No Machine Learning Without Data

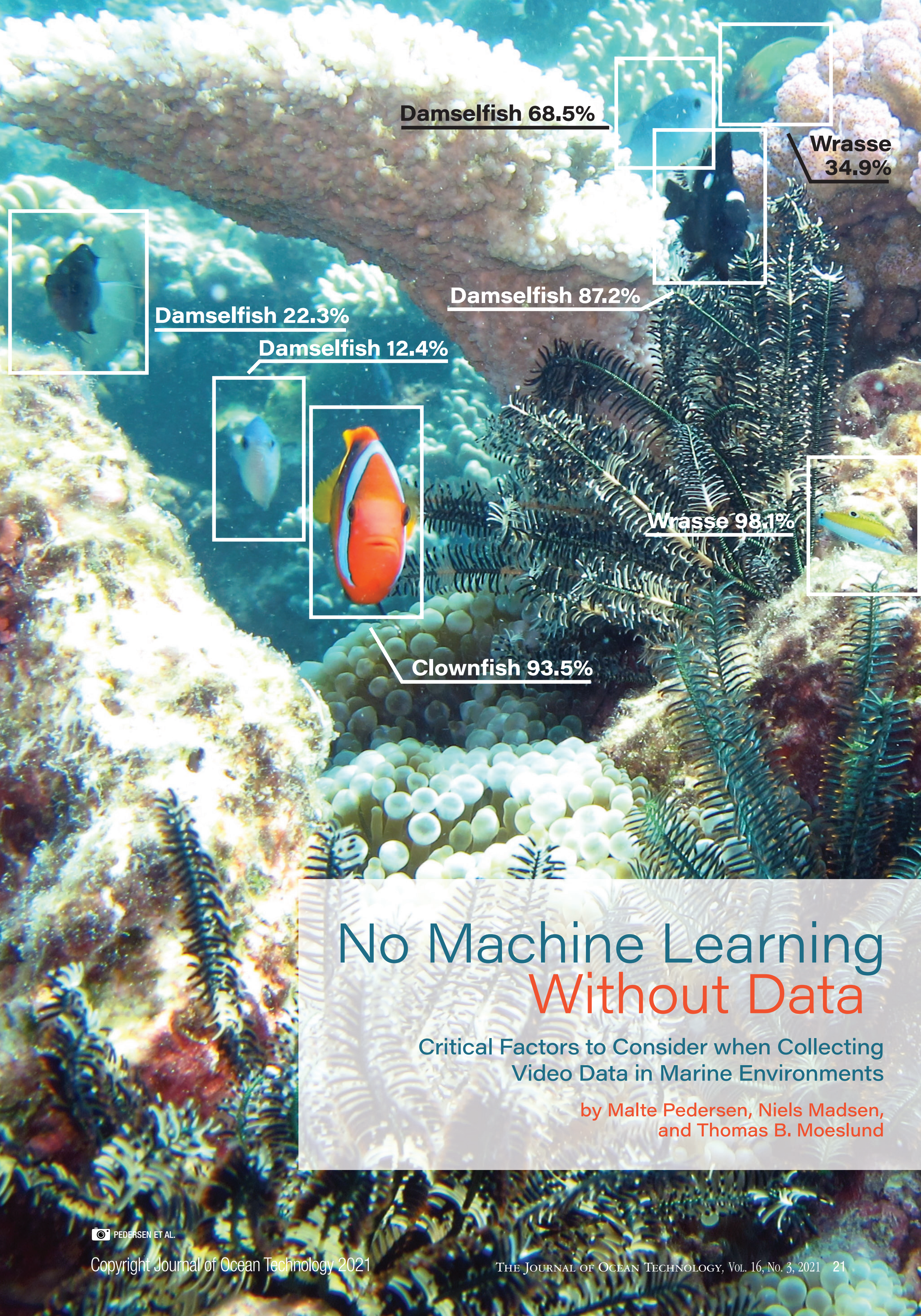
Critical Factors to Consider when Collecting Video Data in Marine Environments

Pedersen, Malte; Madsen, Niels; Moeslund, Thomas B.

Published in:
The Journal of Ocean Technology

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record



No Machine Learning Without Data

Critical Factors to Consider when Collecting Video Data in Marine Environments

by Malte Pedersen, Niels Madsen, and Thomas B. Moeslund

PEDERSEN ET AL.

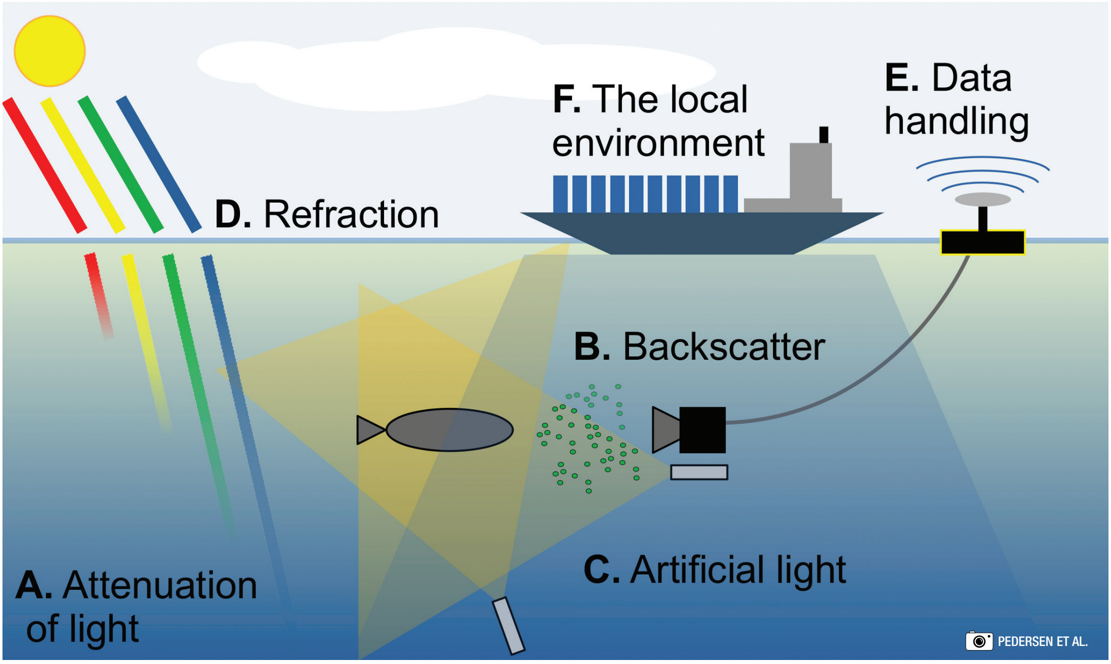


Figure 1: Illustration of six important factors to take into account when collecting video data from underwater environments.

Introduction

An increased political focus on the condition of our marine ecosystems has put researchers under pressure to gather and analyze data at an unprecedented pace. Assessing the impact of global climate change, pollution, and overfishing on the biodiversity and fish stocks are major challenges for fisheries and governments across the world. An increasingly popular tool for gathering biological data in a non-intrusive manner is automated analysis of image and video data using computer vision and machine learning. However, large-scale image-based data collection and automated analysis has not traditionally been common practice among marine researchers.

While images of a given object captured in air looks more or less the same independent of where on the planet you take the photo, this is not the case in marine environments. Images are formed in a camera by capturing the light reflected from objects within the camera's field of view, but marine waters are filled with organic and inorganic matter that absorbs and scatters light. This causes the visibility and colouring to vary widely depending on the location, time, depth, and weather which can make it a challenge to capture high quality video recordings of underwater objects; and without sufficient high quality data, any machine learning algorithm will fail.

Machine learning and computer vision are increasingly used within several fields of biology, but there seems to be a hesitancy when it comes to underwater research areas such as fisheries and marine science. The main reason being that, traditionally, it has been extremely demanding and expensive to capture high quality underwater video footage suitable for automated analysis. However, during the past decade the price of conventional action and underwater cameras has dropped substantially while the image quality has increased. Moreover, the performance of state-of-the-art computer vision and machine learning algorithms has sky-rocketed during the same period, with the introduction of deep neural networks.

Neural networks are machine learning algorithms that learn in a way somewhat similar to the way children learn. They need to see things many times in different settings and be told what they are looking at to be able

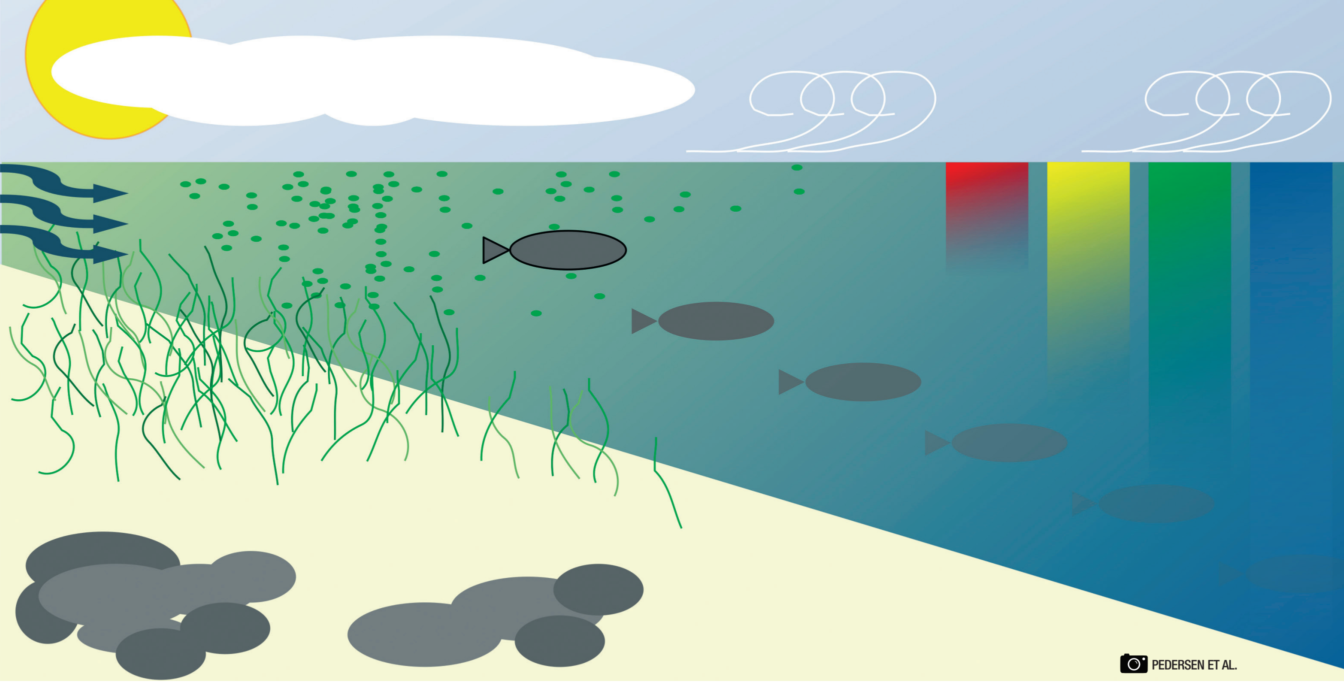


Figure 2: Coastal areas can be especially nutrient rich with high concentrations of phytoplankton that makes the water appear green or brownish. Furthermore, sediments can be resuspended in shallow waters due to the currents and wind. As the depth increases the water gets clearer, but the light intensity is reduced.

to distinguish between them. By presenting a deep neural network with large numbers of fish images, it is possible for the network to learn how to detect and distinguish between species, e.g., mackerel, cod, and tuna. Another network may be designed to track fish through video sequences which can be used for behavioural analysis or for controlling a bycatch release-mechanism inside a trawl. There are many possibilities of using deep neural networks for automating processes in marine settings, but independent of the task at hand or the choice of network, there is a demand for annotated data.

In this essay, we present and discuss the main factors that influence the data capturing process. We hope this will pave the way for other marine researchers to capture high quality data and thereby set the stage for using machine learning algorithms in marine monitoring.

The Six Factors

It is not feasible to create a single protocol for underwater video monitoring due to the extreme variations in marine environments across the globe. However, the six factors illustrated in Figure 1 (attenuation of light, backscatter, artificial light, refraction, data handling, and the local environment) should always be taken into account when capturing data for a machine learning-based marine video monitoring system. In the following, each of the factors are discussed in greater details.

Attenuation of Light

Probably the most significant difference between capturing images in air and water is caused by the attenuation of light. While it is most often not necessary to take attenuation of light in air into account, it is another matter in water. As light enters water, it takes only a few metres before the long wavelength colours in the red spectrum are absorbed by the water. This is followed by the absorption of the yellow, green, and lastly blue wavelength colours.

The exact depths at which the wavelengths of natural light are absorbed in our ocean, estuaries, and rivers vary greatly depending on the intensity of the light, particles in the water, and other factors. However, objects that are observed through natural light will always appear more dim and colourless as the depth increases, as illustrated in Figure 2.

Figure 3: Three photos captured from a stationary camera a few minutes apart in a brackish strait at 9 m depth with artificial light. The images appear brownish due to high numbers of phytoplankton and resuspended sediments. The visibility goes from semi-clear to unclear from left to right. In shallow straits, estuaries, and coastal areas the water can turn unclear rapidly and is rarely clear at any point.

It is particularly difficult to capture high quality recordings in coastal waters and estuaries as the visibility can vary to a large degree. An example of varying degrees of visibility in a shallow strait is presented in Figure 3. Low visibility is often due to resuspension of sediments and eutrophication caused by shallow water, wind, and excessive amounts of nutrients. Ecosystems with high concentrations of phytoplankton appear green or brownish due to the chlorophyll and carotenoid pigments that reflect green and orange-red wavelengths, respectively.

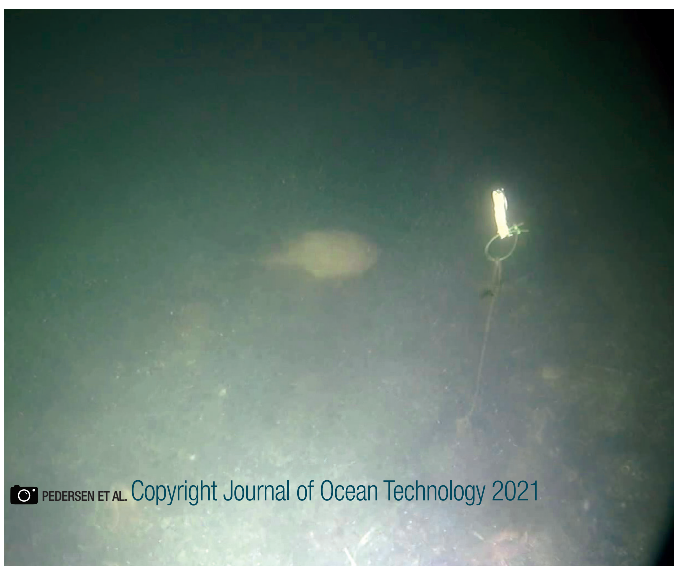
Backscatter

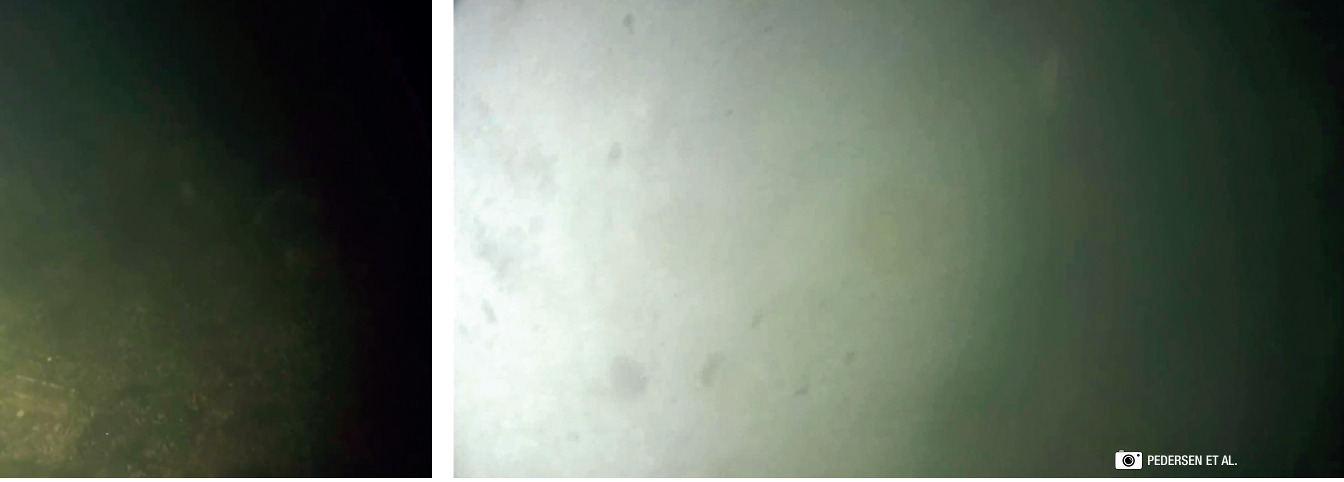
Artificial light can be used to counteract low visibility caused by the attenuation of light. However, there are several things to be aware of when using artificial light in water. The placement and direction of the light can introduce backscatter, which is light absorbed and scattered by small particles in the water between the lens and the object. Backscatter can be the cause of significant noise and it can occur even in seemingly clear water.

The water is semi-clear in Figure 4, but it is difficult to see the seabed due to strong backscatter caused by a single artificial light placed close beneath the camera and pointing into the water column in front of the camera's field of view. The backscatter appears almost like a thick fog with sprinkles due to the varying size of scattering particles in the water. A less severe case can be seen in Figure 5 where the water is also semi-clear and backscatter occurs as sprinkles in the left side of the image. The single light source is placed to the left of the camera and is illuminating the scene in an indirect manner allowing for a clearer view of the seabed.

In environments where the water is mostly unclear, it may not be suitable to use a conventional camera due to the short visual range. Specialized sensors, such as range-gated time-of-flight (ToF) cameras or sonars, can be used to minimize the effect of backscatter and obtain information on objects not seemingly visible. A ToF camera can measure the distance between the camera and the objects in a scene using active

Figure 4: Strong backscatter caused by artificial light positioned close to the camera in semi-clear water.





illumination and measuring the time it takes from when the light is emitted and until it is received by the image sensor. Range-gating allows the ToF camera to only open the shutter and receive light that has travelled a given distance, which is an effective way to see past backscatter. The visual distance of range-gated ToF cameras is, however, still dependent on a certain amount of artificial light penetrating the turbid water and reaching the object. Moreover, the resolution of ToF cameras is most often lower than for conventional cameras. In scenes with very unclear water even range-gated cameras fall short and an alternative can be to use sonar. Sonars are sound-based sensors capable of gathering depth information across long distances independent of the water clarity. However, sonars generally have a very low resolution which makes it nearly impossible to recognize species, count the number of objects, and similar tasks.

Figure 5: Sprinkled backscatter in the left side of the view in semi-clear water caused by indirect artificial light.

Artificial Light

Recording high resolution underwater colour videos is currently only possible using conventional cameras, but as depth increases so does the attenuation of light. The reduction in light causes objects to appear dim and featureless and artificial light can be a necessity. However, it is a non-trivial task to place the light source in an optimal position and it is highly dependent on the environment.

In clear and non-scattering water, it may be possible to illuminate an entire scene satisfactory using a single source of light placed close to the camera; see Figure 6A. However, this setup can be the cause of strong backscatter even in slightly turbid water; see Figure 6B. Backscatter can be minimized by placing the light further away from the camera and on an angle, but this may cause uneven illumination and shadows; see Figure 6C. Multi-directional illumination is a way to combat the problems of uneven lighting and shadows but it requires a larger and more complex setup; see Figure 6D.

It is also possible to reduce backscatter even further by placing the light source very close to the object, but here it is extremely important to take into consideration whether non-uniform lighting, overexposure, and shadows can be a problem, as seen in Figure 7. Generally, the exercise is to find the best trade-off between an even illumination and a minimum amount of backscatter.

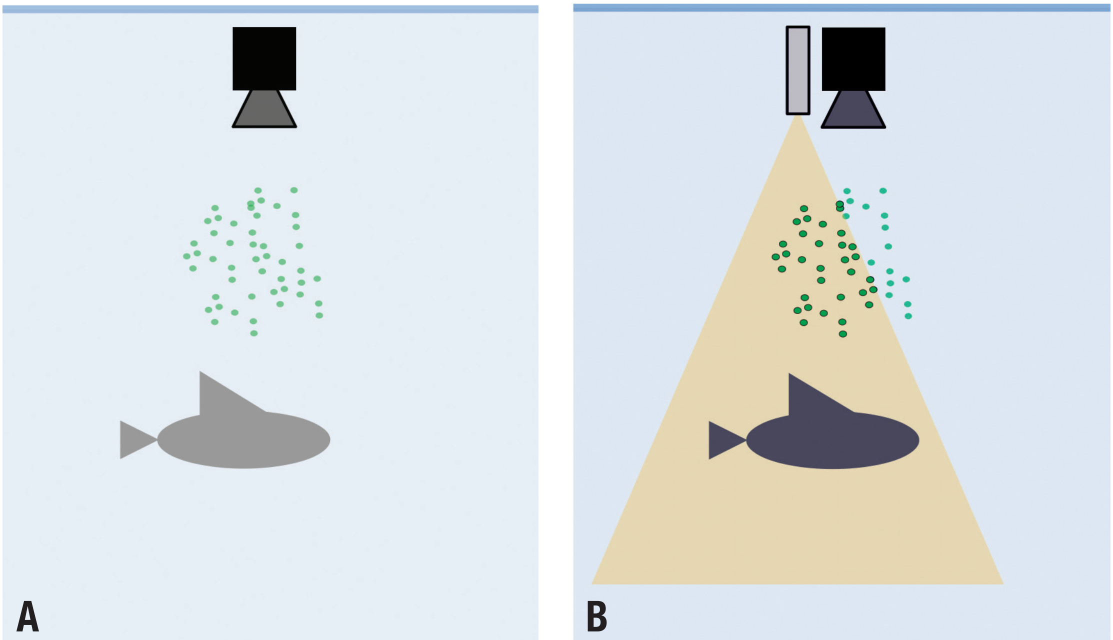


Figure 6: Four light setups for capturing underwater images. (A) No artificial light; the object appears dim and colourless. (B) A single light; even (D) Two light sources; backscatter is minimized and even illumination is achieved.

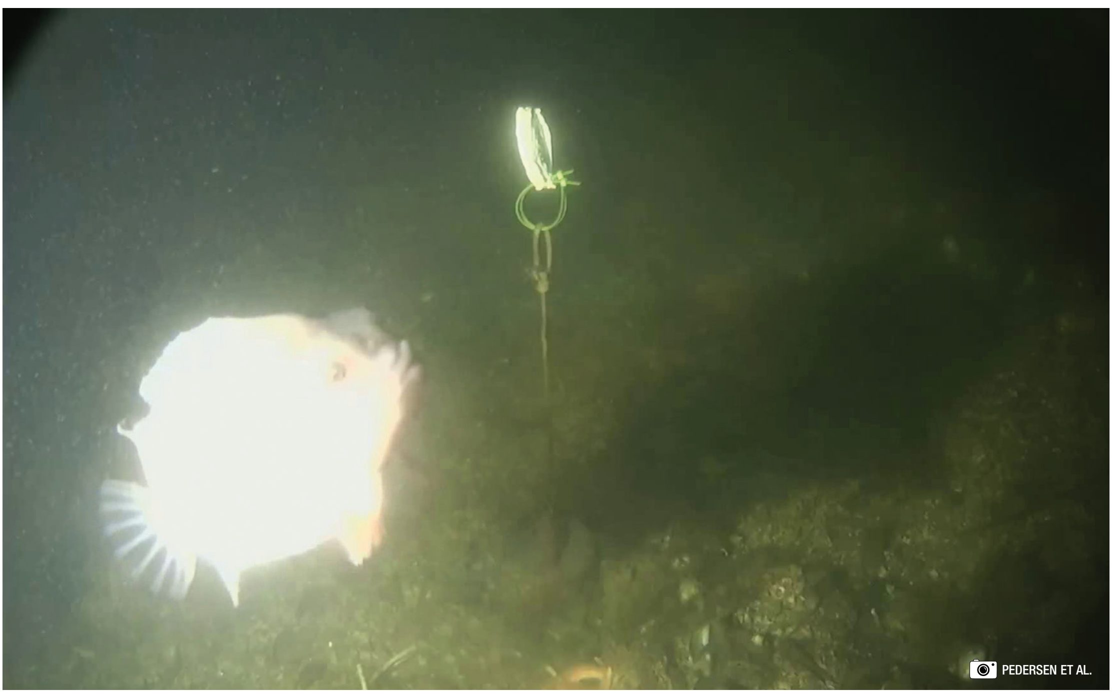
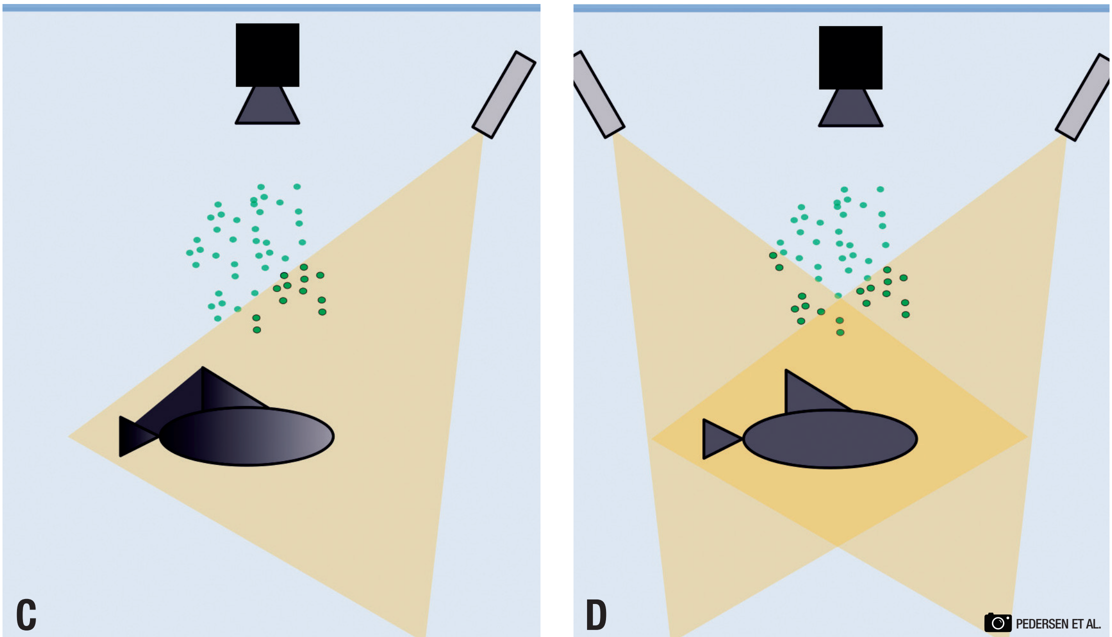


Figure 7: A single light source is placed to the left of the camera at an angle. Some sprinkled backscatter is seen in the left part of the image. The fish is swimming close to the light source causing overexposure and strong shadows on the seabed.



illumination, but significant backscatter. (C) The light is placed at an angle; backscatter is minimized, but shadows and uneven illumination occur.

Refraction
The speed of light varies depending on the medium it travels through. When light moves from air to water the speed is slowed down and this causes an effect known as refraction, where the direction of the light changes with respect to the incident angle and velocity. An example of how refraction affects an image can be seen in Figure 8. The black lines illustrate the rays with no refraction, whereas the dotted lines show the directional change caused by refraction.

Depending on the type of camera, lens, and underwater casing, refraction can significantly decrease the accuracy of measurements as it alters the intrinsic parameters of the camera and distorts the image. A commonly used method to minimize the effect of refraction is to calibrate the camera using a checkerboard or calibration frame. This can provide relatively accurate results if the scene of interest is restricted to a limited space; however, it has been shown that refraction causes single viewpoint models to be invalid and the error grows concurrently with distance and angle to the image sensor. A more robust method is to model the physical attributes of refraction and use ray tracing to adjust for the errors introduced by the perspective model.

Choosing the right calibration method can be difficult and is dependent on the application. Calibration and refraction handling is most often not a necessity for handling relatively simple image processing problems like object detection. However, more complicated machine learning tasks like classification or re-identification may benefit from it. If precise object tracking or 3D reconstruction is the goal, then calibration and refraction handling can be critical.

Data Handling
Recording videos under water can be extremely demanding; therefore, the aim is to get as much out of the recordings as possible. Data storage can be a problem and, whether it is long- or short-term monitoring, a goal is often to keep the storage at a minimum. Therefore, it is important to know what type of image analysis is to be conducted on the data. If the task is to count the number of fish

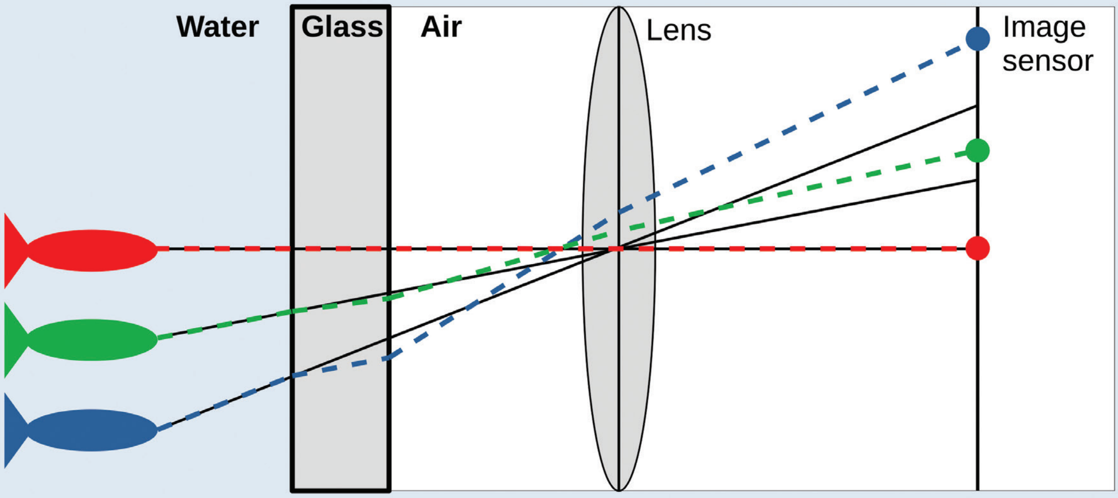


Figure 8: The dotted lines illustrate the refracted light rays, whereas the black lines illustrate the paths the rays would travel through air alone. Notice how the positional displacement on the image sensor increases with the incident angle of the ray.

using object detection, the image resolution and frame rate can be kept relatively low and the videos may even be compressed without significantly influencing the detection rate. Expanding the task to include classification requires a higher resolution and for reliable object tracking a high frame rate is important as marine organisms can move both quickly and erratically. Temporal compression should generally be avoided as it introduces motion blur and amplifies noise, e.g., caused by particles flowing in the water as illustrated in Figure 9.

For long-term monitoring projects, it is important to ensure a steady power supply and a way to retrieve data regularly. Regular

Figure 9: Video compression can reduce the storage size significantly but it also removes information and introduces noise. In this example, small particles draw semi-transparent lines across the image, due to temporal compression, while flowing from left to right.

data retrieval puts less demands on the storage capability of the internal hardware, allows for routine inspection of the data, and serves as a vital backup. Floating stations are common for offshore operations, while cabled land-based stations can be used in some coastal areas.

The Local Environment

The local environment can play a pivotal role in unforeseeable ways. Here we mention common problems that can hinder an otherwise well-structured underwater monitoring setup.

Algae can bloom on the lens within a few days depending on the environment. Algae growth will cloud the view and make the quality of the recordings poor or even useless. In Figure 10 the fish and the surroundings appear green due to algae on the lens and phytoplankton in the water.

Permissions from the local municipality or national maritime authorities may be needed to conduct research in wildlife sanctuaries or close to ship traffic, such as harbours or channels.

Ship traffic can easily destroy a floating surface station; therefore, it is crucial to mark observation spots properly. In cold regions it is also vital to consider potential floating ice.

Flickering from the sunlight when it hits the

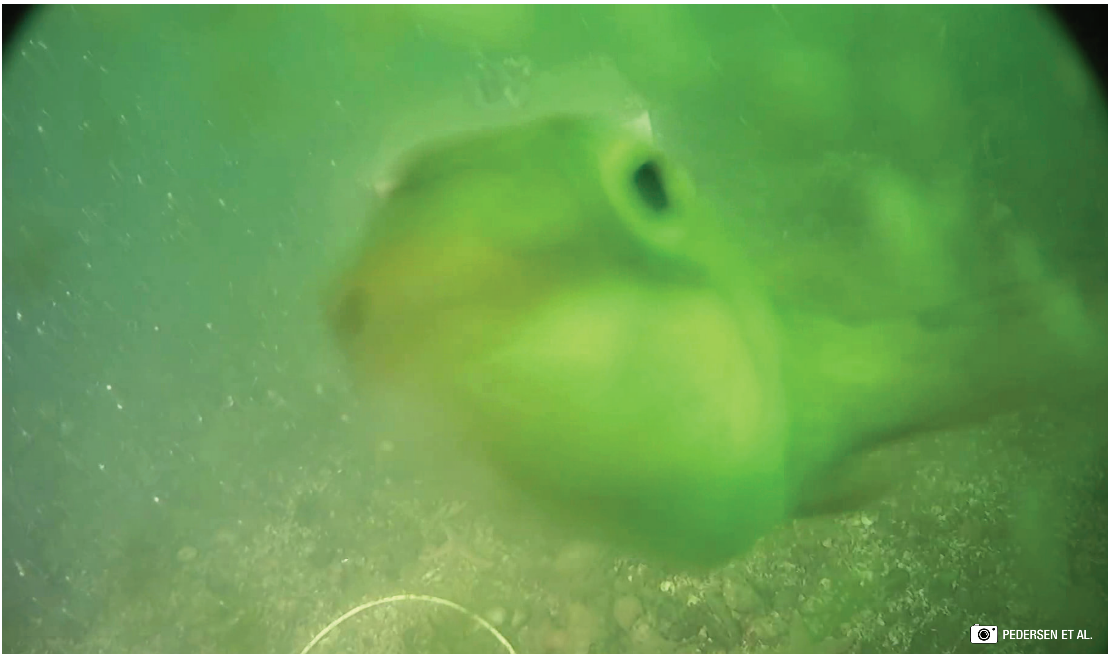


Figure 10: Algae on the lens can make everything appear green and will, in the worst case, block the view entirely. Algae growth varies greatly depending on the local environment.

Figure 11: The behaviour of marine organisms is strongly affected by the presence of artificial light, such as this school of sticklebacks lingering in front of a light source.

waves is especially apparent in shallow waters. Additionally, clouds and boats can cause shadows that will darken the entire scene. **The behaviour** of marine organisms will, in most cases, be affected by the presence of cameras, humans, or vehicles unless well hidden. An example can be seen in Figure 11 where a school of sticklebacks is lingering in front of the camera attracted by a light source. The use of artificial light can both attract and repel animals and alter the local environment around the setup of long-term monitoring projects. Therefore, it is extremely important to take into consideration whether artificial light is necessary for a given setup.

Final Remarks

A requirement of most state-of-the-art machine learning-based computer vision solutions is the need for large annotated image datasets. There are very few available underwater datasets compared to their terrestrial counterparts, and this is a hindrance for the development of dedicated marine vision algorithms. The variability of underwater environments can be extreme and it is, therefore, crucial to have training data from as many regions, environments, and ecosystems as possible to build a strong foundation for the coming generation of marine vision algorithms.

We urge marine researchers to be open-minded about using cameras, marine vision, and machine learning in their research, and sharing their datasets and annotations with the public. Hopefully, the six factors can serve as a stepping stone for many future marine image and video data collection tasks to the benefit of the marine research community.

Acknowledgments

Funded by the Independent Research Fund Denmark under case number 9131-00128B.

Dr. Malte Pedersen received his M.Sc. in computer vision from Aalborg University in 2017. His research interests are focused on solving problems in aquatic environments using computer vision and machine learning. He currently holds a three-year PhD Fellowship at Aalborg University within computer vision in underwater environments and already has several publications within the field.

Dr. Niels Madsen is a professor at Aalborg University with marine biology and technology as main research areas. Over the past 25 years, his research focus was on developing selective and environmentally friendly fishing gears, and several were successfully implemented by legislation in European fisheries. He is editor of the scientific journal *Fisheries Research*, a highly significant journal in fisheries technology.

Prof. Thomas B. Moeslund is currently the head of the Visual Analysis and Perception Laboratory, head of Section for Media Technology, and head of AI for the People Center, all at Aalborg University, Denmark. His overall research interest is building intelligent systems that make sense out of data with a special focus on computer vision and AI.