FishSense: Underwater RGBD Imaging for Fish Measurement

Peter Tueller*‖, Raghav Maddukuri*, Patrick Paxson*, Vivaswat Suresh*, Arjun Ashok†, Madison Bland‡
Ronan Wallace§, Julia Guerrero¶, Brice Semmens**, Ryan Kastner*
*Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA
†Dougherty Valley High School, San Ramon, CA
‡Department of Electrical and Computer Engineering, The College of New Jersey, Ewing, NJ
§Department of Mathematics, Statistics, and Computer Science, Macalester College, Saint Paul, MN
¶Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA
**Scripps Institute of Oceanography, University of California San Diego, La Jolla, CA
‖Corresponding Author: ptueller@ucsd.edu

***Abstract*—There is a need for reliable underwater fish monitoring systems that can provide oceanographers and researchers with valuable data about life underwater. Most current methods rely heavily on human observation which is both error prone and costly. FishSense provides a solution that accelerates the use of depth cameras underwater, opening the door to 3D underwater imaging that is fast, accurate, cost effective, and energy efficient. FishSense is a sleek handheld underwater imaging device that captures both depth and color images. This data has been used to calculate the length of fish, which can be used to derive biomass and health. The FishSense platform has been tested through two separate deployments. The first deployment imaged a toy fish of known length and volume within a controlled testing pool. The second deployment was conducted within an 70,000 gallon aquarium tank with multiple species of fish. A Receiver Operating Characteristic (ROC) curve has been computed based on the detector's performance across all images, and the mean and standard deviation of the length measurements of the detections has been computed.**

***Index Terms*—Underwater Imaging, RGBD Imaging, Machine Learning, Fish Monitoring**

I. Introduction

Our ocean ecosystems are facing a problem of enormous proportions: overfishing. Despite estimates that 3.5 trillion fish are swimming throughout the ocean, ecologists estimate that the planet will be devoid of fish as early as 2048 [19]. This chilling prediction is backed by fishing data as well: approximately 50% of the global fish population is harvested for human consumption *every year* [11]. These figures do not even include the illegal fishing market that comprises an estimated 18% of the yearly catch [1]. Such a drastic loss in biodiversity will not only impact the health and quality of our marine ecosystems, but will also starve nearly three billion people who depend on fish for survival.

In order to combat this daunting problem, scalable, yet simple, sensing systems are desperately needed to measure fish populations; specifically their biomass and health. Measurements are predominantly gathered using direct human observation: divers conduct manual censuses of fish populations and observers survey populations from above. Other solutions, such as catch & release, are invasive, often requiring fish to be removed from the sea for measurement. [16] [18]. These methods greatly affect fish health and stress levels, which ultimately negatively impact their growth rate [4]. Research of non-invasive methods for measuring free-swimming fish has yielded solutions that require underwater imaging and machine learning techniques.

Effective methods for underwater imaging are absolutely vital in retrieving valuable data. Typical methods include stereo vision [13], laser-based systems [15] [3], or acoustic-based systems [9]. These methods can be computationally complex and require intricate computer vision algorithms. There are also methods which have been presented which correct the effects of light scattering using light stripe range scanning and photometric stereo [12]. Alternatively, the Sea-thru method attempted to reconstruct the color in images through a process of removing the water from underwater images [2]. Other methods have used light emitters in conjunction with cameras to estimate depth in images. One such method proposed a scanning with a digital camera and laser that operated along a railed system [3]. Another method proposed included the use of a structured light depth camera used for 3D underwater capture, which also employed a calibration method that accounted for underwater refraction [6], though this system did not have sufficient range for capturing fish. Many of these systems are unwieldy, large, or too difficult for divers to operate, making them problematic for sustainable use.

In this work, we propose FishSense: a new handheld underwater RGBD imaging platform that measures fish length. This system can be used to derive fish biomass [7] and health. The FishSense core technologies include depth cameras, compute platforms, software, and a sleek mechanical design. FishSense utilizes scientific advancements in underwater imaging and machine learning to collect and process data. This integrated technology provides scientists with meaningful data to advance scientific predictions about fish health and populations. Our main contributions are:

1) A low-profile diver-operated platform for free-swimming fish imaging, as seen in Fig. 1;

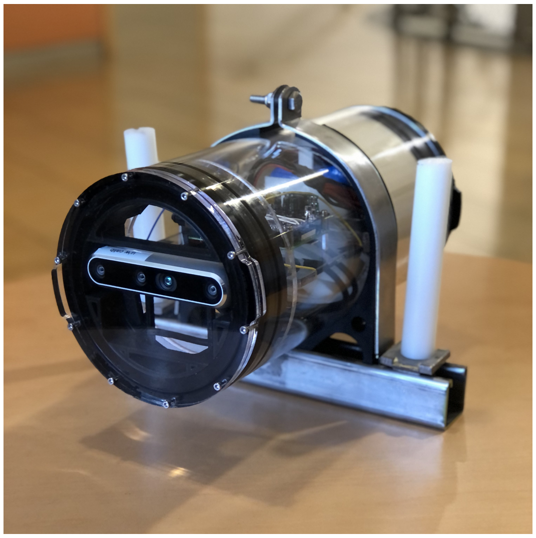


Fig. 1: FishSense imaging platform

2) An implementation of a machine learning detector for identifying fish in RGBD images;
3) An implementation of a fish length measuring algorithm for RGBD images.

The remainder of this paper is organized as follows. Section II provides an overview of the FishSense platform detailing the mechanical and electrical interfaces, as well as the Intel RealSense imaging device. Section III delves into the machine learning application with relation to fish detection and measurement. Section IV addresses the collected data. Finally, Section V provides the conclusion.

II. Platform

The FishSense design is a 12 inch long, 6 inch diameter cylindrical enclosure which stores our depth camera, compute platform, and batteries. The Intel RealSense D455 Depth Camera is mounted on one end, the four battery packs are on the other, and the Raspberry Pi 4 and switch system are located in between (see Figure 2).

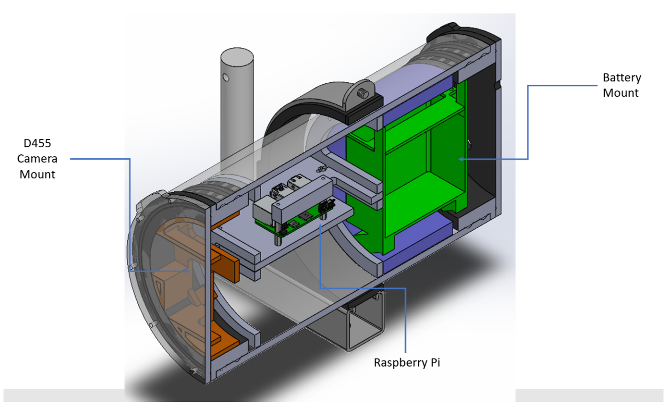


Fig. 2: FishSense mechanical prototype cross-section

A. Mechanical

A small, handheld system allows for easy portability, storage, and modularity. Our design allows for multiple configurations without changes to the housing. For example, the system can be easily converted into a camera trap, underwater vehicle payload, aquaculture cage monitor, or simply as an attachment to another piece of equipment. Mounting our device for any application is made simple and straightforward due to the multiple directions of screw holes.

Our device is fully optimized for underwater functionality. The system is well equipped to handle underwater deployments, with a depth rating of 65 meters, handles for divers, and magnetic switches for power and recording. Even with its sealed, waterproof design, it has four blank ports on the back plate, which enable wired connections into the system. With these ports, users can stream the camera feed to a topside computer, add an extended power bank for longer deployments, or simply charge the batteries without needing to open the housing.

B. Electrical

Our initial prototype uses a Raspberry Pi 4 with the Intel RealSense D455 and a 1TB SSD both connected via USB 3.0. It is powered by four 4S 3500 mAh lithium ion battery packs. With a fully charged pack, RealSense depth and color videos can run at 1-5 fps for approximately 12 h. As seen in Fig. 3, the battery is connected to a power module that distributes power to the reed switches, LEDs, and Pi.

To power on FishSense, a magnet needs to be applied to a reed switch. The walls of the enclosure are thin enough to allow for even a small magnet to activate the switch, which enables the user to power on the FishSense system without the need to open the watertight enclosure. It also allows for faster setup time and makes it more practical for amateur use. Another reed switch is used to start and stop the data collection, and a simple LED array alerts the user to the status of the system.

C. Intel RealSense D455

This project utilizes the Intel RealSense D455 Depth Camera. RealSense camera captures RGB images with a global shutter and a resolution up to 1280 x 800 at frame rate of 30 fps. The stereoscopic depth images have a resolution up to 1280 x 700 and a frame rate up to 90 fps. The RealSense camera uses an infrared emitter to project a pattern so that the stereoscopic depth cameras can easily reconstruct a depth image. The RealSense camera also contains the Intel RealSense Vision Processor D4 for vision processing to align the RGB and depth images together and capture a 3D view of the world. The Intel RealSense D455 camera was selected because of it's ability to perform with dynamic movement and for it's simplicity. Our camera has been calibrated for underwater use, which will allow us to capture and construct the entire scene with the D4 processor. This is important because our system is dependant on the accuracy of the camera, and it needs to construct the 3D view at the correct scale. The RealSense

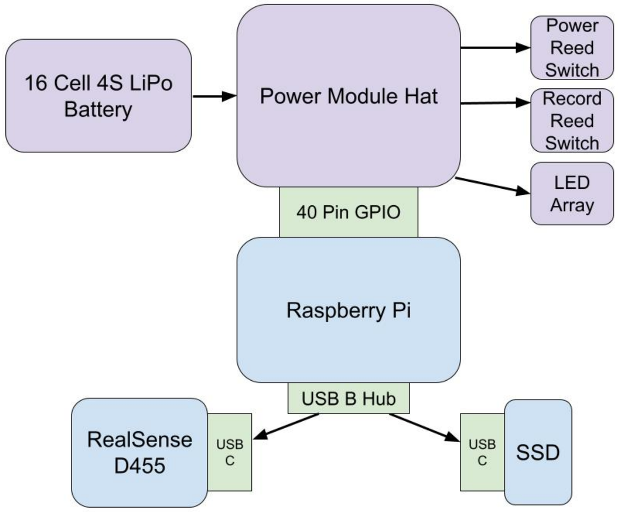


Fig. 3: FishSense wiring diagram

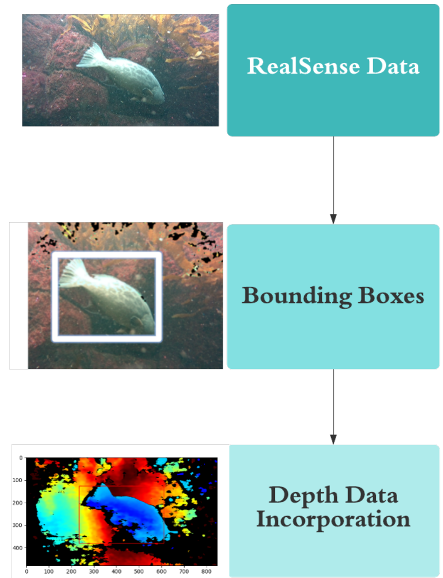


Fig. 4: Process of data interpretation

camera is positioned at one end of the FishSense system to capture images of fish in a desired location. The camera was used in conjunction with the Intel RealSense SDK 2.0.

III. Fish Detection and Measurement

Complementary to underwater imaging, techniques must be utilized to evaluate the collected data for extended use. A binary image classifier was used in [10] where Regions of Interest (RoI) were identified in data. Another source attempts to classify fish via an Artificial Neural Network, a supervised learning algorithm [14]. The machine learning method in [17] utilized the Speeded-Up Robust Features (SURF) and Support Vector Machines (SVM) algorithms along with a bag of features for simplicity. While both algorithms have benefits, the YOLOv4 model has been tested and time-proven to be the best when accounting for both detection speed and accuracy, and as such, it is the model we chose for our system.

The camera captures two pieces of information per frame. One being a standard RGB image, the other being a comma separated value (csv) file containing the distance of each pixel from the camera (i.e. depth image). Initially our machine learning algorithms work on the RGB image, detecting potential fish, and then the pixel values denoting the bounding box around each fish are recorded. From this bounding box, pixels at the tip of the head and the end of the tail are selected, and the corresponding depth associated with those pixels are retrieved from the depth image. These pixels are then projected into 3D space by using the intrinsic parameters of the RealSense depth camera, and the Euclidean distance between the points are measured to determine the length of the detected fish. This process can be seen at a high level in Figure 4.

The performance of machine learning model is tested using Mean Average Precision, as well as ROC curves. The machine learning algorithm uses a test dataset that was 10% of the training dataset. This test set is only used to measure the accuracy of the predicted bounding boxes, against human annotated bounding boxes. The average loss is also calculated, and used along the mAP to ensure learning was accurate. Once the models finish training, they are then tested in a controlled pool with a toy fish, as well as an aquarium with many species of fish to generate ROC curves.

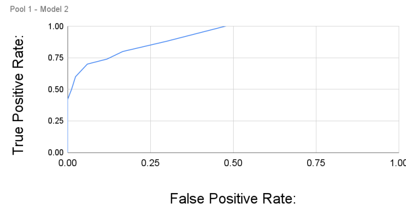
The performance of the length measurement component of our process is assessed by analyzing all the true detections of the toy fish in our controlled pool tests. The average and standard deviation of length measurements are computed and compared to the ground truth length of the toy fish.

IV. Analysis

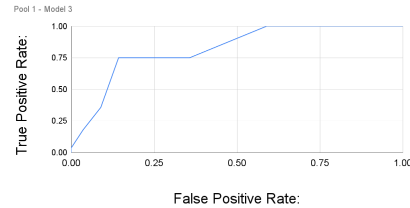
For training the models, three distinct data sets were used. The first set of weights were trained off a subset of fish within Google's Open Image Dataset [8] and the OzFish Dataset [5] that were not colorful or exotic i.e Cuttlefish, Lionfish, Betafish etc. This allowed for the weights to standardize to ”normal” or ”regional” fish, which allows our system to perform much better in mounted scenarios, where we can predict which species will be present. The second set of weights were trained off images of our toy fish in various scenarios. This allows our system to perform better a single species application like aquaculture. The last set of weights were trained off of the entirety of the Open Image Dataset and OzFish Dataset. These weights provided a baseline for our system would perform in open ocean scenarios, where we might encounter a wide variety of fish. All three of these weights were tested against two different data sets. The first being our toy fish in a controlled test pool, and the other being various species of fish in an aquarium.

The ROC curves shown in Figure 5 demonstrate the functionality and trade offs of each set of weights. Model 1 is

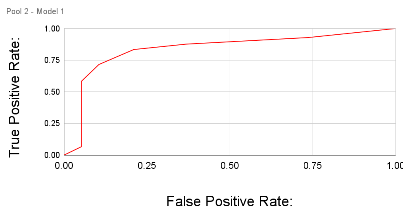
(a) Region based model in Single fish tests



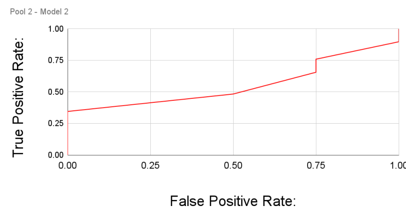
(b) Species specific model in Single fish tests



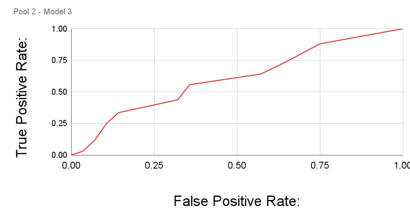
(c) General use model in Single fish tests



(d) Region based model in Multiple fish tests



(e) Species specific model in Multiple fish tests



(f) General use model in Multiple fish tests

Fig. 5: ROC curves for three models over two pool tests

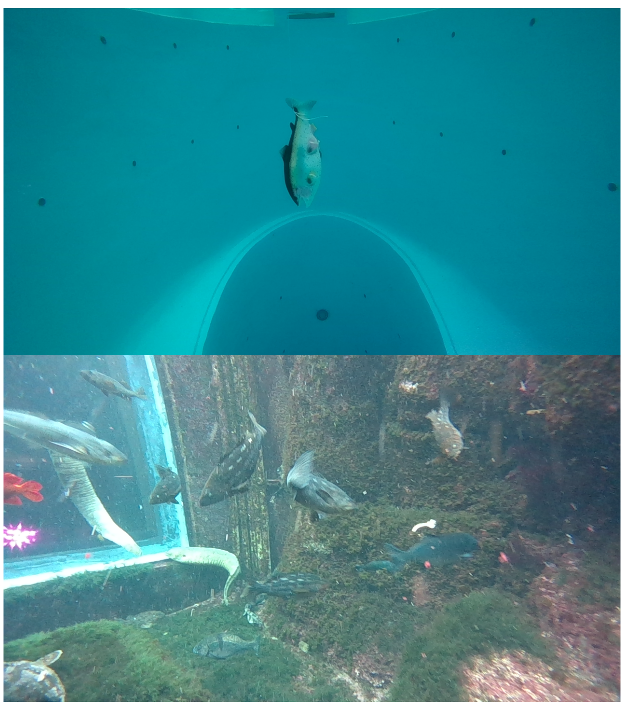


Fig. 6: Sample images from the Pool 1 (top) and Pool 2 (bottom) datasets

the set of weights used for usage in mounted scenarios with predictable "regional" species. Model 2 is the set of weights trained for usage in single species detection. Model 3 is for general use, trained to detect all fishes. Pool 1 had a single toy fish in a controlled test pool, whereas Pool 2 is a 70,000 gallon aquarium that contains many different fish, such as leopard sharks, moray eels, Giant Black Sea Bass, Garibaldi, kelp bass, opaleye, and other fish native to Southern California. Sample images from each of these pool datasets can be seen in Figure 6. Model 1 performed well with a high threshold in both datasets, as despite detecting false positives, it confidently detects real fish with a high deal of accuracy (Figures 5a and 5d). Unsurprisingly, Model 2 performed extremely well in the Pool 1 dataset, as seen in Figure 5b, as it was trained off of the specific fish used in that experiment, but fell short when confronted with many different species in the Pool 2 dataset, as seen in Figure 5e. This tells us that there is a danger to focusing the model too much on a single type of fish and encountering a fish outside of the training dataset. Finally, it can be seen that Model 3, the general model, performed decently in the single fish test (Figure 5c) and performed average in the multiple fish test(Figure 5f). This tells us that a general model can be useful when imaging unknown species of fish, but that there is opportunity to improve detection when there is knowledge of the "regional" type of fish, and especially if there is interest only in a single species of fish.

The toy fish that is used in the Pool 1 deployments was measured by a tape measure and found to be 31.5cm. When using the depth images over all the true positive detections of that fish in the Pool 1 dataset, we found the average length to be 31.83cm, with a standard deviation of 1.950cm. This result is extremely encouraging, and suggests FishSense is a valid instrument for length calculations underwater. We observed the error in the measurement to come from two sources. First, the depth images occasionally had blur, which made the fish seem longer or wider than it actually was, or on rare occasions, missing chunks, which affected the quality of point selection. The second source of error was the orientation of the fish in the image. If the fish was not in clear profile, then it was harder to pick the same points on the fish consistently.

V. Conclusion

In this work, we have outlined FishSense, an underwater RGBD imaging platform for fish detection and fish length measurement. We have experimentally validated our system with three detection models and two underwater datasets, and have shown it to accurately detect and measure fish.

Although our initial prototype has demonstrated feasibility in many of our intended applications, future iterations will improve the efficiency of the system and allow it to extend into new applications. To achieve longer deployments, an increase in computing power, storage capacity, and battery count are essential. Integrating a higher resolution color camera allows for better image analysis and species classification. Also, the addition of an LCD screen for depth map display, system diagnostics, and a settings menu will make underwater use easier and open up other possibilities during underwater operations.

Our current compute platform is limited by the bandwidth of its USB controller. More specifically, it cannot simultaneous read from the RealSense and write to the SSD at 30 fps using the same USB controller. As such, we performed our experiments at a reduced framerate, but to solve this problem, future compute platforms will utilize more powerful hardware to perform full throughput video recording. More computing power will additionally enable the possibility of real-time data processing on-board the device. Currently, an NVIDIA Jetson TX2 processor serves as the catalyst for these solutions. With a higher bandwidth USB controller and SATA compatibility, we expect to save data at the full frame rates the RealSense offers, along with exploring more advanced real-time processing.

Another opportunity for improvement lies with the detection model. It currently performs detections with RGB data independent of depth data; however, by incorporating the depth data into the detections to create particular regions of interest (ROI), detection accuracy should increase dramatically. Since most fish will be in within some range of the camera, by setting boundaries and parsing through the depth image, various ROIs could be created. These regions might include a lot of false positives, as it would pickup rocks, kelp etc. However, if these ROIs are combined with detectors, and even with the actual neural network architecture these false positives can be accounted for. Likewise, the depth data could also be incorporated into preprocessing techniques to counteract the hazing that occurs in water. By using the RealSense SDK this is a lot more intuitive, and can bolster detector accuracy as well the increase the viability of above water data.

References

[1] David J. Agnew, John Pearce, Ganapathiraju Pramod, Tom Peatman, Reg Watson, John R. Beddington, and Tony J. Pitcher. Estimating the worldwide extent of illegal fishing. *PLOS ONE*, 4(2):1–8, 02 2009.

[2] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[3] C Almansa, L Reig, and J Oca. The laser scanner is a reliable method to estimate the biomass of a senegalese sole (solea senegalensis) population in a tank. *Aquacultural engineering*, 69:78–83, 2015.

[4] Paul J. Ashley. Fish welfare: Current issues in aquaculture. *Applied Animal Behaviour Science*, 104(3-4):199–235, 2007.

[5] Australian Institute Of Marine Science. OzFish Dataset - Machine Learning Dataset For Baited Remote Underwater Video Stations, 2020. Type: dataset.

[6] Sundara Tejaswi Digumarti, Gaurav Chaurasia, Aparna Taneja, Roland Siegwart, Amber Thomas, and Paul Beardsley. Underwater 3d capture using a low-cost commercial depth camera. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.

[7] TW Fulton. The rate of growth of fishes. twentysecond annual report. *Fisheries Board of Scotland. Retrieved from https://ci. nii. ac. jp/naid/10030255872*, 1904.

[8] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, et al. The open images dataset v4. *International Journal of Computer Vision*, 128(7):1956–1981, 2020.

[9] D-Q Lin, H Zhang, M Kang, and Q-W Wei. Measuring fish length and assessing behaviour in a high-biodiversity reach of the upper yangtze river using an acoustic camera and echo sounder. *Journal of Applied Ichthyology*, 32(6):1072–1079, 2016.

[10] Simone MariniVimal, Emanuela Fanelli, Valerio Sbragaglia, Ernesto Azzurro, Joaquin Del Rio Fernandez, and Jacopo Aguzzi. Tracking fish abundance by underwater image recognition. *Scientific Reports*, 8, 2018.

[11] Alison Mood and Phil Brooke. Estimating the number of fish caught in global fishing each year. *Fishcount.¡ En línea*, 2010.

[12] Srinivasa G. Narasimhan and Shree K. Nayar. Structured light methods for underwater imaging: Light stripe scanning and photometric stereo. *Proceedings of OCEANS 2005 MTS/IEEE*, 3, 2005.

[13] Daniel Pérez, Francisco J Ferrero, Ignacio Alvarez, Marta Valledor, and Juan C Campo. Automatic measurement of fish size using stereo vision. In *2018 IEEE international instrumentation and measurement technology conference (I2MTC)*, pages 1–6. IEEE, 2018.

[14] Ricardus Anggi Pramunendar, Sunu Wibirama, and Paulus Insap Santosa. Fish classification based on underwater image interpolation and back-propagation neural network. *International Conference on Science and Technology*, 5:1–6, 2019.

[15] Austin A Rizzo, Stuart A Welsh, and Patricia A Thompson. A paired-laser photogrammetric method for in situ length measurement of benthic fishes. *North American Journal of Fisheries Management*, 37(1):16–22, 2017.

[16] NJC Strachan. Length measurement of fish by computer vision. *Computers and electronics in agriculture*, 8(2):93–104, 1993.

[17] M. Vimal Raj and S. Sakthivel Murugan. Underwater image classification using machine learning technique. *International Symposium on Ocean Technology*, pages 166–173, 2019.

[18] Darren J White, C Svellingen, and Norval JC Strachan. Automated measurement of species and length of fish by computer vision. *Fisheries Research*, 80(2-3):203–210, 2006.

[19] Boris Worm, Edward B. Barbier, Nicola Beaumont, J. Emmett Duffy, Carl Folke, Benjamin S. Halpern, Jeremy B. C. Jackson, Heike K. Lotze, Fiorenza Micheli, Stephen R. Palumbi, Enric Sala, Kimberley A. Selkoe, John J. Stachowicz, and Reg Watson. Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314(5800):787–790, 2006.