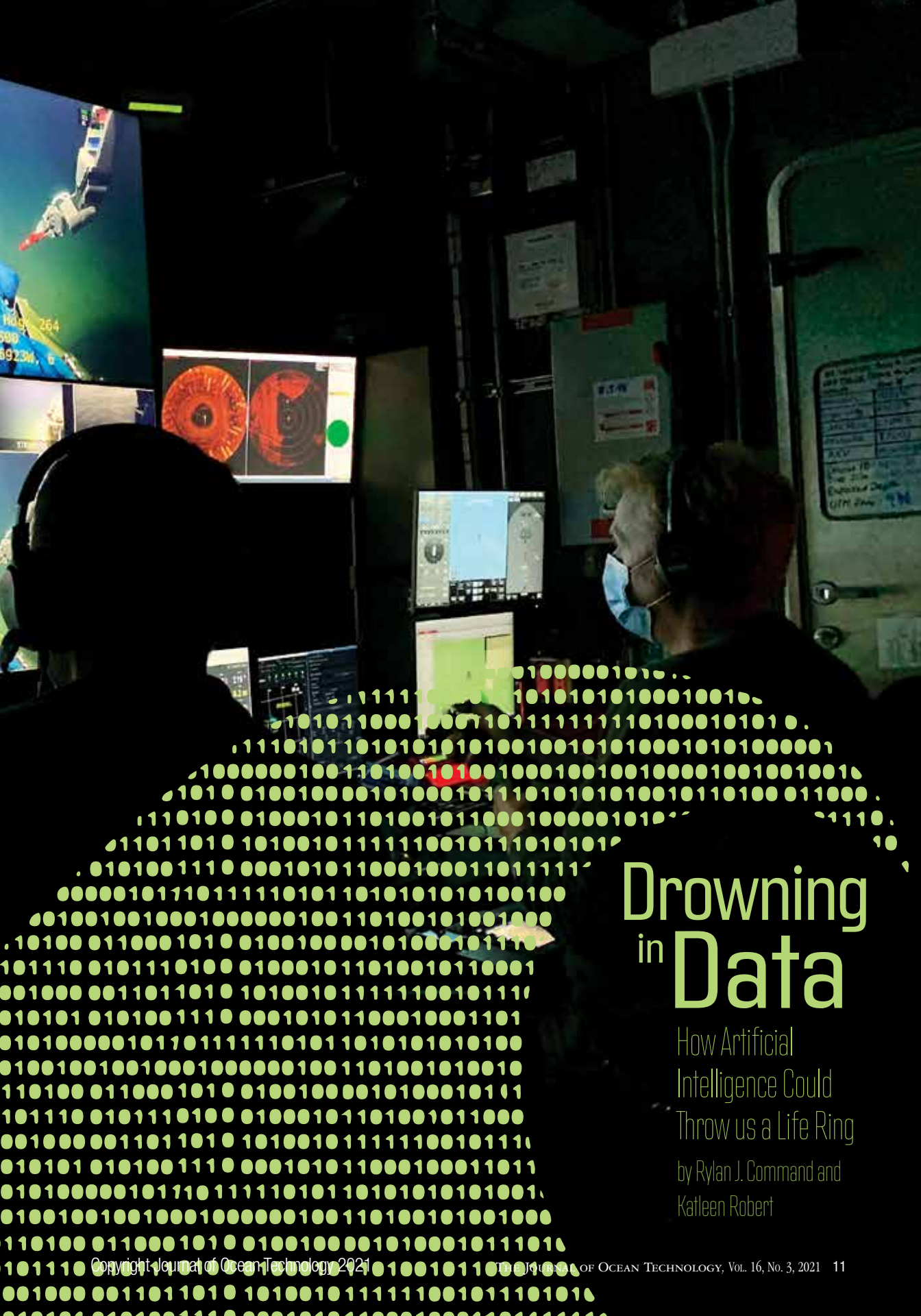

Drowning in Data

How Artificial Intelligence Could Throw us a Life Ring

by Rylan J. Command and Katleen Robert

Introduction

The amount of daily data collected by ocean observing systems is massive. The National Oceanic and Atmospheric Administration (NOAA) in the United States collects around 20 terabytes of ocean data every day, and projects that its archives will store over 250 petabytes (PB) of ocean data by 2030. To put that in perspective, 200 PB is the estimated volume of all written material that has ever been printed. Ever. In addition to the sheer volume, ocean data comes in many different flavours covering the biological, chemical, and physical nature of the ocean. The instrument types needed to collect this information are numerous, including temperature loggers, oxygen sensors, fluorometers, turbidity monitors, acoustic Doppler current profilers, hydrophones, satellites, and many more. All of these data types present their own challenges for interpreting and understanding oceanography from surface to seafloor. Free, open-source software has been developed to assist in processing oceanographic data to elucidate spatial and temporal patterns in the ocean. However, this wealth of collected environmental data is only the tip of the iceberg.

Advances in camera and image capture technologies and infrastructure for long-term deployments have led to a revolution in underwater data collection. From opportunistic time-lapse imagery to near real-time, high-definition video, our ability to collect high quality time-series of the seafloor has never been greater. However, this revolution has generated a looming question: who is going to watch it all? There is so much biological information contained in a single image – what species are present? How many species are there? What are they doing? What sort of habitat do they live in? How do they interact with their habitat? Furthermore, video and imagery data files are large, requiring massive storage capacity, which is energy intensive and expensive.

As the saying goes, a picture is worth a thousand words. To give a real-world example, a seven-year oxygen sensor dataset collected at one observation/second takes up around 14 GB when saved as a series of comma-separated values. In contrast, a single five-minute, high-resolution video clip (fast becoming the industry standard in seafloor observing systems) is around 180 MB. Seven years of these videos collected hourly is about 11,000 GB – or 785 times larger than the oxygen text files. How many words are those pictures worth?

The Need for Automated Video Analysis

Because each video or image contains so much information, it is incredibly time-consuming and labour intensive to process. Video and imagery collected from seafloor observatories are commonly used to identify the organisms present in a particular area. Marine plants and animals provide numerous ecosystem services from which humans benefit, such as producing food, filtering water, recycling nutrients, and producing the oxygen we breathe. Data collected on the diversity of marine life helps inform fisheries management, marine spatial planning, ecosystem health assessments, and fosters a connection with an otherwise unseen environment. Analyzing these data often requires expert knowledge or training to reliably identify the species of interest. Manual identification of organisms in video is time-consuming and represents a large bottleneck.

During co-author Command's master's degree studies at the Marine Institute of Memorial University of Newfoundland and Labrador, he had the opportunity to work with Ocean Networks Canada's (ONC) seafloor observatories off the coast of British Columbia, and the Marine Institute's newly installed cabled observatory in Holyrood, Newfoundland. At both sites, he studied temporal trends in the abundance of large seafloor organisms (Figure 1). As an example, 2,228 videos (~389 GB of data) required three months of work to count a single species of sea urchin. To put that in perspective, ONC alone collects roughly 10 GB of underwater videos daily.

Figure 1: Images from Ocean Networks Canada's cabled observatories co-author Command used for his master's degree research. Left: The pink urchin *Strongylocentrotus fragilis* at the Barkley Canyon Upper Slope node of the NEPTUNE cabled observatory off Vancouver Island, British Columbia. Right: The northern sunstar, *Solaster endeca*, and the sea cucumber, most likely *Psolus phantapus*, at the Holyrood Bay Underwater Network in Conception Bay, Newfoundland and Labrador. Distance between the lasers in the left image is 10 cm.

Advances in artificial intelligence are already making computer vision tasks faster and more efficient, and more accurate than ever. Pre-trained models, open access software and code, large libraries of labelled data, and large communities are working together to share research and push the cutting edge of artificial intelligence. From particle counts and object tracking, to automated classification of organisms using complex algorithms modelled on the human brain, combinations of machine learning and computer vision approaches are rapidly improving our ability to process video and imagery data.

To Supervise or not to Supervise?

Two main categories of machine learning are used to automate image analysis – supervised and unsupervised learning – which lends themselves to different tasks. Supervised learning requires the model to be trained on a labelled dataset to learn, for example, which image belongs to which category (i.e., fish versus crab). Once the model has been trained, it is put to the test on a batch of previously unseen images to see how well it can recognize and correctly classify each image (i.e., how often does the model say fish = fish versus fish = crab). To accomplish this, an expert must first create the labelled dataset by assigning categories to pixels within a series of images collected from an observatory, for example. By putting in the time upfront, a researcher can train a supervised learning model to recognize species, and (hopefully) correctly classify unseen images into the correct categories.

One major limitation of supervised learning is the time-consuming nature of creating the labelled dataset in the first place. The minimum amount of data needed for a high-performing model depends on many factors, but a general rule of thumb is to make sure the amount of training data is at least 10 times the number of features used for training. For underwater images, there are often dozens of features extracted from an image relating to size, shape, texture, and colour that can inform classification, so the minimum training data size can be on the order of 1,000-10,000 images. Additionally, the model can only identify images and make predictions for the labels (i.e., species) for which it was trained. Expanding the labelled dataset to include a new species and retraining the model is time-consuming.

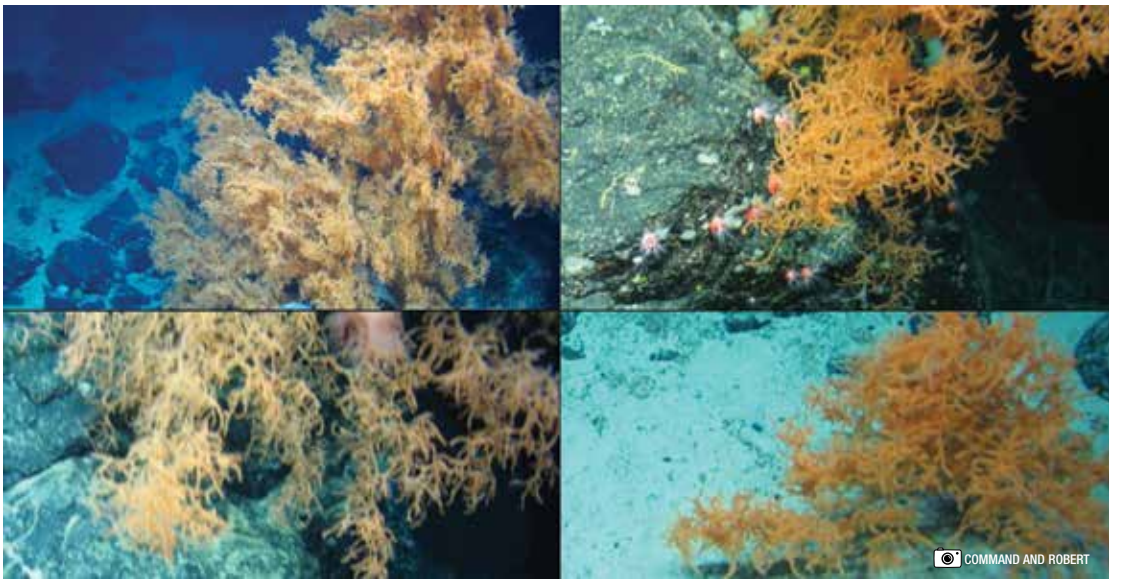

Figure 2: Images of black coral *Leiopathes sp.* taken from the ROV Holland I on the 2018 Tectonic Ocean Spreading at the Charlie-Gibbs Fracture Zone (TOSCA) expedition to the Charlie-Gibbs Fracture Zone on the Mid-Atlantic Ridge. Each of these images likely contains different species of the genus Leiopathes; however, distinguishing them in images may be impossible, requiring genetic techniques.

Where supervised learning requires a labelled dataset to learn to assign labels, unsupervised learning looks for patterns in an unlabelled dataset and clusters similar items together based on some criteria. The major challenge with unsupervised learning is defining the criteria with which clusters should be separated. This may be done by extracting features from an image, such as colour, shape, and texture, and separating images into groups based on how (dis)similar they are. Automated feature extraction using computer vision software (i.e., opencv in C++ or Python) is becoming more accessible and more efficient, and unsupervised learning is particularly useful for pattern recognition tasks.

There are several drawbacks with unsupervised learning for images. First, there is the computational complexity of working with very large, unlabelled datasets that are required to produce the desired outcome. Second, clustering algorithms are often difficult to interpret as there is sometimes limited or zero capacity to determine the criteria by which a particular clustering decision was made. Finally, clusters created with unsupervised learning may be inaccurate and often require extensive validation by experts or with available "ground truth" data. This is especially problematic when different species have very similar features, making it difficult to accurately distinguish between two different clusters of organisms that may have different ecological niches (i.e., consume different foods, are active at different times of day, etc.) but similar features in an image (Figure 2). This last point becomes especially challenging in underwater imagery, where variable lighting conditions may reduce image quality and make it more difficult to differentiate features among similar species.

The Power of the (Artificial) Brain

Artificial neural networks (ANNs; Figure 3) are one promising avenue of machine learning. These algorithms are based on the architecture of the human brain, and consist of layers of nodes (or neurons) connected by weighting functions (or synapses). An input is fed into the network, and passes through a series of weighting functions in each layer that determines if the information should be passed to the next layer, or stop. The output of each layer is determined by an activation function, which is influenced by the relative importance

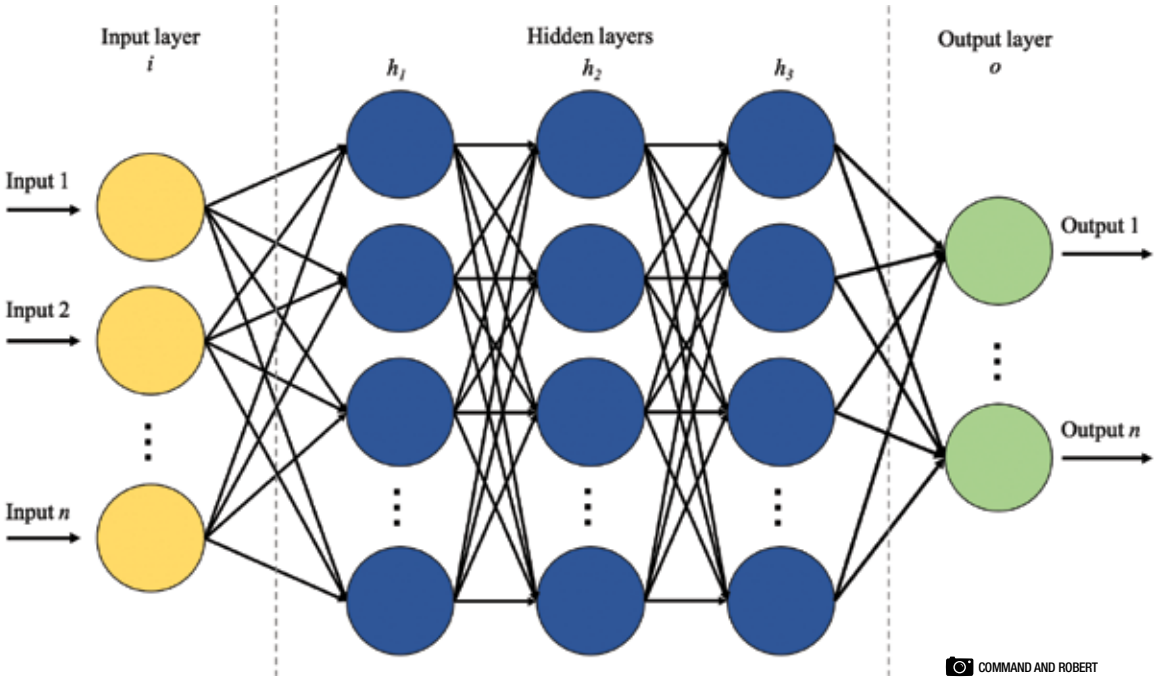

Figure 3: Structure of an artificial neural network (ANN). Input variables are fed into the network at the input layer and a calculation is performed. A series of weights and mathematical functions are applied, and errors are calculated to determine if the output, or signal, of a layer continues to the next layer. The weights and decisions made are not directly observable from the input and output of the network, so these are referred to as "hidden layers." The output layer takes as inputs the result of the hidden layers and calculates the output of the network.

of each input, and becomes the input for the next layer. The network continues in this way to identify features based on thresholding criteria at each layer and produces an output – either a classification probability or a value. The weights for each node are determined through training and backpropagation, where the result is fed backwards through the network to fine-tune the weights and improve the model – this process is repeated numerous times to adjust weights to produce the expected output. This is computationally costly and time-consuming, but with the help of hardware improvements over the past decade, has become more feasible and accessible.

Convolutional neural networks (CNNs; Figure 4) are a subset of ANNs that are particularly suited to deriving features from images. CNNs make use of convolution and pooling, steps that feed into and train a layered ANN to extract information and classify images or objects in a video. The convolution step defines what the important features are in an image, such as edges. The pooling step shrinks the information in a given frame by taking averages or finding maxima of nearby pixels for a sample window of a certain size, which is applied across an entire frame – essentially keeping only the important features as defined by the convolution step. By combining these steps of finding important features and removing the noise, information is passed through the network; convolution and pooling steps are repeated multiple times to condense the information contained in a frame and flatten the input image to be fed into the ANN. Most, if not all, modern computer vision models incorporate CNNs.

The possible applications for this technology are virtually limitless given the amount of video data collected daily by ocean observing systems. These models have been applied to fish identification for fisheries management and stock assessment, classification of algae species from plankton tows to monitor phytoplankton blooms, and environmental

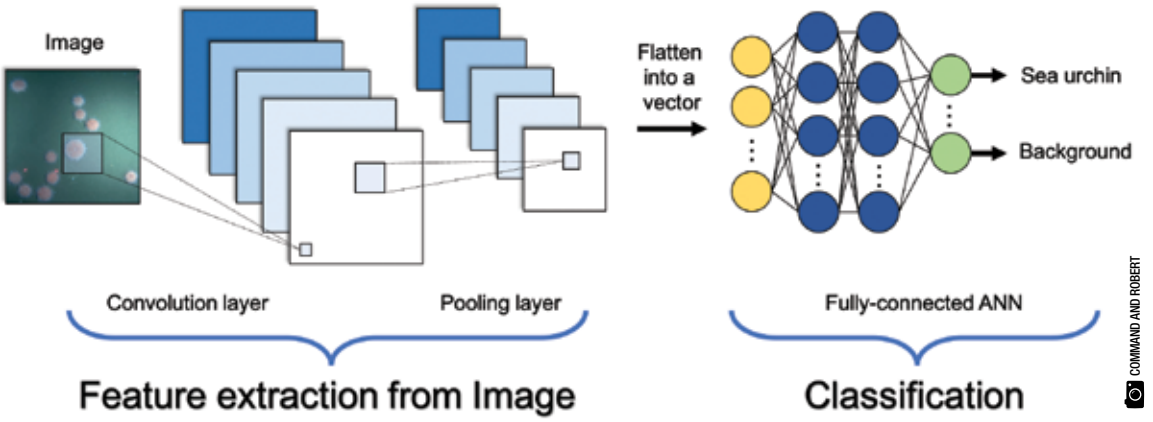

Figure 4: Diagram of a convolutional neural network (CNN). Convolutions are a class of mathematical functions that describe the amount of overlap between two functions as one is shifted over the other (i.e., it "blends" the functions). CNNs use a series of convolution and pooling steps to filter and reduce the dimensions of an input matrix, and are particularly well suited to extracting features from images. The result of a CNN is often fed into the input layer of an artificial neural network (ANN) that provides the classification step.

monitoring of marine energy projects such as wind or oil and gas platforms. These models are usually purpose-built but could be trained on similar datasets to address other problems from basic ecology to assessing responses to climate change.

This approach has been applied to the identification of pink urchins (*Strongylocentrotus fragilis*) from underwater video with 99% accuracy – the same species for which co-author Command manually sifted through 2,228 videos over the course of three months to count. This species lends itself well to computer vision tasks as it is easily distinguished from the background and other nearby organisms by its pink colour and unique morphology.

Computer Vision Underwater

Computer vision software has been optimized for object tracking and identification of day-to-day objects and situations, such as traffic and faces, to make light work of everyday situations (think self-driving cars and facial recognition). However, unique challenges exist for computer vision tasks in the marine environment. Variable or uneven lighting conditions due to high turbidity and particle loading create distorted or hazy images. Camera lenses can also become obscured through fouling by algae. Light attenuation with depth and the use of artificial lighting can also produce variable lighting conditions, with optical backscatter and colour fading making species identification more difficult. Addressing these challenges requires careful pre-processing of video and image data. Image processing pipelines to deal with each of these challenges are needed, but so far are only available on a case-by-case basis.

One possible way to improve predictions for some species would be to incorporate known life-history traits and ecologies into machine learning models. Spatiotemporal data about diel vertical migration and burrow emergence, or seasonal spawning and migration patterns using video timestamps and GPS tracking of instruments could help a model learn which organisms are likely to be present at a given site at a given time, and improve accuracy.

You Can Help

Machine learning models and CNNs can help information flow faster by automating some of the processing tasks, like object detection and classification, but only after training on a large dataset of labelled images. Labelling

these images is likely the most time-consuming part of the analysis, and cannot be automated without having a large, labelled dataset in the first place. How do we overcome this paradox of image analysis? This is where you come in.

Over the course of your internet use, you most likely have come across Google's image CAPTCHA. This usually shows up as the final "security check" before submitting an online form. For this check, Google shows you a grid of images and asks you to select only images with a stop sign or a crosswalk, or something similar. Have you ever wondered why these images are always related to traffic? Google is making use of perhaps the most efficient way to create a massive, labelled dataset – crowd sourcing annotations – to train its driverless car algorithms.

A few similar applications exist for identifying species in underwater images, and studies have been done to compare the performance of experts, machine learning algorithms, and trained volunteers (i.e., "the crowd"). Ocean Networks Canada's "Digital Fishers" is a crowd-sourced ocean science observation game where players label deepsea videos from ONC's cabled observatories and remotely operated vehicle dives. The player progresses through the game with each level unlocking more information about organisms and asking for more complex annotations. These crowd-sourced annotations directly contribute to ONC's database, creating a labelled dataset to train machine learning algorithms to automatically identify deepsea organisms from which scientists will be able to extract valuable information.

The fields of computer vision and artificial intelligence have the potential to unlock the bottleneck on underwater image and video processing and analysis. Given the current rate of ocean data collection, and the upcoming investment in ocean research brought by the United Nations Decade of Ocean Science for Sustainable Development, we certainly have our work cut out for us.

Rylan Command is a graduate student in the 4D Oceans Lab in the School of Ocean Technology; he is in the M.Sc. Fisheries Science and Technology program at the Fisheries and Marine Institute of Memorial University of Newfoundland and Labrador. His research focuses on temporal trends in the abundance, behaviour, biodiversity, and distribution of seafloor megafauna using underwater cabled observatories. His goal is to measure the response of benthic megafauna to rapid change over time to understand the role of anomalies, like marine heatwaves, and seasonal changes, like the spring phytoplankton bloom, in structuring marine communities. He is also interested in fisheries and food systems, and the ways in which humans use and distribute marine resources globally. He hopes to use available ocean data to engage broadly with both scientific and non-scientific audiences.

Dr. Katleen Robert in an assistant professor within the School of Ocean Technology at the Fisheries and Marine Institute of Memorial University where she holds a Canada Research Chair in ocean mapping. Her research aims at developing quantitative and repeatable approaches to map seafloor habitats. Her focus has been on examining fine-scale species-environment relationships using benthic imagery and multibeam sonars to build full coverage predictive maps. She is also interested in how benthic environments change temporally, and has been dappling with data from seafloor cabled-observatories since she started as a graduate student. Her hope is to one day merge these two research streams to produce maps showing the spatio-temporal heterogeneity of benthic habitats.