International Journal of Computer Applications (0975 – 8887)
Volume 186 – No.22, May 2024

Underwater Marine Animal Classification using a Convolutional Neural Network

Christen Loyola, Kishore P, Jerin Thomas
Student
Kumaraguru College of Technology,
Chinnavedampatti, Coimbatore

Chandrakala D
Professor
Kumaraguru College of Technology,
Chinnavedampatti, Coimbatore

ABSTRACT
Many diverse living forms found in aquatic habitats are critical to ecological balance and scientific advancement. The classification and identification of aquatic life, including fish, marine animals, and other species, is critical to understanding and protecting these ecosystems. This project uses machine learning to classify aquatic life automatically. The system collects underwater data, preprocesses it to improve quality, and then employs models trained on an annotated dataset of aquatic species. These models can identify and classify underwater animals with high accuracy, and they support both real-time and batch processing. A novel application of Convolutional Neural Networks (CNNs), a type of deep learning, is used to automate the classification of aquatic life from images and videos in this study. It is impossible to overestimate the importance of aquatic ecosystems in maintaining biodiversity and ecological balance. The various species that live in these environments are vital to the food chain, nutrient cycling, and ecosystem stability. As human activities such as pollution, overfishing, and habitat destruction continue to have an impact on aquatic ecosystems, there is a growing need to understand, monitor, and protect these environments. Traditionally, aquatic species classification has been labor-intensive, relying on manual observation and taxonomic expertise. However, the scalability of this approach is limited, particularly when dealing with vast and diverse aquatic ecosystems. This study addresses this issue by using machine learning, specifically Convolutional Neural Networks, to automate the species classification process.

General Terms
Image classification

Keywords
Convolutional Neural Networks, Machine Learning, Aquatic life, Classification

1. INTRODUCTION
Convolutional Neural Networks (CNNs) are regarded as a towering paradigm in the field of deep learning, respected for their unrivaled performance in image categorization tasks. CNNs emerge as crucial friends when it comes to identifying aquatic life because of their exceptional ability to autonomously learn and extract complex information from photos, eliminating the need for human intervention. However, the unique attribute of CNNs goes beyond mere autonomy; it is their ability to deal with large datasets that truly separates them. This characteristic is especially important in the world of aquatic life, where the diversity of aquatic creatures manifests itself in a plethora of shapes, sizes, and colors, making manual classification prohibitively labor-intensive. There are several advantages to using CNN to classify aquatic life. CNNs are particularly useful for image classification tasks because they can learn and extract features from images without the need for human intervention. Furthermore, CNNs can handle large amounts of data and can learn complex patterns in data that humans may not notice. Another advantage of using CNN for aquatic life classification is that it can be trained to recognize subtle differences between species that humans may find difficult to distinguish. Finally, CNNs can be used to improve the accuracy and efficiency of underwater species identification. This could have serious consequences for conservation efforts and environmental monitoring.

What truly distinguishes CNNs is their astonishing ability to detect the smallest differences between species, a feat that may elude human observers. CNNs appear as a beacon of accuracy at the depths of the aquatic environment, where small differences in morphology or coloration can be the only identifiable features. No nuance is lost in the hands of these neural networks, and the combination of their keen eye and massive data handling skills accelerates the science of aquatic life classification to new heights. Nonetheless, the data is the foundation of CNN's success in this domain. Creating a rich and well-curated dataset of aquatic life is the foundation upon which CNN's performance is built. The model's effectiveness is intrinsically tied to the quality and quantity of data; clean, abundant data yields superior categorization results. This data gathering phase is critical since it is the sustenance that feeds the neural network's learning process.

The advantages of using CNNs for aquatic life classification are numerous. As previously stated, their ability to extract critical features from photos relies on their intrinsic ability to alleviate the strain of human feature extraction. Furthermore, CNNs' data-handling capabilities are nothing short of astounding; they can sift through massive amounts of data and detect intricate patterns that humans may miss. It's not just about identifying species; it's about deciphering the intricate relationships and nuances that exist between them. The importance of CNNs in this arena is even greater. These neural networks can be painstakingly taught to detect details in species distinction that human specialists may find difficult to detect. Their capacity to detect even subtle differences between species can be a great asset in study, conservation, and environmental monitoring.

The implications of using CNNs to classify aquatic species go beyond scientific curiosity. Their presence has the potential to transform underwater species identification, improving both accuracy and efficiency. This is a game changer for environmentalists and environmental experts. CNNs can accelerate more successful conservation policies and enable real-time monitoring of aquatic ecosystems by permitting faster and more accurate species identification, increasing our ability to safeguard and preserve the delicate balance. Furthermore, CNNs have the potential to not only transform aquatic life classification but also to advance our understanding of marine ecosystems. Their ability to effectively evaluate large amounts of picture data opens the door to in-depth ecological investigations. This new capability will aid researchers in identifying trends, tracking population dynamics, and assessing the impact of environmental changes on aquatic species. CNNs are an essential tool for scientific research, whether it is tracking the migration patterns of marine mammals, investigating the distribution of coral reef invertebrates, or assessing the health of aquatic ecosystems.

2. OBJECTIVE OF THE PROJECT

The main purpose of the project is to use Convolutional Neural Networks (CNNs) to create a reliable and effective system for categorizing marine species underwater is a major project with wide-ranging consequences. Fundamentally, this goal involves applying cutting-edge deep learning methods to long-standing problems in environmental monitoring, conservation, and marine biology. Our goal is to overcome the drawbacks of conventional manual categorization techniques, which are frequently labor-intensive, time-consuming, and sensitive to subjective biases, by utilizing CNNs.

3. Literature Survey

Techniques for classifying and detecting underwater images have advanced because of several investigations. CNN models have been widely used for effective image and video classification since Liu et al. [1] presented a unique technique that improved feature extraction from low-quality underwater photos by integrating the convolutional block attention module (CBAM) into the YOLOv5 backbone network. Similarly, Han et al. [2] highlighted the usefulness of CNNs for efficient classification tasks among different algorithms by proposing a new underwater CNN recognition technology targeted at optimising detection programmes for underwater remotely operated vehicles (ROVs) to detect and classify marine organisms. Additionally, because of their lightweight parameter sizes and strong feature extraction capabilities, Alaba et al. [3] used the MobileNetv3-large and VGG16 backbone networks as feature extractors.

Advances in accuracy and efficiency in object detection have been suggested by researchers in recent works that address related difficulties. In the paper by Tsung-Yi Lin et al. [4] establishes Focal Loss, reducing the class imbalance during dense detector training and paving the way for the development of RetinaNet. Similarly, Sun et al. [9] address the shortcomings of traditional object detection methods in "Sparse R-CNN," introducing a novel approach based on a fixed, sparse collection of learned object proposals. Furthermore, in "Cascade R-CNN: Delving into High-Quality Object Detection," Zhaowei Cai et al. [5] present Cascade R-CNN, a multi-stage object identification architecture that gradually gets more discriminating against false positives. On the COCO dataset, Cascade R-CNN performs better than single-model object detectors, demonstrating its usefulness. The paper by Zhuang et al. [6] explains utilising cutting-edge deep learning algorithms, the research seeks to improve the precision and effectiveness of marine animal identification systems. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two complex deep learning architectures. Like semantic segmentation, Xiu Li et al. [8] describe their paper which functions in a per-pixel prediction manner and achieves remarkable average precision without compromising simplicity. It does this by doing away with the requirement for pre-defined anchor boxes. Ahmad Salman et al. [7] uses deep learning approaches to investigate fish species classification in unrestricted underwater habitats. The project attempts to address the difficulties caused by the complexity and variety of underwater landscapes for precise fish species identification by utilising deep learning techniques.

To improve performance, Hai Huang et al. [10] looks into the use of Faster R-CNN for the detection and identification of marine organisms. The paper addresses the difficulties in identifying and classifying marine species from underwater imagery by expanding on the Faster R-CNN framework, a well-liked object identification model. The study intends to enhance model robustness and accuracy in recognising a variety of marine species by implementing data augmentation techniques, advancing automated monitoring and conservation activities in aquatic environments. Rich feature hierarchies are introduced by Ross Girshick et al. [11] for precise object detection and semantic segmentation. The paper presents a deep learning framework that greatly increases the accuracy of object detection and semantic segmentation tasks by using convolutional neural networks (CNNs) to extract hierarchical features from input photos.

Z. Zhao et al. [12] addresses the difficulties of automatic fish detection and identification in underwater films by putting forth the innovative Composited FishNet framework. Anderson et al. [13] highlights the importance of fish species identification in the Pantanal region for the maintenance of the ecology, human habitation, and tourism. Using convolutional neural networks (CNNs), the paper presents a novel approach that aims to improve recognition accuracy, particularly for species that share comparable traits. It suggests a three-branch CNN design for fish species, families, and orders. The goal of this work from N. Jayachandra et al. [14] is to create a machine vision system that can recognise different kinds of marine life in films taken underwater. An ideal training set is produced for every marine animal by assembling image sets from frames in underwater movies with varying resolutions. To identify moving sea animals, an automated frame fragmentation and picture set creation technique is suggested. Videos from a variety of underwater cameras are used, such as those taken by scuba divers' high-resolution cameras and remotely operated vehicles (ROVs).

Abdelouahid Ben Tamou et al. [15] presented a deep learning-based approach for underwater live fish recognition. They used transfer learning and the convolutional neural network AlexNet to extract information from foreground fish photos found in underwater datasets. Their method, which combines a linear SVM classifier for classification with pretrained AlexNet networks—either with or without fine-tuning—produced encouraging outcomes. Moreover, the trial carried out on the Fish Recognition Ground-Truth dataset demonstrated the efficacy of their suggested approach. This study adds to the increasing corpus of research investigating deep learning approaches to underwater video processing, opening new possibilities for more effective and impartial fish population research.

The hierarchical book recommendation system that displays books according to readers' preferences in different genres, search history, and context, requires additional data about the context of book consumption, which is not easy to collect [14]. By using various similarity metrics such as Pearson correlation, cosine similarity, Kendall's Tau correlation, Jaccard similarity, Spearman Rank Correlation, Mean-squared distance, etc., a neighborhood-based approach was proposed and concluded that the Spearman Correlation Coefficient method works best, with a mean absolute error of less than 1.2. A recommendation method based on opinion mining that proposes top-ranked books on different disciplines of computer science by collecting users' requirements and ratings, categorized the features for the books, assigned weights to the categorized features, and accordingly provided ranks is limited to only computer science books and recommends only ten books for a particular query [18]. Text based Classification techniques are also suggested to integrate with the recommendation systemas such filtering methods are executed in many social media nowadays to filter the offensive words [19].

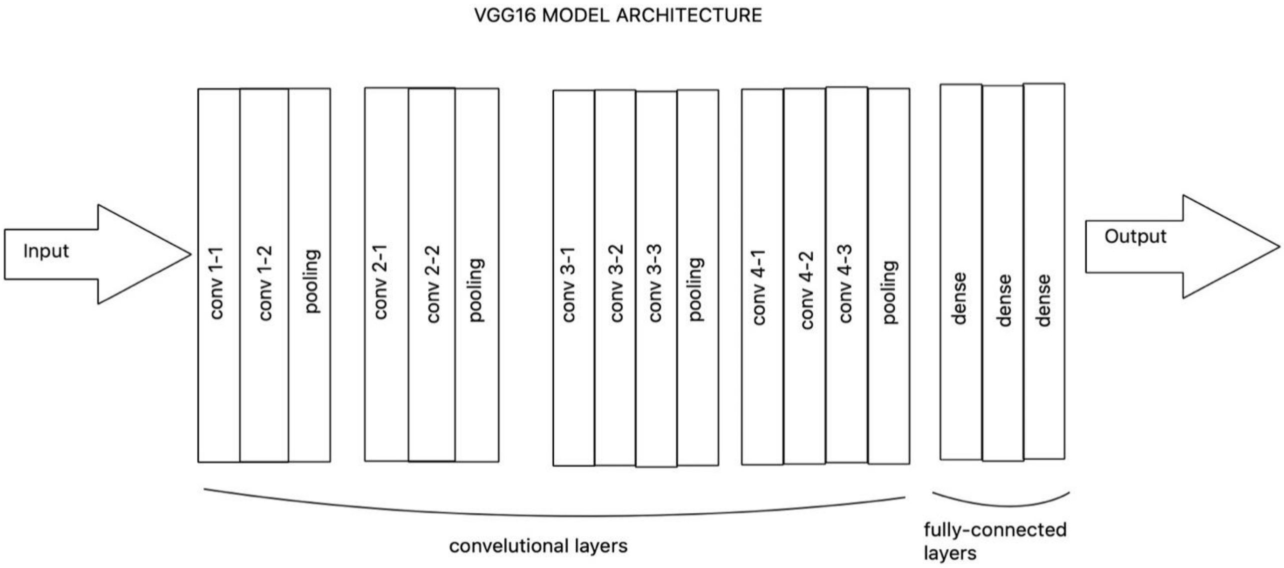

Figure 1. VGG16 Prebuilt Model

The efficiency and performance of a recommendation system is dependenton the performance of the algorithm used to generate therecommendations. To increase the speed and accuracy of the recommendation system, hybrid recommendation systems are used, which combine multiple filtering techniques to generate more efficient recommendations. This paper will explore the effectiveness of the hybrid filtering method in comparison to other filtering techniques for book recommendation systems.

4. METHODOLOGY

4.1 Data Collection

The first steps in creating a solid model for classifying aquatic species are gathering and preparing data. First, compiling a varied collection of underwater photos entails obtaining pictures featuring a variety of aquatic animals, including fish and dolphins. To guarantee that the model can be applied to a wide range of settings, these photos should show diversity in terms of species, positions, backdrops, lighting, and underwater environments. To train a model that can correctly classify aquatic species in real-world scenarios, it is imperative to capture this diversity. The next step is to label the photos with the appropriate class labels after the dataset has been gathered. For the model to be efficiently trained to recognise and classify distinct classes, each aquatic species must have accurate labelling.

The dataset is usually partitioned into training, validation, and testing sets after annotation. This partitioning guarantee accurate assessment of the model's performance in both the training and validation phases and offers a dispassionate appraisal of its generalisation skills. The validation set is used to check performance and adjust hyperparameters while the training set is used to train the model. The testing set, which is kept apart from the validation and training sets, acts as a stand-alone standard to assess how well the model performs on unobserved data. This method of dividing the dataset allows for a thorough evaluation of the model's generalisation to new underwater image instances, which aids in the creation of a trustworthy and accurate system for classifying aquatic species.

4.2 Model Architecture

Choosing the right model architecture is essential to developing an effective CNN-based underwater aquatic life picture categorization system. Selecting an appropriate architecture that can efficiently learn and extract significant characteristics from the input photos is necessary to correctly categorise various aquatic life classifications.

CNN designs that are in demand include Inception, VGG, ResNet, and their variations. These architectures are well-known and have proven effective in a variety of image classification applications. One well-known example of a uniform structure and simplicity is VGG, which consists of several convolutional layers followed by max-pooling layers. Using parallel convolutional layers with varying filter sizes, initiation architectures enable multi-scale feature extraction, improving the model's capacity to identify a variety of patterns and structures in images.

In this project VGG16 was the model used to train and test the dataset. The VGG16 architecture is a convolutional neural network (CNN) architecture that is commonly employed for image categorization applications.

Karen Simonyan and Andrew Zisserman of Oxford University's Visual Geometry Group Lab proposed it in 2014 VGG16's architecture has 13 convolutional layers, 5 max-pooling layers, and 3 fully linked layers

The network receives an image with dimensions (224, 224, 3). The first and second layers each have 64 channels with 3x3 filter sizes and the same padding. Following a max pool layer of stride (2, 2), two layers with convolution layers of 128 filter size and filter size (3, 3) are added. This is followed by a stride (2, 2) max-pooling layer that is the same as the preceding layer. There are then two convolution layers with filter sizes of (3, 3) and 256 filters. After that, there are two sets of three convolution layers and a max pool layer. Each has 512 filters of the same size (3, 3) with the same padding. This image is subsequently sent into the first of two convolution layer stacks.

A collection of neurons, each connected to every other neuron in the layer above it, make up each fully connected layer. The network can capture complex interactions between the target classes and the learned properties because of its dense connectedness. Complex patterns and feature combinations are gradually reduced as the data passes through these fully connected layers, leading to a representation that is ultimately discriminative enough to distinguish between various classes of objects—in this case, aquatic life forms like fish and dolphins.

A SoftMax layer comes after the fully linked layers in the VGG16 model. Based on the learnt features, the SoftMax layer assigns probabilities to each output class and computes the probability distribution over the classes. The output class is then anticipated to be the one with the highest likelihood. This last stage is essential for transforming the unprocessed output of the fully linked layers into insightful predictions, which helps classify input photos into different aquatic life categories.

4.3 Model Evaluation

A crucial first step in determining how well the trained model performs and how accurate it is in classifying photos of unseen aquatic life is model evaluation. In this section the method of evaluating the model for this project is elaborated. To make sure the model can generalise successfully to new and unexplored data instances, the assessment step entails thoroughly testing the model's capabilities on a separate testing dataset. To compute several assessment metrics to measure the model's classification performance across diverse aquatic life classes the measurements that is used are F1-score, recall, accuracy, and precision.

The percentage of correctly categorised photos in the testing dataset, relative to the total number of images, is called accuracy. It offers a broad indicator of how accurate the model is generally at categorising photos of aquatic life. By calculating these assessment metrics, learning about the model's advantages and disadvantages when it comes to categorising various aquatic life classes is done.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Fig 2. F-1 Score and formula.

4.4 Model Deployment and Integration

An important first step in turning research efforts into real-world applications for underwater picture classification is deploying the trained model into production contexts. A thorough description of the deployment process is given in this section, together with information on the user interface design, scalability and reliability considerations, integration with current systems, and performance assessment in real-world settings. To make interacting with the deployed model easier, an intuitive user interface was created. Through the interface, users can upload new underwater photos for classification and receive real-time species predictions for the aquatic life seen in the photos.

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For three authors, you may have to improvise.

5. RESULTS AND DISCUSSIONS

The project used a Streamlit application that allowed users to input image file locations for categorization to classify over thirty kinds of aquatic life. Following a thorough training and assessment process, the model attained an accuracy level of above 70% overall. This part includes a presentation and discussion of the outcomes as well as learnings from the categorization performance.

Table 1. Accuracy of Model

Validation Loss	0.8674425482749939
Validation Accuracy	0.7727272510528564

A promising level of performance in classifying aquatic life species is shown by the achieved accuracy of over 70%. By contrasting the model's predictions with ground truth labels from the testing dataset, this accuracy was determined. It shows that the model can differentiate between a variety of aquatic species based on input photos, even though it is not quite accurate. Variations in classification performance are found among the various aquatic life classes upon closer examination. Higher accuracy rates in some classes can mean that the model is capable of successfully differentiating between these species based on the visual characteristics seen in the photos. On the other hand, some classes might be difficult, which would result in poorer accuracy rates. Comprehending these disparities can yield significant understanding of the model's merits and demerits in categorising aquatic life kinds. During the assessment process, several significant findings surfaced. The model showed a surprising level of ability in differentiating visually various classes of aquatic life, including species of fish with characteristic anatomical traits or unique colour patterns. Furthermore, the model's robustness in practical applications was demonstrated by its consistent performance under various illumination settings and climatic circumstances.

Apart from assessing the performance of individual models, the study also investigated the overall accuracy and performance metrics of several models and algorithms used in the project. Through methodical testing and experimentation were done to determine which model would work best for the project's particular goals. Numerous models were subjected to computation and analysis of various metrics, including accuracy, precision, recall, F1 score. This thorough assessment made it easier to make well-informed decisions on optimisation techniques and model selection, which in turn helped choose the best model to meet the project's goals with the best possible performance and efficiency.

Table 2. Comparisons of Models and Algorithms

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.65	0.67	0.64	0.65
Support Vector Machine	0.45	0.47	0.44	0.45
VGG 16	0.78	0.75	0.80	0.77
Naïve Bayes	0.55	0.57	0.54	0.55

Even with the promising outcomes, there were several obstacles and restrictions with our project. One significant issue was the existence of classes with comparable visual traits, which confused and misclassified the model. Lower accuracy rates for certain classes may also have been caused by uncommon or underrepresented classes in the training sample. Strategies like data augmentation, class balancing methods, or fine-tuning model architectures to increase classification performance for difficult cases may be needed to address these issues.

6. CONCLUSION

Convolutional neural networks (CNNs), a type of machine learning technique, are used in this study to automatically classify and identify aquatic species, which is a significant advancement in autonomous classification and identification. Maintaining biodiversity and ecological balance depends heavily on aquatic habitats, hence it is imperative to comprehend and safeguard these settings. When working with large and diverse ecosystems, the human observation-based, labor-intensive classification of aquatic species has historically presented scaling issues. However, this study has shown that it is feasible to automate the process of species categorization with high accuracy by using machine learning models that have been trained on annotated datasets of aquatic species.

A noteworthy accomplishment is the creation of a Streamlit application that can accurately classify more than 30 types of aquatic life with an accuracy rate higher than 70%. This programme offers a scalable and effective way to identify and track aquatic species, which has the potential to transform ecological study, conservation efforts, and ecosystem monitoring. Furthermore, the automatic classification of photos and videos using CNNs creates new opportunities for the in-depth study and investigation of aquatic environments.

Although the results are encouraging, it is recognised that there are still issues to be resolved, such as dealing with misclassifications, strengthening the model's resilience in a variety of environmental circumstances, and improving user engagement and feedback systems.

7. FUTURE SCOPE

The suggested web application has great potential for the identification of aquatic species and environmental monitoring in the future since it uses Convolutional Neural Networks (CNNs) to automatically categorise fish data. Although the project's initial focus is on using CNNs to categorise fish species based on provided photos, there are many opportunities for development and improvement to increase the application's functionality and effect. Adding more data sources and modalities to enhance the classification process is one possible direction for future improvement.

For the time being, the web application mostly uses static graphics to identify species. To improve the classification process, additional data sources, such as video footage, underwater sonar data, or environmental sensor readings, might offer supplementary information. The programme has the potential to provide more thorough insights into aquatic ecosystems and enhance the accuracy of species identification in a variety of situations by merging several data modalities.

8. REFERENCES

[1] Lin Z, Zhuang Y, Jia P, Wu C, Xu H, Lin Z, A Novel Underwater Image Enhancement Algorithm and an Improved Underwater Biological Detection Pipeline, Journal of Marine Science and Engineering, 2022; 10(9):1204, http://doi.org/10.3390/jmse10091204

[2] Alaba, S.Y.; Nabi, M.M.; Shah, C.; Prior, J.; Campbell,M.D.; Wallace, F.; Ball, J.E.; Moorhead, R. Class-Aware Fish Species Recognition Using Deep Learning for an Imbalanced Dataset. Sensors 2022, 22, 8268. https://doi.org/10.3390/s22218268

[3] Alaba SY, Nabi MM, Shah C, Prior J, Campbell MD, Wallace F, Ball JE & Moorhead R. (2022) Class-Aware Fish Species Recognition Using Deep Learning for an Imbalanced Dataset.

[4] Tsung-Yi Lin; Priya Goyal; Ross Girshick; Kaiming He & Piotr Dollár (2018) Focal Loss for Dense Object Detection.

[5] Zhaowei Cai & Nuno Vasconcelos (2018) Cascade R-CNN: Delving into High Quality Object Detection.

[6] Zhuang, P., Xing, L., Liu, Y., Guo, S., & Qiao, Y. (2017). Marine Animal Detection and Recognition with Advanced Deep Learning Models. Conference and Labs of the Evaluation Forum.

[7] Ahmad Salman, Ahsan Jalal, Faisal Shafait, Ajmal Mian, Mark Shortis, James Seager & Euan Harveyx (2016). Fish species classification in unconstrained underwater environments based on deep learning.

[8] Xiu Li, Min Shang, H. Qin and Liansheng Chen (2015) Fast accurate fish detection and recognition of underwater images with Fast R-CNN

[9] Peize Sun; Rufeng Zhang; Yi Jiang; Tao Kong; Chenfeng Xu; Wei Zhan; Masayoshi Tomizuka & Zehuan Yuan (2023) Sparse R-CNN: An End-to-End Framework for Object Dete

[10] Hai Huang, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi, Ai-Yun Zang (2019) Faster R-CNN for marine organisms' detection and recognition using data augmentation.

[11] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation

[12] Z. Zhao, Y. Liu, X. Sun, J. Liu, X. Yang and C. Zhou (2023) Composited FishNet: Fish Detection and Species Recognition from Low-Quality Underwater Videos

[13] Anderson Aparecido dos Santos, Wesley Nunes Gonçalves (2019) Improving Pantanal fish species recognition through taxonomic ranks in convolutional neural networks.

[14] N.Jayachandra and Dr. A. R. Nadira Banu Kamal (2017) A Novel Framework for Automated image set preparation for moving objects in underwater videos.

[15] Abdelouahid Ben Tamou, Abdesslam Benzinou, Kamal Nasreddine, Lahoucine Ballihi (2018) Underwater Live Fish Recognition by Deep Learning