REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

Object Recognition in Underwater Environments Using AI Computer Vision Techniques

Harshit Goyal[1*], Priyank Sirohi[2]

[1*]M.Tech CSE, SCRIET, Chaudhary Charan Singh University, Meerut, India. Harshitg503@Gmail.Com
[2]Assistant Professor, SCRIET, Chaudhary Charan Singh University, Meerut, India. Priyanksirohi01@Gmail.Com

1. ABSTRACT

Towards this end, we make use of the Semantic Segmentation of Underwater Imagery (SUIM) dataset, which consists of over 1.5k densely annotated photos from eight different item categories that have ground-truth examples. Vertebrate fish, invertebrate reefs, aquatic plants, robots, human divers (like me!), and even the seafloor are among the more than 2,000 categories. This dataset reminds us of organized synthesis by gathering data from multiple ocean expeditions and collaborative experiments with both humans (unmanned)and robots. The same authors' team published a more current work that included a thorough performance benchmarking using cutting-edge global representations that are easily downloadable as open-source code. When it comes to underwater Inspection, Maintenance, and Repair (IMR) duties, the assessed methods facilitate the use of Autonomous Underwater Vehicles (AUVs) for autonomous interventions. A selection of test objects was made that is indicative of the types of applications that use IMR and whose shapes are usually known in advance. As a result, in realistic settings, CAD models produce virtual representations of these things when noise is added, and resolution is decreased. We validated our approach through extensive testing on both simulated scans and real data obtained using an AUV combined with an in-house rapid laser scanning sensor. Additionally, testing was done underwater in areas where shifting terrain caused by an unstable bed may have altered the contour of items being followed. To show how it broadens the scope, the research goes deeper into evaluating the performance of cutting-edge semantic segmentation algorithms using recognized measures. Finally, we present a fully convolutional encoder-decoder model which is tailored for competitive performance and computational efficiency. The model achieved 88% accuracy which is very high as far as underwater image segmentation goes. This study shows how the model could be put to practical use in various tasks from visual serving, saliency prediction and complex scene understanding. Importantly, the ESRGAN utilization improves images quality that enriches the soil on which our model succeeds. It lays a strong foundation for forthcoming research in the field of underwater robot vision through formulation, modeling, and introduction to benchmark dataset.

2. INTRODUCTION

Over the past few years, a plethora of techniques for object detection and recognition have emerged in the literature. The increasing requirement for autonomous systems capable of interacting with chaotic, ill-organized, and poorly structured real-world scenarios has motivated this development effort. Over the past ten years, there has been a significant advancement in object identification for mobile robots. Robotic kitchen environments represent an application situation where a promising level of performance has been attained. Using color and depth photographic equipment, robots can recognize common objects like bowls, plates, and cups so they can locate and automatically pick them up. The primary goal of NVIDIA's recently founded artificial intelligence robotics research lab is to teach a robotic arm how to recognize various utensils and navigate an IKEA kitchen. Using stereo vision systems for object recognition and grasping, robots tried to precisely identify portions of the item from images and determine the appropriate grabbing areas. A plethora of other uses have been developed for mobile robot recognition in interior contexts. These include advanced driver-assisted systems, industrial and farming applications, and home support for the elderly or persons with disabilities. The two primary application scenarios—automotive autonomy and interior service robotics—are somewhat to blame for the surge in effort on object detection and recognition. In both circumstances, humans work alongside robots, whose activities may pose a threat to human safety. In this regard, there has been a push to make the recognition process more robust and quicker. Many complementary sensory modalities, such as color cameras, laser scanners, Light Detection and Ranging (LIDAR), depth sensors based on texture projection, and more, can be applied to improve the endurance of land robots. However, because of payload restrictions and environmental factors that are unfavorable to these kinds of sensors, in certain application settings, such underwater robots, the use of complementing sensors may become extremely limited or impractical. When it comes to sensing in general and object perception in particular, the atmosphere of water is among the most difficult. Owing to the rapid attenuation and scattering of light and other electromagnetic waves, optical sensing is limited to object detection and recognition at very close ranges, typically a few meters. Much longer sensing distances are possible with acoustic propagation, but the resulting coarse-resolution and noisy object representations make it impossible for autonomous object grabbing to precisely identify and locate objects. Compared to its above-water cousin, comparatively fewer underwater object identification applications were documented. These include the recognition of various geometric shapes, such as cubes and cylinders, pipeline tracking and identification based on acoustic imaging, cable identification, and pipeline inspections in seabed survey activities using acoustic imaging cameras. In order to be

able to grip and manipulate such items in the future, we are involved in investigating techniques that are appropriate for underwater object recognition in this work. Among the many long-term possible application scenarios are the following:
- The oil and gas industries are frequently responsible for the inspection, maintenance, and repair of offshore facilities.
- Finding and classifying marine species in order to learn more about their natural surroundings.
- The secure and safe examination of potentially harmful, polluting, and inaccessible aquatic resources, including the lookout for objects that have been created.
- Avoiding submerged collisions by employing systems to find and identify an alternative obstacle, as in the early examination of accident scenes.

3. RELATED WORK

Machine learning techniques have long been used for applications related to underwater recognition and detection. Traditional methods in this field heavily relied on manually designed features for the detection of underwater objects, which included characteristics like shape, color, and texture. In order to distinguish between different underwater coral scales, the authors used a combination of texture and color data in addition to Support Vector Machines (SVMs). Chuang et al. used texture features retrieved via the phase Fourier transform for fish detection, On the other hand, Kim and colleagues introduced a technique that utilizes color-based image segmentation and metatemplate object selection. In certain cases, algorithms even employed more complex features, like the Scale Invariant Feature Transform (SIFT) and the Histogram of Oriented Gradients (HOG). For an extensive span of time, these approaches were regarded as the most accomplished in the field of underwater object detection.

However, the applicability of these hand-crafted features had limitations. Firstly, their task-specific design limited their ability to generalize; characteristics designed for low-light scenes might not be appropriate for well-lit underwater scenes or situations where the objects to be recognized significantly alter. Second, as Villon et al. showed when they utilized HOG features with SVM for fish classification, performing less well than end-to-end deep learning frameworks due to the fragmented nature of feature extraction and classification. Proposing and validating useful handcrafted features would also require a high level of competence.

On the other hand, features from big datasets can be separately extracted using supervised deep learning algorithms. Deep learning is a specialized branch of machine learning that analyzes data using layered structures modeled after biological neural networks. It needs a large amount of training data in order to extract meaningful and discriminative features with the least amount of human assistance. Deep learning architectures easily extract characteristics from input data, in contrast to typical machine learning models that are task-specific and frequently require human tweaks. In a variety of computer vision applications, such as object identification, object tracking, image segmentation, and image classification, deep learning networks have demonstrated outstanding performance. Deep learning has been used extensively in underwater item detection. Choi employed convolutional neural networks (CNNs) to classify fish species, while Villon et al. used fast-RCNN framework, Faster-RCNN to improve fish detection speed, and a deep learning model to identify coral reef fishes. Yang and colleagues met the real-time detection criteria by using the YOLOv3 framework for underwater object detection. There are also issues with deep learning-based detection algorithms even with their advantages over conventional machine learning models. Noisy data and class imbalance can cause deep learning models to have trouble identifying small objects, which increases the number of false positives and false negatives. Therefore, more effort is needed to address these challenging issues in deep learning-based underwater object detection.

4. SEMANTIC SEGMENTATION

One of the most difficult computer vision tasks is semantic segmentation for underwater item detection, which entails accurately classifying and defining different objects and areas inside underwater imagery. Underwater robotics, environmental monitoring, marine research, and ocean exploration are just a few of the fields in which it finds extensive use. Understanding the intricate underwater environment is crucial.

In the context of underwater object detection, semantic segmentation attempts to partition an input underwater picture into distinct semantic regions, each of which is labeled with an item category. Semantic segmentation gives each pixel a meaningful label, allowing for a pixel-by-pixel examination of the underwater environment Object detection, on the other hand, focuses on locating and recognizing particular things inside an image. Deep learning methodologies are the state-of-the-art for accomplishing semantic segmentation for underwater object detection. These approaches are based on Convolutional Neural Networks (CNNs), which can automatically extract hierarchical characteristics from photos.

For this purpose, Fully Convolutional Networks (FCNs) are often used because They offer end-to-end learning and are specifically designed for dense pixel-wise predictions. The first step in the semantic segmentation process is gathering a sizable and varied dataset of underwater photos. Then, ground-truth labels corresponding to the several item categories—fish, corals, rocks, sand, and other marine animals or structures—are manually added to each image at the pixel level.

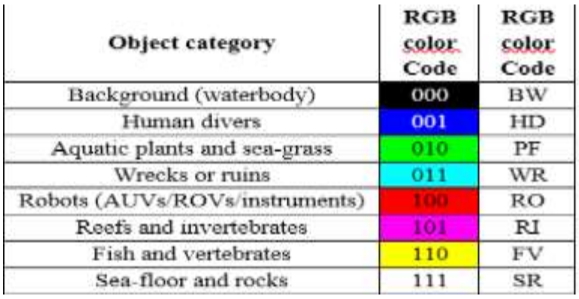
The annotated data is fed into the deep learning model during training so that it may learn to recognize pertinent characteristics that define each item category. With a focus on minimizing the pixel-wise classification loss, the model guarantees precise predictions for the semantic label of every pixel.

The trained model is used on fresh, unexplored underwater photos during the inference stage. After processing the input image, the model produces a pixel-by-pixel probability map, in which each pixel is linked to the probability that it belongs to a particular object category. To get the final segmentation mask, a thresholding step is frequently used, in which the label of the most likely item category is allocated to each pixel.

However, there are a number of difficulties with semantic segmentation due to the complexity of underwater photography. Absorption, scattering, and color attenuation can cause underwater photographs to deteriorate, resulting in decreased visibility and image quality. Furthermore, distinct underwater phenomena like noise and backscatter might make it more difficult to detect objects accurately.

5. THE SUIM DATASET

The SUIM dataset contains a comprehensive and diverse set of objects that are essential for the semantic analysis of underwater images. Background waterbody (B.W), human divers (H.D), aquatic plants/flora (P.F), wrecks/ruins (W.R), robots and instruments (R.O), reefs and other invertebrates (R.I), fish and other vertebrates (F.V), and seafloor and rocks (S.R) are the categories that are visually represented using a 3-bit binary RGB color coding scheme. Table 1 thoughtfully outlines this scheme.



Object category	RGB color Code	RGB color Code
Background (waterbody)	000	BW
Human divers	001	HD
Aquatic plants and sea-grass	010	PF
Wrecks or ruins	011	WR
Robots (AUVs/ROVs/instruments)	100	RO
Reefs and invertebrates	101	RI
Fish and vertebrates	110	FV
Sea-floor and rocks	111	SR

Table 1: Object Categories and Associated Color Codes in the Suim Dataset.

The SUIM dataset includes 1,525 RGB images in total for training and validation. Additionally, a freely supplied set of 110 test photos is included to help with the benchmark assessment of semantic segmentation models. The resolutions of these pictures vary widely; some of the measurements are 1906 × 1080, 1280 × 720, 640 × 480, and 256 × 256. These photos were carefully chosen from a large collection that was amassed during underwater scientific expeditions and cooperative experiments incorporating humans and robots in a variety of underwater settings.

Moreover, to introduce a wide range of the natural underwater scenes and experimental configurations that are suitable for human-robot cooperation, we judiciously incorporated a smaller subset of images sourced from established Extensive datasets, particularly EUVP, USR 248, and UFO 120, were drawn upon. These datasets contributed to the variety of object categories, their associations, and the subtleties in RGB channel that intensify values within the SUIM dataset and are vividly illustrated in the captivating visual representation featured in Figure 1.

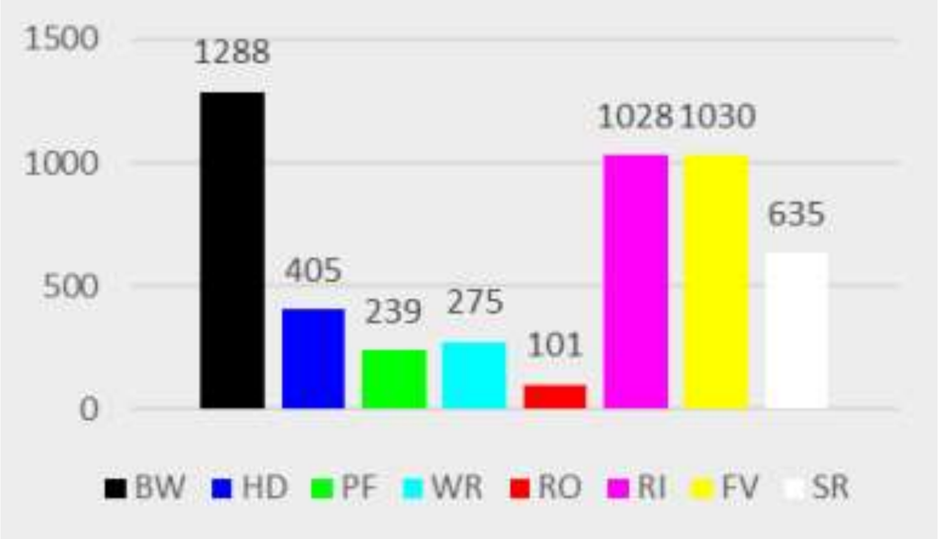


Figure 1: SUIM Dataset's Object Category Statistics.

The SUIM dataset stands as a testament to the meticulous work of seven human annotators who dedicated themselves to the intricate task of pixel-level annotations. Figure 2, showcasing these annotations alongside sample images, unequivocally showcases the dataset's exceptional quality and precision.

The paramount objective of this annotation endeavor was to establish consistent object classification throughout the dataset, particularly when faced with potentially confounding distinctions like those between plants/reefs and vertebrates/invertebrates. This stringent approach serves as a guarantee of the dataset's unwavering reliability and its broad applicability in the realms of computer vision and image analysis.

In the pursuit of this precision, we diligently adhered to the guidelines delineated in references. The standards were of great importance in guaranteeing the precision and consistency of object categorization in the dataset, hence enhancing its scholarly and practical significance.

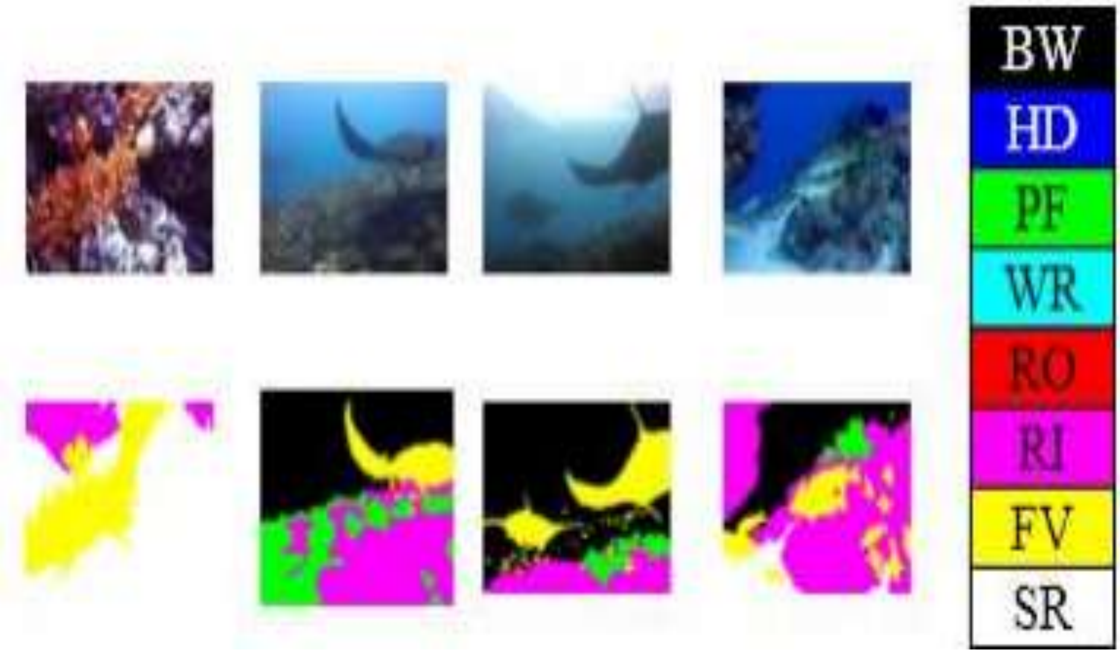


Figure 2: Sample Pictures from the SUIM Dataset with Associated Pixel Annotations.

6. PRE-PROCESSING

Color attenuation, scattering effects, and low contrast can occur in images taken in different or unequal lighting circumstances, which can result in a loss of information value. To address this problem and retain lost information, Schettini and Corchs presented an overview of earlier research on underwater picture augmentation. Of the different dimensions of degradation, classification performance is highly impacted by contrast loss. We have included several pre-processing sub-steps as follows to guarantee consistent image quality and to improve contrast:

A. Image Super-Resolution using ESRGAN.

To improve the resolution and quality of the low-resolution photographs, image super-resolution is an essential preprocessing step in underwater imaging. There are many obstacles in underwater photography that lead to low-resolution and low-quality photos. Allow me to present Enhanced Super-Resolution Generative Adversarial Networks, also known as ESRGAN, a cutting-edge deep learning technique tailored for super-resolution photos. Its operation is based on a Generative Adversarial Network, or GAN, which is made up of a discriminator network that distinguishes between high-resolution images that are generated and ground truth images. The generator network produces high-resolution images. Leveraging ESRGAN for underwater image super resolution begins with the collection of a substantial dataset featuring high-quality underwater images, which serves as the basis for model training. The ESRGAN model learns the intricate mapping from low-resolution to high-resolution underwater images using this dataset. It's significant that it considers the particularities of underwater photography, such as difficulties with light scattering and blur caused by absorption. By doing so, it produces visually pleasing and informative high-resolution images that are well-suited for underwater applications. After training, the ESRGAN model can be used to improve the resolution of fresh underwater photos, which greatly enhances a range of underwater applications, especially those that depend on object recognition and categorization. For a visual representation of the ESRGAN architecture, please refer to Figure 3.

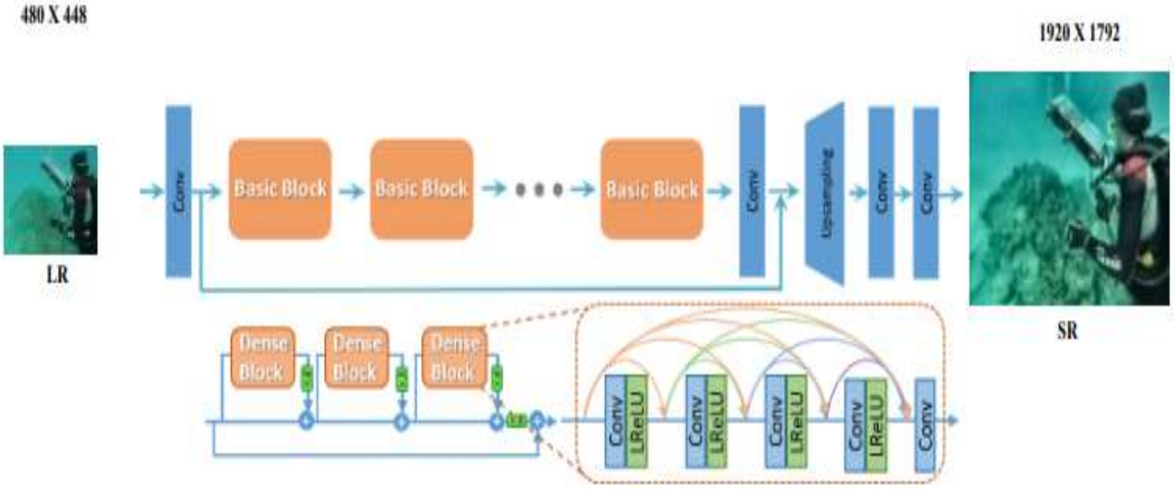


Figure 3: Better Super-Resolution Generative Adversarial Network Architecture (ERSGAN).

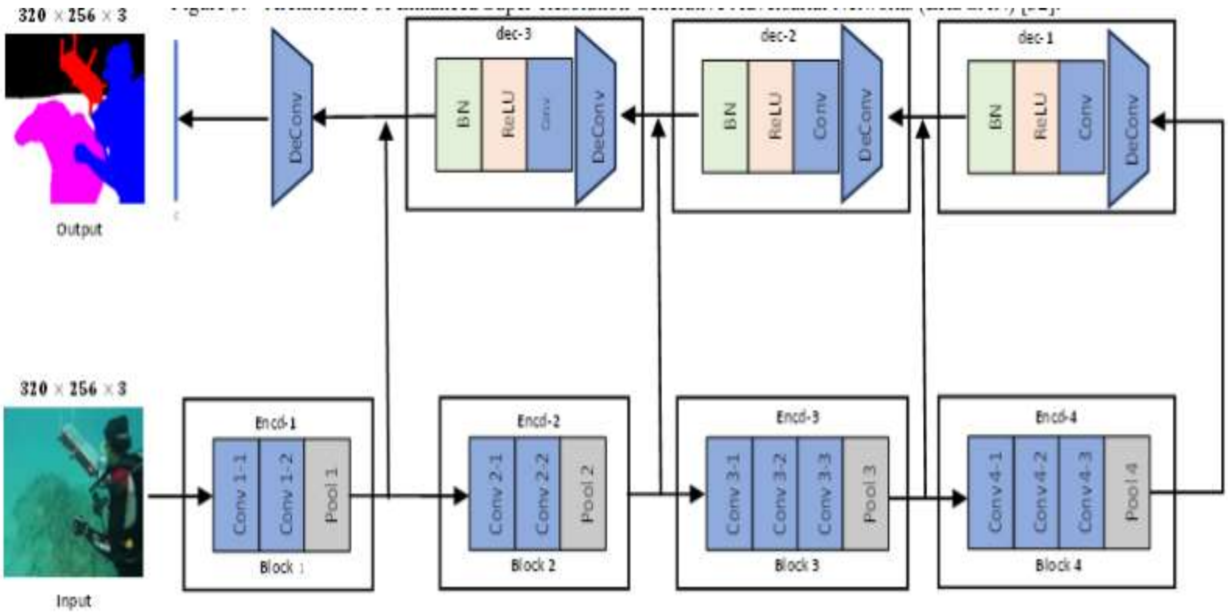


Figure 4: The following is the design of the suggested end-to-end model for semantic segmentation of underwater images. The first four blocks of a pre-trained VGG 16 model are used by the model to encrypt. Three mirrored decoder blocks and a deconvolution layer are then used to decode and produce the semantic segmentation map.

7. METHODOLOGY

A. Network Architecture

Our primary objective is to enhance the functionality of our model, which employs a neural network with twelve encoding layers that has already been trained. A graphic depiction of the architectural features is shown in Figure 4. Our work's primary goal is to apply this paradigm to produce better outcomes.

The approach we have outlined is depicted in Figure 5 and has been developed through a comprehensive review of relevant literature in addition to a detailed analysis of current methods and models. This extensive examination of the literature included a comparative study of various models related to picture segmentation, image contrast enhancement, and prominent object detection. To accomplish our study goals, the suggested methodology consists of several thoughtful steps:

7.1. First Preprocessing for Super-Resolution Underwater Images: The first phase of our methodology focuses on enhancing the resolution of underwater images, which often suffer from low quality and resolution. In this regard, we explored several super-resolution models, conducting a thorough evaluation to identify the most suitable approach for our specific needs. Our extensive evaluation led us to select the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) as the optimal solution for our super-resolution process.

7.2. Model Implementation with Convolutional Encoder-Decoder Architecture: In the next step, we proceeded with the implementation of our model. Our model design is based on a fully convolutional encoder-decoder architecture with skip relationships between mirrored aggregate layers. This architecture is integral to our approach as it plays a crucial role in the extraction and reconstruction of high-resolution information from low-resolution input.

7.3. Comparative Assessment of Proposed Solution: We performed a comparative analysis with other models that handle related problems in order to verify the effectiveness of our suggested methodology. This step allows us to quantitatively measure the performance and effectiveness of our approach to other solutions available in the field.

It is important to emphasize that the efficacy of our suggested methodology is based on a careful fusion of preprocessing steps, architectural design decisions, and the particular elements of our model. These components are carefully combined to make sure that our model performs well and produces results that are consistent with the main goal of our study. Our goal is to provide a strong and effective solution for underwater image enhancement and associated applications by giving careful consideration to these factors.

7.4. Microsoft COCO: Another large-scale dataset, Common Objects in Context, contains over 330 thousand annotated photos. It has fewer categories but more instances of the same classes than ImageNet. 2.5 million occurrences in those images have labels. In a situation where they might be located, the dataset attempts to provide some photos with partially occluded objects. Three other issues in scene understanding research are also addressed: non-canonical view detection (most datasets exhibit the items in clear, unobstructed view) and accurate 2D localization, where the labels are a more or less precise segmentation mask.

7.5. BENTHOZ-2015: This dataset is comprised of thousands of expertly annotated images of the seafloor off the coast of Australia. The images were collected as part of the Australian government's marine research program, the Integrated Marine Observing System (IMOS). Researchers studying benthic habitats and the organisms that inhabit them will find this dataset interesting Because it is georeferenced (every image has a GPS coordinate linked with it) and incorporates additional sensor data, such depth, height, temperature, and salinity, it may also be used to construct 3D maps and develop or test the Visual SLAM algorithm. The Internet 2 provides free access to the dataset, and it comes with Squidge, an annotation tool that is helpful for handling, examining, and annotating photos, videos, and large-scale mosaics.

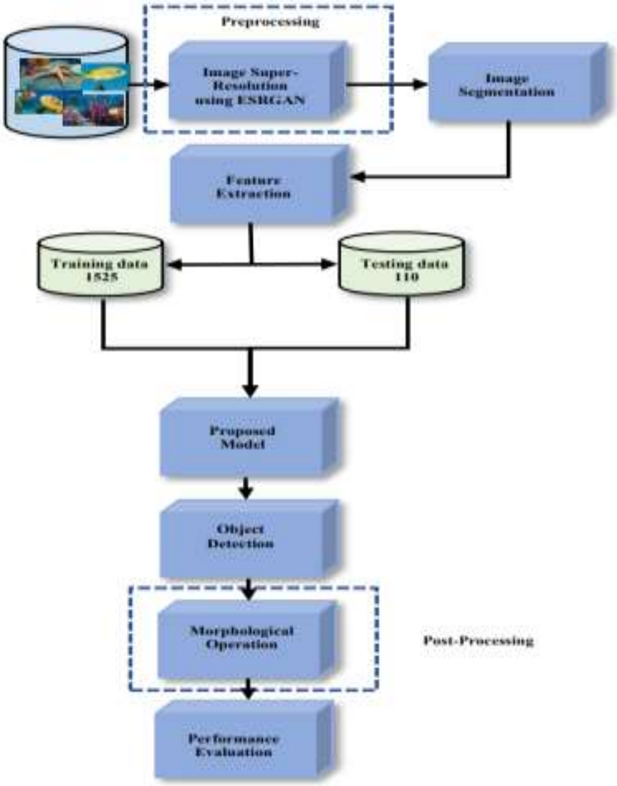


Figure 5: Block Diagram for Detecting Underwater Images and Objects.

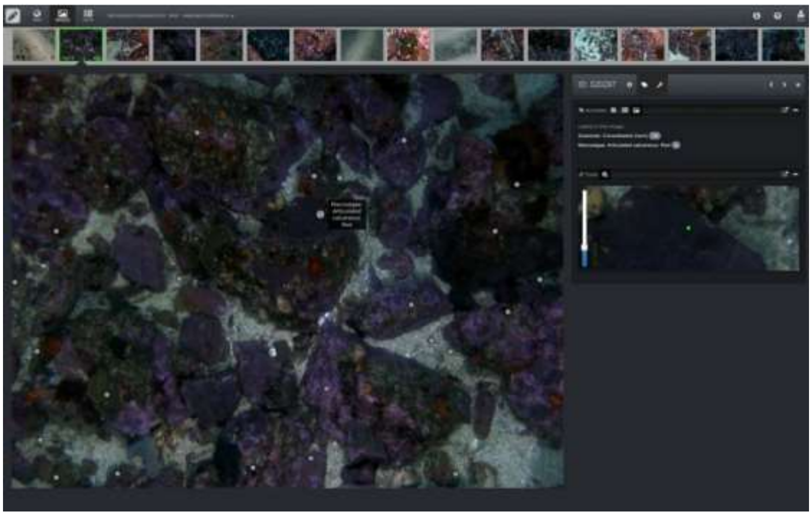


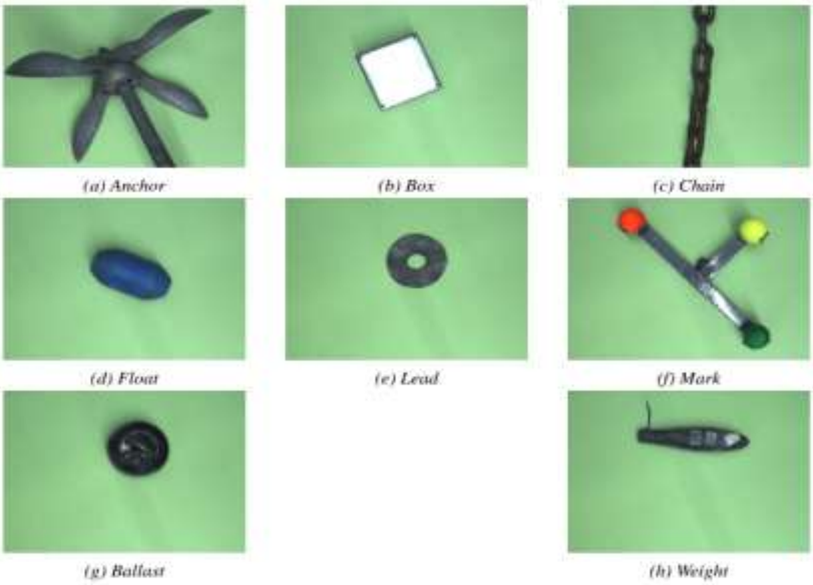
Figure 6: Annotated image sample from the BENTHOZ-2015 collection.

7.6. Marine Underwater Environment Database (MUED): A collection of 8600 photos of 430 distinct underwater objects, each having one or more items in a complex setting with variations in object posture, turbidity, and illumination, is provided by Jian et al. (2019). After evaluating salient-object recognition methods solely on this dataset, the authors concluded that a great deal of state-of-the-art algorithms do not adapt well to complex underwater environments. Unfortunately, because this dataset lacks images of items out of water, it cannot be utilized to compare how well two environments perform in an object classification job.

8. PROPOSED DATASET: HEIMDACA

This thesis aims to compare the performance differences between the same image classification method applied in the two contexts (under and above water) with respect to the same objects or classes. Instead of doing large-scale image classification, since a lot of work has already been done in this field. Since no dataset could be located for this use, it was necessary to take pictures of many items under two distinct circumstances:

The custom dataset HEIMDACA 3 contains images of eight items (classes) in two different settings: underwater (aquatic) and above the water's surface (aerial). The dataset consists of the following classes: Figueiredo et al. (2016) pioneered the artificial marker mark for UAV navigation and localization. Other items are weight, a round epoxy object with a ring on top, used as ballast in small water vehicles, float, a floater used in pool lane separators, and lead, a circular lead disk. Hybrid Environment Image Dataset for Applications in Classification.



(a) Anchor *(b) Box* *(c) Chain*

(d) Float *(e) Lead* *(f) Mark*

(g) Ballast *(h) Weight*

Figure 7: Instances of items gathered within the aerial domain.

This dataset was gathered in two locations: An underwater tank at the Department of Electrical Engineering and Computers at the University of Porto's Faculty of Engineering, and an aerial picture collection facility run by CRAS. All photos were taken using an L-shaped structure, with the camera-object distance preset for each set from a top-down viewpoint.

There were two sets of photos captured in each environment:
1. Aerial: Without backdrop and with backdrop (Fig.7), photos captured using a green cardboard piece that helps segmentation masks be created.
2. Aquatic: Photographs were taken in an area of the tank where the bottom was mostly intact, giving a clear background. Rough Background: images captured in the more texturally rich part of the tank.

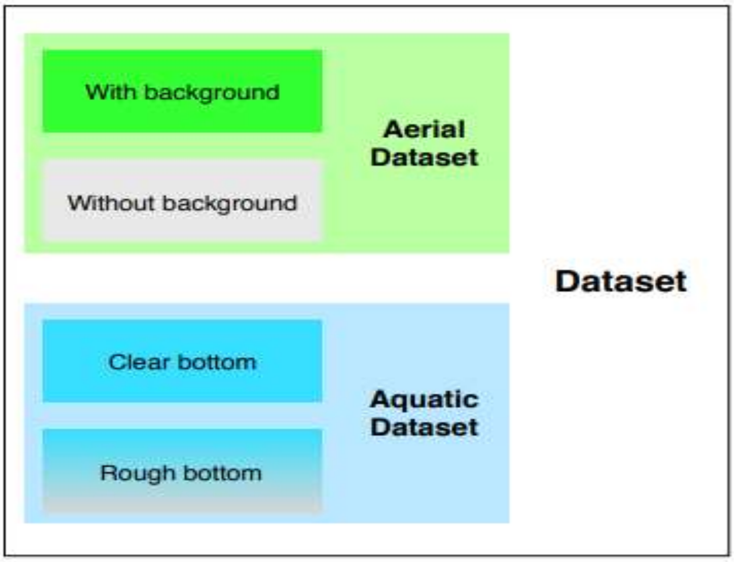


Figure 8: Dataset Division.

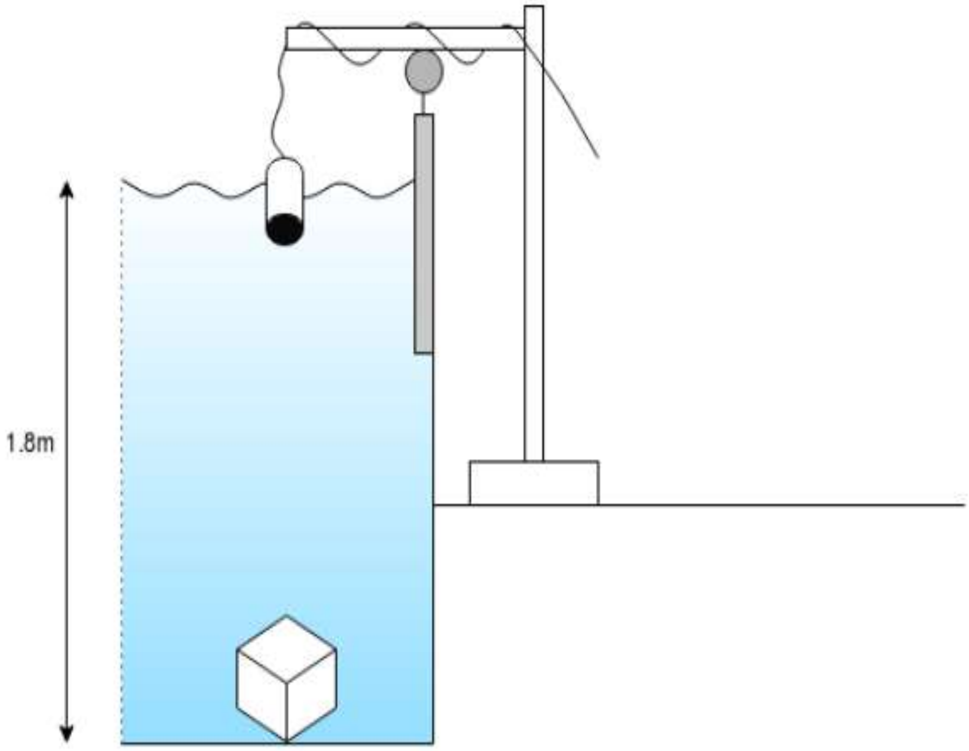


Figure 9: Underwater image capture setup.

With the exception of a plastic enclosure that made the camera waterproof so it could be used safely in an underwater setting, an Allied Vision 4 MAKO G-125C camera was utilized to capture both aerial and aquatic datasets. The flat-pane window in the enclosure may have caused distortion because of the water-glass interaction, however following examination of the collected frames, no appreciable deterioration in distortion was discovered.

Pictures were taken in RGB8Packed pixel format, with a resolution of 1292 by 964 pixels. The camera's built-in algorithm was used to automatically alter the exposure, gain, and white balance settings when it was used in the chosen setting and there were no objects in its field of vision. The necessary values were kept constant for each set of circumstances, after they stabilized, and Table provides a summary of them.

	Aerial		Aquatic	
	W/Background	W/o Background	Clear	Rough
Distance to object	0.5 m	1.0 m	0.9 m	1.8m
Exposure	26.5 ms	18.0 ms	28.7 ms	25.5 ms
Gain	13 dB	9 dB	0 dB	0 dB
White Balance	Red: 1.60	Red: 1.60	Red: 3.0	Red: 3.0
White Balance	Blue: 2.70	Blue: 2.70	Blue: 3.0	Blue: 3.0

Ten frames were chosen for each object in the aerial and marine photos after many images were taken. The selected photographs attempted to present the thing from several angles. Then, using Fiji, a program that enables speedy creation and testing of image processing techniques, black and white segmentation masks were created. Because of their consistent background, color segmentation was used to construct the masks for the aerial photos. The image was first segmented, and after that, morphological dilation was used to create a mask that included the whole item. Masks for the underwater photos were made by hand with the "Polygonal Select" tool.

These ten selected frames were used to create 450 images per object by random combinations of rotations, scale changes from 50% to 100% of the original image size, and horizontal and vertical mirroring. By using this method, it was possible to prevent overfitting and a lack of model generalization while also adding diversity to the training sets and speeding up the data collection process. To simulate variations in the camera viewpoint and assess how resilient the models are to such variation, perspective transformation was applied to images of aquatic subjects.

9. Conclusion

The material in this section included a specifically produced dataset for the thesis. Instead of classifying a huge number of items in two different settings on an eight-item set, its goal is to evaluate and quantify the influence of the undersea environment on the performance of picture classification algorithms. To add more variation to the data, data augmentation was applied to a portion of the photos.

Final Thoughts Computer vision has been used more and more for fish detection, monitoring, and management as image technologies and artificial intelligence techniques have advanced (Figure 3). Numerous studies covering a broad spectrum of applications can be found in the literature. Good results published in the literature can give the impression that the "mission accomplished" is achieved, but in reality, most studies have serious constraints that make it difficult to apply the suggested procedures in practice. These restrictions are typically related to the difficulty of gathering high-quality imaging data on fish, particularly when using an underwater setup. Although overcoming the obstacles covered in this review will take time and effort, doing so is necessary to make possible technology that can enhance the management and exploration of fish resources.

Knowing the issues that require suitable answers and the actual maturity of techniques and related technologies is essential for scientists, researchers, and entrepreneurs wanting to investigate the possible market to prevent mediocre products and services. Many technology-based businesses and startups have failed in other economic sectors where computer vision and artificial intelligence have been studied for longer because of the entry of immature technologies into the market. Beyond the financial damage brought about by those failures, subpar products frequently tarnish prospective buyers' impressions of a certain technology, which makes it harder for future competitors to succeed—even if their offerings are sound. Hopefully, the fish industry can steer clear of this predicament.

It is difficult to forecast the direction of research, especially given how swiftly and dynamically computer vision and artificial intelligence have developed in the recent past. On the other hand, it appears more likely that deep learning techniques and the application of data fusion principles will continue to gain popularity in combining the information generated by different data sources. The tighter the gap is between academic findings and practical needs, the more representative datasets that are collected and made accessible.

10. References

1. Álvarez Ellacuría, A.; Palmer, M.; Catalán, I.A.; Lisani, J.L. Image-based, unsupervised estimation of fish size from commercial landings using deep learning. ICES J. Mar. Sci. 2020, 77, 1330–1339. [CrossRef]
2. Banno, K.; Kaland, H.; Crescitelli, A.M.; Tuene, S.A.; Aas, G.H.; Gansel, L.C. A novel approach for wild fish monitoring at aquaculture sites: Wild fish presence analysis using computer vision. Aquac. Environ. Interact. 2022, 14, 97–112.[CrossRef]
3. Saleh, A.; Sheaves, M.; Rahimi Azghadi, M. Computer vision and deep learning for fish classification in underwater habitats: A survey. Fish Fish. 2022, 23, 977–999. [CrossRef]
4. Ditria, E.M.; Sievers, M.; Lopez-Marcano, S.; Jinks, E.L.; Connolly, R.M. Deep learning for automated analysis of fish abundance: The benefits of training across multiple habitats. Environ. Monit. Assess. 2020, 192, 698. [CrossRef] [PubMed]
5. Ditria, E.M.; Lopez-Marcano, S.; Sievers, M.; Jinks, E.L.; Brown, C.J.; Connolly, R.M. Automating the Analysis of Fish Abundance Using Object Detection: Optimizing Animal Ecology With Deep Learning. Front. Mar. Sci. 2020, 7. [CrossRef]
6. Shafait, F.; Mian, A.; Shortis, M.; Ghanem, B.; Culverhouse, P.F.; Edgington, D.; Cline, D.; Ravanbakhsh, M.; Seager, J.; Harvey, E.S. Fish identification from videos captured in uncontrolled underwater environments. ICES J. Mar. Sci. 2016, 73, 2737–2746. [CrossRef]
7. Noda, J.J.; Travieso, C.M.; Sánchez-Rodríguez, D. Automatic Taxonomic Classification of Fish Based on Their Acoustic Signals. Appl. Sci. 2016, 6, 443. [CrossRef]
8. Helminen, J.; Linnansaari, T. Object and behavior differentiation for improved automated counts of migrating river fish using imaging sonar data. Fish. Res. 2021, 237, 105883. [CrossRef]
9. Saberioon, M.; Gholizadeh, A.; Cisar, P.; Pautsina, A.; Urban, J. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. Rev. Aquac. 2017, 9, 369–387. [CrossRef]
10. Salman, A.; Siddiqui, S.A.; Shafait, F.; Mian, A.; Shortis, M.R.; Khurshid, K.; Ulges, A.; Schwanecke, U. Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. ICES J. Mar. Sci. 2020, 77, 1295–1307.
11. Kim D, Lee D, Myung H, Choi H-TJISR. Artificial landmark-based underwater localization for AUVs using weighted template matching. 2014;7:175-84. https://doi.org/10.1007/s11370-014- 0153-y.
12. Chuang M-C, Hwang J-N, Williams KJIToIP. A feature learning and object recognition framework for underwater fish images. 2016;25(4):1862-72. https://doi.org/10.48550/arXiv.1603.01696.
13. Alaba, S. Y., Nabi, M. M., Shah, C., Prior, J., Campbell, M. D., Wallace, F., ... & Moorhead, R. (2022). Class-aware fish species recognition using deep learning for an imbalanced dataset. Sensors, 22(21), 8268. https://doi.org/10.3390/s22218268.
14. Villon S, Chaumont M, Subsol G, Villéger S, Claverie T, Mouillot D, editors. Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between Deep Learning and HOG+ SVM methods. Advanced Concepts for Intelligent Vision Systems: 17th International Conference, ACIVS 2016, Lecce, Italy, October 24-27, 2016, Proceedings 17; 2016: Springer. https://doi.org/10.1007/978-3-319-48680-2_15.
15. A. K. Gupta, A. Seal, M. Prasad, and P. J. E. Khanna, "Salient object detection techniques in computer vision—A survey," vol. 22, no. 10, p. 1174, 2020.
16. N. Chen, W. Liu, R. Bai, and A. J. A. I. R. Chen, "Application of computational intelligence technologies in emergency management: a literature review," vol. 52, pp. 2131-2168, 2019.
17. H. Qin, X. Li, J. Liang, Y. Peng, and C. J. N. Zhang, "Deep Fish: Accurate underwater live fish recognition with a deep architecture," vol. 187, pp. 49-58, 2016.
18. H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, and A.-Y. J. N. Zang, "Faster R-CNN for marine organisms' detection and recognition using data augmentation," vol. 337, pp. 372-384, 2019.
19. LeCun Y, Bengio Y, Hinton GJn. Deep learning. 2015;521(7553):436-44. http://dx.doi.org/10.1038/nature14539.
20. Aruna, S. K., Deepa, N., & Devi, T. (2023, May). Underwater Fish Identification in Real-Time using Convolutional Neural Network. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 586-591). IEEE. https://doi.org/10.1109/ICICCS56967.2023.10142531.
21. Zhao, D., Yang, B., Dou, Y., & Guo, X. (2022, November). Underwater fish detection in sonar image based on an improved Faster RCNN. In 2022 9th International Forum on Electrical Engineering and Automation (IFEEA) (pp. 358-363). IEEE. https://doi.org/10.1109/IFEEA57288.2022.10038226.
22. Han, G., Huang, S., Ma, J., He, Y., & Chang, S. F. (2022, June). Meta faster r-cnn: Towards accurate few-shot object detection with attentive feature alignment. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 1, pp. 780-789). https://doi.org/10.1609/aaai.v36i1.19959.
23. Yang, H., Liu, P., Hu, Y., & Fu, J. (2021). Research on underwater object recognition based on YOLOv3. Microsystem Technologies, 27, 1837-1844. https://doi.org/10.1007/s00542-019-04694-8.
24. Shen, L., Tao, H., Ni, Y., Wang, Y., & Stojanovic, V. (2023). Improved YOLOv3 model with feature map cropping for multiscale road object detection. Measurement Science and Technology, 34(4), 045406. http://dx.doi.org/10.1088/1361-6501/acb075.

25. Bosse, S., & Kasundra, P. (2022). Robust Underwater Image Classification Using Image Segmentation, CNN, and Dynamic ROI Approximation. Engineering Proceedings, 27(1), 82. https://doi.org/10.3390/ecsa-9-13218.
26. Chen, Z., Wang, Y., Tian, W., Liu, J., Zhou, Y., & Shen, J. (2022). Underwater sonar image segmentation combining pixel-level and region-level information. Computers and Electrical Engineering, 100, 107853. https://doi.org/10.1016/j.compeleceng.2022.107853.
27. Wang, J., He, X., Shao, F., Lu, G., Hu, R., & Jiang, Q. (2022). Semantic segmentation method of underwater images based on encoder-decoder architecture. Plos one, 17(8), e0272666. https://doi.org/10.1371/journal.pone.0272666.
28. Liu Z, Tong L, Chen L, Zhou F, Jiang Z, Zhang Q, et al. Canet: Context aware network for brain glioma segmentation. 2021;40(7):1763-77. https://doi.org/10.1109/tmi.2021.3065918.
29. Alavianmehr, M. A., Helfroush, M. S., Danyali, H., & Tashk, A. (2023). Butterfly network: a convolutional neural network with a new architecture for multi-scale semantic segmentation of pedestrians. Journal of real-time image processing.
30. Dakhil, R. A., & Khayeat, A. R. H. (2022). Review On Deep Learning Technique For Underwater Object Detection. arXiv preprint arXiv:2209.10151. https://doi.org/10.48550/arXiv.2209.10151.