UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Master of Science Course in Environmental Engineering

Innovative methodological approaches for marine communities assessments

Relatore:

Prof.ssa EVA SAVINA MALINVERNI

Correlatore:

Prof. CARLO CERRANO

Prof.ssa EMANUELA FANELLI

Tesi di Laurea di:

SANITÀ MARSIA

Anno Accademico 2019/2020

List of Figures

Contents

1. Introduction

More than 70% of the Earth is covered by water. Our oceans are like lifeblood of Mother Nature. They are responsible for the production of more than half of the oxygen and for the absorption of most of the carbon from the environment [1]. Given the magnificence of the oceans, much of the global warming depends on the warming of ocean waters. This heat stored in the oceans causes a variation in water temperatures. The consequence is that the marine ecosystem begins to change. Migrations of tropical marine living beings take place in new, previously colder areas. Below the surface of the seas, hundreds of kilometres from the mainland, there is a world of giants and predators, ancient life forms and lost cities. In the same way, also the existence of seagrass and coral reefs is very important. Thus is very important for our environment maintaining the oceans in good health. It is estimated that almost a quarter of all the carbon dioxide produced by the human activities, emitted in the last two decades, has been captured by the waters. It is as if our destiny and the destiny of the oceans are strictly tied. Therefore helping the oceans is necessary for our survival.

The Mediterranean Sea covered less than 1% of the water, but it hosts about 8% of known marine species. This means thousands of creatures that are part of marine biodiversity to safeguard [2]. Multiple factors are causing the crisis of Mediterranean Sea: the most important are the overfishing and pollution. They are threating our precious ecosystem. Recent studies claim that over the 93% of the fish stocks analysed are overfished, consequently in the last 50 years the Mediterranean has lost about 41% of marine mammals. The Mediterranean Sea can be seen as an area of very high biodiversity. Unfortunately, this scenario could very soon change. Many of its species are endangered, due to the destruction of their habitat, falling victim to illegal fishing or dying as by-catch in fishing nets [3]. Concentration of the waste, in particular plastic waste, is another big problem. These micro plastic particles are harmful to marine fauna because are ingested instead of real food. Sometimes becomes death traps [4].

Figure 1 - A turtle trapped in a net [5].

Figure 2 - Plastic effects on a turtle [5].

A threat, of which there is much investigation, are the microplastics, which come from many different sources, because, are a very real risk for the marine ecosystem. Very often, these fragments, in the order of 5 mm, by other fishes are mistaken for jellyfish or squid. In this way, fishes eats plastics, and as the consequences, these kind of plastics ends to be ingested by other bigger fishes, or by humans, or in the worst of the hypothesis, fishes and turtles end to dead. The sources are more than one. Microplastics derive from industrial wastewater, from toothpastes, from body scrubs, synthetic fibers and mostly by the run-off. A recent study "Invisible plastics" realized by Orb Media, highlights a global contamination also of drinking water, since 83% of drinking water samples taken in all parts of the world are contaminated [6].

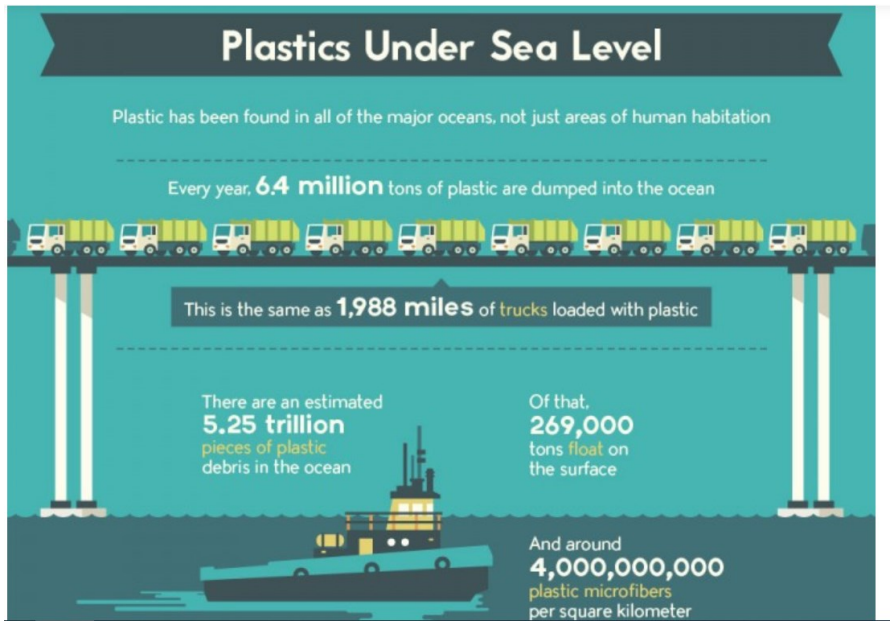

Figure 3 - Plastics Under Sea Level (a) [7].

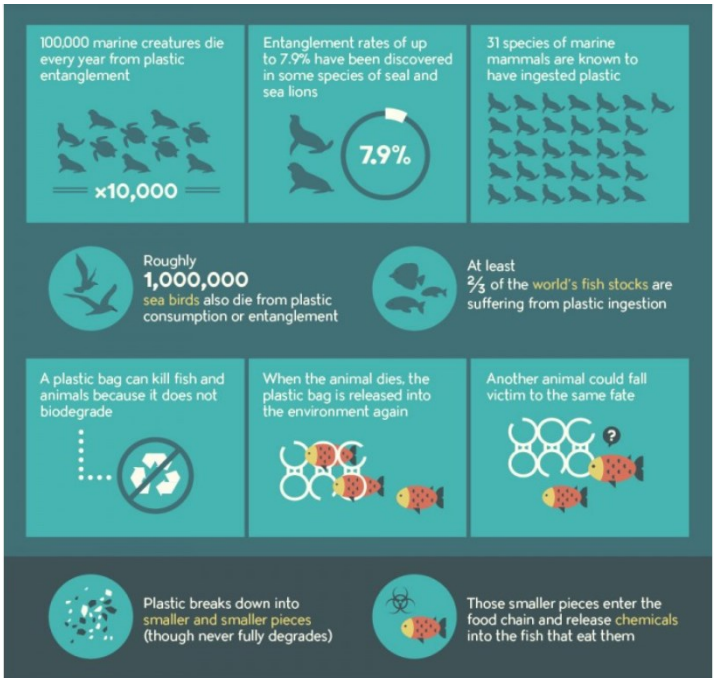

Figure 4 - Plastics Under Sea Level (b) [7].

Figure 3-Figure 4 represent how the plastic is nowadays part of the feed chain, determining physical and toxicological effects on marine organisms and human health. A new approach is proposed in [8]: the authors apply a fluorescent dye that is adsorbed onto plastic surfaces and renders them fluorescent when irradiated with blue light. Using an orange filter, and using image-analysis process, the fluorescent particles are easily identified and then counted. The researches proposes cross-validation by comparison with infrared (IR) microscopy. Microplastics of different sizes could be detected and counted in marine sediment samples.

Within decades, a million terrestrial and marine species could become extinct. It is therefore very important to understand the causes and monitor what happens underwater. Marine biologists on Crete warn that dramatic changes are already taking place and could spread across the entire Mediterranean Sea [9]. These scientists note that biodiversity is changing rapidly especially in shallow ecosystems. They describe a situation in which the seabed is increasingly depopulated. The method of analysis foresees a monitoring of these changes, and the best methodology is to dive and observe the species and quantize them. In this way, after a decade they will be able to know how the situation has changed from before to after. The underwater caves of clay are rich in sponges and corals, which can be subject to stress. Images taken underwater are analysed using special software to identify and classify the species. Mediterranean Sea is used as an example, due to its high human activities and it tourist rate, in order to replicate the analysis also for others sea. It should be borne in mind that if a species is affected by the change, the entire marine balance will also be compromised.

A great threat to the benthos community are the powerful of the meteo-marine events, closely related to global climate changes [10], that are drastically changing the biological richness and structure of the community. Examples of global climate changes are tsunamis, storms, cyclones and other physical disturbances. The big problem is because these kind of episodes are more and more frequent. For example in this paper work, the authors take reference on Liguria Sea. It has been a strong destructive storm in fall 2018, accompanied by South East wind exceeding 130 kmh^{-1} that generate 10-m high waves. This complexity of event hitting the coast of Ligurian Sea, devastated and damaged, creating a profound changes to the coastal morphology. Being the cliffs of the Portofino an ideal habitat-forming gorgonian, soon after the storm it was very interesting to surveys the area. In this way, being this area explored in the past, in particular in 1997, 2002 and 2016, therefore it was possible to compare its health status before and after the impact of the storm.

Another element that affects the marine environment are seaports. Sometimes is possible that the new species, hooked to the hull of boats, they enter the new place, establish there and later compete with native species. To monitoring this problem, the researches uses specialized plates in many European locations. Passing months, the plates are colonized by little marine species, and this is useful to collect data.

Another reason for which is important to monitor the deep sea is that it is unknown at 99%. Claire Nouvian at TEDxAUCollege conference says that a new specie is discovered each two weeks [11]. Several questions must to be answered related to deep-sea animals. It must to be considered that many species do not need to light and so they are still unknown. Great consideration must also be given to corals, which act as real underwater trees on whose branches fish eggs are deposited (Figure 5) [12].

Figure 5 - Deposited fish eggs [12].

The ships, from 1980 to 1990, increase their fishing and went deeper to the ocean meeting animals that have never met before. People starts to eat fishes which are never been eat before. Everything is done in order to not safeguard the deeper marine environment. They use to fish with trawl nets bringing up corals and not considering the environmental damage that they are causing [11]. Most of the time the fishermen do not keep the corals with them, but only select the fish that get entangled between the corals and the net and then throw the coral back into the sea. Dump the corals from the boat causes their breakup (Figure 6) [13].

Figure 6 - Overfishing example [13].

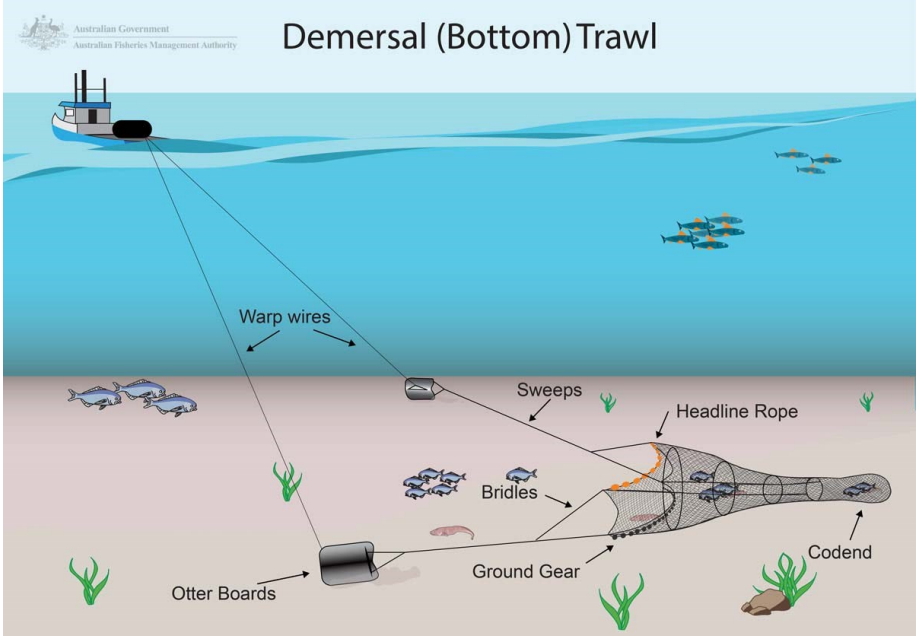

Figure 7 – Demersal (Bottom) Trawl [14].

As it is seen in Figure 7, the Demersal (Bottom) trawl, crawl on the bottom and collect everything they find on the sea floor [14]. The net has an opening, for the entry of fish, armed with weights with the function of moving the sediment and making fish and other animals come out. The ship must travel at a speed of 2.5 to 4 knots in order to operate the system. For what concern the bentonite nets

there are several types for example the most common is the “paranza”, instead the smallest one is the “sogliolara” used for mussels, clams and flat fishes. An even smaller net is “Gangamella” which is trailed nightly in the Oceanic Poseidon to catch shrimp. The notoriety of this type of fishing is due to its bad environmental impact. The nets destroy and collect everything. In Italy, this kind of fishing is avoided.

Another one of the most serious threats to the sustainability of the species that populate the seas is the overfishing. Each years from 11 to 26 tons of fishes are illegally fished. This typical problem affects the Mediterranean Sea that is going into decline. In order to guarantee the future of the Mediterranean sea and its ecosystem is important to reduce the fishing mortality, regulate the period of fishing in based on fishes sexual maturity, avoid fishing for unwanted species and increase effective surveillance and control operations [15].

In Italy it was deciding to avoid the bottom trawl specially under the coast where the more sensible species creatures live [14]. To preserve the protect area it is suggested to use several stratagems. One of these is the use of concrete blocks in areas of particular biological interest. They are blocks with steel hooks having a double function. The first one is to damage fishing nets and gear; the second is to create support for benthic organisms by increasing their biodiversity. It is also important to know that, in order to safeguard marine fauna, the reproduction periods of the main organisms being fished must be respected. Connected to this point is also the illegal fishing of the mussels [16]. Breaking the rocks where the bivalves lives causes permanent changes in the substrate characteristics. In addition others effects can be the removal of benthic species, passing from a highly structural habitats to a simplest one. That is what threats the Mediterranean Sea, because illegal fishing is widely applied in that area. In particular, in the area of Sorrento and Salento peninsula are suffering a great loss in terms of natural capital and ecosystem services (Figure 8) [16].

Figure 8 - Removal of a mussel [16].

Another big problem is the non-selectivity during the fishing activity. The bottom trawl collects everything even non-marketable fish. This kind of non-selective fishing is called bycatch and refers to valuable species and inedible organisms. Furthermore, it is preferable to choose nets with larger meshes so as not to fish too young subjects, preserving the natural reproduction of the species.

The Adriatic Sea becomes everyday less fishy and in this case the reason can be environmental [17]. Generally, the Adriatic Sea has always been full of fish but ultimately, especially in the upper part, there is a decreasing of the caught. It should be borne in mind that in the northern part of the Adriatic Sea there is the discharge of the river Po that is rich in nutrients, pollutants and freshwater flowrate. The overproduction of mussels and clams is due to the presence of the nutrients. The presence of eutrophic conditions, due to the high rate of nutrients, constitutes an excellent substrate for the growth of blue fish, particularly in winter, autumn and summer. Eutrophic phenomena when they develop during the summer period, accompanied by periods of anoxia and hypoxia, cause suffering to living organisms. It is therefore very important to plan and manage fishing activities in a sustainable way. The solution is a management plan, a biological stop period, fishing stop identification and preparation of areas reserved for restocking. It is also important to define at European level, what are the benchmarks and the performance indicators to establish the level of improvement has been reached. These indicators have been activated in order to monitor the human impact on the environmental system. Is in this way very important to verify the level of biodiversity conservation and the maintenance of reproductive capacity of fish stocks [18].

In the recent years, many people have been worrying about the health of the oceans, promoting various initiatives. The most important is the Agenda 2030. The Agenda 2030 for Sustainable Development is an action program for people, the planet and prosperity, signed in September 2015 by the governments of the 193 member countries of the United Nations [19]. Agenda 2030 is based on Sustainable Development, which is defined as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” [20]. It incorporates 17 Sustainable Development Goals (SDGs) - into a large action program for a total of 169 'targets' or milestones. Its beginning is marked in 2016 and it guides the world on the way to go for the next 15 years. For over two years of intensive public consultation and engagement with civil society and other stakeholders around the world, giving particular attention to the voice of the poorest and most vulnerable, in order to obtain as result the Goals and the targets [21]. The challenge is to achieve these goals by 2030. 2030 Agenda aims to end poverty in the world, social inequalities, guarantee education for all, safeguard the environment in all its forms, build sustainable development models, guarantee resilient infrastructures, promote peaceful societies and finally propose sustainable development.

This thesis work refers to Goal 14. Within 2020 the targets are the sustainably manage, protect and strengthen the marine and coastal ecosystems in order to avoid negative impacts. Regulate the harvest and put an end to overfishing, fishing illegal, undeclared, unregulated, and destructive fishing practices. Rebuild fish stocks in order to be capable of producing the maximum sustainable yield by their biological characteristics. Eliminate subsidies that contribute to illegal fishing. Improve the conservation and sustainable use of the oceans and their resources through the application of international law, which is reflected in UNCLOS (United Nations Convention on the Law of the Sea [22]), which provides the framework legal framework for the sustainable use and conservation of the oceans and their resources [23]. The commitment is to build a sustainable and resilient future for people and the planet. Environmental protection is one of the three elements to be harmonized together for the well-being of living beings and the planet. The other two are economic growth and social inclusion. To support national efforts, a revitalized global partnership at the global level is needed. This is recognized in 2030 Agenda. The 17 Sustainability Development Goals (SDGs) and 169 targets, at the global level, will be monitored and reviewed using a set of global indicators (Figure 9) [20].

Figure 9 - Sustainable Development Goals [20].

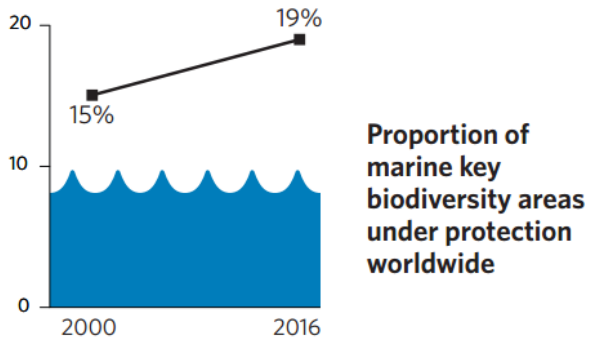

Figure 10 - Proportion of marine key biodiversity areas under protection worldwide [20].

Conserve and sustainably use the oceans, seas and marine resources for sustainable development. This Goal seeks to promote the conservation and sustainable use of marine and coastal ecosystems and to prevent marine pollution and increase the economic benefits to small island resources. People who live in coastal communities are the 37 per cent of the global population in 2010, so the marine resources are particularly important for them. An increase, from 15 per cent to 19 per cent, of protected areas, has been from 2000 to 2016, in order to preserve the marine biodiversity (Figure 10).

An aspect of fundamental importance for evaluating marine biodiversity is the populousness of living beings. It is also possible to valuate a comparison between the same species but in different areas. Growth is one of the most important indicators for assessing the health of individuals or the community as a whole. Due to changes in temperature, poor water quality, low availability of food resources and high population density, fishes sacrifice somatic growth. At the contrary, when the fishes are in good health, share high growth rates because they are enabled by ideal environmental conditions. Growth rate gives also information about the minimum catch size or type of fishing gear allowed. This allows you to verify the effectiveness of the survey over time.

Underwater video are used to study marine ecosystem mainly for observing animal abundance and behaviour in the context of environmental monitoring [24]. This kind of video provides detail information in a form that is very easy to be understand. Sometime is very necessary a level of automation to process large videos. Computer vision and machine learning provide automatic monitoring surveillance.

In order to monitor what happens in the underwater world, what are the variations both on a morphological and physical level, a great help is given by geomatics. Aerial digital photogrammetry, laser scanning and multibeam bathymetry can be a good support by geomatics to morphological reconstruction [25]. Underwater photogrammetry does not differ much from aerial or terrestrial ones. The only difference that could exist is due to the disturbances caused in particular by the refraction of the water glass dioptre and the presence of the case [26]. Due to the presence of biofouling and turbidity of the water, the operators who dive must work with their chambers at a very close distance ranging from 0.5 to 2 meters.

Since it works on large image datasets, it is often necessary to resort to very fast image recognition systems. Faster than the human eye and human brain can be only a machine equipped with artificial intelligence. The most useful technique for this cause may be object detection. This technique represents the possibility to identify an object inside a video or image, thanks to the identification in the box and a label associated with it. It also then possible to classify the objects in group of objects

with the same labelling and if sometime there is a classification hierarchy, there is a chance to do a hierarchical tree.

In this thesis work, innovative methodological approaches for marine communities assessments have been studied and applied on real case studies. A multidisciplinary methodology has been applied, including geomatics, biology and computer science aspects. Geomatics allowed a reliable design of diving surveys oriented to underwater datasets creation; furthermore, thanks to geomatics theoretical aspects, a critical evaluation of existing datasets has been performed. Thanks to biology, marine communities have been correctly identified. Finally, computer science allowed the processing of the obtained underwater datasets and the targeting of pilot results that open a wide multidisciplinary action field.

1.1 Thesis Organization

The present thesis work has been organized as follows: Chapter 1 reports the introduction, focusing on critical aspects related to underwater environment; the motivation of this thesis work is highlighted. The main theoretical concepts applied in the current work have been reported in Chapter 2: an overview of the geomatics, biology and computer science main theoretical aspects has been proposed. In Chapter 3, an extended literature review related to underwater geomatics, to underwater monitoring and to underwater computer science applications has been reported. Finally, the main research topics assessed in the thesis work have been presented. Case studies and results have been presented in Chapter 4, while Chapter 5 reports conclusion and future work. A brief description of the data repository, software and devices that have been used in this thesis work has been reported in the Appendix.

2. Theoretical Background

In this chapter, a theoretical background on geomatics, biology and computer science is reported.

2.1 Geomatics

A large amount of data elements must be inevitably organized, processed, handled and then used for a correct representation of the territorial situation; both marine and terrestrial. These data must be processed in an interdisciplinary and interoperation manner, and the geomatics is the discipline able to connect these requirements. Observing the etymology of the word "geomatics": geos (Earth) and matics (informatics). This discipline is born with the purpose of to increase the potential of computing in order to revolutionize surveys and representation sciences that requires the treatment of huge amounts of data. The revolution refers on giving a geographical location of each object on our planet. In other words Geomatics can be defined as a systemic, multidisciplinary, integrated approach to selecting the instruments and the appropriate techniques for collecting, storing, integrating, modelling, analyzing, retrieving at will, transforming, displaying and distributing spatially georeferenced data from different sources with well-defined accuracy characteristics, continuity and in a digital format [27].

Some of disciplines and techniques that constitute geomatics are:

- Computer science: to represent and process applicable information through the development of technological instruments (i.e. hardware) and of methods, models and systems (i.e. software).
- Geodesy: to determine the shape and size of the Earth; it defines on the one hand the surface of reference in its complete form, the geoid, as well as in its simplified form, the ellipsoid, and on the other hand the external gravitational field as a function of time.
- Topography: started with and part of geodesy, this is a combination of procedures for direct land survey. Topography is a combination of methods and instruments to comprehensively measure and represent details of the Earth's surface:
    - Planimetry: to determine the relative positions of the representation of points on the Earth's surface with respect to the same reference surface.
    - Altimetry: to determine the height of the points on the Earth's surface with respect to the geoid surface.
    - Tachymetry: for the planimetric and altimetric survey of the Earth's surface zones.
    - Land surveying: to measure areas, moving and rectify borders, levelling zones of the Earth physical surface.

- Cartography: to supply a possible description of the shape and dimension of the Earth and its natural and artificial details, by means of graphical or numerical representation of more or less wide areas, following fixed rules.
- Photogrammetry: to determine the position and shapes of the objects by measuring them on photographic images.
- Remote Sensing: to remotely acquire territorial and environmental data and to combine methods and techniques for subsequent processing and interpretation (this definition also fits digital photogrammetry).
- Global Positioning System (GPS): to provide the three-dimensional (3D) position of fixed or moving objects, in space and time, all over the Earth's surface, under any meteorological conditions and in real time.

Geomatics uses terrestrial, marine, airborne and satellite-based sensor to acquire spatial and other data. Data are the basis of information and in general its represents the measure of the external world. More of the time the two terms data and information are used as synonymously but in reality, there is a deep difference between them. According to some rules, an expert, both human or not, is able to convert data into information. The acquisition of information goes via a cognitive process based on data [27].

2.1.1 Photogrammetry

Photogrammetry is defined as the process of deriving metric information about an object through measurements of the object made on photographs, leaving to photointerpretation (by human visual analysis) the task to obtain qualitative information (human experience remains a determinant factor) [27]. In the photogrammetry technique, having two images captured from two different points in the place, is possible to reconstruct the shape and the dimensions of an object, according to the 3D coordinates points [28]. In other word, Photogrammetry guarantees the geometric correspondence between the real object and its representation based on photographic images [27]. The measurements allowed by photogrammetry, can be performed in two and three dimensions (2D and 3D), exploiting both photograms acquired by traditional photogrammetric cameras and digital imagery. Remote sensing (RS) is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft). Sometimes the term RS is used instead of photogrammetry because is considered as the first remote sensing technology.

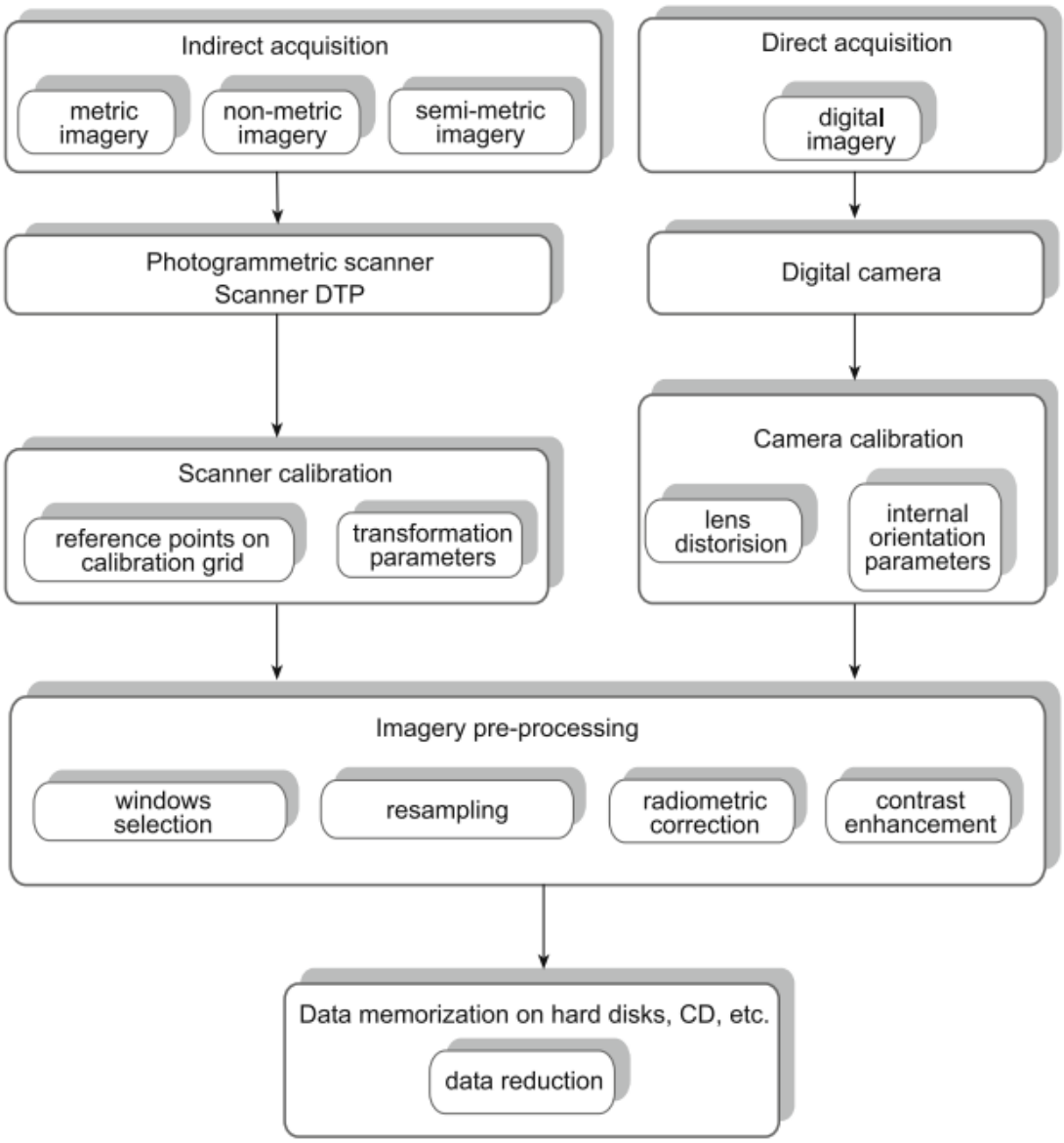

Figure 11 - Acquisition unit of a digital photogrammetric system [27].

When a computer manages the imagery processing steps, photogrammetry is defined as "digital photogrammetry" (Figure 11). Is an example the case in which frame images are generated by digital camera. In digital photography, the image is formed from an analogical electric signal, which is transformed into a pixel by an analogical–digital converter through a sampling process. The pixels are assigned three different values of Digital Number (DN), corresponding to red (R), green (G) and blue (B). A processor's screen uses RGB values to reproduce the color of each pixel [27]. Some of the advantages of digital photographic acquisition is quicker availability of the image with the same acquisition time and the digital processing accessibility. Advantages of digital photogrammetry are for example the reduction of instruments in size and cost and the reduction of work time due to the automation of some phases. On the other hand, the cost of the software is very important and is the

same for the requirements of the instrumentation. If the images are acquired from a camera, is the case of direct image acquisition.

What is essential for the acquisition of the primary metric information object is the camera. Metric accuracy depends on the camera because it is responsible of the metrics information letting their utilization without any distortions [28]. For example, if the final result is a cartography, the accuracy depends on method of implementing the taps, flight altitude, aircraft speed and many other "primary acquisition" parameters. With the introduction of technologies for image data acquisition from space in a wider region of the electromagnetic spectrum, the meaning of photogrammetry and photo-interpretation has been extended to comprise remote sensing, moving from classical photo-interpretation to the use of digital image processing in addition to human interpretation, and applying computer analysis techniques to imagery besides photography [27]. The positioning systems applied to points on the Earth's surface allow the 3D positioning of static or moving objects in space and time, in every place on the globe, under all meteorological conditions and continuously. They are based on the reception of radio frequency signals emitted from artificial telecommunication satellites. The ground station must be equipped with an antenna and a receiver: their degree of complexity and cost depends on the measurement level of accuracy needed to determine geocentric coordinates (WGS84) of any point on the Earth's surface [27]. The term Spatial Data Information was officially introduced as information related to the terrestrial Globe in 3D space. The common definition Geographic Information (GI) is used to indicate everything concerning the 3D positioning and georeferencing of objects on our planet. Spatial data is playing an increasingly important role in 3D scene and geological modelling. There are great differences in various research fields regarding data acquisition techniques, morphological descriptions, and application purposes of spatial objects [29].

Instrumental ground observation permits non-destructive and proximal sensing. Using traditional photographic instruments, digital cameras, video cameras in the visible and near infrared, thermal cameras for studies in the thermal bands, radiometers and spectroradiometers, it is possible to obtain images and measures of the objects and study their condition in real or near real time [27].

Image transformation can be on single band or multiple bands (multispectral). Single-band transformation is based on algorithms (e.g. algebraic operations) independently applied to each one of the bands acquired by a sensor. Other single-band transformations are the operations carried out in digital image processing techniques, such as the modification of the histograms enhancing the contrast in image visualization, and the digital operators, or filters. In multispectral transformations, bands can be combined to produce, from the originals, synthetic bands, also called pseudo-bands.

These transformations can generate new images with higher information used in the interpretation phase or as processing before an automatic classification [27].

2.1.2 Photogrammetry in marine environment

The beginning of the underwater photogrammetry started at the end of the XIX century. A maritime engineer, tried an inverse camera, completely immersed and with an air lens; (tesi 5) the development operations turned out to be conducted without difficulty, as if the photograph had been taken on land. The innovation that most contributed to the development of underwater photogrammetry was the ease given to the diver by SCUBA (Self Container Underwater Breathing Apparatus) useful for human respiration (Figure 12) [30].

Figure 12 – SCUBA equipment [30].

In the past, the photogrammetry was only for war uses and cartography. Going on, it is widely used for the topographic survey and in other sectors as the engeneering, the architectural, and the geological one. Although apparently photography seems sufficient in providing information, the difference is that photography only gives a qualitative descriptive approach. If you wanted to know, for example, the distance between two points, you would need metric information thet is a qualitative metric approach. Such an approach can be found in the art photogrammetric, which then combines both these two approaches. In photogrammetry, the first phase is the survey, the second is the plotting and the last one is the field completion. The plotting phase is the process of elaboration of the three-dimensional model starting from frames taken in the taking phase. It consists in reproducing the path of the rays in the opposite lights direction that impressed the sensitive emulsion, through projectors geometrically identical to the camera used and arranged in the same positions who had the camera at the time of shooting. For the reconstruction of the optical model it is therefore necessary to restore the orientation inside and outside of the photographs, it is necessary to place the frames in the identical

ones locations where the film or sensor was at the time of shooting. Internal orientation is carried out by orienting the frame in the device of projection in the identical position it had with respect to the lens of the camera. By external orientation, on the other hand, we mean operations aimed at identifying position of the frame with respect to the photographed object, orienting the projector as the video camera; for this operation, it is necessary to know the coordinates of the gripping point. The tool used for processing the collected data in order to obtain information descriptive metrics of the detected object, takes the name of restitutor or stereorestitor [30].

2.2 Artificial Intelligence and Machine Learning

Alan Turing, a famous mathematician, proposed the interesting question "Can Machines Think?" (early 1950s). This question, proposed by Alan Turing in the paper "Computing Machinery and Intelligence" represented the main idea of artificial intelligence [31]. The simulation of human intelligence processes by machines is called artificial intelligence: it can represent the ability of the computer intelligence to learn itself to solve problems. Marvin Minsky, John McCarthy and two senior scientists, Claude Shannon and Nathan Rochester of IBM, organized the Dartmouth Conference of 1956. At this conference, the expression "Artificial Intelligence" was officially coined as the title of the field [32].

In the year 1959, Arthur Samuel, one of the experts in machine learning, defined machine learning as "The field of science that gives computers the ability to learn without being explicitly programmed" [33]. In more simple terms, Machine learning can be defined as a field of science involving algorithms to give computers an ability to learn and behave like humans by feeding the algorithms with relevant data. The main goal of a machine learning algorithm is to learn the features from the provided training data and generalize beyond the data which has been used for training [34], [35], [36].

2.2.1 Deep Learning and Neural Networks

Deep learning is a sub-field of machine learning that teaches computers what to do by replication human learning process i.e. learning by example [37], [38]. It mainly deals with developing algorithms, which are inspired by the functioning and structure of a human brain [34], [35], [36]. Even though deep learning was initially theorized in the 1980s, in recent times deep learning is gaining a lot of attention because it is achieving higher accuracy than ever before sometimes also exceeding human performance; furthermore, deep learning requires a large amount of labeled data for training. Finally, it requires large computing power. Nowadays, GPUs and cloud computing clusters are being used to substantially reduce the training time [34], [35], [36]. These deep learning methods usually follow neural network architecture so deep learning algorithms are often called as

deep neural networks [34], [35]. Deep learning includes a set of mathematical models that are very efficient for the solution of complex problems [36].

The logistic regression represents the training algorithm that is the substratum of a neural network (Figure 13).

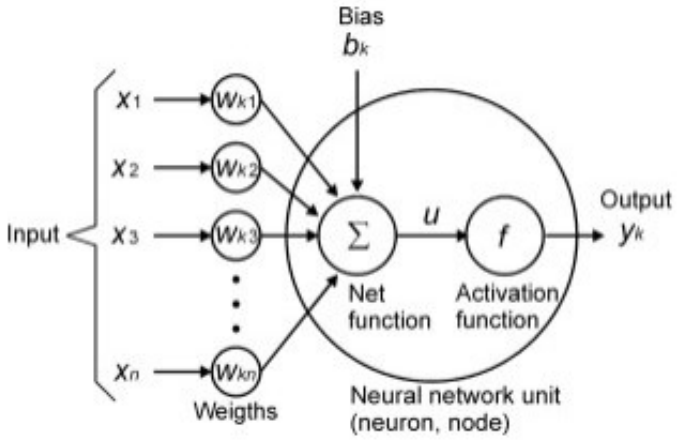

Figure 13 - Logistic regression structure.

Given the input vector x, its features are multiplied by the weights w, adding a bias b to the result. The output value of the logistic regression is obtained applying the activation function to the previous result. The activation function performs a regularization of the result.

An example of a problem that can be solved through a logistic regression is: given the real result y of an experiment, compute an estimation $\hat{y}$ through the model $\hat{y} = W^T x + b$. x represents an input vector (known) and W, b represent the model parameters.

A training procedure is needed in order to tune the model parameters on x, y data (usually called labeled data). The training procedure is composed by two phases: a forward propagation and a backward propagation. In the forward propagation, the $\hat{y}$ value of the current iteration is obtained, while in the backward propagation the obtained estimation error is processed in order to adjust the model parameters for the future iterations (using the derivative concept). The optimization problem stops when the estimation error is negligible, i.e. the model has reached a satisfactory training level. The learning rate parameter regulates the update speed of the model terms. In the training procedure, generally also y is known. The trained model can be then applied on data that are not previously classified.

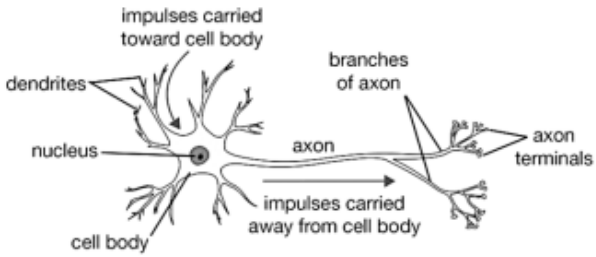

Figure 14 - Biological structure of a neuron [34].

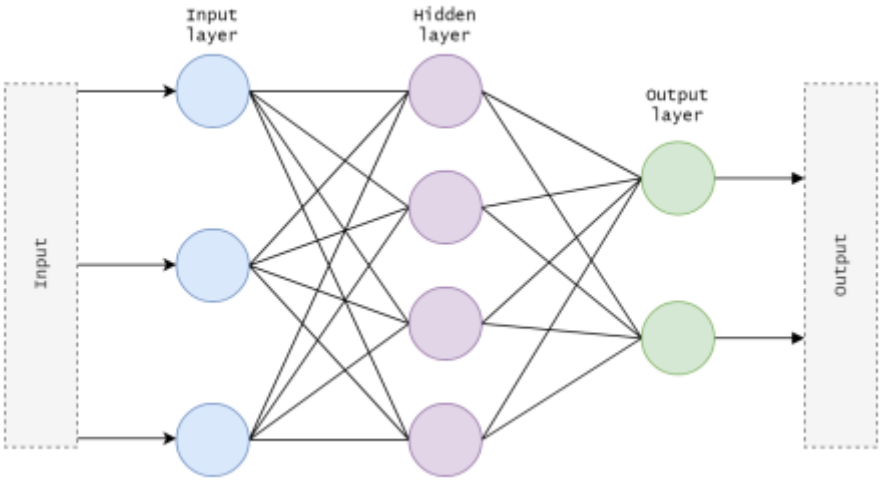

Figure 15 - Neural network architecture example [34].

In the biological structure of a neuron (Figure 14), the dendrites bring the signal to the cell body and then all the received signals are summed. If the sum value is greater than a threshold value, the neuron will fire the signals that means that the signals are carried away from the cell body on to the next one. In a similar fashion, when a computational neural network is being trained the nodes (neurons) will fire whenever it learns specific patterns from the input data [39].

A neural network can be defined as the composition of many logic units (Figure 15). The architecture of a neural network is divided in different levels. Each level has a defined number of logic units, an activation function and its specific model parameters (W, b). The Input Layer is responsible for transferring the input information to the next layers. All the necessary computations are performed in the Hidden Layers and then the signals or the information is passed on to the following layers. In the Output Layer an activation function is used that will convert the signals or information received in the desired format. An activation function is generally used to define an output of a certain node on a set of inputs [36].

Based on the problem to be solved and to the computational needs, appropriate structures exist: there are neural networks with a complexity that develops mainly in the cardinality of the logic units of each level or in the number of levels [36].

As previously explained, the training procedure consists in the adjustment of the parameters of each logic unit. In particular, the training takes place in the backward propagation phase. Different training procedure approaches exist [36], [40]:

- Supervised Learning: the algorithms are fed in with input data and correct output data (correct answer). The training data containing both input/output pairs are termed as labeled data. Classification and regression problems are Supervised Learning problems [34], [35].
- Unsupervised Learning: unlike supervised learning, in this approach, the training data is not labeled. The algorithms look for hidden patterns and structures in the training data [34], [35]. Clustering and association problems are Unsupervised Learning problems [34], [35].
- Semi-Supervised Learning: this approach considers as inputs only partially classified data. Generative neural networks and recurrent neural networks (RNNs) adopt this approach [36].
- (Deep) Reinforcement Learning: it focuses on taking suitable actions based on a certain situation in an environment to maximize the rewards. The training data in reinforcement learning does not have any correct answer key like supervised learning, the algorithm itself makes decisions intelligently to complete the given task [34], [35].

Several types of neural networks exist:

- "Standard" Neural Networks: they are characterized by a simplified architecture and they are exploited for the classification/elaboration of structured data.
- Convolutional Neural Networks (CNNs): Computer Vision adopts CNNs; in these networks, convolution function is used as model.
- Recurrent Neural Networks (RNNs): they elaborate data derived from previous computations.

2.3 Computer Vision and Object Detection

The automated extraction of information from images is called computer vision. The target is represented by grouping and searching image content exploiting information, e.g. camera position, object detection, object recognition, 3D models [41]. Computer vision can be defined as a field where algorithms are studied in order to help computers see, learn and understand the content of digital data [42]. Computer vision includes different problems, e.g. image coding, recognition, segmentation, detection: during the years, different techniques have been developed for the solution of each problem [43], [44], [45].

One of the main problems faced by computer vision is the object detection. Object detection consists in identifying instances of objects within an image and classifying them as belonging to a certain class (such as humans, animals or cars) [46]. The goal is to develop computational techniques and

models that provide one of the basic elements necessary for computer vision applications: answer to the question "what objects are here?". Many applications for computer vision (segmentation, image captioning, object tracking, …) depend on object detection problem.

2.3.1 Object Detection Models

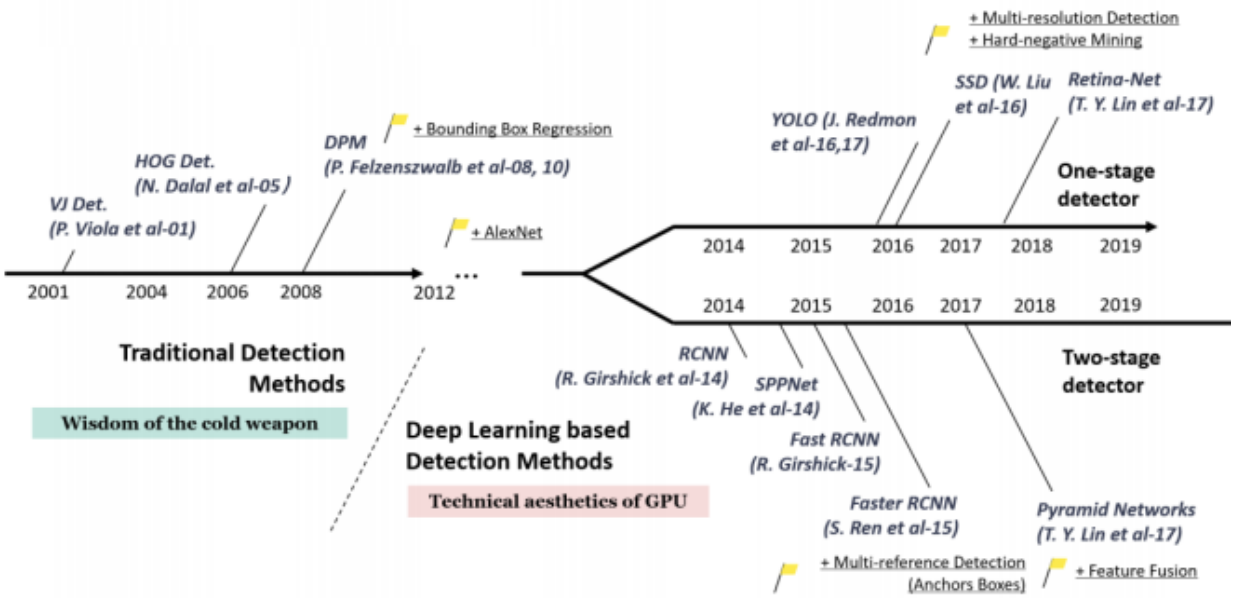

Figure 16 - Timeline of the evolution for object detection models [36], [46].

Object detection models can be divided into two-stage and one-stage detectors. Two-stage detectors adopt a “coarse-to-fine” policy, since they perform the detection in multiple phases. One-stage detectors try to complete the detection in a single step using a single network. As shown in Figure 16, exist different object detection models for each of the two categories.

Region-based CNNs models (R-CNN) first select several proposed regions from an image (for example, anchor boxes are one type of selection method) and then label their categories and bounding boxes (e.g., offsets) [47]. Then, they use a CNN to perform forward computation to extract features from each proposed area. Afterwards, we use the features of each proposed region to predict their categories and bounding boxes (through a Support Vector Machine (SVM)) [48].

Spatial Pyramid Pooling Networks (SPP-Net) have a convolutional neural architecture that employs spatial pyramid pooling to remove the fixed-size constraint of the network [49]. Specifically, an SPP layer on top of the last convolutional layer is added. The SPP layer pools the features and generates fixed-length outputs, which are then fed into the fully-connected layers (or other classifiers). In other words, some information aggregation at a deeper stage of the network hierarchy (between convolutional layers and fully-connected layers) is performed to avoid the need for cropping or warping at the beginning [50].

As in the R-CNN detector, the Fast R-CNN detector also uses an algorithm like edge boxes to generate region proposals [51]. Unlike the R-CNN detector, which crops and resizes region proposals, the Fast R-CNN detector processes the entire image. Whereas an R-CNN detector must classify each region, Fast R-CNN pools CNN features corresponding to each region proposal. Fast R-CNN is more efficient than R-CNN, because in the Fast R-CNN detector, the computations for overlapping regions are shared [52].

The Faster R-CNN detector adds a region proposal network (RPN) to generate region proposals directly in the network instead of using an external algorithm like edge boxes [53]. The RPN uses anchor boxes for object detection. Generating region proposals in the network is faster and better tuned to the input data [52].

YOLO (You Only Look Once) detector has been the first deep learning one-stage detector [54]. The YOLO authors do not use the pre-existing two-stage paradigm "proposal detection + verification": YOLO applies a single model to the whole image. The image is divided into different regions and then bounding boxes are predicted; for each bounding box, YOLO determines the probability that the considered bounding box belongs to a certain class. The whole procedure is executed based on a single network. Updated YOLO versions have been the created, obtaining an accuracy improvement while maintaining a high computational speed [55], [56].

The main drawback of YOLO is the reduced accuracy in the objects localization when compared to two-stage detectors (especially on the little objects). The successive versions of YOLO and Single-Shot Detector (SSD) [57] tried to solve this problem. Other types of recently developed detectors are Feature Pyramid Networks (FPN) [58] and Retina-Net [59]. SSD and Retina-Net are one-stage detectors while FPN are two-stage detectors [36].

3. Literature Review

During the last years, many technologies have been developed in order to monitoring the seabed of the oceans. One of the reasons why people are nowadays interested to monitor the marine fauna is to protect the marine ecosystem and to predict fishes productivity for human community and to improve the conservation and management strategies to ensure their sustainability [60]. In order to explore and maintain the wide biodiversity and life of underwater ecosystems, monitoring and subsequent analysis of the information collected is necessary [61].

At the same time, the development of artificial intelligence oriented to image treatment for animal counting, classification and other computer vision problems requires the development of analysis tools for ecological annotation, as well as semantic infrastructures for combining datasets and recognition and classification algorithms [61]. Machine learning is a subset of artificial intelligence that enables to take decision based on data. Artificial intelligence makes possible to integrate machine learning capabilities into data driven modelling systems in order to bridge the gaps and lessen demands on human experts in oceanographic research. Machine learning algorithms have proven to be a powerful tool for analysing oceanographic and climate data with high accuracy in efficient way. The main application of machine learning in oceanography is prediction of ocean weather and climate, habitat modelling and distribution, species identification, coastal water monitoring, marine resources management, detection of oil spill and pollution and wave modelling [62].

Photogrammetric methods can be applied in order to ensure optimal trade offs between experiments cost, experiments invasion and goodness of the acquisitions. Large amounts of images or movies cannot all be manually processed, and autonomous routines for recognizing the relevant content, classification, and tagging are urgently needed [61]. It is therefore necessary to use an automatic process that can drastically reduce the time of detection respect to human detection. To do this, scientists and biologists use artificial intelligence techniques.

Object detection problem has drawn considerable attentions in both marine engineering and aquatic robotics. However, unfortunately, to better investigate, it is necessary consider a pre-processing phase using enhancement methods. Underwater environment has complicated conditions due to the bad lighting conditions and due to the presence of biofouling. These reduce visibility and limit many practical applications of underwater images and videos in marine investigations. So many underwater image enhancement (UIE) algorithms are involved in order to improve the performance of underwater object detection [63].

A thing that is common for state-of-the-art detection methods and deep learning methods in particular,

is that they are in need of relatively large amounts of training data. During the years, the realization of complex and reliable datasets became an additional challenging research target [64].

3.1 Image Enhancement

Underwater image enhancement has been attracting much attention due to its significance in marine engineering and aquatic robotics. Numerous underwater image enhancement algorithms have been proposed in the last few years. Both image processing and underwater vision exploit image enhancement capabilities. The underwater effects reduce visibility, decrease contrast, and even introduce color casts, which limit the practical applications of underwater images and videos in marine biology and archaeology, marine ecological, to name a few. To solve this problem, earlier methods rely on multiple underwater images or polarization filters, while recent algorithms deal with this problem by using only information from a single image [65].

A significant role on underwater images enhancement has been played by the iMARECULTURE project [66], [67]. Along the same water column, there is a disomogeneous propagation on light that is attenuated differently according to its wavelength, and to the distance between the object and the point of view. It is noted that a blue/green colour dominance the underwater imagery. It is surely possible to improve the light condition by using an artificial light, but at the same way, it does not illuminate uniformly producing bright spots in the images. iMARECULTURE project has the aims to develop new tools and new technologies to improve the public awareness of underwater cultural heritage. A software tool has been developed for enhancing underwater images. It is very useful for automatically processing a dataset of underwater images. It uses a set of image enhancement algorithms, in particular it implements five algorithms (ACE, CLAHE, LAB, NLD and SP) that propose different approaches for the resolution of the underwater image enhancement problem. The first algorithm is the Automatic Colour Enhancement (ACE) algorithm. It is a very complex technique that the authors employed using a faster version. They adjusted two parameters to tune the algorithm. In particular, they take care of an α parameter related to the strength of the enhancement: the larger is the parameter, the stronger is the enhancement. The second is the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm that is an improved version of the Adaptive Histogram Equalization (AHE) algorithm. The oldest and newest version has the same aim to improve the standard histogram equalization. The improvement is that the CLAHE prevents the over-amplification of noise that at the contrary the AHE can generate. Two parameters are provided in order to control the output of this algorithm: the tile size and the contrast limit. The third one is the Lab Color Correction (LAB) algorithm. It is based on the assumptions of the grey world and uniform illumination of the scene. It increases the contrast by performing histogram cut-off and stretching and

the it covert the image back to the RGB space. NLD (Non-Local Image Dehazing) method is based on the assumption that colours of haze-free image can be very well approximated by few hundred distinct colours. A Screened Poisson equation has been applied to the last algorithm (SP) that means Screened Poisson Equation for Image Contrast Enhancement. In this case, the Screened Poisson Equation is applied separately to each colour channel, together with the simplest colour balance with a variable percentage of saturation as parameters.

Other types of enhancement methods also exist. Fuzzy Set Theory Image Enhancement Method is one of these [68]. This refers for example when a pixel should become darker or brightener respect to how it is already.

A big challenge is the caustics removal effect [69]. The authors propose a new novel solution based on two small and easily trainable CNNs: SalienceNet and DeepCaustics. This solution is implemented in shallow water, with deep variations between 0.5 and 1.5 meters. Major problems such as limited natural illumination is present mostly for deep water. On the other hand, shallow water poses and additional problem for divers during the photogrammetric acquisition phase, in particular it refers to the greater difficult to control the buoyancy. Must not be forgotten how the waves can affect the stability of the divers and camera.

A different approach in terms of image enhancement is carried out in 2019 by the authors of [70]. Their approach is a bit different from the previous above because they did not use a software but tried to apply a CNN called UWCNN. Being a light-weight network structure and effective training data, the UWCNN can be easily extended to underwater videos for frame-by-frame enhancement. Is taking place a wide use of underwater vehicles able to explore, recognize and interact with the marine environment. The drawback is the fact that underwater images and video seldom meet the expectation concerning the visual quality. Despite comparing this new approach to the oldest is evident the improvement of the UWCNN, it is also true that the goodness of this deep-learning based underwater image and video enhancement methods does not match the same success of the deep learn-based solution such as classification, segmentation, recognition.

3.2 Underwater Photogrammetry

Underwater photogrammetry increases its use thanks to the possibility to use fast and quite inexpensive software applications [71]. In this paper, the authors tried to show the effect of underwater ports in front of the photographic lens. Being the possibility to choose between two ports, an investigation on flat and dome port has been carried out, by the 3D modelling application. The experiments have been conducted in a semi submerged industrial structure. The results, after having

tested both the ports, proves that the flat port performs worse than the dome. The flat port provide high image residuals, lower precision, and lower accuracy in object space.

At the same time, the use of camera underwater is very different respect to the terrestrial use [72]. In underwater use, it is involved the water presence that present a pressure housing in front of the camera as if is and additional optical element. For underwater photogrammetry, the spherical dome port is suggested. This port is very expensive, but it is also the most useful because it keeps the main geometric characteristics of the lens unchanged.

The authors of [73] put the attention on the metric evaluation of different off-the-shelf camera systems for making high resolution and also high accuracy measurements, in order to monitoring along the time, how the coral reef varies and trying to measure these variations in the range of few cm per years. A comparison between high quality and low-cost systems has been made. The same it has been done between the dome ports and flat ports or between fisheye to moderate wide angles. To assess the accuracy of the photogrammetric models, tests are repeated at different camera to object distance in order to investigate how the distance induces errors. Within a test area of 5 meters x 5 meters, 9 points have been established. A photogrammetric coded target has been placed on the top of a 30 cm high pole; in this way, it is possible to be visible during the image recording process. It is also useful for an automatic recognition and image coordinate measurement. The image data processing has been conducted with two free network self-calibrating bundle adjustment approach, using Agisoft Metashape [74] and DBAT [75].

In order to establish a good performance for the flat port, a group of scientists tried to carry out some experiments both in sea and in the pool [76]. Starting from the point that using a flat port means having a clearly degradation of the accuracy, they tried to solve this negative aspect. The authors tried to mitigate the change of the characteristics of the images due to the presence of the curvilinear distortions and optical aberrations. The use of their stochastic model given benefits in terms of accuracy improving up to 50%. They wanted to demonstrate that homogeneous weighting of image observations might be not the right choice. Probably, varying the image weight, can achieve a better interpretation of the stochastic model. In order to demonstrate that a different weighting is beneficial, they use seven different datasets.

The authors of [77] propose a multi-temporal underwater photogrammetric survey located in French Polynesia, in order to detect the coral growth estimated to be 10-15 mm/year. For the authors is very important to have a high level of accuracy and a good resolution, ensuring the possibility to the surveys, having the same reference system, to repeat the same survey over the time. The final accuracy to which the authors are reached is in the order of 1 cm and few millimetres for 2017 and

2018 monitoring campaigns, respectively. Sometimes working in 2D models is not enough to study coral reefs because there are some parameters that need the 3D models. This complex model refers to the case of the morphometric characteristics such as rugosity, colony shape volume and surface area. In the recent years the acquisition of images by unmanned underwater vehicles (UUVs) and divers, are obtained thanks to the use of Structure-from-Motion (SfM) photogrammetry. It is a useful technique for the reconstruction of geomorphic features, archaeological sites and benthic communities. The authors establish a reference frame installing an underwater control network that is composed by benchmarks in fixed position. This let to have ground control points (GCPs) coordinates at subcentimeter level of accuracy, used as reference markers. A geodetic reference network is defined by means of several well identifiable points as GCPs. Each point is associated to its coordinate that, used in photogrammetry processing, provides metric content to generate a 3D models. A GCP must respect some simple rules. It should cover, horizontally and vertically, the investigated area, in a more homogeneous way possible. It must to be visible in a high number of images. It should allow topographic measurements of the network. The most common geodetic instruments are for example a GNSS (Global Navigation Satellite Systems) receivers or a total station.

In [78], the authors propose an image quality improvement in low-cost underwater photogrammetry. They claim that the underwater photogrammetry is a widespread cheap method for modelling underwater areas at different scales and at different levels of accuracy. The quality of camera sensor and the different possibility of its components for example lenses and domes let the photogrammetry be a technique applicable for different scenario. An alternative to an expensive camera are the GoPro and the Sony, that thanks to their underwater protection covers are easily comparable to a full-frame Digital single-lens reflex (DSLR) with dedicated underwater housings. The authors tried to correct a sever aberration seen in underwater images of GoPro cameras. They used two models to correct the blue and the red channel with respect to the green. They worked using two different dataset, one taken by GoPro and the second taken by a Panasonic Lumix GH4, in a test field of 25 square meters. Thanks to the CloudCompare software [79], the two dense point clouds are compared.

According to [78], also in [80] a cheap cost-effective method is proposed to generate a 3D point cloud. The authors present a work using the commercial software Agisoft PhotoScan [74], thanks to its easily, quickly and low cost equipment. The innovation is in the feature segmentation achieved based on colour information previously captured from the photographs. The feature segmentation separates an object into named features.

A long-term observation in benthic ecosystems in Antartctica using underwater photogrammetry has been carried out in [81]. Having information about community structure to be detected and quantified

is a priority in Antarctica. The diffusion of the underwater photogrammetry in the last period is justified by being a non-destructive and low-cost method for high-resolution topographic reconstruction. The authors wanted to show the validity and utility of this harmless technique in order to describe benthic community for a long period recording videos during standard SCUBA surveys. They did a first mission in 2015 when did their first transect survey, and the revisited the same area in 2017 [82]. By the photogrammetric procedure, they obtained a 3D model of the seafloor and inhabiting organisms. In total the area of investigation, cover more or less 200 square meters. Comparing the two tasks, the scientists noted some differences along the time. Firstly increase the number of urchins and secondly a completely disappearance of some sponges. In any case, leaving out the gravity of the results, the authors were able to demonstrate the efficacy of photogrammetry for monitoring.

Another big trend is the underwater mapping, but also in this case, a good technique is the underwater photogrammetry. It lets to use a non-metric camera in order to obtain a 3D reconstruction of coral reefs able to produce photomaps up to in 3D [83]. In this case, the test area is a rectangular area of 25 meters for 2 meters, in a depth of 5 meters. A geodetic measurement is carried out in the modelling process. Data was processed using 822 photos with an overlapping of more than 75%. The scale map resulted is 1:100 and the Ground Sampling Distance (GSD) is 0,395 meters. The image collection has been created using a GoPro Hero 4 Silver 12 MP. Because the coral on a reef has a tree shape with tree branch, the presence of a hole in the modelling still exists, although the massive corals create a model shape relatively very simple. The presence of the cavity require an increase of the number of the photos in order to obtain a complete 3D model.

3.2.1 3D Modelling

3D modelling represents an important step in underwater photogrammetry. Different techniques can be applied.

In order to estimate the density and the maximum height of the population structure and to obtain information about the maximal height, maximal width, planar surface and 3D surface, classification and segmentation techniques have been used in [84]. The classification has been performed using the CANUPO classifier available as plugin for CloudCompare software [79]. It is then possible to generate point clouds according to distinguish the background matrix and the gorgonians point clouds. The “gorgonian” point cloud has been useful for the computation of the colony abundance, density and morphometries. CANUPO processed the point cloud twice. The first time, it counts the gorgonian’s fans by approximating them to elementary planar objects and it calculates the maximum height, the maximal width and planar fan surface. In a second time, it divides the point cloud into a

mosaic having facets with maximal dimension of 5cm x 5cm. Starting from these facets, the 3D canopy surface was estimated as the sum of the individual facet surfaces. Gorgonian are very important because are indicators of the climatic anomalies on the coaralligeneous community. This paper work has been carried out to improve current predictions of the status of the coralligeneous habitat in Mediterranean Sea. This approach has been very relevant for the monitoring of slow growing and threatened species as gorgonian.

In [85], the authors propose the exploitation of a Structure-from-Motion (SfM) technique oriented to glass sponge monitoring in Northeast Pacific Ocean. The authors find as a challenge the possibility to work in different SfM workflows, based on the combined impact that dark, cold and turbid waters have on the veracity of the SfM. This paper work refers to the design, the development, the test and the deployment of an innovative underwater SfM workflow in order to generate high-resolution 3D models.

In [86] the authors work on a comparison between different software, in order to evaluate the performance of each of them in terms of 3D reconstruction models of submerged archaeological sites. In particular, they focused on Agisoft Photoscan [74], VisualSFM [87], SURE [88], 3D Zephyr [89] and Reality Capture [90] software. The particularity of this study is that the researchers do not focusing on the algorithms, but they evaluate the final results as the generation of 3D point clouds. Their dataset refers to an ancient shipwreck at 45 meters below the sea level, and is composed by 19 images, with a small camera to object distance of 1 meter, and 42 images with higher camera to object distance of 3 meters. In order to achieve their scope, the authors did a comparison between the numbers of total points, cloud to cloud distances, evaluating also the surface roughness, the surface density of the point cloud. As the final step, they evaluate the 3D metric in a way to establish which one performs better.

The authors of [91] carry out an example of a 3D modelling with a previous enhancement images processing. Divers collected the dataset and the scientists handled it and use it for geovisualization purposes. A combination of effects from manual editing of images radiometry (captured at shallow depths) and the selection of parameters for point cloud definition and mesh building (processed in 3D modelling software) are evaluated. The 3D models have been created at three different depths: one at 3.4 meters, one at 10 meters and the last one at 14 meters. Different have been also the set in which these images have been captured; in particular on the seafloor, on a part of a wreck and on a small boat's wreck.

The authors of [83] took place to a larger project relative to 3D measurements in order to evaluate the temporal change of coral cover in tropical waters. In particular, in the last years, scientists have

conducted several studies in order to evaluate the accuracy in reconstructing a 3D object model using a simple and low cost camera system. In this particular case, the authors propose a comparison of the accuracy between a 3D point clouds generated by using a camera acquired from a system camera mounted in underwater housing and a popular GoPro camera respectively. By a working distance of about 1.5 meters, the GoPro reaches an accuracy of 1.3 mm in the air and 2 mm in the water. On the other hand, the more complex system achieves an accuracy of 1.8 mm in the water that is higher respect to the GoPro.

In [77], the authors in order to obtain the construction of a 3D models of submerged objects and structures, suggest a more rapid and efficient procedure in terms of data acquisition and data processing. They present the possibility to use natural control points, signalized markets and tie points. The algorithms rely on the Structure from Motion (SfM) technique. As downside is, the limit tied to the high accuracy that at level of cm cannot be reached. In particular in the case of which are the corals the object to be investigated, it is necessary having a high accuracy, in order to monitoring the coral growth. On the other hand is very useful to monitoring structures with a length of ten meters or more, such as coral barriers, wrecks and long walls.

The authors of [92] use recent advances in computer vision and machine learning for the measuring of coral reefs structural complexity. The authors propose an approach to automatically classify 3D reconstructions of reef sections and assessed the accuracy of this approach. 3D reconstructions of reef sections were generated using commercial Structure-from-Motion software with images extracted from video surveys. To generate a 3D classified map, locations on the 3D reconstruction were mapped back into the original images to extract multiple views of the location. Several approaches were tested to merge information from multiple views of a point into a single classification, all of which used convolutional neural networks to classify or extract features from the images, but differ in the strategy employed for merging information. All approaches performed a high classification accuracy, proofing their suitability for many ecological applications.

3.3 Underwater Monitoring

The monitoring of the sponges in order to obtain information about ecosystem change is proposed in [93]. The peculiarity of the sponges is because they are morphologically diverse and lack indicators of annual growth. The authors use emerging technologies in a way to measure volume and surface area. They take care of 16 sponge species in Caribbean Sea. The volume has been determined by photogrammetry that is a good solution to avoid the traditional destructive method that requires the removal of the organism from the substratum.

In [85], the authors propose the exploitation of a Structure-from-Motion (SfM) technique oriented to glass sponge monitoring in Northeast Pacific Ocean.

The authors of [94] propose a new monitoring tool for micro and macro fauna in large areas and for long time. Their paper describes in particular the results obtained from automatic fish recognition thanks to the use of images acquired by the GUARD1 device. The Argo project fleet consists of 3000 free-drifting profiling floats capable of measuring physical-chemical parameters such as salinity, temperature and surface speed of the water. In addition to these parameters, it also acquires and image every ten minutes. The GUARD1 device processes and transmits information on the surface but has reduced computational capabilities. The device sends the number of species identified and the acquisition time to the surface in a textual manner. The transmission of the images is more complicated because they must be compressed and the authors have studied a protocol for this. Their paper describes an algorithm capable of recognizing and counting fish, and is compared with a count done manually. The experiments were conducted in the Adriatic Sea during the year 2017, from February to March. The image acquisition is carried out once every ten minutes in order to include different light conditions throughout the day, both in condition of calm and high turbidity. Overall, the dataset consists of 12331 images. Training and validation were performed on 20% of the dataset (1233 images). In fact, in conditions of turbidity the GUARD1 drastically reduces its field of vision and consequently the fishes that are not so close to it vanish. The image recognition methodology proposed in [69] suggest a combination of segmentation and machine learning coupled with K-fold Cross-Validation framework. The segmentation phase aims to separate the fish from the bottom. From here, the images taken at instant t are subtracted from the images taken at instant $t-1$ and $t+1$ respectively. The images obtained from the two subtractions are then processed with CLAHE algorithm (see 3.1), to emphasize the differences. The reason why these subtractions are applied is that they help to reduce the presence of biofouling. The purpose of the learning process is to generate a binary classifier such as to define as:

- 1 means presence of at least one fish in the image;
- 0 means the total absence of fishes in the image.

The algorithm is based on a Supervised Machine Learning classifier that makes use of a generic programmer. The performances of the binary classifier, evaluated on 90% of the dataset, has not been involved in training and validation phases. In order to evaluate the automatic recognition performance, fishes have been manually counted and the resulting time series was compared with that generated automatically by the binary classifier. The Pearson index between the two time series is 0.976.

In [95] the authors use a method based on image analysis and genetic programming to evaluate the temporal dynamics of fish abundance. In this case, 20000 images acquired every 30 minutes, were processed continuously for two years, day and night. The images were acquired at the OBSEA-EMSO testing-site. OBSEA is located at a depth of 20 meters 4km from Villanova i la Geltrù (Catalonia, Spain). The classifier was tested on images acquired in 2013. 10961 images were manually scored based on the degree of turbidity and biofouling present. Regions of Interest (ROIs) were selected and then they were automatically extracted from dataset images for training and validation. Finally, they were manually labelled, to identify the positive and the negative example. One of the problem of [95] is the overlapping fish in large schools. In this case, segmentation is not possible. In this experiment, large schools of fish were split into different ROIs and the magnitude of the abundance of fish was captured. Here, too, an image classifier of the 10-Fold Cross-Validation Framework type was used for the training and validation phase. The innovation described in this paper concerns the development of a new automatic method of recognizing and estimating the number of fish, not on images but directly on video. All without any kind of discrimination. The training and validation dataset was extrapolated from that of 2012, about 10% (1191). On these images, the ROIs were automatically extracted and then they were labeled manually. For the testing phase, the numbers of fish counted in the images of 2013 were taken as ground truth. The correlation is also made in this case with the Pearson index. Some correlations are also evaluated in relation to different effects caused by the underwater environment.

Another big challenge has been conducted by the authors of [96], who tried to develop a differentiate method of tracking and counting fish in low quality unconstrained underwater videos. Their videos are taken automatically and continuously by underwater video-surveillance camera near the ken-Ding Taiwan sub-tropical coral reef waters. Even in this case, the experiments and the video are taken in uncontrolled conditions and unconstrained area. This means high variability of the luminosity and water flow according to the external weather and the time of the day. The video collected are part of the on-going efforts of the Tawainese EcoGrid project. Before the advent of these new technologies of object detection, the marine biologists used to determine the existence and used to quantify the different fish species using casting nets. This kind of method was extremely dangerous because was often responsible of habitat damage and fish death, in addition to the high cost of its application. Their work of video processing consists of three parts:

- Video texture analysis;
- Fish detection;
- Tracking modules.

The second part is based on two algorithms computed in independent way. Then their respectively results are combined in order to have a more accuracy of the outcome. On the contrary, the tracking was carried out by the application of the CamShift algorithm. Its peculiar is that of be able to track objects whose numbers may vary over time. In total, the authors approach has been tested on 20 videos, achieving an overall accuracy as high as 85%. An interesting aspect is the first part of the video processing. They detect on the average texture and colour properties of each frame. In conclusion, what is suggested by the authors is to combine the best algorithm of each one of the three phases, creating a new architecture with the best algorithm.

In [97] it is shown how visual sampling techniques represent a valuable resource for a rapid, non-invasive data acquisition for underwater monitoring purposes. Long-term monitoring projects usually requires the collection of large quantities of data, and the visual analysis of a human expert operator remains, in this context, a very time consuming task. Strategies for the automatic recognition of benthic communities are required to effectively exploit all the information contained in visual data. Supervised learning methods, the most promising classification techniques in this field, are commonly affected by two recurring issues: the wide diversity of marine organism, and the small amount of labeled data. This paper discusses the advantages offered by the use of annotated high resolution ortho-mosaics of seabed to classify and segment the investigated specimens, and several strategies to obtain a considerable per-pixel classification performance although the use of a reduced training dataset composed by a single ortho-mosaic are suggested by authors. The proposed methodology can be applied to a large number of different species, making the procedure of marine organism identification an highly adaptable task.

The accurate classification and 3D mapping of benthic habitats in coastal ecosystems are vital for developing management strategies for these valuable shallow water environments. However, both automatic and semiautomatic approaches for deriving ecologically significant information from a towed video camera system are quite limited. In [98], a semiautomated framework for high-resolution benthic habitat classification and 3D mapping using Structure from Motion and Multi View Stereo (SfM-MVS) algorithms and automated machine learning classifiers is proposed. The semiautomatic classification of benthic habitats was performed using several attributes extracted automatically from labeled examples by a human annotator using raw towed video camera image data. Three machine learning classifiers (k-nearest neighbor (k-NN), support vector machine (SVM), and bagging (BAG)) were trained by using these attributes, and their outputs were assembled by the fuzzy majority voting (FMV) algorithm. The correctly classified benthic habitat images were then geo-referenced using a differential global positioning system (DGPS). Finally, SfM-MVS techniques used the resulting classified geo-referenced images to produce high spatial resolution digital terrain models and

orthophoto mosaics for each category. The framework was tested for the identification and 3D mapping of seven habitats (e.g. corals and seagrass) in a portion of the Shiraho area in Japan. Using the FMV algorithm, an overall accuracy of 93.5% in the semiautomatic classification of the seven habitats has been obtained.

3.4 Underwater Datasets

A thing that is common for state-of-the-art detection methods and deep learning methods in particular, is that they are in need of relatively large amounts of training data. During the years, the realization of complex and reliable datasets became an additional challenging research target.

One of the most popular underwater datasets for fish detection and species classification is the Fish4Knowledge (F4K) dataset [99]. It was recorded from 10 cameras between 2010 and 2013 in Taiwan and it has been used for multiple detection and classification algorithms [100], [101]. The F4K dataset is large and consists of videos and images with complex scenes, various marine species and many annotations making it an obvious benchmark dataset.

Another large dataset is the Jamstec E-Library of Deepsea Images (J-EDI) [102], which consists of videos and images of deep sea organisms captured by Remotely Operated underwater Vehicles (ROV). The images of the J-EDI dataset are annotated on an image level.

It is not only fish that are of interest within marine vision. Another critical field is monitoring of benthic organisms, such as scallops and corals. The HabCam dataset consists of 2.5 millions annotated images of mainly scallops, but also fish and starfish [103]. The images have been captured along the continental shelf off the east coast of the USA.

The Australian benthic dataset (BENTHOZ-2015) is a benthic dataset recorded along the coasts of Australia and used for classifying coral [104]. BENTHOZ-2015 dataset consists of an expert-annotated set of georeferenced benthic images and associated sensor data, captured by an autonomous underwater vehicle (Sirius AUV) around Australia. This type of data is of interest to marine scientists studying benthic habitats and organisms. AUVs collect georeferenced images over an area with consistent illumination and altitude, and make it possible to generate broad scale, photo-realistic 3D maps. The complete data set described here has been made available on Squidle [105]. Squidle is a new web-based framework that facilitates the exploration, management and annotation of marine imagery. It provides a user-friendly interface that integrates spatial map-based data management tools with an advanced annotation system. The online annotation system permits scientists to easily collaborate on both the labeling and use of their data. It will in future also provide a platform for using and testing machine learning and computer vision algorithms on marine imagery.

A table of expert annotations, sensor data, geolocation and image metadata for BENTHOZ-2015 is available for download from figshare [106], the online scientific data repository.

Another dataset is the Brackish Dataset, an underwater dataset with annotated image sequences of fish, crabs, and starfish captured in brackish water with varying visibility [64].

3.5 Underwater Machine Learning and Deep Learning

Recent methods, dedicated to automatic identifications in underwater environment, have been a drastically increase over the last decades [60]. Most of these methods have been performed in standardized light and fish position conditions. The main problems for fish identification are:

- colour and brightness of the image;
- the complexity of the 3D architecture of the image;
- unusual position of the fishes;
- fishes are very often hidden behind other fishes or corals;
- high variability parameters of the acquisition camera.

Because the monitoring of fish biodiversity, in space and time, on coral reef is a big challenge in marine ecology, is very necessary to perform an automatic and accurate algorithm to identify fishes quickly.

The authors of [60] suggest a new deep learning method for accurate and fast identification of coral reef fishes in underwater images. In addition to the complexity of the operation, due to the presence of the dust (biofuling), another issue is the high cost of the underwater operations. It is a very difficult performance and very time-consuming. The authors, with the purpose to identify and counting fish individuals on photos and videos, firstly tested the performance of a CNN trained with different photographic databases, in order to identify 20 species. The novelty consists of comparing the performance of species identification by CNN and the performance of human identification. The CNN was trained on 900000 images including:

- whole fish bodies;
- partial fish bodies;
- environment (reef bottom or water).

In conclusion, the rate of correct identification, was much higher in the CNN identification (94.9%), respect to the human identification (89.3%). A positive aspect, responsible for the best performance, is due to the ability of the CNN, to identify fish individuals partially hidden behind corals or other fishes. On the contrary, what has emerged to be better in human identification, has been the ability to

identify fishes in unusual position, for example twisted body. We can therefore say that deep learning can thus perform efficient fish identification and so it promises a new cheaply and effectively protocols for monitoring fish biodiversity.

Another kind of application of the deep learning is the use of YOLO algorithm in the field of water power application. This is due to the fact that these new technologies like tidal and in stream turbines can affect negatively to the fish habitat and to wildlife in general in underwater environment. In [24], the authors propose a solution to save time to process large video, because human analysis of video takes a lot of time. In addition, in this case the big challenge are:

- Illumination changes;
- Unequal spectral propagation;
- Low contrast;
- Clutter in the form of floating vegetation;
- Change in visibility due to turbidity.

The authors refers to the Fish4Knowledge project [99] whose dataset consists mostly on coral reef fishes.

The authors of [107] have used a deep architecture composed by principal component analysis (PCA) that is used in two convolutional layers. To extract information to large poses is used a spatial pyramid pooling (SPP). To conclude, for the classification, a linear support-vector machines (SVM) classifier is used. They propose a framework for underwater live fish recognition in unconstrained natural environment. They used to capture images by underwater camera in the open sea, capturing in this way a dataset. Their code is available on GitHub [108]. The authors using this method are able to reach an accuracy of 98.64%. The dataset used has been made by the Fish4Knowledge (F4K) project [99]. It is composed by underwater live video fish captured in open sea. In total 27370 fish images have been verified, divided in 23 clusters, whose each cluster is presented by a representative species. Following instructions from marine biologists, a manual labelling of the species has been made. A wide variation in fish position, in scale, in orientation, in colours and even in texture, is present in the dataset for each species. Once, the labelling has been made, it is the time to eliminate the background of the fish image. The total images are divided into:

- 5/7 for training;
- 1/7 for validation;
- 1/7 for test.

In conclusion, the author notes that has to be improved only the rotation, because the foreground

extraction procedure and foreground resizing make the fish unified in size, the translation and the zoom would not be necessary. Another experiment that has been conducted by the authors provide a decrease of training set from 5/7 to 3/7, but in this case, the result is a decreasing of the accuracy.

In [61], the authors propose a prototype implementation of the analysis pipeline provided by deep-sea videos taken by one of the fixed cameras at the LoVe Ocean Observatory network of Lofoten Islands (Norway) at 260 m depth. The results showed good classification on an independent dataset, having an accuracy value of 76.18%.

In [109], an alternative approach to the training of the deep models that can be applied to interpret real-world underwater imagery. The authors highlight that deep learning models require large annotated datasets for model training and these are typically unavailable for underwater imaging applications. To demonstrate the proposed approach, the specific problem of biofouling detection on marine structure is analysed. A dataset of annotated synthetic images is created using a rendering procedure in a virtual underwater environment under a wide variety of conditions and feature biofouling of various size, shape, and colour. Each rendered image has a corresponding ground truth per-pixel label map. A contemporary deep encoder–decoder network, termed SegNet, is trained on the created dataset and tested in the segmentation of new real world images. The proposed approach achieved a satisfactory accuracy and a low computation time.

The authors of [64] use the created Brackish Dataset to fine tune and test the YOLOv2 and YOLOv3 CNNs in order to compute a baseline. The YOLOv2 object was pre-trained on Imagenet and fine-tuned to fish datasets and it was obtained from the VIAME (Video and Image Analytics for Marine Environments) toolkit [110], [111]. The YOLOv3 detector was the original version pretrained on the Open Images dataset [112]. The dataset is split randomly into 80 % training, 10 % validation and 10 % test data. Each network is trained for 30,000 iterations using their original training regime, only adjusting the batch size and setting the input size to 416 × 416, and with the earliest layer weights frozen. The mean Average Precision (mAP) metric is used to evaluate the object detectors performance.

In [113], authors first perform image enhancement to produce depth information and benefit many vision algorithms and advanced image editing. Authors try to develop a novel underwater fish detection and tracking strategies combining YOLOv3 algorithm and parallel Correlation Filter. In particular, YOLOv3 fast and accurate detection properties are used, together with the high tracking accuracy of parallel Correlation Filter. The proposed approach has been implemented on the NVIDIA Jetson TX2 for online fish detection and tracking, enabling a fast system and rapid experimentation. It has been shown in the experiments that the developed scheme of this paper achieves consistent

performance improvements on online fish detection and tracking for ocean observatory network. F4K dataset has been exploited for the simulation experiments [99].

In [114], the authors, in order to meet the requirements of fast detection and classification of underwater targets during intelligent underwater robot operation, propose an improved YOLOv3 algorithm named YOLOv3-UW algorithm to improve the detection accuracy and detection speed. Compared with the YOLOv3 algorithm, the YOLOv3-UW algorithm improves the clusters algorithm of data sets, optimizes the network structure, and improves the residual module. The final experimental results show that detection speed and detection accuracy of the YOLOv3-UW algorithm are higher than the YOLOv3 algorithm.

In [115], the authors describe an integrated system for vision-based counting of wild scallops in order to measure population health, particularly pre- and post-dredging in fisheries areas. Sequential images collected by an autonomous underwater vehicle (AUV) are independently analyzed by a convolutional neural network based on the YOLOv2 architecture.

3.6 Main Research Topics

The target of the thesis work has been represented by the application of classification and object detection methods in underwater environment. The work has been oriented to obtain marine communities assessments. The interested marine communities are mussels, sponges and fishes.

The thesis work has been conducted through the investigation on three underwater fields:

- Underwater geomatics.
- Computer science.
- Underwater biology.

Through the methodologies proposed by geomatics, the terrestrial approach has been suitably extended to underwater environments. Reliable design procedures of underwater diving surveys have been approached in order to obtain the maximum yield on the creation of underwater datasets from surveys conducted during the thesis work. DISVA (Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona (AN), Italy) divers conducted two underwater surveys during the thesis work, oriented to mussels and sponges data acquisition. To the obtained datasets, other existing datasets related to mussels and fishes have been analyzed and used.

Geomatics has been also applied for a classification procedure on mussels: initially, a photogrammetric approach has been exploited.

In order to include computer science methodologies, innovative deep learning software has been used for object labelling and detection. For this purpose, a contribution has been given by DII (Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Ancona (AN), Italy).

4. Case study

4.1 Mussels

The mussels' case study consisted on the processing of two image datasets: MUSSEL1 and MUSSEL2. The research target consisted in the classification and in the object detection on the analyzed underwater datasets.

MUSSEL1 dataset has been obtained through the processing of a video performed within a pilot project of DISVA (Dipartimento di Scienze della Vita e dell'Ambiente) of Università Politecnica delle Marche (Ancona (AN), Italy).

MUSSEL2 dataset has been obtained from Squidle [105] repository.

4.1.1 MUSSEL1 Dataset

The video (02:25 (mm:ss)) has been realized in 2020 at "Passetto" location in Ancona (AN), Italy (Adriatic Sea).

The sea depth of the explored area was about 5-6 meters. The video has been realized by a diver with a Sony RX100V camera (Appendix C - Devices). During this diving survey, the absence of GCPs is justified by the fact that this was only an explorative mission oriented to take information about the investigated area. Consequently, a second more accurate survey from a geomatics and photogrammetry point of view has not been performed due to the Covid-19 pandemic. The geomatics aspects that have been analyzed for the second survey design have been reported in the following paragraphs.

Data Preprocessing

The first 25 seconds of the video have been considered. 1498 frames have been extracted using Agisoft Metashape (Appendix B - Software) [74]. Each frame is 1920x1080 pixel.

An image enhancement procedure has been performed through Image Enhancement Process Tool [66]. In the following, some examples on significant frames have been proposed, using the five enhancement algorithms provided by the tool (ACE, CLAHE, LAB, NLD, SP).

In Figure 17, the comparison of the five algorithms provided by the software on three frames has been reported. It can be noted how the enhanced frames differ from the original frame.

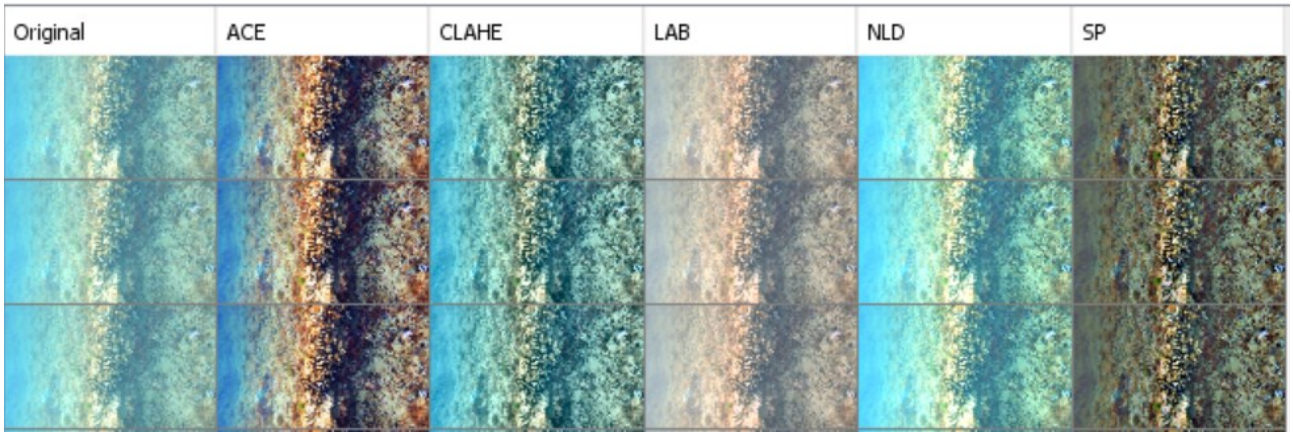

Figure 17 - Comparison of the five enhancement algorithms provided by Image Enhancement Process Tool on three frames (MUSSEL1 dataset).

3D Modelling

Starting from the enhanced frames, Agisoft Metashape software has been used again (Appendix B - Software) [74]. For each of the enhanced frames, the 3D model has been created (Figure 18-Figure 22).

The first step has been the camera alignment. A high accuracy has been operated; a key point limit of 40000 and a tie point limit of 4000 have been set. The second step has been the generation of the dense point cloud with a medium quality. Then, the mesh surface has been created and finally the texture has been obtained. The texture obtained from each frame has been reported in Figure 18-Figure 22.

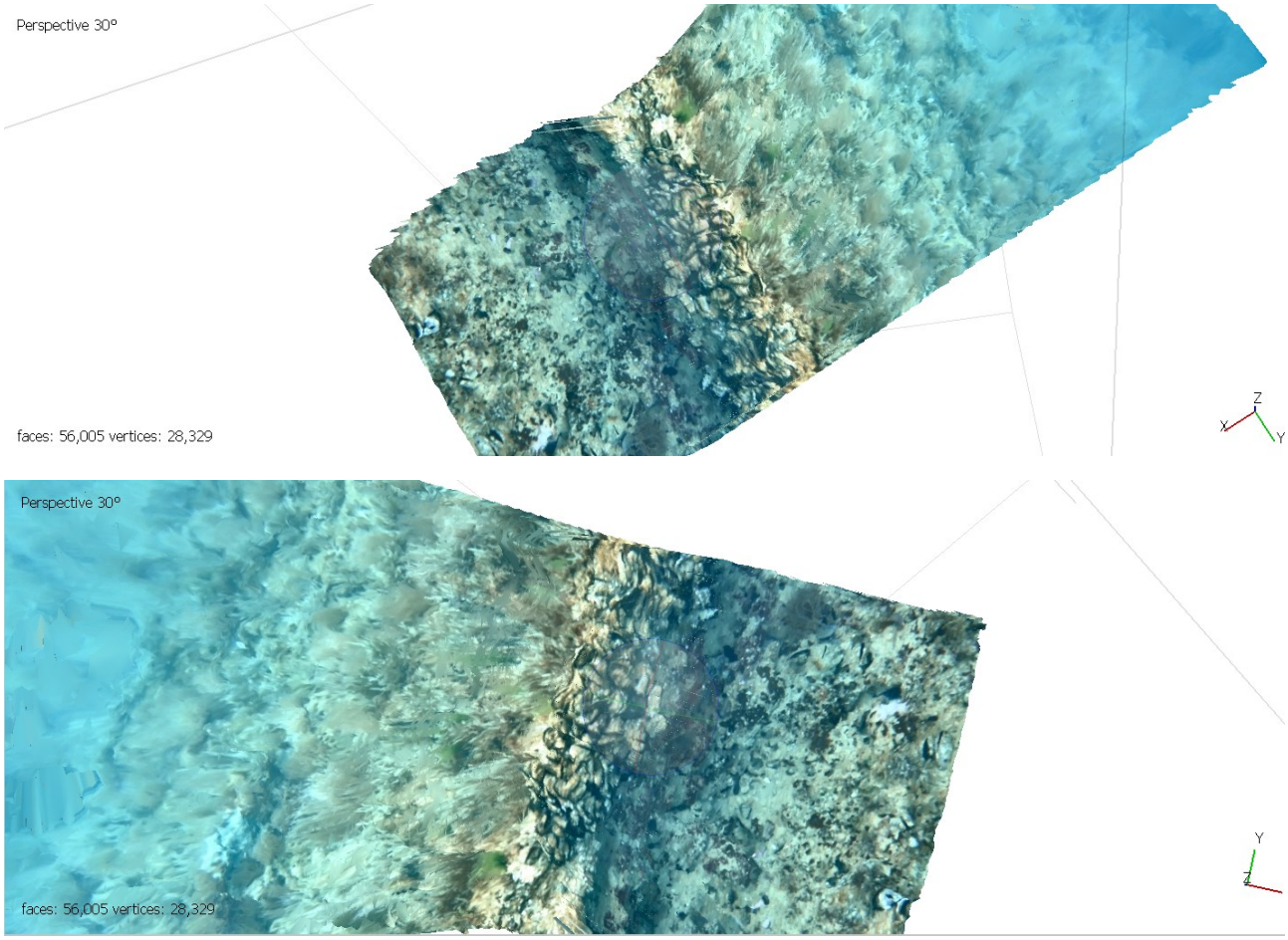

Figure 18 - 3D model of the three CLAHE enhanced frames (MUSSEL1 dataset).

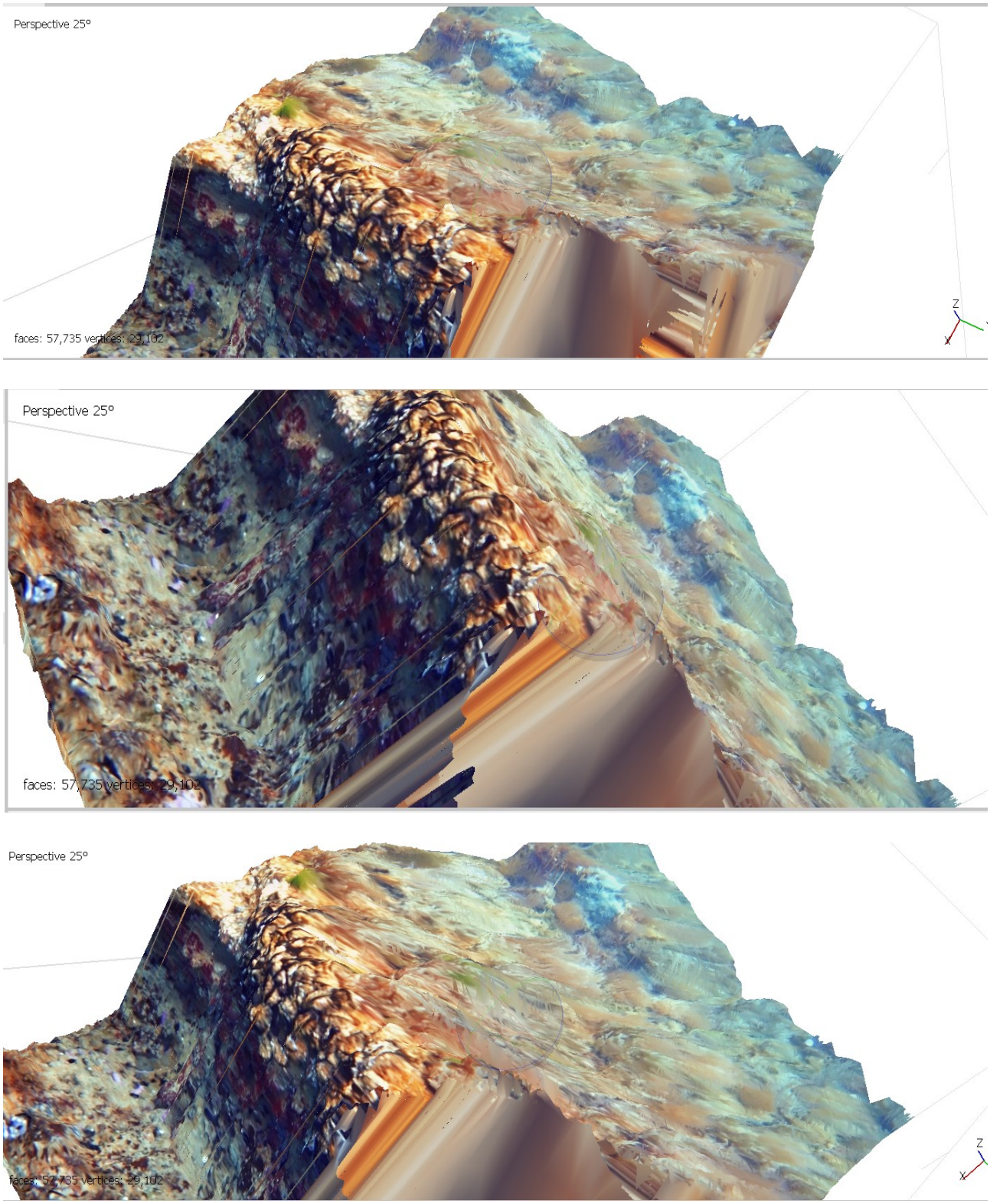

Figure 19 - 3D model of the three ACE enhanced frames (MUSSEL1 dataset).

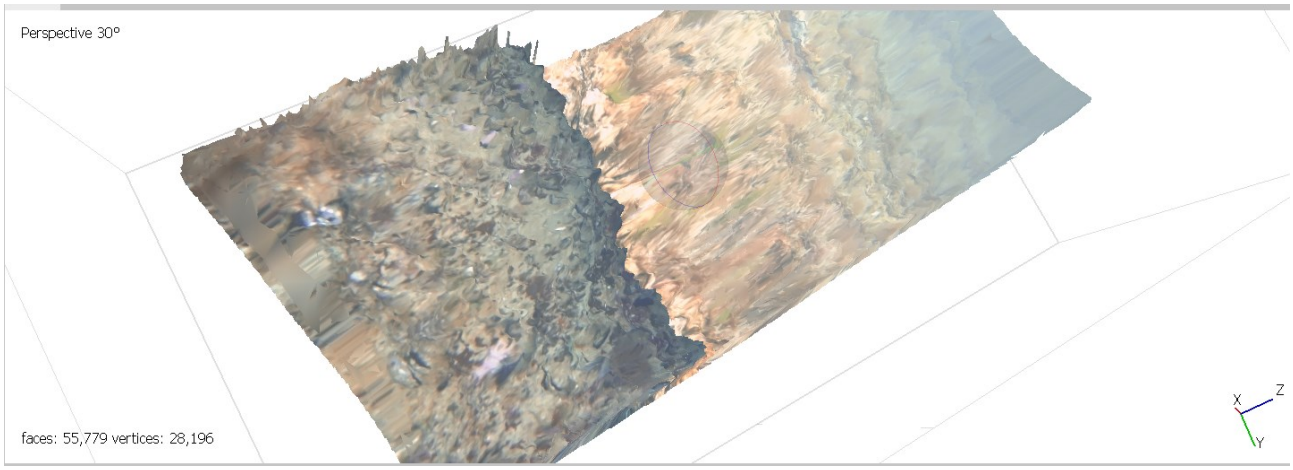

Figure 20 - 3D model of the three LAB enhanced frames (MUSSEL1 dataset).

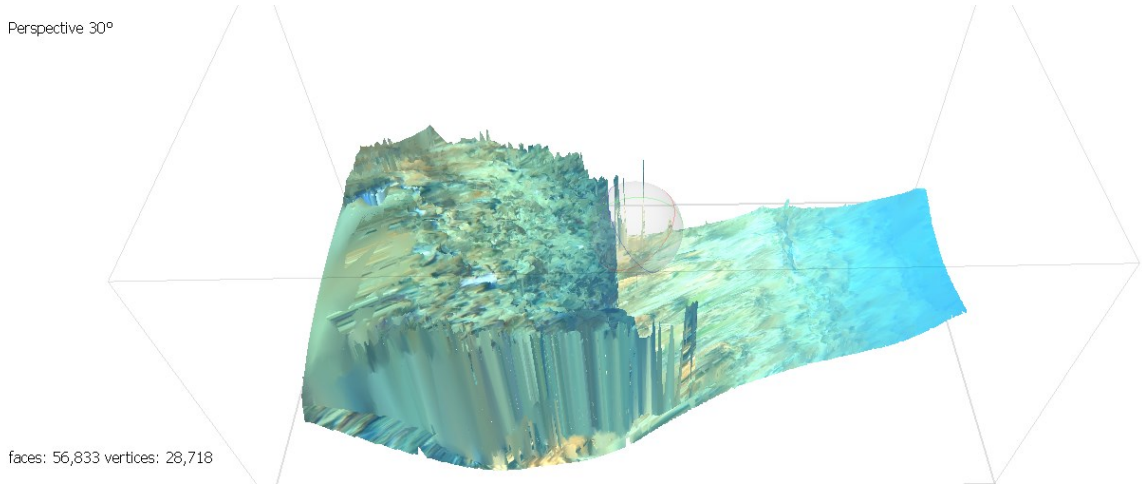

Figure 21 - 3D model of the three NLD enhanced frames (MUSSEL1 dataset).

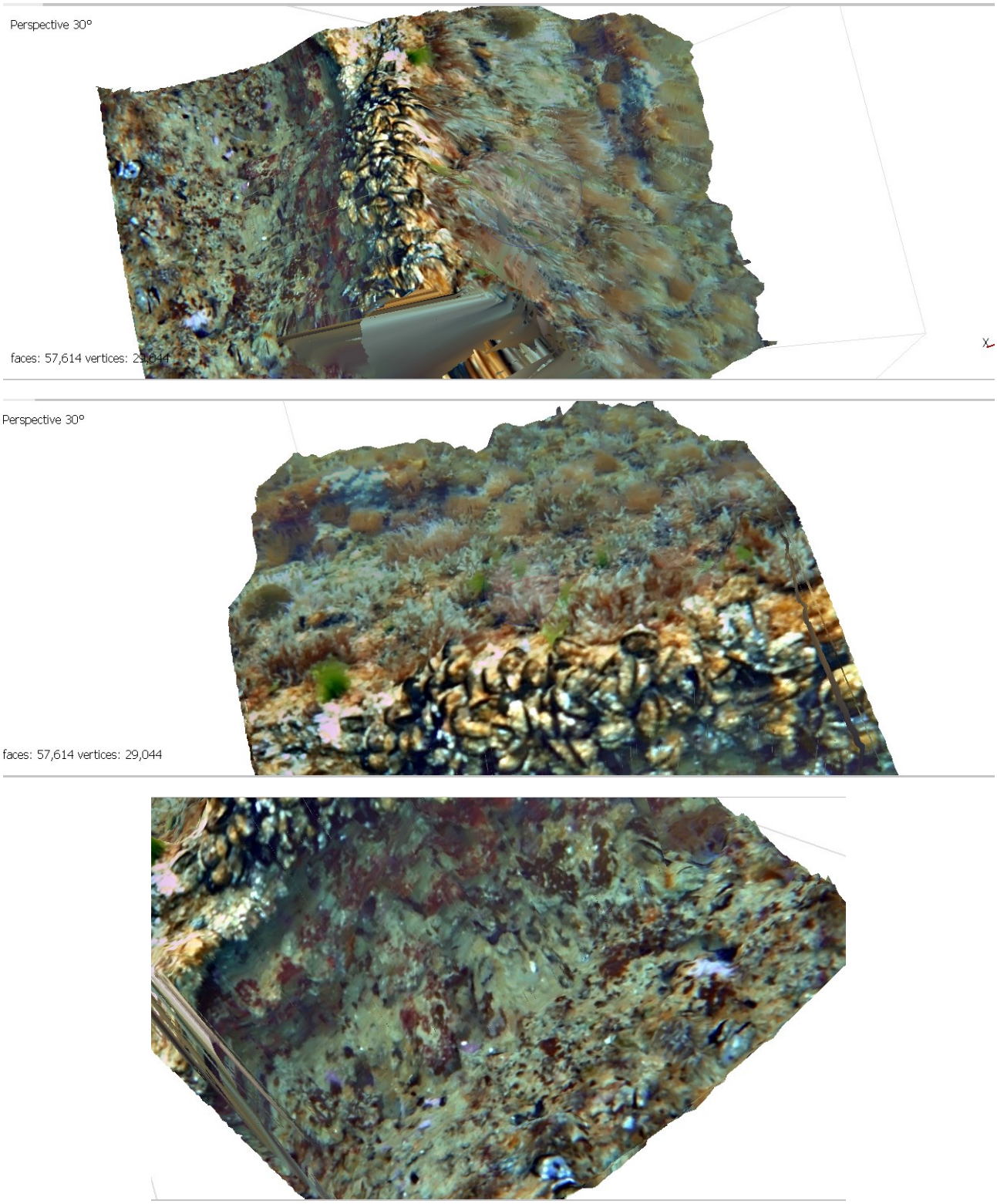

Figure 22 - 3D model of the three SP enhanced frames (MUSSEL1 dataset).

Classification

Among the obtained textures, the 3D model related to the ACE enhanced images has been selected. The evaluation has been made based on an optical point of view (human eye).

The next step has been a classification experiment in order to predict if an observed underwater object is a mussel. Classification is a supervised learning problem that predicts that the data belongs to a particular category (the mussel in the case at issue).

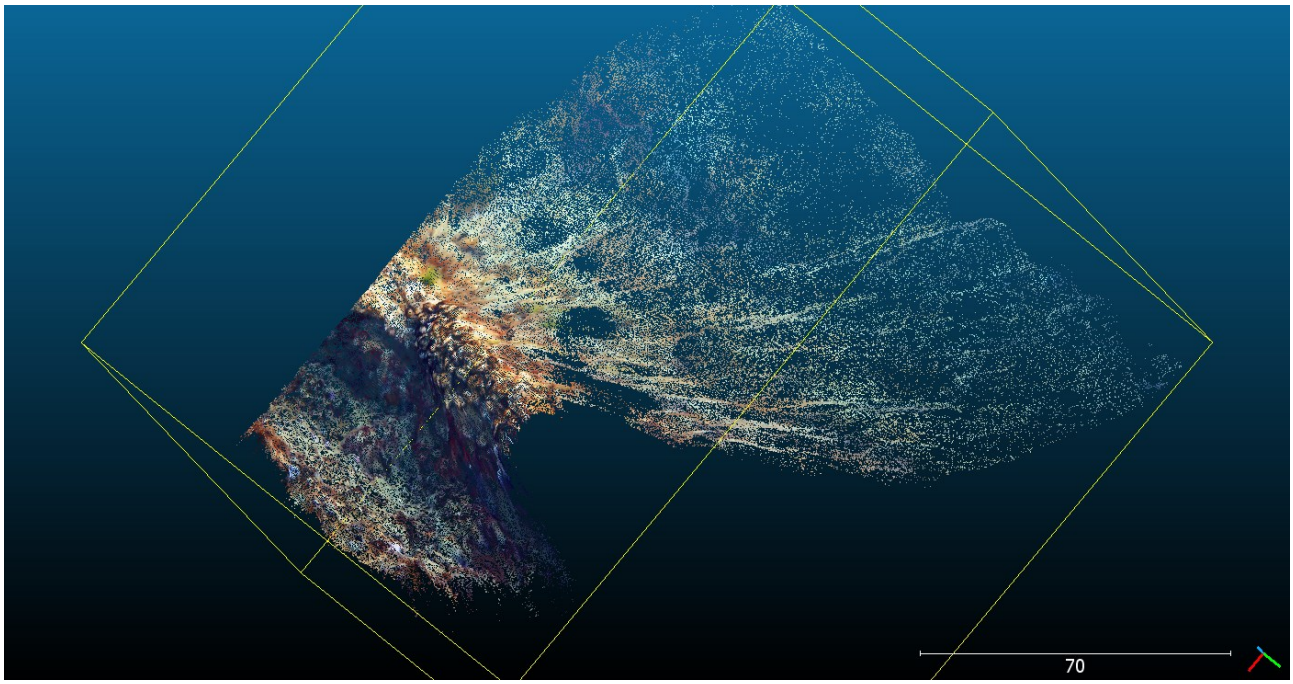

Figure 23 - Point cloud obtained from ACE enhanced frames (MUSSEL1 dataset).

The ACE point cloud (Figure 23) has been exported from Agisoft Metashape (".ply" file) and it has been loaded on CloudCompare software (Appendix B - Software) [79].

Through the "Interactive Segmentation Tool" (also called "scissors tool") of CloudCompare, two polygons have been created: the first polygon represents the mussels (Figure 24) and the second polygon represents the seagrass (Figure 25).

The two polygons represent the classes used for the classification (mussels and seagrass). The classification has been performed through CANUPO plugin, which performs an automatic classification of a point cloud [116]. A classifier has been created (".prm" file).

The obtained classifier has been tested on the same point cloud used for the creation of the classifier (training). Figure 26 reports the obtained CANUPO.class, while Figure 27 represents the obtained CANUPO.confidence.

The CANUPO.class shown in Figure 26 reports how the classifier classifies each single point of the point cloud. In the proposed example, the red color represents the seagrass, while the blue color represents the mussels.

The CANUPO.confidence shown in Figure 27 provides a performance metric of the classifier action on the input point cloud.

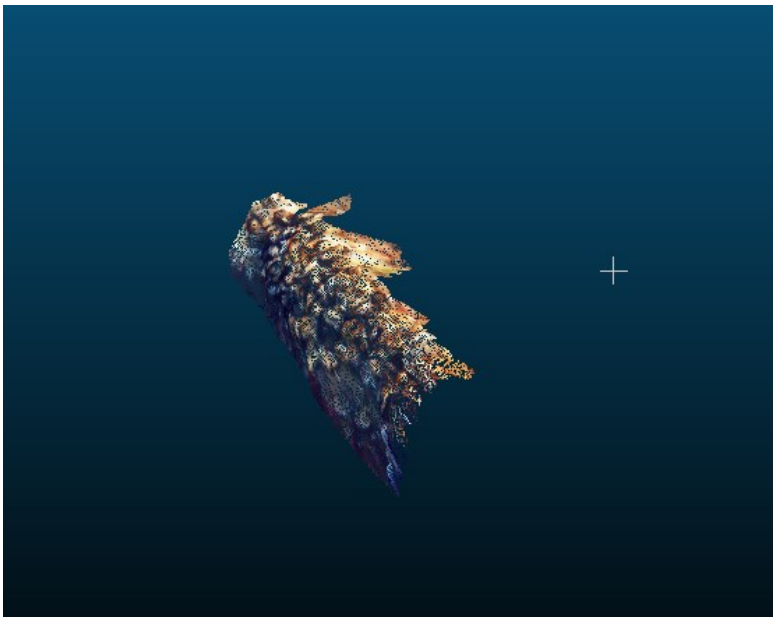

Figure 24 - Mussels polygon (MUSSEL1 dataset).

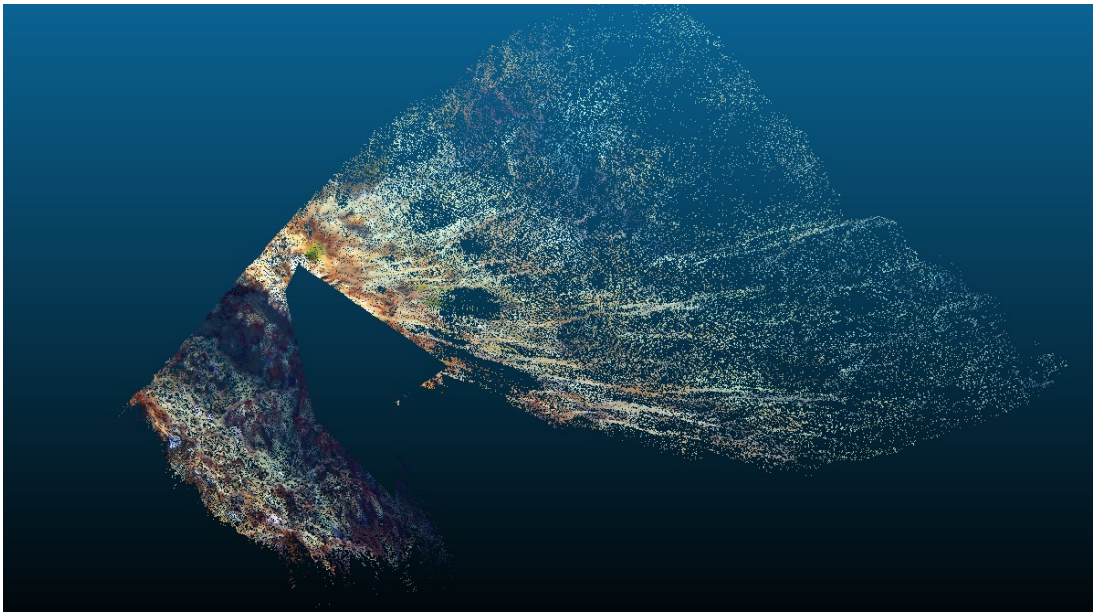

Figure 25 - Seagrass polygon (MUSSEL1 dataset).

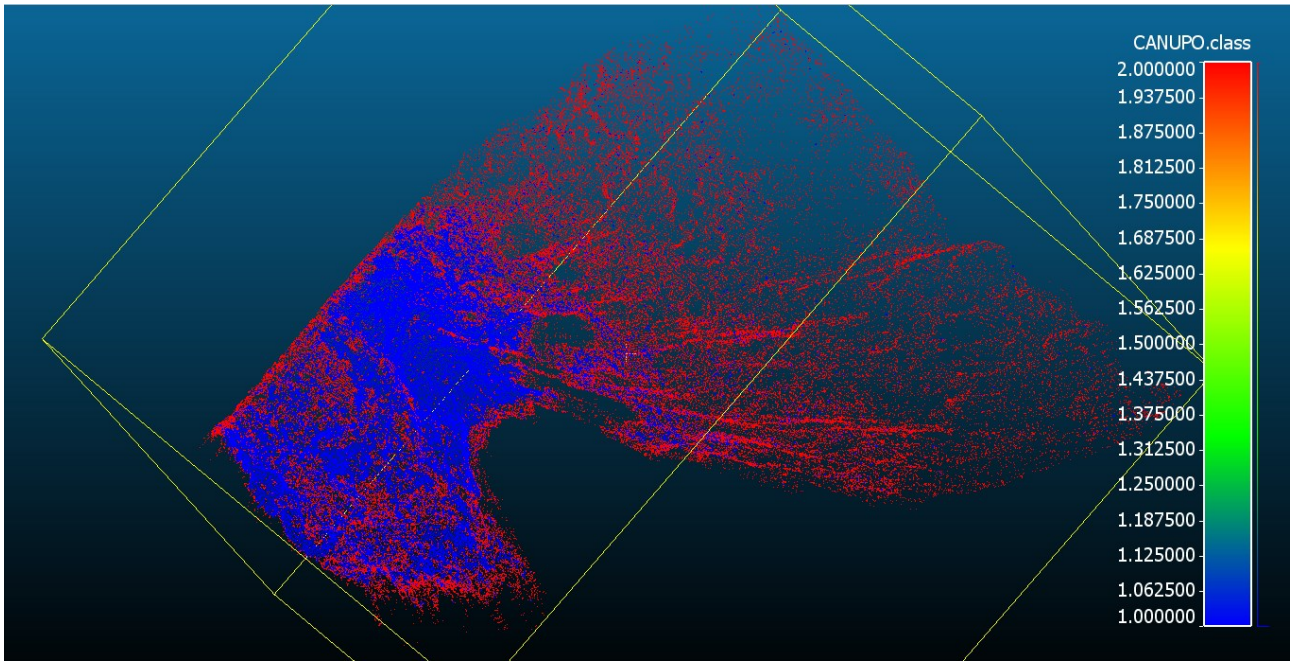

Figure 26 - (MUSSEL1 dataset) Point cloud classification (blue=mussels, red=seagrass).

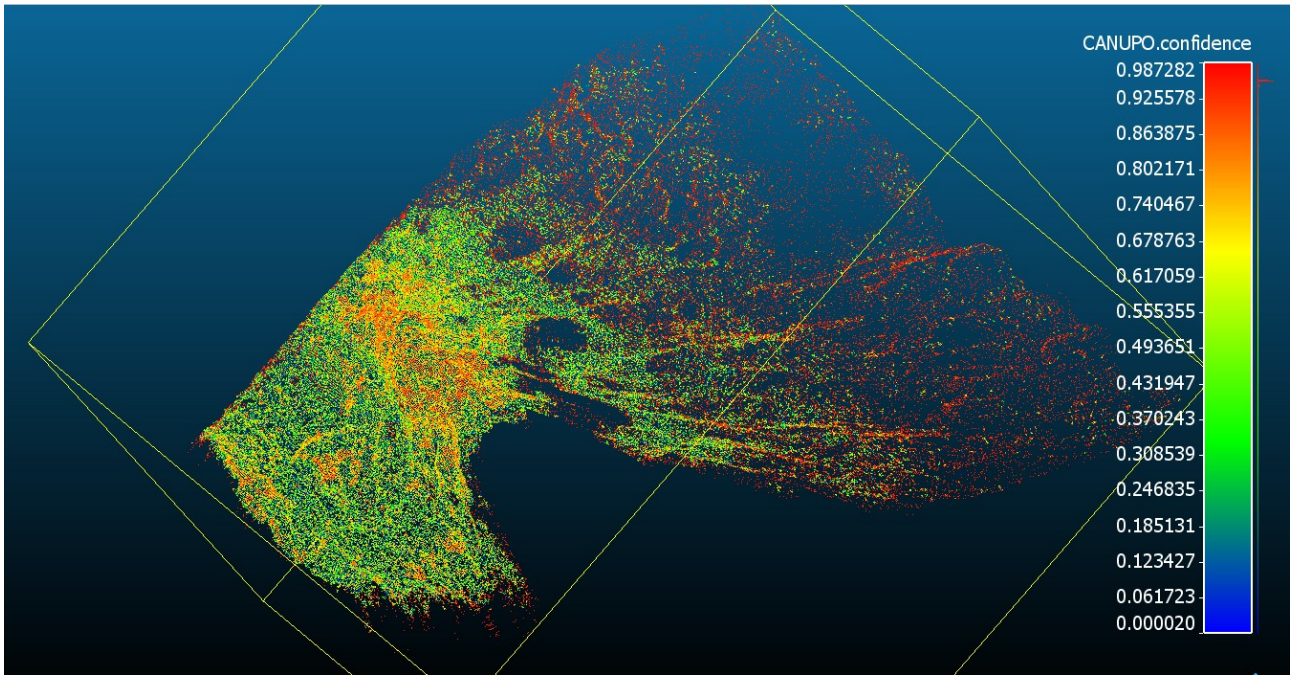

Figure 27 – Point cloud classification confidence (MUSSEL1 dataset).

Design of a second underwater survey

In order to design the new survey, some theoretical geomatics aspects have been considered.

GSD (Ground Sampling Distance) defines the spatial resolution that is a very important parameter in a photogrammetric survey. As GSD means the distance, measured on the ground, between two closed images pixels. In a more simple way, it is the pixel size on the field or in another way can be defined as the distance between the central points of two closed pixels of a digital image. In this way, it is

easy understandable that the less is the GSD the greater is the detail of the image. The calculation in done by the following equation:

$$GSD = (D \cdot d)/f \tag{1}$$

where D is the distance between camera and field instead d is the side size of a pixel (that has a square form). There are also parameters that influence the GSD computation: resolution of the camera, focal length (f), perspective and flying height. The focal length is a physical property of a lens. It is the distance between the centre of the lens (nodal point) and the lens focus. The focal length is measured in mm. For better understand, a more curved lens has a shorter focal length than a flatter one. What does the difference, increasing the focal length, changes the optic:

- 16 mm is a super wide angle;
- 24 mm e 35 mm are wide angle;
- 50 mm is a normal optic;
- 70 mm, 150 mm, 300mm are telephoto lens.

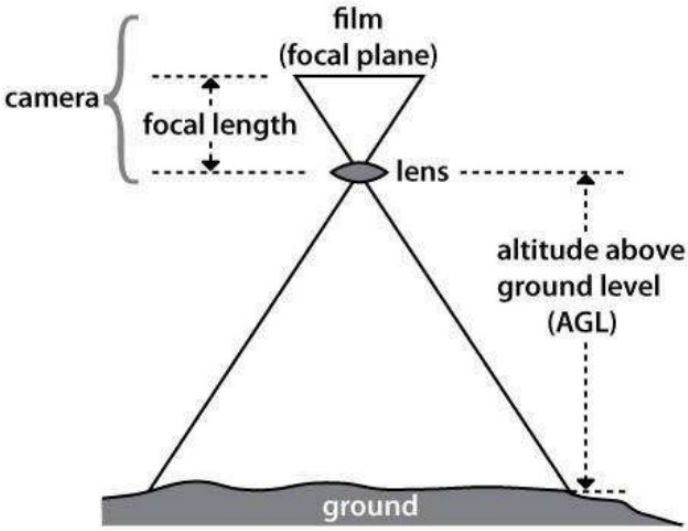

Figure 28 - GSD illustration [117].

The perspective can be relative or absolute. As relative perspective means how the points are placed relatively with each other. The absolute perspective is instead the position of the points in the reconstructed model and their effective position on the ground.

To switch between equivalent focal length (fe) to real focal length (fr), it is necessary knowing or the photo sensor size or knowing the "Crop Factor" (CF). It is inversely proportional to the sensor size. Once the crop factor is known, it can be easily computed the real focal length by:

$$fr = fe/CF \tag{2}$$

In order to having the absolute accuracy it is also necessary the presence of the Ground Control Points (GCPs). They are points clearly visible on the field, of which the coordinates have high accuracy because measured with professional devices. It is very important that the GCPs are measured with an accuracy higher than the pixel size in the field. It is also very useful having between 10 and 15 GCPs, in order to allow a good reconstruction of the 3D model [117].

Another key point to obtain a reliable 3D model reconstruction is the calibration; it lets to compute the internal camera calibration parameters and the images distortions due to the lens. Three kinds of calibration exist: the pre-calibration, the calibration in loco and the self-calibration. The pre-calibration is done with objects having their dimensions known, and it is done in the office where the environmental conditions are different. On the other hand, the calibration in loco is executed directly on the field, setting up a test area. As the last type of calibration, there is the self-calibration where for the camera calibration, the acquired dataset is used.

The considered camera has been a Sony RX100V camera (Appendix C - Devices), with an equivalent focal length of 35 mm, and a crop factor 2.73. The real focal length is 12.89 mm. The available images are 1920x1080 so the size of a pixel is 0.0032 mm over a pixel. The GSD is 0.3 mm per pixel. Hypothetically, it has been considered an overlapping of 80 %, so for a test area with a length of 50 meters, it has been computed the number of images to be captured (519). It has been considered 1 meter as the distance between the camera and the object.

4.1.2 MUSSEL2 Dataset

In parallel to the methodological workflow depicted in the previous paragraph related to MUSSEL1 dataset, a mussels' benchmark dataset in the Adriatic Sea has been searched among the existing datasets. All the most popular available datasets have been checked (see 3.4), together with Squidle [105] (Appendix A - Repository) and fighsare [106] (Appendix A - Repository) repositories.

In particular, it has been searched a dataset belonging to an underwater environment similar to the environment characterizing the MUSSEL1 dataset. The only available mussels' dataset has been found on Squidle repository. This dataset has been obtained in 2012 through a Reef Life Survey at "Portonovo" location in Ancona (AN), Italy (Adriatic Sea). The images containing mussels are about an hundred (600x800). The mean depth is about 5 meters. There is no information about the acquisition mode.

Figure 29 - MUSSEL2 dataset frame.

Data Annotation

No annotations have been found on the Squidle dataset, so a manual annotation procedure has been executed. For this target, VIAME (Video and Image Analytics for Marine Environments) toolkit has been exploited (Appendix B - Software) [110], [111]. Three classes (also called "tracks") have been created: mussel, algae and seabed. The seabed represents all the objects that are not mussels or algae. About 1000 annotations have been completed.

Figure 30 - VIAME dataset upload GUI (MUSSEL2 dataset).

Figure 30 shows the VIAME Graphical User Interface (GUI) that allows uploading all the images contained in a folder.

```
mussel
algae
seabed
```

Figure 31 - ".txt" file containing the labels (VIAME).

The labels related to the classes have been defined in a ".txt" file (Figure 31). Box-Level annotations have been executed (Figure 32-Figure 34).

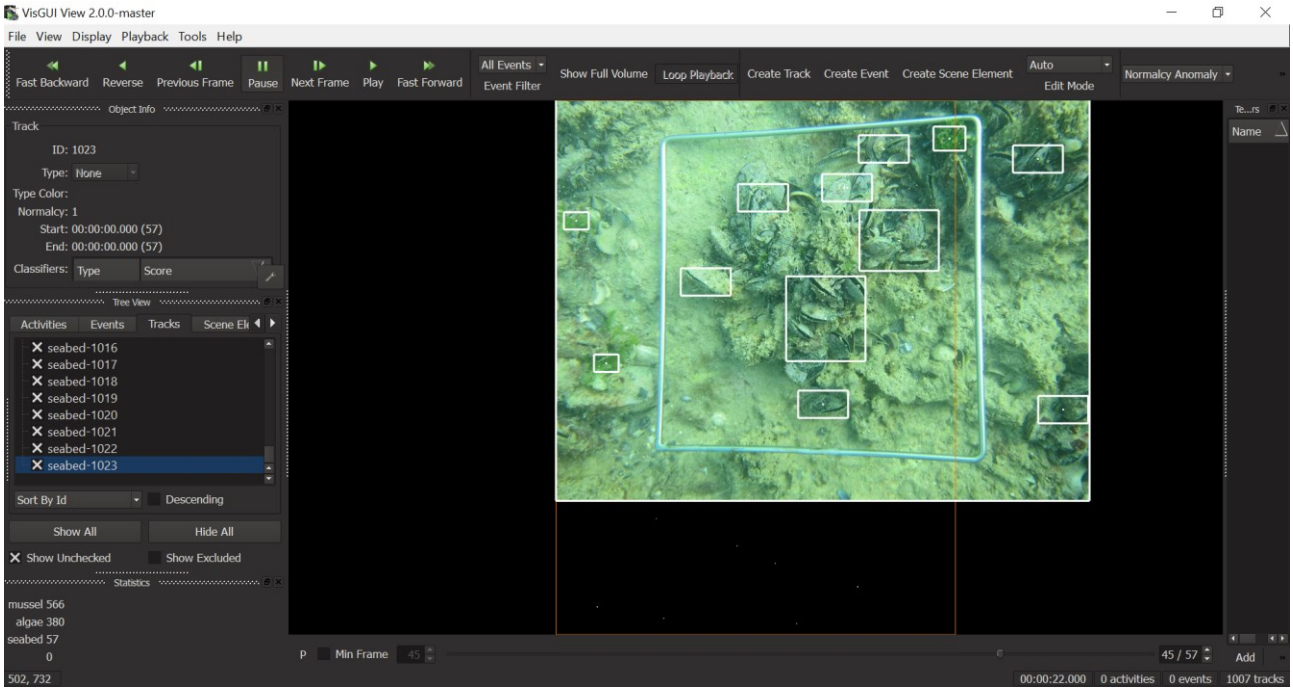

Figure 32 - VIAME Box-Level annotation (MUSSEL2 dataset) (a).

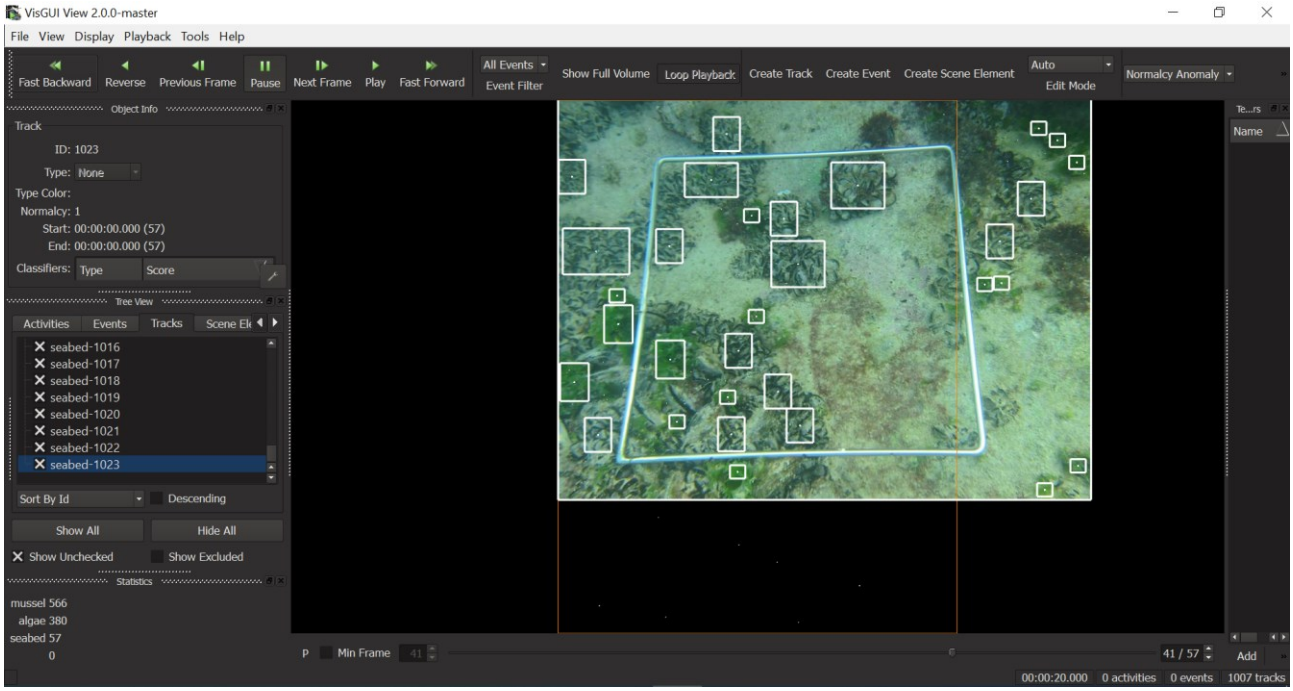

Figure 33 - VIAME Box-Level annotation (MUSSEL2 dataset) (b).

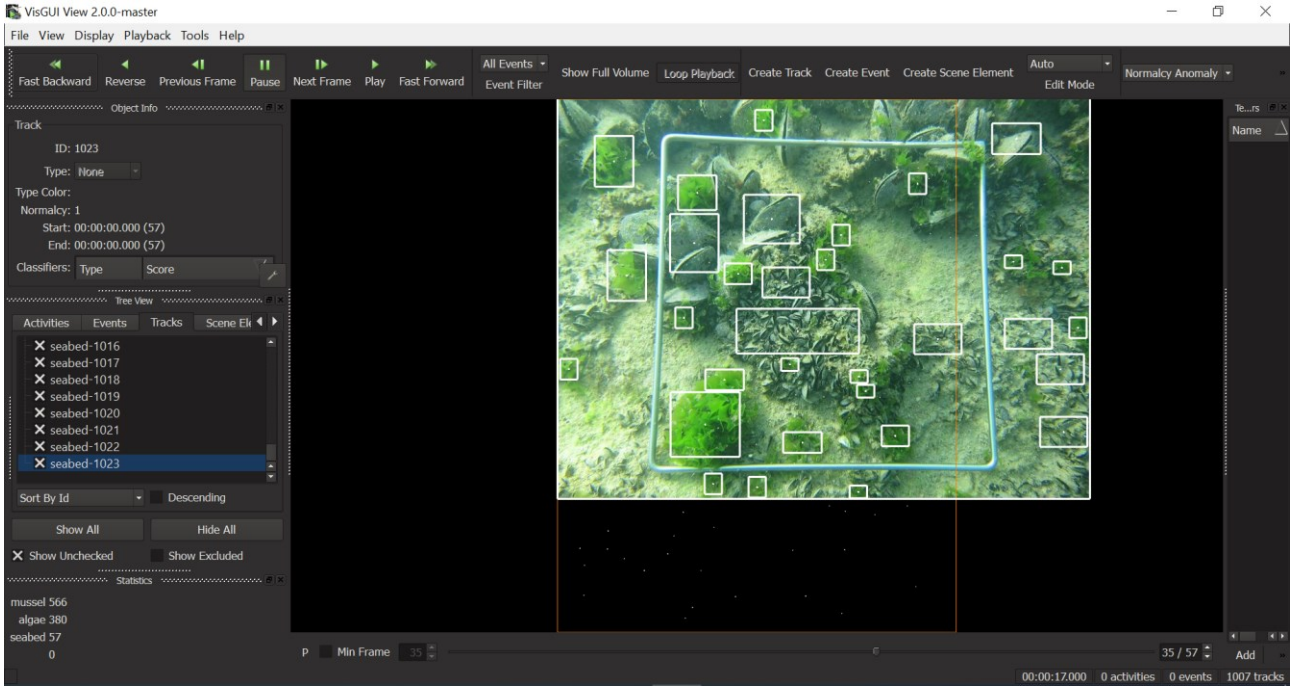

Figure 34 - VIAME Box-Level annotation (MUSSEL2 dataset) (c).

Each obtained annotation (corresponding to the rectangular boxes that can be noted in Figure 32-Figure 34) can be automatically saved in a ".csv" file (Figure 35). This characteristic of VIAME is a very user friendly feature because it allows re-uploading in the future of the started project.

```
1  0,figure 24.jpg,0,209,216,359,368,1,-1,mussel,1
2  1,figure 24.jpg,0,427,175,558,298,1,-1,mussel,1
3  3,figure 24.jpg,0,122,376,252,478,1,-1,mussel,1
4  4,figure 24.jpg,0,461,395,573,499,1,-1,mussel,1
5  5,figure 24.jpg,0,416,318,487,406,1,-1,algae,1
6  6,figure 24.jpg,0,368,193,415,253,1,-1,algae,1
7  7,figure 24.jpg,0,95,265,164,332,1,-1,mussel,1
8  8,figure 24.jpg,0,304,455,373,522,1,-1,mussel,1
9  9,figure 25.jpg,1,348,201,425,281,1,-1,mussel,1
10 10,figure 25.jpg,1,57,119,134,200,1,-1,mussel,1
11 11,figure 25.jpg,1,112,411,184,480,1,-1,algae,1
12 12,figure 25.jpg,1,320,344,369,396,1,-1,algae,1
13 13,figure 25.jpg,1,459,531,508,583,1,-1,algae,1
14 14,figure 25.jpg,1,450,318,525,410,1,-1,mussel,1
15 15,figure 25.jpg,1,149,241,174,274,1,-1,algae,1
16 16,figure 25.jpg,1,330,414,355,447,1,-1,algae,1
17 17,figure 25.jpg,1,279,673,316,726,1,-1,algae,1
18 18,figure 25.jpg,1,561,184,598,237,1,-1,algae,1
19 19,figure 25.jpg,1,566,451,603,504,1,-1,algae,1
20 20,figure 25.jpg,1,434,594,471,625,1,-1,algae,1
21 21,figure 25.jpg,1,390,762,427,793,1,-1,algae,1
22 22,figure 25.jpg,1,531,402,568,433,1,-1,mussel,1
23 23,figure 25.jpg,1,102,210,136,277,1,-1,mussel,1
```

Figure 35 - ".csv" file containing the annotations (VIAME).

Model Training

There are several detection and classification models in VIAME toolkit, e.g.:

- YOLOv2, Gridded YOLOv2, YOLOv3, Gridded YOLOv3, YOLO-WTF;
- Cascade Faster RCNN, Mask Faster RCNN.

Model training has been attempted using YOLOv3 provided by VIAME, but the procedure has not been successfully completed on a laptop computer with a CPU. In fact, to exploit all VIAME features, a NVIDIA GPU equipped with at least 8GB tailored memory is needed. Furthermore, Deep Learning CUDA libraries are needed [110], [111].

Hence, a suitable hardware has been searched: DII (Dipartimento di Ingegneria dell'Informazione) of Università Politecnica delle Marche (Ancona (AN), Italy) allowed the use of Google Colab platform (Appendix B - Software) [118]. An attempt of upload VIAME software on Google Colab platform has been performed: VIAME software installation and use need of a GUI interaction, but this feature is not available on Google Colab platform through a remote connection.

4.2 Sponges

The sponges' case study consisted on the processing of an image dataset. The research target consisted in the object detection.

The sponges' dataset has been obtained through the processing of a video performed within a project of DISVA (Dipartimento di Scienze della Vita e dell'Ambiente) of Università Politecnica delle Marche (Ancona (AN), Italy). The project is called MedSpon: it consists in the characterization of new antibiotic principles against WHO priority pathogens of sustainable produced marine sponges for pharmaceutical applications. The MedSpon project focuses on two Mediterranean sponge species: *Chondrosia reniformis* and *Axinella polypoides*. The target of the project, based on the knowledge from mariculture, is to optimize an aquaculture processes for the target species by successfully breeding sponge fragments to serve as a source of secondary metabolites. In particular, considering the production of secondary metabolites can change during life cycle of the species like reproductive period and food availability, strong wild populations of target species will be studied by project partner Università Politecnica delle Marche at Italian field sites to assess the optimal habitat conditions and provide information for mass production in land-based recirculation aquaculture system (RAS). Under controlled conditions in RAS different abiotic factors and types of food are tested in parallel approaches to create effective rearing conditions for bioactive contents [119].

The video has been realized in 2020 at “Passetto” location (Figure 36) in Ancona (AN), Italy (Adriatic Sea). In particular, *Chondrosia reniformis* sponge population has been investigated inside and outside a cave. The sea depth of the explored area was about seven meters. The video has been realized by a diver with a Sony RX100V camera (Appendix C - Devices).

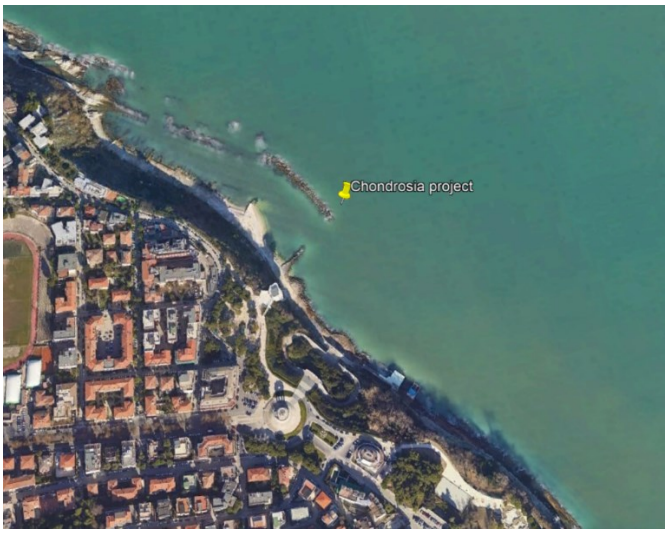

Figure 36 - Sponges diving survey location.

Data Annotation

About 5400 frames have been extracted using Agisoft Metashape (Appendix B - Software) [74]. Each frame is 1920x1080 pixel. A manual annotation procedure has been executed. For this target, the annotation tool created by DII (Dipartimento di Ingegneria dell’Informazione) of Università Politecnica delle Marche (Ancona (AN), Italy) has been exploited. One class has been created: *chondrosia*. A very large number of annotations has been created.

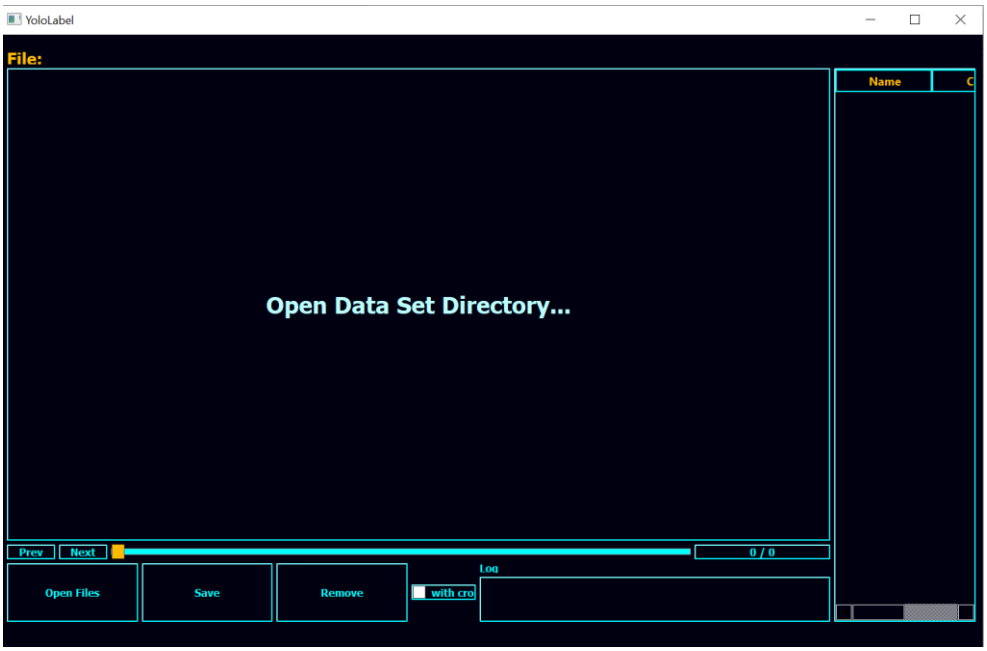

Figure 37 - DII annotation tool dataset upload GUI.

Figure 38 - DII annotation tool frame visualization (sponges dataset).

Figure 37-Figure 38 shows the DII annotation tool GUI that allows uploading all the images contained in a folder.

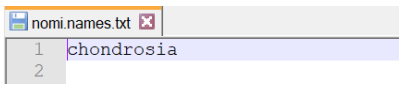

Figure 39 - ".txt" file containing the labels (DII annotation tool).

The labels related to the classes have been defined in a ".txt" file (Figure 39). Box-Level annotations have been executed. Figure 40-Figure 42 report annotation examples on frames that have been used for model training purposes.

Each obtained annotation (corresponding to the rectangular boxes that can be noted in Figure 40-Figure 42) can be automatically saved in a ".txt" file (Figure 43). On each row of the ".txt" file, the class number (start from 0), the normalized $x - y$ axis center and the normalized width and height of the box can be noted. The ".txt" file of DII annotation tool is a very user friendly feature because it allows re-uploading in the future of the started project.

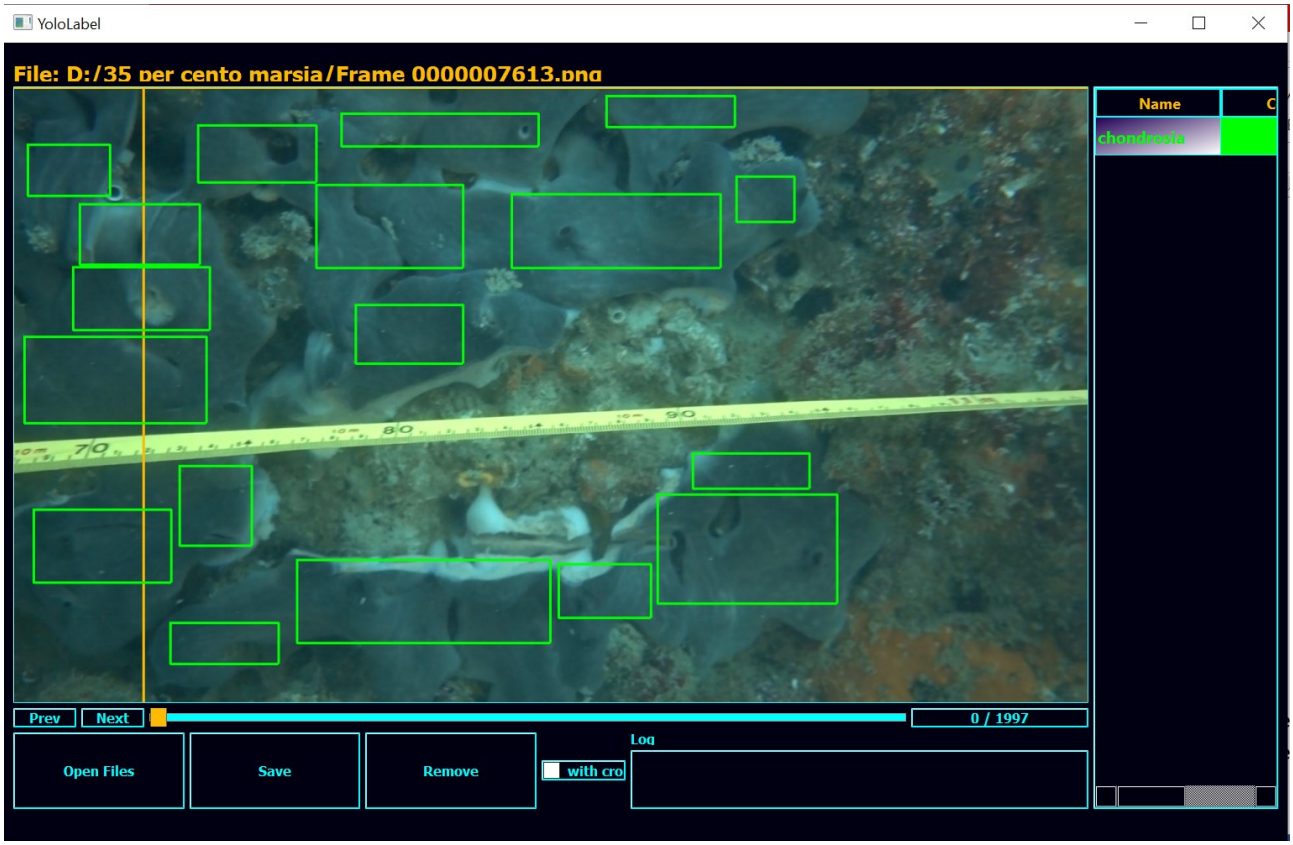

Figure 40 - DII annotation tool Box-Level annotation (sponges dataset) (a).

Figure 41 - DII annotation tool Box-Level annotation (sponges dataset) (b).

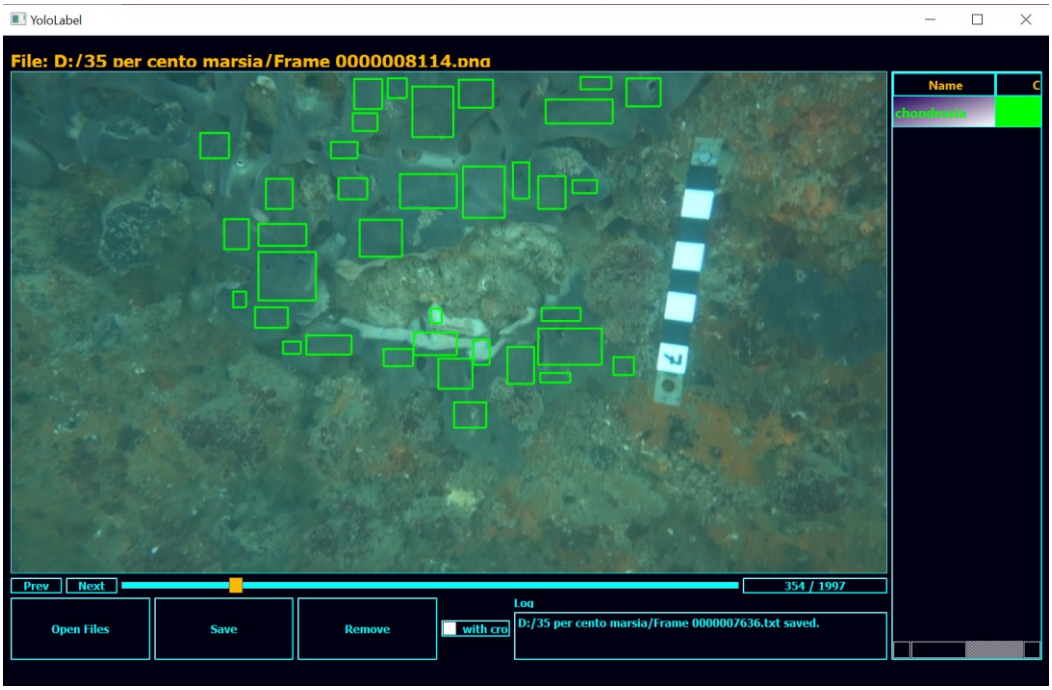

Figure 42 - DII annotation tool Box-Level annotation (sponges dataset) (c).

```
0 0.349609 0.225784 0.136719 0.135061
0 0.560547 0.233288 0.194531 0.120055
0 0.367969 0.401091 0.100000 0.095498
0 0.682813 0.750341 0.167187 0.177353
0 0.686328 0.623465 0.108594 0.057299
0 0.381250 0.835607 0.235937 0.135061
0 0.082031 0.745566 0.128125 0.118690
0 0.094141 0.475443 0.169531 0.140518
0 0.118359 0.343111 0.127344 0.102319
0 0.116797 0.238745 0.111719 0.098226
0 0.226172 0.107776 0.110156 0.092769
0 0.050781 0.134379 0.076563 0.083220
0 0.396484 0.068895 0.183594 0.053206
0 0.611328 0.038881 0.119531 0.050477
0 0.550000 0.818554 0.085938 0.087312
0 0.195703 0.903820 0.100781 0.066849
0 0.187500 0.680082 0.067188 0.129604
0 0.699609 0.181446 0.053906 0.073670
```

Figure 43 - ".txt" file containing the annotations of each frame (DII annotation tool).

Model Training

Using Google Colab platform, DII developed a model training workflow oriented to YOLOv3 detector training. The access to the developed workflow has been made possible through Google Drive. YOLOv3 algorithm has been imported through GitHub [108], [120], [121].

YOLOv3 has been selected as reference detector because of its improved speed and accuracy features documented in several research papers (see 3.5).

A ".zip" folder has been uploaded on Google Drive. The folder contained 252 elements: 126 selected images among the dataset images and the corresponding 126 ".txt" annotation files. In order to have

the option of using all the uploaded images for the training phase, all annotation files have been uploaded. Obviously, all the annotation files of the images that have been effectively selected for the training phase are really used by the software.

On the Google Colab model training workflow, the first step is represented by the automatic creation of three frame subfolders:

- A training subfolder (80 elements).
- A validation subfolder (20 elements).
- A test subfolder (26 elements).

The training frames allow the model parameters fitting (e.g. weights of connections between neurons), while the validation frames allow the hyperparameters (e.g. the number of hidden units (layers and layer widths) in a neural network) tuning [122].

Using the training and validation subfloders, the training procedure has been started defining 1000 epochs. A computation time of about six and a half hours has been registered.

The performance metrics of the training phase have been resumed in Figure 44.

GIoU (Generalized Intersection over Union) represents the attempt of the algorithm to maximize overlap area between the predicted boxes and the ground truth. Objectness, also called confidence loss, represents a metric that takes into account the Intersection over Union (IoU) and the probability that a given object is in the considered bounding box [56]. As can be noted in Figure 44, with the progress of the epochs (on the x axis), GIoU and Objectness losses decrease. GIoU and objectness refer to training data, while val GIoU and val Objectness refer to validation data.

Precision, recall and F1 metrics represents relationships between true positive (TP), false positive (FP) and False Negative (FN) [34].

mAP@0.5 represents a metric that corresponds to the mean Average Precision (mAP): mAP@0.5 is the average of AP with IoU=0.5. AP is the IoU average [56].

The obtained model parameters (weights) from the training phase have been saved in a ".pt" file and the training results have been saved in a ".txt" file. The weights saving option is a very user friendly feature because it represents an updated starting point from which the training phase can be continued or the test phase can be started.

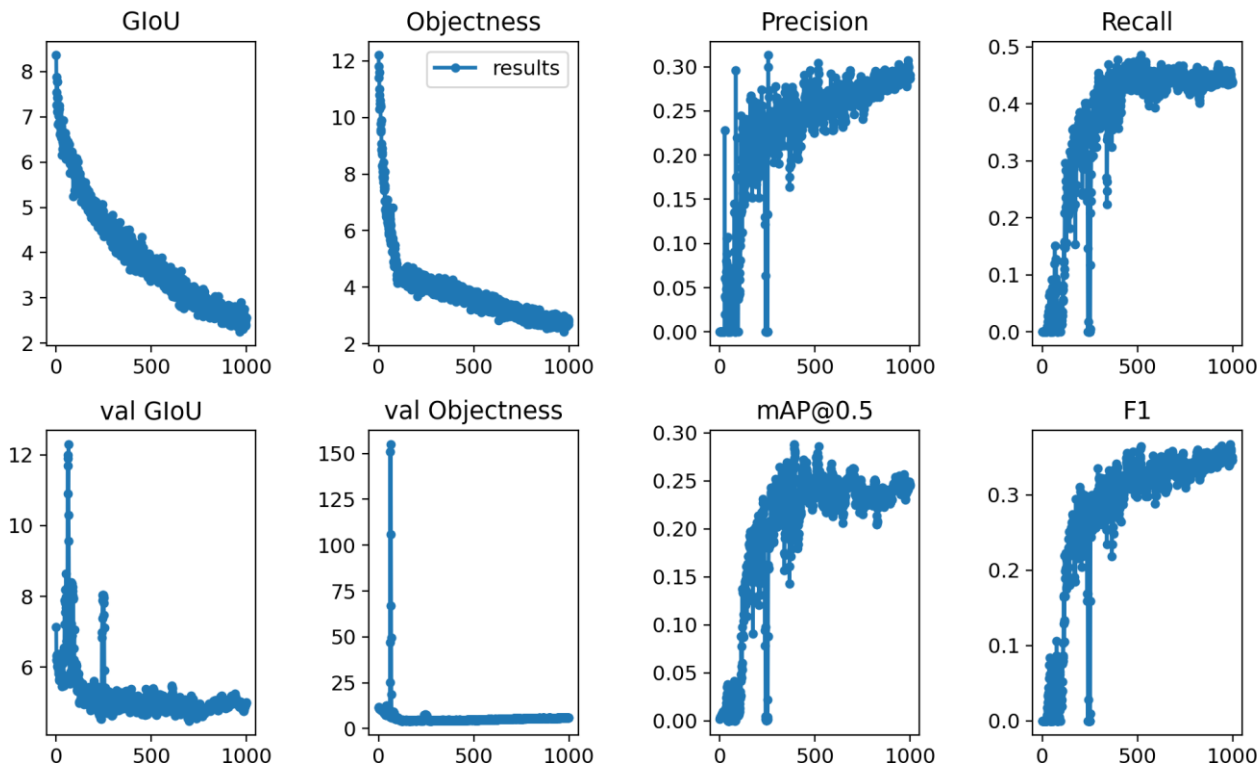

Figure 44 - Performance metrics of the training phase.

Test

Using Google Colab platform, DII developed a test workflow oriented to YOLOv3 detector test. The access to the developed workflow has been made possible through Google Drive.

The test phase has been performed on the test subfolder, using the obtained model parameters of the training phase.

Figure 45-Figure 48 show examples of the test phase on images that have not been used for the training phase. The label *chondrosia*, together with the related rectangular box and the related confidence score (similar to the objectness), are automatically generated by the tool and they can be noted in Figure 45-Figure 48. The rectangular box and the confidence score are outputs of YOLOv3 algorithm: the confidence score informs how certain YOLOv3 algorithm is that the predicted bounding box actually encloses some object. The potential of YOLOv3 algorithm is described by its outputs: in the object detection, YOLOv3 algorithm identifies if an object is present in the image and the class of the object (classification) and predicts the co-ordinates of the bounding box around the object when an object is present in the image (localization).

A comparison between automatic detection and manual detection has been reported in Figure 49-Figure 52.

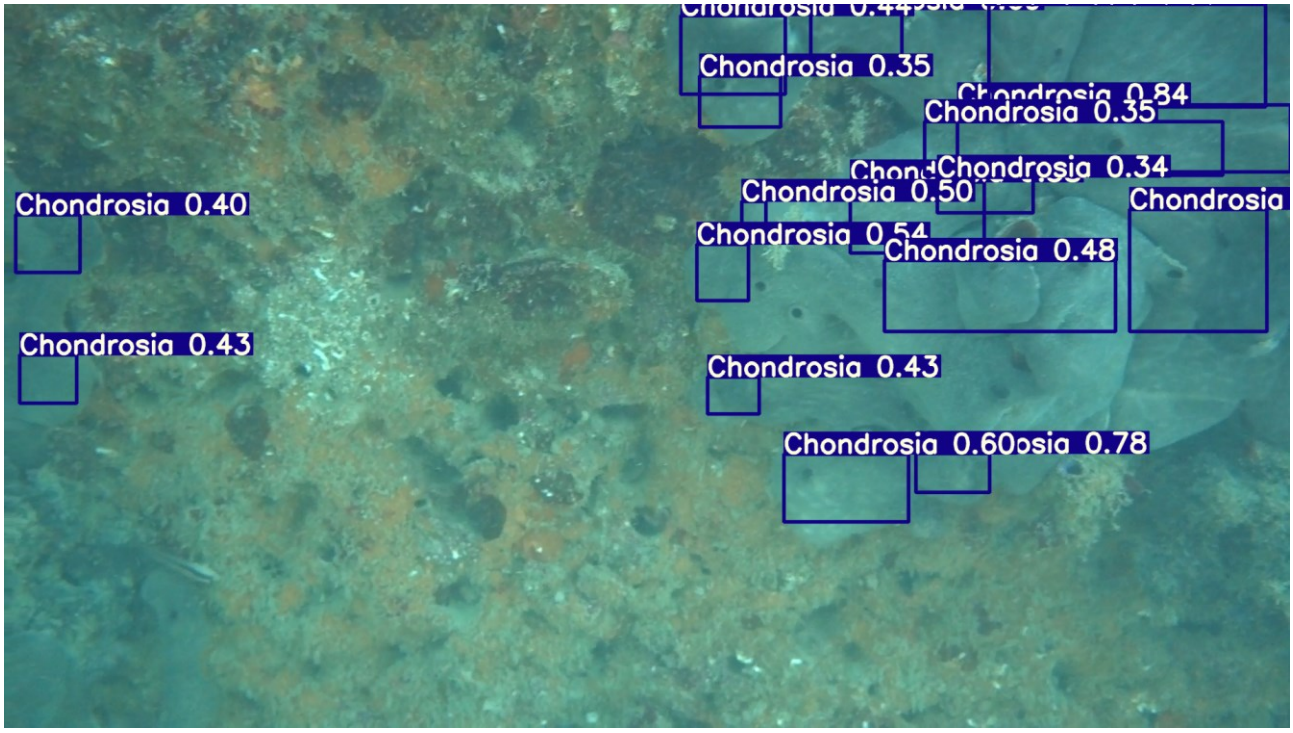

Figure 45 - YOLOv3 test (DII tool), automatic detection (sponges dataset) (a).

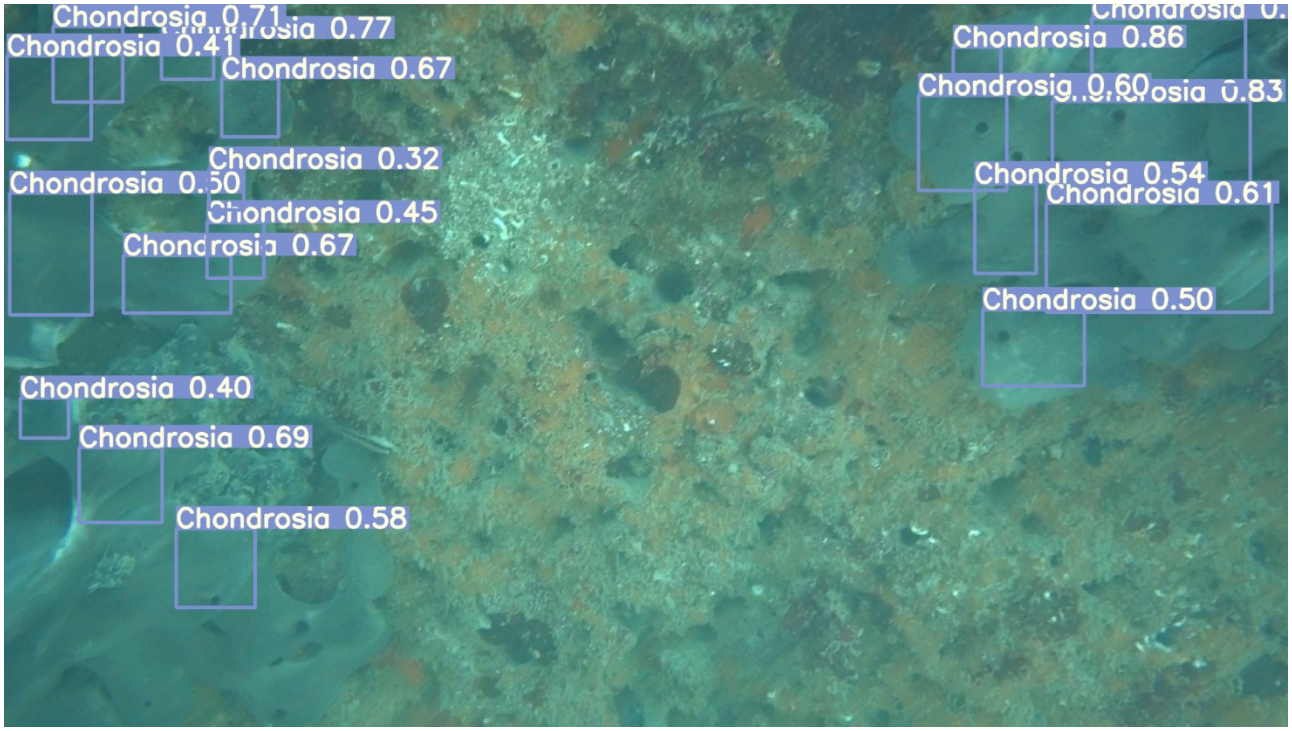

Figure 46 - YOLOv3 test (DII tool), automatic detection (sponges dataset) (b).

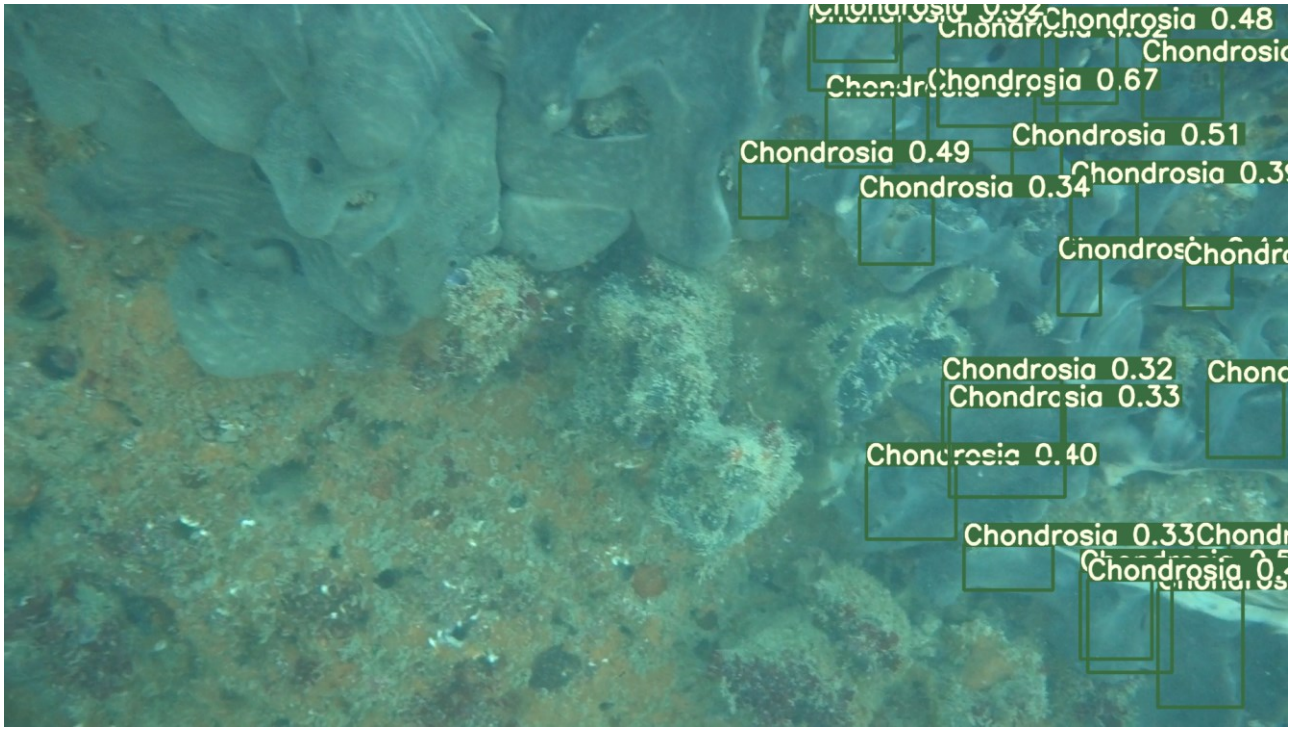

Figure 47 - YOLOv3 test (DII tool), automatic detection (sponges dataset) (c).

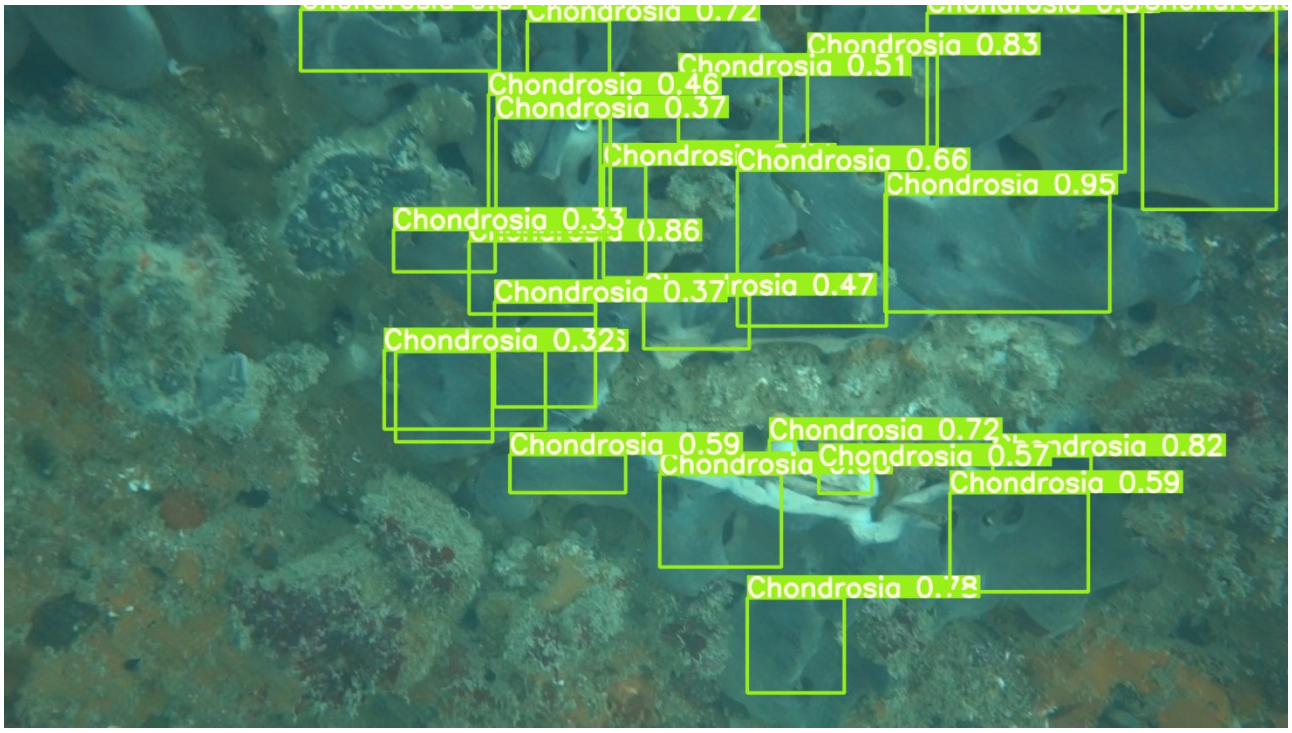

Figure 48 - YOLOv3 test (DII tool), automatic detection (sponges dataset) (d).

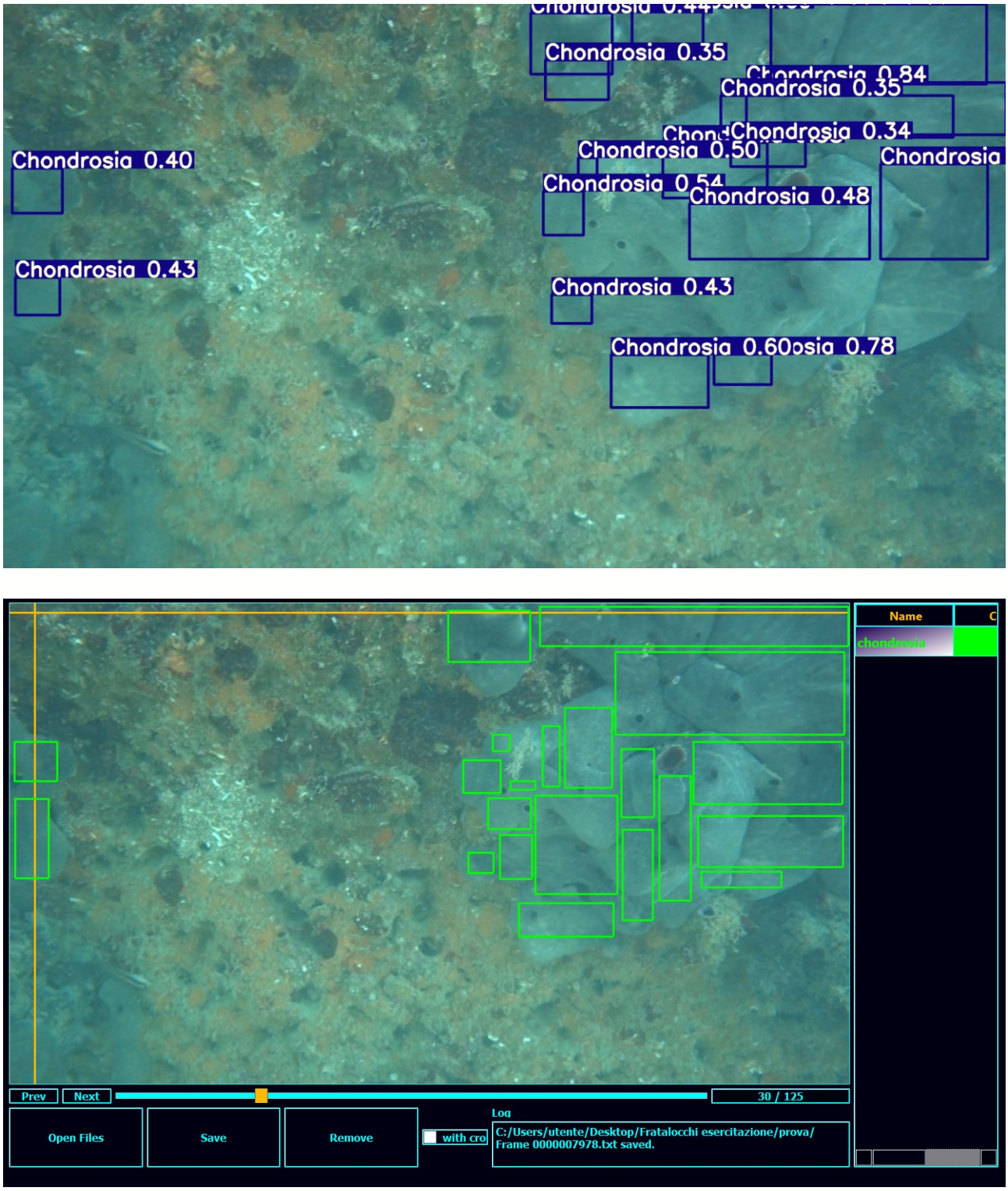

Figure 49 - YOLOv3 test (DII tool), automatic vs manual detection (sponges dataset) (a).

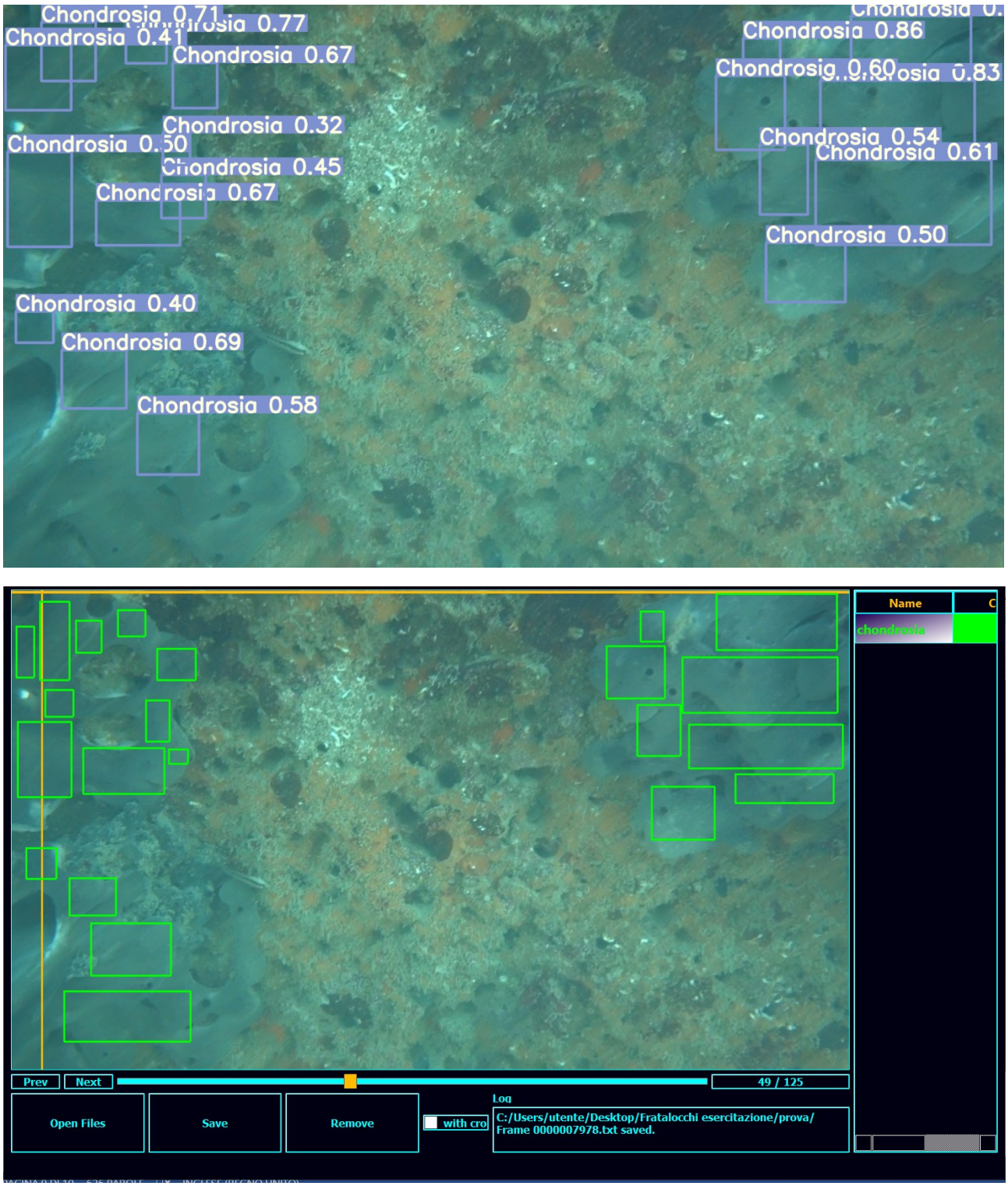

Figure 50 - YOLOv3 test (DII tool), automatic vs manual detection (sponges dataset) (b).

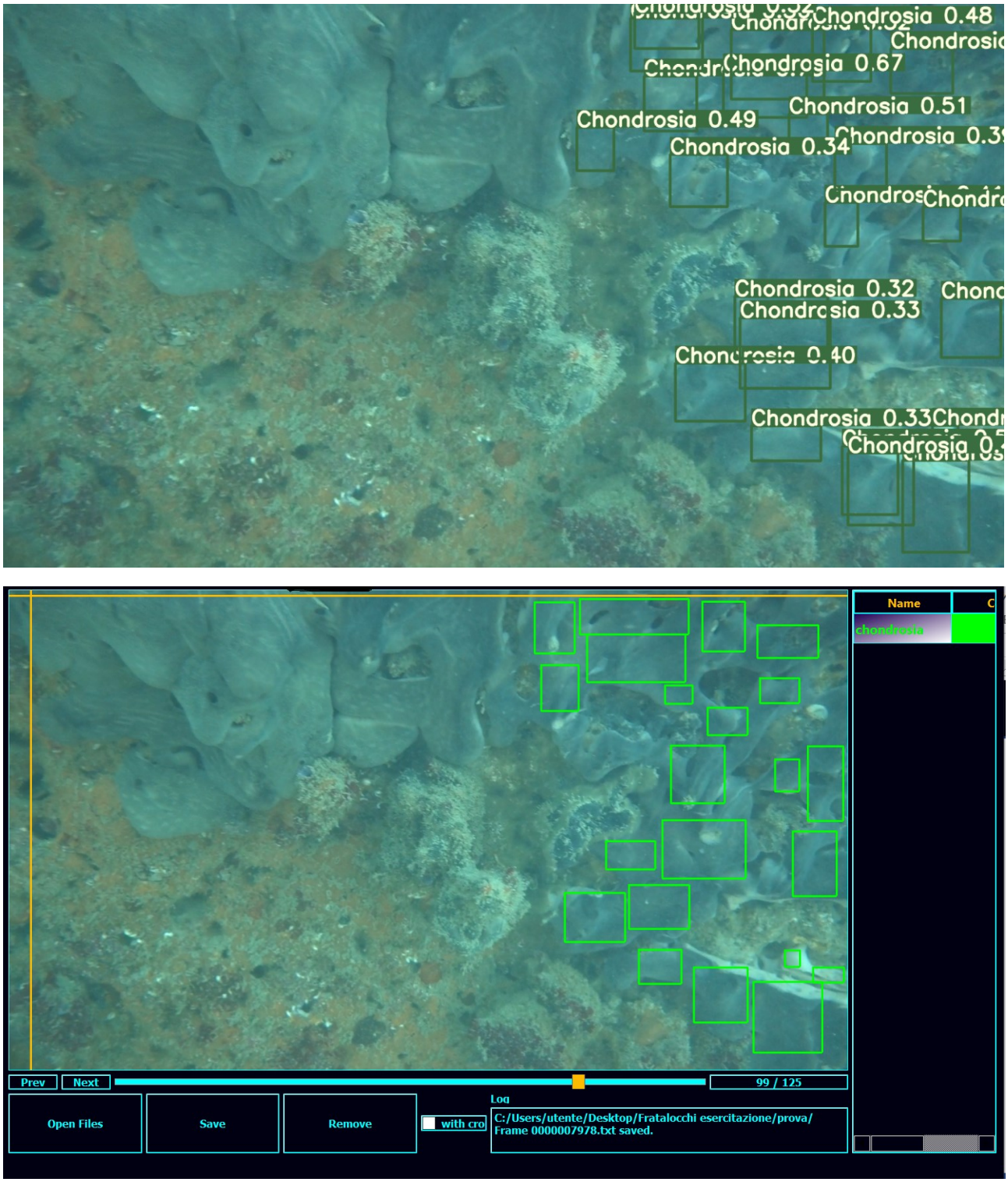

Figure 51 - YOLOv3 test (DII tool), automatic vs manual detection (sponges dataset) (c).

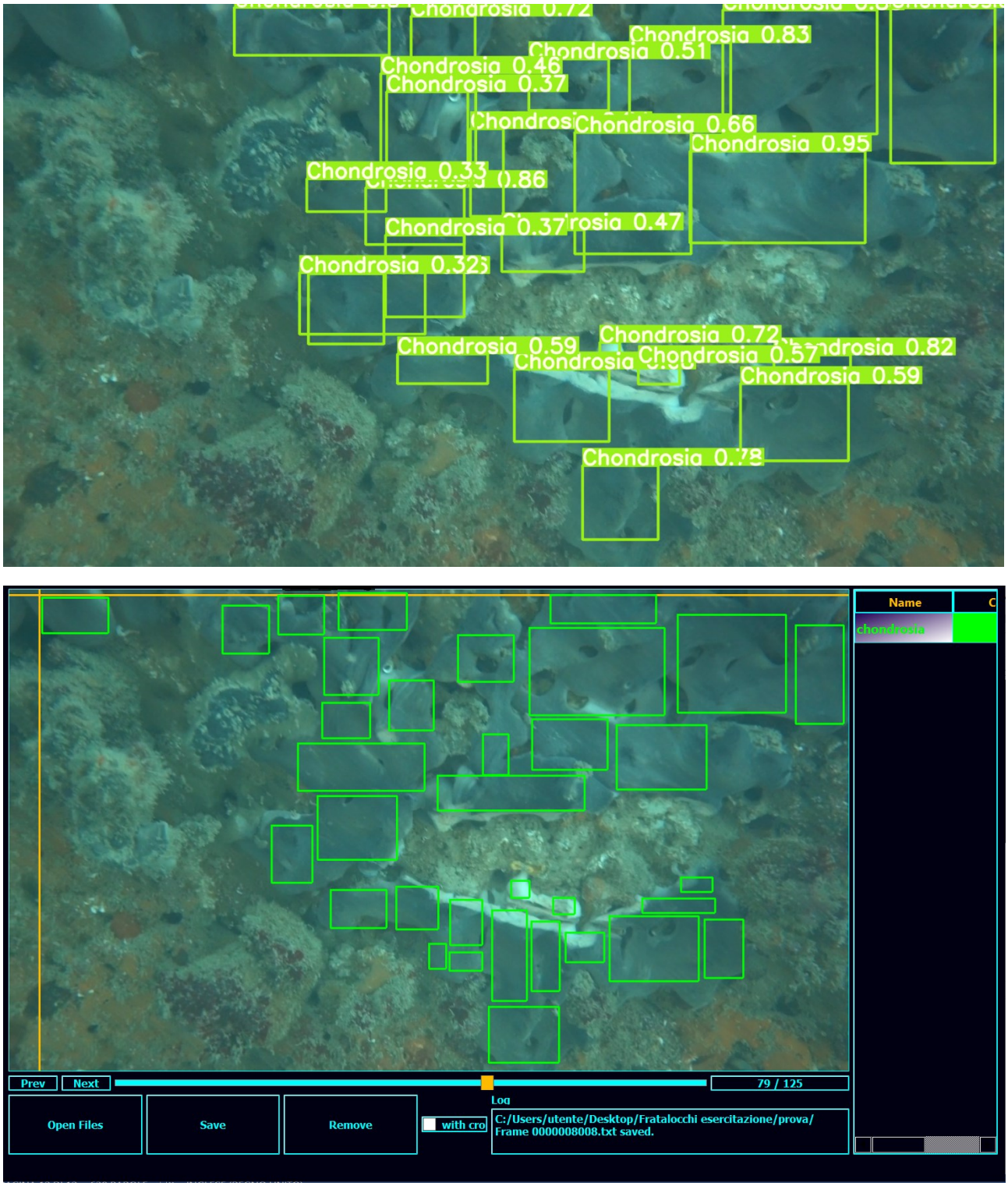

Figure 52 - YOLOv3 test (DII tool), automatic vs manual detection (sponges dataset) (d).

4.3 Fishes

The fishes' case study consisted on the processing of an image dataset. The work on the fishes' dataset will be the object of a challenge proposed to the Ocean Hackathon event [123]. The research target could be oriented not only on the object detection, but also on the object tracking problem.

A further methodological improvement has been obtained in the fishes' detection. With respect to the mussels and to the sponges, the following key aspects must be taken into account:

- The fishes are dynamic objects to be investigated, while mussels and sponges are static objects.
- Fishes swim in schools, so a fish is usually covered by another fish or by another underwater object (overlapping).

The fishes' dataset has been obtained through the processing of some videos provided by DISVA (Dipartimento di Scienze della Vita e dell'Ambiente) of Università Politecnica delle Marche (Ancona (AN), Italy). The videos are related to the SmartBay Observatory. The SmartBay Observatory in Galway Bay is an underwater observatory, which uses cameras, probes and sensors to permit continuous and remote live underwater monitoring. It was installed in 2015 on the seafloor 1.5 km off the coast of Spiddal, Co. Galway, Ireland at a depth of 20-25m. Underwater observatories allow ocean researchers unique real-time access to monitor ongoing changes in the marine environment. The Galway Bay Observatory is an important contribution by Ireland to the growing global network of real-time data capture systems deployed in the ocean. Data relating to the marine environment at the Galway Observatory site is transferred in real-time through a fiber optic telecommunications cable to the Marine Institute headquarters and then made publicly available on the internet. The data includes a live video stream, the depth of the observatory node, the water temperature and salinity, and estimates of the chlorophyll and turbidity levels in the water, which give an indication of the volume of phytoplankton and other particles, such as sediment, in the water. Maintenance take place on the observatory every 18 to 24 months. Video data is streamed in near-real-time from the observatory and this dataset describes the video footage that is available for download [124].

A Kongsberg High Definition Pan & Tilt Zoom Camera (Appendix C - Devices) is mounted on the observatory, which can be panned, tilted, zoomed and focused in real-time to give a unique view of subsea life in the vicinity of the observatory. The data is streamed live to the internet [125].

Figure 53 reports the set of fishes imaged at the SmartBay cabled observatory [126].

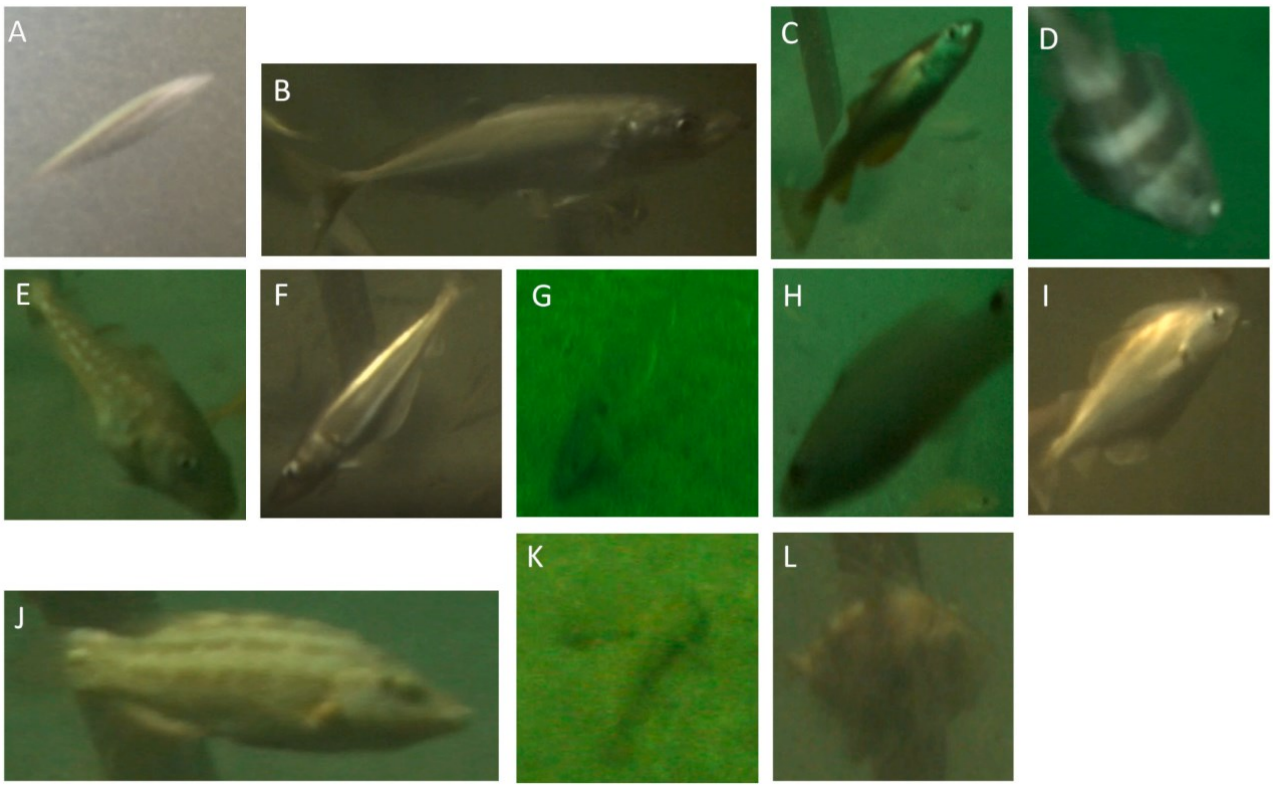

Figure 53 - List of fishes imaged at the SmartBay cabled observatory, as (A) Clupea harengus, (B) Trachurus trachurus, (C) Trisopterus minutus, (D) Trisopterus luscus, (E) Gadus morhua, (F) Pollachius pollachius (G) Melanogrammus aeglefinus, (H) Ctenolabrus rupestris, (I) Merlangius merlangus, (J) Labrus rupestris, (K) Chelidonichthys lucerna, and finally (L) Zeus faber [126].

Data Annotation

Among the available videos, 11 videos (02:00 (mm:ss)) related to 2018 have been selected, monitoring both daytime and night periods in different seasons (see Table 1).

Table 1 - Overview of the selected periods of 2018 (SmartBay Observatory).

Night		Daytime	
Date	**Hour**	**Date**	**Hour**
03.01.2018	00:00	02.01.2018	13:30
03.03.2018	00:00	03.03.2018	13:30
03.05.2018	01:00	-	-
02.08.2018	23:43	03.09.2018	13:52
02.10.2018	00:01	08.11.2018	13:59
03.12.2018	23:59	03.12.2018	13:55

About 11000 frames have been extracted using Agisoft Metashape (Appendix B - Software) [74]. Each frame is 1280x720 pixel.

A manual annotation procedure has been executed. For the fishes' annotation procedure, the two

following rules have been adopted:

- A fish must be annotated only if it is covered by other objects no more than 10% of its surface.
- A fish must be annotated only if its class is evidently identifiable.

For the annotations execution, the annotation tool created by DII (Dipartimento di Ingegneria dell'Informazione) of Università Politecnica delle Marche (Ancona (AN), Italy) has been exploited.

Twelve classes have been created (Figure 53). A very large number of annotations has been created. Box-Level annotations have been executed. Figure 54-Figure 57 report annotation examples.

Figure 54 - DII annotation tool Box-Level annotation (SmartBay dataset) (a).

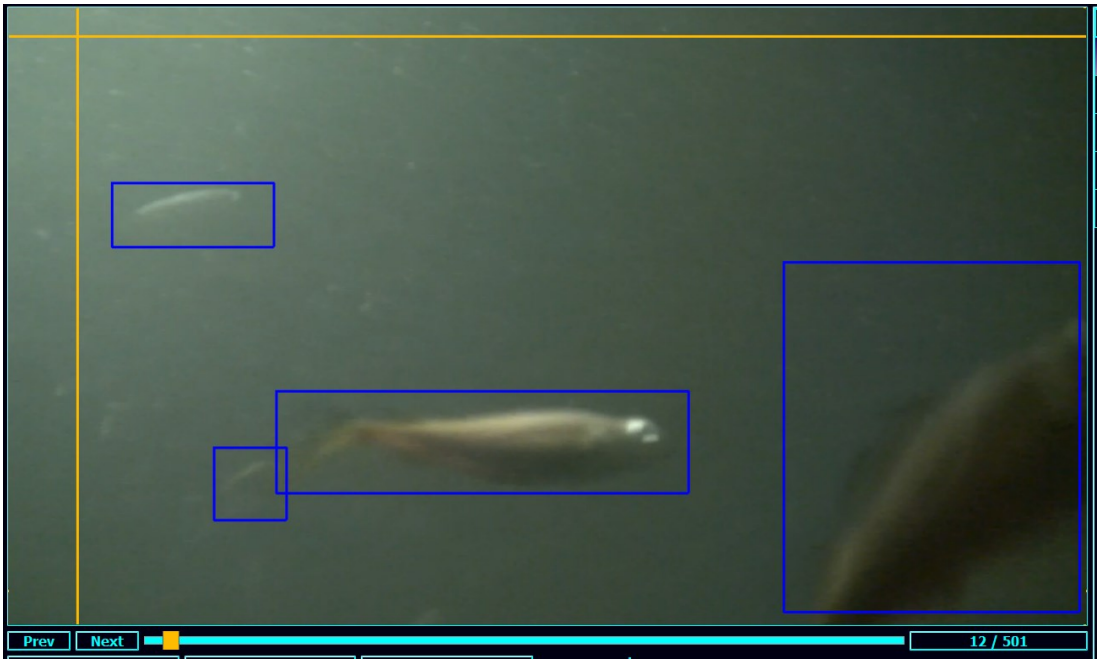

Figure 55 - DII annotation tool Box-Level annotation (SmartBay dataset) (b).

Figure 56 - DII annotation tool Box-Level annotation (SmartBay dataset) (c).

Figure 57 - DII annotation tool Box-Level annotation (SmartBay dataset) (d).

5. Conclusion and future work

In the recent years, topics like sustainable development and biodiversity monitoring related to marine ecosystem have attracted the attention of researchers of different fields.

The necessity to develop spatially widespread monitoring networks that allow to understand the ideal marine environment is due to the complexity of the marine ecosystem. This complexity leads to the need of analyzing a large amount of data that must be labelled and classified. This process is crucial for capture significant information and let the researchers to survey the underwater environment that is still almost unknown.

In this thesis work, innovative methodological approaches for marine communities assessments have been studied and applied on real case studies. In particular, the target of the thesis work has been represented by the application of classification and object detection methods in underwater environment. The work has been oriented to obtain marine communities assessments. The interested marine communities have been mussels, sponges and fishes.

In the thesis development, underwater geomatics, underwater biology and computer science theoretical and practical ideas have been merged, in order to obtain a multidisciplinary methodology.

The identification of the selected marine communities and their labelling has been obtained thanks to biology. Existing underwater datasets have been evaluated thanks to geomatics key points. At the same time geomatics theory has been applied in the design of diving surveys oriented to underwater datasets creation. Mussels existing datasets have been used; furthermore, the start of a proprietary mussels and sponges dataset has been started, using also image enhancement methods.

Geomatics concepts have been also applied for the development of a photogrammetric approach to the object classification (mussels), exploiting 3D modelling and CANUPO classifier for point clouds. Encouraging preliminary results have been obtained in the mussels' classification.

In order to test the potential of computer science when tied to underwater geomatics and biology, innovative deep learning software has been used for object labelling and detection. In particular, YOLOv3 model has been employed. The application has interested the sponges and the fishes. Also in these applications, satisfying pilot results have been obtained and other evaluations are under development.

A gradual methodological evolution has been obtained in the thesis work: geomatics concepts have been step by step enriched by biology and computer science ideas.

The future work will be addressed on further studies on geomatics and deep learning methods for

object detection in different environments. For example, different deep learning models will be compared on the created datasets in order to find the best one based on the required specifications. Additional underwater surveys will be conducted, in order to expand the created datasets and to improve the methodology that has been applied, both in geomatics and computer science fields.

The thesis work has been developed through a cooperation between DICEA (Dipartimento di Ingegneria Civile, Edile e Architettura), DISVA (Dipartimento di Scienze della Vita e dell'Ambiente) and DII (Dipartimento di Ingegneria dell'Informazione) of Università Politecnica delle Marche (Ancona (AN), Italy).

Bibliography

[1] M. Moniruzzaman, S. M. S. Islam, M. Bennamoun and P. Lavery, "Deep Learning on Underwater Marine Object Detection: A Survey," in *Advanced Concepts for Intelligent Vision Systems. ACIVS 2017. Lecture Notes in Computer Science*, 2017.

[2] "https://www.greenpeace.org/italy/storia/5197/ecco-perche-dobbiamo-trattare-bene-gli-oceani/#:~:text=Ma%20non%20%C3%A8%20solo%20una,difese%20contro%20i%20cambiamenti%20climatici," [Online].

[3] "https://www.wwf.ch/it/stories/le-minacce-del-nostro-mediterraneo-7-fatti-sorprendenti," [Online].

[4] "https://www.repubblica.it/ambiente/2019/06/14/news/_con_queste_reti_salveremo_migliaia_di_tartarughe_in_italia_-228760350/," [Online].

[5] "https://www.wwf.it/?40340%2FSalviamo-le-tartarughe-anche-dalla-plastica," [Online].

[6] "https://orbmedia.org/stories/Invisibles_plastics/," [Online].

[7] "https://www.reteclima.it/plastiche-microplastiche-mari-oceani-infografica-sintesi/#:~:text=La%20presenza%20delle%20microplastiche%20in,forma%20di%20microplastica%20o%20di," [Online].

[8] H. Qin and al., "DeepFish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing,* 2015.

[9] "https://it.euronews.com/2020/03/31/proteggere-la-biodiversita-marina-per-proteggere-noi-stessi," [Online].

[10] F. Betti, G. Bavestrello, M. Bo, F. Enrichetti and R. Cattaneo-Vietti, "Effects of the 2018 exceptional storm on the Paramuricea clavata (Anthozoa, Octocorallia) population of the Portofino Promontory (Mediterranean Sea)," *Regional Studies in Marine Science,* vol. 34, February 2020.

[11] "https://www.youtube.com/watch?v=62PxxzRIOYI&ab_channel=TEDxTalks," [Online].

[12] "https://www.google.com/imgres?imgurl=https%3A%2F%2Fthumbs.dreamstime.com%2Fb%2Fcoralli-molli-ed-allegati-delle-uova-della-seppia-dei-bambini-lavate-ancora-via-sulla-riva-di-mare-allegate-su-corallo-molle-e-143100014.jpg&imgrefurl=https%3A%2F%2Fit.dreamsti," [Online].

[13] "https://www.lescienze.it/news/2018/07/31/news/pesca_illegale_transhipping_trasbordo_satellite-4062228/?rss," [Online].

[14] "https://www.fiorital.com/strascico-di-fondale/," [Online].

[15] "https://www.wwf.it/il_pianeta/impatti_ambientali/pesca_eccessiva_e_illegale/," [Online].

[16] A. Colletti, B. Savinelli, G. Di Muzio, L. Rizzo, L. Tamburello, S. Fraschetti, L. Musco and R. Danovaro, "The date mussel Lithophaga lithophaga: Biology, ecology and the multiple impacts of its illegal fishery," *Science of the Total Environment,* vol. 744, 2020.

[17] "DIMINUIRE LA PRESSIONE DELLA PESCA IN ADRIATICO," *ECOSCIENZA,* 6 2014.

[18] G. Pedicillo, Indici di struttura ed accrescimento standard per le principali specie ittiche autoctone nel bacino del fiume Tevere, Università degli Studi di Perugia, Dottorato di Ricerca in Biologia e Ecologia, 2009/2010.

[19] "https://www.ingenio-web.it/24111-che-cosa-e-lagenda-2030-per-lo-sviluppo-sostenibile," [Online].

[20] "https://www.un.org/sustainabledevelopment/development-agenda/," [Online].

[21] U. Nations, "Resolution adopted by the General Assembly on 25 September 2015," 2015.

[22] U. Nations, "https://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf," [Online].

[23] A. I. p. l. S. S. (ASVIS), "Trasformare il nostro mondo: l'Agenda 2030 per lo sviluppo sostenibile," 2015.

[24] W. Xu and S. Matzner, "Underwater Fish Detection using Deep Learning for Water Power Applications," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018.

[25] M. FABRIS, M. ANZIDEI and P. BALDI, "ANALISI DI MODELLI DIGITALI DEL TERRENO DELL ISOLA DI PANAREA (ISOLE EOLIE) ESTRATTI CON DIFFERENTI METODOLOGIE," in *Atti 12a Conferenza Nazionale ASITA*, L'Aquila, 2008.

[26] P. Drap, L. Long and D. Scaradozzi, "PROGETTO EUROPEO VENUS: RILIEVO AUTOMATICO E REALTÀ VIRTUALE, EVOLUZIONI E RISULTATI DALLA SUA CONCLUSIONE," in *Convegno Conoscenza e Tutela del Patrimonio Sommerso*, Pisa, 2012.

[27] M. A. Gomarasca, Basics of Geomatics, Springer, 2009.

[28] A. Manzino, Lezioni di Topografia, 2000.

[29] D. Pan and al., "3D scene and geological modeling using integrated multi-source spatial data: Methodology, challenges, and suggestions," *Tunnelling and Underground Space Technology,* vol. 10, 2020.

[30] L. C. Kapotas, *Fotogrammetria subacquea in archeologia,* 2008/2009.

[31] A. M. Turing, "COMPUTING MACHINERY AND INTELLIGENCE," in *Mind*, vol. 59, October 1950, pp. 433-460.

[32] J. A. Perez, F. Deligianni, D. Ravi and G. Z. Yang, *Artificial Intelligence and Robotics,* Robotics and Autonomous Systems (RAS), 2016.

[33] A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development,* p. 210–229, July 1959.

[34] S. N. M. Venkata, *Detection and Tracking of Humans in an Underwater Environment Using Deep Learning Algorithms,* 2019.

[35] H. R. Pamuluri, *Predicting User Mobility using Deep Learning Methods,* 2020.

[36] E. KOHMANN, *Tecniche di deep learning per l'object detection,* 2019.

[37] "https://machinelearningmastery.com/what-is-deep-learning/," [Online].

[38] "https://it.mathworks.com/discovery/deep-learning.html," [Online].

[39] "https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc," [Online].

[40] M. Z. Alom and al., "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," *arXiv.org,* 2018.

[41] J. E. Solem, Programming Computer Vision with Python: Tools and algorithms for analyzing images, Sebastopol, CA: O'Reilly Media, June 2012.

[42] J. Brownlee, A Gentle Introduction to Computer Vision, March 2019.

[43] M. Sung, S. C. Yu and Y. Girdhar, "Vision based realtime fish detection using convolutional neural network," in *IEEE OCEANS 2017-Aberdeen*, 2017.

[44] R. Szeliski, Computer Vision: Algorithms and Applications, London ; New York: Springer, November 2010.

[45] Y. Xu and al., "Underwater image classification using deep convolutional neural networks and data augmentation," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2017.

[46] Z. Zou and al., *Object Detection in 20 Years: A Survey,* IEEE, 2019.

[47] K. E. Van de Sande and al., "Segmentation as selective search for object recognition," in *2011 IEEE International Conference on Computer Vision (ICCV)*, 2011.

[48] "https://d2l.ai/chapter_computer-vision/rcnn.html," [Online].

[49] K. He and al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *European conference on computer vision*, 2014.

[50] "https://paperswithcode.com/method/spp-net#:~:text=SPP%2DNet%20is%20a%20convolutional,of%20the%20last%20convolutional%20layer.," [Online].

[51] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015.

[52] "https://it.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.html," [Online].

[53] S. Ren and al., "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, pp. 1137-1149, 2017.

[54] J. Redmon and al., "You Only Look Once: Unified, Real-Time Object Detection," *arXiv,* 2016.

[55] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[56] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv,* 2018.

[57] W. Liu and al., "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016.

[58] T.-Y. Lin and al., "Feature pyramid networks for object detection," *Review of Scientific Instruments,* 2017.

[59] T.-Y. Lin and al., "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence,* 2018.

[60] S. Villon, D. Mouillot, M. Chaumont, E. Darling, G. Subsol and al., "A Deep learning method for accurate and fast identification of coral reef fishes in underwater images," *Ecological Informatics,* vol. 48, pp. 238-244, 2018.

[61] V. Lopez-Vazquez, J. Lopez-Guede, S. Marini, E. Fanelli, E. Johnsen and J. Aguzzi, "Video Image Enhancement and Machine Learning Pipeline for Underwater Animal Detection and Classification at Cabled Observatories," *Sensors,* vol. 20, no. 3, 2020.

[62] H. Ahmad, "Machine learning applications in oceanography," *Aquatic Research,* vol. 2, no. 3, pp. 161-169, 2019.

[63] L. Chen, L. Tong, F. Zhou, Z. Jiang, Z. Li, J. Lv, J. Dong and H. Zhou, "A Benchmark dataset for both underwater image enhancement and underwater object detection," *Computer Science,* 2020.

[64] M. Pedersen, J. B. Haurum, R. Gade and T. Moeslund, "Detection of Marine Animals in a New Underwater Dataset with Varying Visibility," in *CVPR Workshops 2019*, 2019.

[65] C. Li and al., "An Underwater Image Enhancement Benchmark Dataset and Beyond," *IEEE Transactions on Image Processing,* vol. 29, pp. 4376-4389, 2020.

[66] "https://imareculture.eu/," [Online].

[67] M. Mangeruga, M. Cozza and F. Bruno, "Evaluation of Underwater Image Enhancement Algorithms under Different Environmental Conditions," *J. Mar. Sci. Eng.,* 2018.

[68] E. Irmak and A. H. Ertas, "A review of robust image enhancement algorithms and their applications," in *2016 IEEE Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, 2016.

[69] P. Agrafiotis, D. Skarlatos, T. Forbes, C. Poullis, M. Skamantzari and A. Georgopoulos, "UNDERWATER PHOTOGRAMMETRY IN VERY SHALLOW WATERS: MAIN CHALLENGES AND CAUSTICS EFFECT REMOVAL," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2018.

[70] C. Li, S. Anwar and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition,* vol. 98, 2020.

[71] F. Menna, E. Nocerino and F. Remondino, "FLAT VERSUS HEMISPHERICAL DOME PORTS IN UNDERWATER PHOTOGRAMMETRY," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2017.

[72] E. Nocerino, F. Menna, F. Fassi and F. Remondino, "UNDERWATER CALIBRATION OF DOME PORT PRESSURE HOUSINGS," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2016.

[73] E. Nocerino, F. Neyer, A. Gruen, M. Troyer, F. Menna, A. Brooks, A. Capra, C. Castagnetti and P. Rossi, "Comparison of Diver-Operated Underwater Photogrammetric Systems for Coral Reef Monitoring," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2019.

[74] "https://www.agisoft.com/," [Online].

[75] N. Börlin and P. Grussenmeyer, "Bundle adjustment with and without damping," *The Photogrammetric Record,* vol. 28, no. 144, pp. 396-415, 2013.

[76] F. Menna, E. Nocerino, P. Drap, F. Remondino, A. Murtiyoso, P. Grussenmeyer and N. Börlin, "IMPROVING UNDERWATER ACCURACY BY EMPIRICAL WEIGHTING OF IMAGE OBSERVATIONS," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2018.

[77] P. Rossi, C. Castagnetti, A. Capra and al., "Detecting change in coral reef 3D structure using underwater photogrammetry: critical issues and performance metrics," *Appl Geomat,* pp. 3-17, 2020.

[78] F. Neyer, E. Nocerino and A. Grün, "Image Quality Improvements in Low-Cost Underwater Photogrammetry," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 42, 2019.

[79] "http://www.cloudcompare.org/," [Online].

[80] D. W. James, F. Belblidia, J. E. Eckermann and J. Sienz, "An Innovative Photogrammetry Color Segmentation based Technique as an Alternative Approach to 3D Scanning for Reverse Engineering Design," *COMPUTER-AIDED DESIGN & APPLICATIONS,* vol. 14, pp. 1-16, 2017.

[81] P. Piazza, V. Cummings, A. Guzzi and al., "Underwater photogrammetry in Antarctica: long-term observations in benthic ecosystems and legacy data rescue," *Polar Biol,* vol. 42, p. 1061–1079, 2019.

[82] P. Piazza, V. J. Cummings, D. M. Lohrer and al., "Divers-operated underwater photogrammetry: Applications in the study of antarctic benthos," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 2018.

[83] T. Guo, A. Capra, M. Troyer, A. Gruen, A. J. Brooks, J. L. Hench, R. J. Schmitt, S. J. Holbrook and M. Dubbini, "Accuracy assessment of underwater photogrammetric three dimensional modelling for coral reefs," in *23rd International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Congress, ISPRS 2016*, 2016.

[84] M. Palma, M. Rivas Casado, U. Pantaleo, G. Pavoni, D. Pica and C. Cerrano, "SfM-Based Method to Assess Gorgonian Forests (Paramuricea clavata (Cnidaria, Octocorallia))," *Remote Sens.,* vol. 10, p. 1154, 2018.

[85] I. Lochhead and N. Hedley, "3D MODELLING IN TEMPERATE WATERS: BUILDING RIGS AND DATA SCIENCE TO SUPPORT GLASS SPONGE MONITORING EFFORTS IN COASTAL BRITISH COLUMBIA," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2020.

[86] M. Vlachos, L. Berger, R. Mathelier, P. Agrafiotis and D. Skarlatos, "SOFTWARE COMPARISON FOR UNDERWATER ARCHAEOLOGICAL PHOTOGRAMMETRIC APPLICATIONS," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2019.

[87] "http://ccwu.me/vsfm/," [Online].

[88] "https://www.nframes.com/," [Online].

[89] "https://www.3dflow.net/it/software-di-fotogrammetria-3df-zephyr/," [Online].

[90] "https://www.capturingreality.com/," [Online].

[91] I. Sarakinou, K. Papadimitriou, O. Georgoula and P. Patias, "UNDERWATER 3D MODELING: IMAGE ENHANCEMENT AND POINT CLOUD FILTERING," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2016.

[92] B. Hopkinson, A. King, D. Owen, M. Johnson-Roberson, M. Long and S. Bhandarkar, "Automated classification of three-dimensional reconstructions of coral reefs using convolutional neural networks," *PLoS ONE,* vol. 15, no. 3, 2020.

[93] L. Olinger, A. Scott, S. McMurray and al., "Growth estimates of Caribbean reef sponges on a shipwreck using 3D photogrammetry," *Sci Rep,* vol. 9, 2019.

[94] S. Marini, L. Corgnati, C. Mantovani, M. Bastianini, E. Ottaviani, E. Fanelli, J. Aguzzi, A. Griffa and P.-M. Poulain, "Automated estimate of fish abundance through the autonomous imaging device GUARD1," *Measurement,* vol. 126, pp. 72-75, October 2018.

[95] S. Marini, E. Fanelli, V. Sbragaglia and al., "Tracking Fish Abundance by Underwater Image Recognition," *Sci Rep,* vol. 8, 2018.

[96] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan and R. Fisher, "Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos," in *Proc. 3rd Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, 2008.

[97] G. Pavoni, M. Corsini, M. Callieri, M. Palma and R. Scopigno, "SEMANTIC SEGMENTATION OF BENTHIC COMMUNITIES FROM ORTHO-MOSAIC MAPS," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2019.

[98] H. Mohamed, K. Nadaoka and T. Nakamura, "Towards Benthic Habitat 3D Mapping Using Machine Learning Algorithms and Structures from Motion Photogrammetry," *Remote Sens.,* vol. 12, 2020.

[99] B. J. Boom, J. He, S. Palazzo, P. X. Huang, C. Beyan, H.-M. Chou, F.-P. Lin, C. Spampinato and R. B. Fisher, "A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage," *Ecological Informatics,* vol. 23, pp. 83-97, 2014.

[100] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnol. Oceanogr. Methods,* vol. 14, pp. 570-585, 2016.

[101] A. Salman, S. A. Siddiqui, F. Shafait, A. Mian, M. R. Shortis, K. Khurshid, A. Ulges and U. Schwanecke, "Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system," *ICES Journal of Marine Science,* vol. 77, no. 4, pp. 1295-1307, July-August 2020.

[102] "http://www.godac.jamstec.go.jp/jedi/e/," [Online].

[103] T. R. and al., "Evolution of a benthic imaging system from a towed camera to an automated habitat characterization system," in *OCEANS 2008*, 2008.

[104] M. Bewley, A. Friedman, R. Ferrari and al., "Australian sea-floor survey data, with images and expert annotations," *Sci Data,* vol. 2, 2015.

[105] "https://squidle.org/," [Online].

[106] "https://figshare.com/," [Online].

[107] H. Qin, X. Li, J. Liang, Y. Peng and C. Zhang, "DeepFish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing,* vol. 187, pp. 49-58, 2016.

[108] "https://github.com/qinhongwei/deepfish-release," [Online].

[109] M. O'Byrne, V. Pakrashi, F. Schoefs and B. Ghosh, "Semantic Segmentation of Underwater Imagery Using Deep Networks Trained on Synthetic Imagery," *J. Mar. Sci. Eng.,* vol. 6, 2018.

[110] M. Dawkins and al., "An Open-Source Platform for Underwater Image and Video Analytics," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017.

[111] "https://github.com/VIAME/VIAME," [Online].

[112] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv,* 2018.

[113] L. S. and al., "Embedded Online Fish Detection and Tracking System via YOLOv3 and Parallel Correlation Filter," in *OCEANS 2018 MTS/IEEE*, 2018.

[114] T. Shi, Y. Niu, M. Liu, Y. Yang, C. Wang and Y. Huang, "Underwater Dense Targets Detection and Classification based on YOLOv3," in *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2019.

[115] C. Rasmussen, J. Zhao, D. Ferraro and A. Trembanis, "Deep Census: AUV-Based Scallop Population Monitoring," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017.

[116] "http://www.cloudcompare.org/doc/wiki/index.php?title=CANUPO_(plugin)," [Online].

[117] "https://3dmetrica.it/," [Online].

[118] "colab.research.google.com," [Online].

[119] "https://www.awi.de/en/science/special-groups/aquaculture/aquaculture-research/projects/medspon.html," [Online].

[120] "https://github.com/ultralytics/yolov3," [Online].

[121] "https://pjreddie.com/darknet/yolo/," [Online].

[122] "https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets," [Online].

[123] "https://www.cnr.it/it/evento/16852/ocean-hackathon-48-ore-per-decodificare-il-mare," [Online].

[124] "http://data.marine.ie/geonetwork/srv/eng/catalog.search#/metadata/ie.marine.data:dataset.3880," [Online].

[125] "http://smartbay.marine.ie/," [Online].

[126] J. Aguzzi, D. López-Romero, S. Marini, C. Costa, A. Berry, R. Chumbinho, T. Ciuffardi, E. Fanelli, N. Pieretti, J. Del Río, S. S., L. Mirimin, J. Doyle, C. Lordan and P. Gaughan, "Multiparametric monitoring of fish activity rhythms in an Atlantic coastal cabled observatory," *Journal of Marine Systems,* vol. 212, 2020.

[127] "https://www.sony.it/," [Online].

[128] "https://pdf.nauticexpo.com/pdf/kongsberg-maritime/oe14-522-high-definition-pan-tilt-zoom-patz-camera/31233-81610.html," [Online].

[129] "https://www.marktorr.com/deep-learning/," [Online].

Appendix A - Repository

figshare

figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner. figshare is characterized by different tools, e.g. a desktop uploader and an API [106].

The desktop uploader allows quick and easy upload of research outputs, straight from the desktop. All files are uploaded into a private space on figshare, where users can choose whether to make them public or manage them privately.

The figshare API allows to push data to figshare, or pull data out. This is an updated version that allows to manage your figshare data, create collections out of public content or build applications on top of the functionality.

Squidle

Squidle provides a roadmap for online image and video annotation software. Squidle is among new and improved online and in-field platforms for exploration, management and annotation of georeferenced images & video [105]. The core features of the base platform are:

- Flexible data storage.
- Flexible annotation schemes.
- Collaborative/automated labeling.
- In-field data annotation.
- Education & outreach.
- "Media object" annotation.

Appendix B – Software

Agisoft Metashape

Agisoft Metashape is a stand-alone software product that performs photogrammetric processing of digital images and generates 3D spatial data to be used in GIS applications, cultural heritage documentation, and visual effects production as well as for indirect measurements of objects of various scales [74].

Agisoft Metashape allows core photogrammetry processing workflows (e.g. photogrammetric triangulation, dense point cloud generation and editing, 3D model generation and texturing, georeferenced DSM / DTM generation, true / DTM-based orthomosaic generation, ground control and check points support, georeferencing using flight log and / or GCPs) and source data from various sensors (e,g. frame / fisheye camera support, video data import).

CloudCompare

CloudCompare is a 3D point cloud (and triangular mesh) processing software. It has been originally designed to perform comparison between two dense 3D points clouds (such as the ones acquired with a laser scanner) or between a point cloud and a triangular mesh. It relies on a specific octree structure dedicated to this task. Afterwards, it has been extended to a more generic point cloud processing software, including many advanced algorithms (registration, resampling, color/normal/scalar fields handling, statistics computation, sensor management, interactive or automatic segmentation, display enhancement, etc.) [79].

CANUPO is a CloudCompare plugin [116]. The CANUPO suite is a simple yet efficient way to automatically classify a point cloud. It lets you create your own classifiers (by training them on small samples) and/or apply one classifier at a time on a point cloud so as to separate it into two subsets. It also outputs a classification confidence value for each point so that you can quickly identify the problematic cases (generally on the classes borders).

Google Colab

Colaboratory, or "Colab" for short, allows to write and execute Python in a browser, with zero configuration required, free access to GPUs and easy sharing. Data science and machine learning

algorithms can be run on Google Colab platform [118].

Image Enhancement Processing Tool

This software implements five state of the art algorithms aimed to enhance the quality of underwater images. It is capable to automatically process a set of images within a directory, with all algorithms or just a subset of them. The Image Enhancement Tool comes with a simple GUI and it can be executed on Windows 64 bit operating systems (Windows 10 recommended) [66].

The software is downloadable under request at https://imareculture.eu/downloads/project-tools/image-enhancement-process-tool/.

VIAME

In cooperation with the National Oceanic and Atmospheric Administration's (NOAA) Automated Image Analysis Strategic Initiative (AIASI), Kitware and its partners have developed Video and Image Analytics for Marine Environments (VIAME). VIAME is an open source computer vision software platform designed for do-it-yourself artificial intelligence (AI). It is an evolving toolkit that contains many workflows used to generate different object detectors, full-frame classifiers, image mosaics, rapid model generation, image and video search, and methods for stereo measurement.

Originally targeting marine species analytics, it now contains many common algorithms and libraries, and is also useful as a generic computer vision library outside of underwater image and video. VIAME is available as a desktop or a web application.

VIAME is a computer vision application designed for do-it-yourself artificial intelligence including object detection, object tracking, image/video annotation, image/video search, image mosaicing, stereo measurement, rapid model generation, and tools for the evaluation of different algorithms. Originally targetting marine species analytics, it now contains many common algorithms and libraries, and is also useful as a generic computer vision library. The core infrastructure connecting different system components is currently the KWIVER library, which can connect C/C++, python, and matlab nodes together in a graph-like pipeline architecture. Alongside the pipelined image processing system are a number of standalone tools for accomplishing the above. Both a desktop and web version exists for deployments in different types of environments.

The software is downloadable at https://www.viametoolkit.org/.

Appendix C – Devices

Sony RX100V camera

Figure 58 - Sony RX100V camera [127].

Image Sensor		
Sensor Type		1.0-type (13.2mm x 8.8mm) Exmor RS CMOS sensor, aspect ratio 3:2
Number of Gross Pixels		Approx. 21.0 Megapixels
Number of Effective Pixels		Approx. 20.1 Megapixels
Lens		
Lens Type		ZEISS® Vario-Sonnar T* Lens, 10 elements in 9 groups (9 aspheric elements including AA lens)
F-number (maximum aperture)		F1.8 (W)-2.8 (T)
Iris Diaphragm		7 blades
Focal Length (f=)		f=8.8-25.7mm
Focal Length (f=) 35mm format equivalent		[Still image 3:2] f=24-70mm [Still image 16:9] f=26-76mm [Still image 1:1] f=30.5-89mm [Still image 4:3] f=25-73mm [Movie 16:9] f=25.5-74mm (SteadyShot Standard), f=30-86mm (SteadyShot Active), f=33.5-95mm (SteadyShot Intelligent Active), [Movie 4K 16:9] f=28-80mm (SteadyShot Standard), [HFR 960fps] f=38-110mm (Quality Priority), f=52-150mm (Shoot Time Priority), [HFR 480 fps] f=26-75mm (Quality Priority), f=37-105mm (Shoot Time Priority), [HFR 240 fps] f=26-75mm (Quality Priority), f=26-75mm (Shoot Time Priority)
Focus Range (from the front of the lens)		AF (W: Approx. 5cm (0.17 ft.) to infinity, T: Approx. 30cm (0.99 ft.) to infinity)
Optical Zoom		2.9x
Clear Image Zoom	Still Image	Approx. 5.8x (including Optical Zoom)
	Movie	
Digital Zoom*1	Still Image	Approx. 11x (including Optical Zoom)
	Movie	
Screen		
Screen Type		7.5cm (3.0 type) (4:3) / 1,228,800 dots / Xtra Fine / TFT LCD
Brightness Control		Manual (5 steps) / Sunny Weather mode
Adjustable Angle		Up by approx. 180 degrees, down by approx. 45 degrees
Display Selector (Finder/LCD)		Auto / EVF / Monitor
MF Assist Magnification		5.3x, 10.7x

Figure 59 - Sony RX100 camera features [127].

OE14-522 - High Definition Pan and Tilt Zoom (PATZ) Camera (Kongsberg)

Figure 60 - OE14-522 - High Definition Pan and Tilt Zoom (PATZ) Camera (Kongsberg) [128].

Standard Features

Electrical

Horizontal Resolution	800 TV Lines per picture height
Picture Elements Video	1920 (H) x 1080 (V)
Light Sensitivity (limiting)	0.1 Lux (faceplate)
Scene Illumination (limiting)	2 Lux assuming 100% scene reflectance
Sensor Type	$^1/_3$" CMOS sensor with colour mosaic filter
Signal to Noise Ratio	>54dB weighted
Power Input	Constant Voltage 16V-24vdc - 1A max.
Video Output	HD Component Y/Pb/Pr, HD-SDI - coax or fibre optic
Video Output Options	Composite Video
Optical Output Option	CWDM (Coarse Wave Division Multiplex) Wavelength options
Standard Video Formats	1080i 50Hz or 60Hz Component, 720P 50Hz or 60Hz Component, PAL or NTSC Composite (Selected by IR or GUI)
White Balance	Switch between 3,200110K, 5,600K and Auto using the Remote Control
Back Light Compensation	Switch on and off using the Remote Control
Control	Analogue, RS485, RS232 or Pelco-D
Electro-Magnetic Compatibility	BS EN 61000-6.3 2007 Emission BS EN 61000-6-1 2007 Immunity

Optical

Standard Lens	5.1mm to 51mm, F1.8
Iris Control	Automatic, Manual using the GUI
Focus Control	Front Port to infinity (wide angle) 800mm to infinity (narrow angle)
Zoom	10x Optical
Angle of View in Water	5.3° to 50° Diagonal 4.7° to 45° Horizontal 3° to 29° Vertical
Pan & Tilt	±110° Pan & Tilt

Mechanical

Diameter	140 mm (Main Body)
Length	226 mm (excl. connector)
Dome diameter	170 mm
Weight	6.0 Kg in Air, 4.0 Kg in Water
Standard Housing	Titanium Alloy 6AL/4V ASTM B3 48
Connector	Burton 5506-2013 (Y, Pb, Pr), 5506-1508 (HD-SDI)
HD-SDI Output	Seacon Opticon Fibre DG O'Brien Coax Other connectors are available, including fibre or coax hybrid

Figure 61 - OE14-522 - High Definition Pan and Tilt Zoom (PATZ) Camera (Kongsberg) features [128].