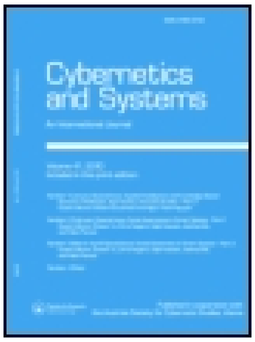

Cybernetics and Systems
An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucbs20

A Deep Learning Approach to Detecting Objects in Underwater Images

Kalaiarasi G, Ashok J, Saritha B & Manoj Prabu M

To cite this article: Kalaiarasi G, Ashok J, Saritha B & Manoj Prabu M (2023): A Deep Learning Approach to Detecting Objects in Underwater Images, Cybernetics and Systems, DOI: 10.1080/01969722.2023.2166246

To link to this article: https://doi.org/10.1080/01969722.2023.2166246

Published online: 23 Jan 2023.

CYBERNETICS AND SYSTEMS: AN INTERNATIONAL JOURNAL
https://doi.org/10.1080/01969722.2023.2166246

A Deep Learning Approach to Detecting Objects in Underwater Images

Kalaiarasi G[a], Ashok J[a], Saritha B[b], and Manoj Prabu M[c]

[a]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, Tamilnadu, India; [b]Department of Electronics and Communication Engineering, Jai Shriram Engineering College, Tirupur, Tamilnadu, India; [c]Department of Bio Medical Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamilnadu, India

ABSTRACT

A deep learning approach, also known as deep machine learning or deep structure learning, has recently been found to be successful in categorizing digital images and detecting objects within them. Consequently, it has rapidly gained attention and a reputation in computer vision research. Aquatic ecosystems, especially seagrass beds, are increasingly observed using digital photographs. Automatic detection and classification now requires deep neural network-based classifiers due to the increase in image data. The purpose of this paper is to present a systematic method for analyzing recent underwater pipeline imagery using deep learning. There is a logical organization of the analytical methods based on the recognized items, as well as an outline of the deep learning architectures employed. Deep neural network analysis of digital photographs of the seafloor has a lot of potential for automation, particularly in the discovery and monitoring of underwater pipeline images.

KEYWORDS

Deep learning; underwater pipeline image; object detection; aquatic ecosystem

1. Introduction

Underwater image recognition is incredibly useful and beneficial for a variety of uses, including pipeline maintenance, mining, marine life monitoring, and military applications. Light moves at a pace of 20 m in clear water and slightly slower in turbid or coastal water, Veiga et al. (2022). Underwater light visibility is rather limited, and the speed of entry into the water drops rapidly. Poor quality underwater photos make it harder to recognize items. Underwater photography is done at great depths and might be of poor quality, Mathias et al. (2021). Because underwater photography is frequently done in deep water, an autonomous underwater vehicle (AUV) is employed to inspect the photographs.

The artificial light in the AUV intensifies the brightness in the shot, but it generates a haze and sounds like noise as it moves across the water. To

CONTACT Kalaiarasi G kalaiibe@gmail.com Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, Tamilnadu, India.

increase recognition, underwater images must be preprocessed. The goal of image preprocessing is to increase image quality by increasing distortion and visual qualities. Much research has been conducted on this subject, but it appears to be challenging. Object identification is a computational visual concept for recognizing things in photos and movies, including identifying comparable targets in another image, Simonyan et al. (2014). The purpose of object recognition is to identify items in photographs and detect objects in the same way that people do. Figure 1 illustrates underwater image detection.

Images are perceived from several viewpoints, including front, side, and rear. When an object is partially hidden, viewers experience it in multiple shapes and sizes, Yan et al. (2013). It recognizes letters, faces, lanes, and voices, among other things. Water covers more than two-thirds of the world's land area, but few tools for studying marine life have been devised. Marine security, including shipwrecks and naval warfare, is a significant component of object detection in marine life surveillance, Marini et al. (2018). The detection of maritime items consists of two steps: feature extraction and classification. Four geometric features were produced artificially before recognizing surface vessels, Li et al. (2015). Explicitly learning sea life can aid in the resolution of maritime challenges such as disaster avoidance, target detection, emergency rescue, and tracking and detection, Jalal et al. (2014).

Deep cleaning techniques are utilized in marine systems for data reconstruction, categorization, and prediction. As part of a deep learning system, it has two learning layers, a conventional learning layer and a fully conventional learning layer, comprehending complex layers through conversation and promoting object characterization, Lee et al. (2016). Traditional vector machines outperform ordinary neural networks and CNNs in object detection and classification because of their hyperparameters, Yang et al. (2021). When Constrained Boltzmann Machines or RBM were integrated with

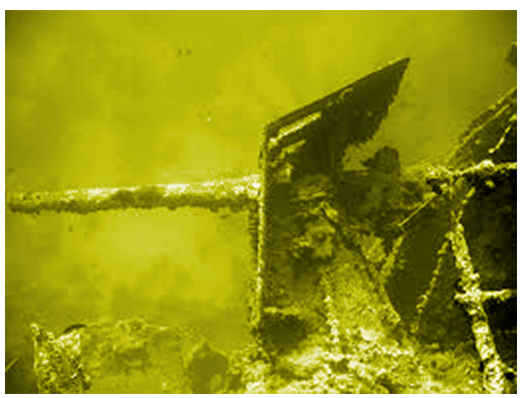

Figure 1. Underwater image detection.

deep learning approaches, learning frameworks grew tremendously. If they have access to strong GPUs with huge memory capacities in the computer, it can use the DBN method to supplement deep learning strategy.

When researching deep learning architectures, they merged CNN, RNN, and AIin the object detection process, yielding good results. A robust training method with no fake elements enables object detection. Obtain nonlinear information by utilizing CNN capabilities such as weight sharing, pooling, local connections, and multi-layers. AlexNet's deep learning success in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) is tied to the CNN technique based on VGGNet and ResNet. As a result, deep CNN findings can provide excellent accuracy as well as data regarding the deep learning process. DNNs have proven to be successful in the industrial and academic deep learning sectors, according to research. Specific article recognition in 3 D, as well as visual and object segmentation recognition, were all processed. Many deep learning and marine life architectures have been proposed in recent years, but their merits and weaknesses remain undefined. As a result, major difficulties in identification persist. This study's deep learning mechanisms of marine creatures are time-based and have both theoretical and practical value in the marine engineering field.

2. Underwater Object Detection

For many years, the UOD method has been employed in marine ecological study. Strachan et al. employed color and form descriptors to identify fish on a conveyor belt monitored by a digital camera. It exhibited a vision system for recognizing fish in real-time movies, which included object detection and tracking techniques. Ravanbakhsh and others To recognize reef fish, a Histogram of Oriented Gradients (HOG) & Support Vector Machine (SVM) method was applied. However, the foregoing approaches rely largely on artisan features, which restricts their ability to express them. With the rapid evolution of deep learning techniques, a special deep learning-based UOD technique was recently described. A fast R-CNN-based approach for identifying and identifying fish species from underwater images is given.In addition, underwater pictures of plankton items were classified using a deep residual network model. A lightweight deep neural network for fish detection based on concatenated ReLU, Inception, and HyperNet is also reported. In addition, by incorporating multi-scale features and extra contextual information, a single-shot feature aggregation deep network for UOD was presented. The YOLOv2 and YOLOv3 Deep models were refined and tested on a brackish water dataset that included annotated image sequences of fish, crabs, and starfish collected in diverse viewpoints.

Furthermore, UOD's dataset is quite limited, limiting the development of his deep learning-based UOD algorithm.

2.1. Deep Learning in Fish Detection and Classification

Prior to 2015, there had been few attempts to apply deep learning to fish identification. Ravanbakhsh et al. [13] used the haar classifier to classify shape features. Principal component analysis was used to model the characteristics (PCA). It employs a moving average method to balance the accuracy and processing speed of underwater fish recognition. Both systems have disadvantages when it comes to analyzing big numbers of underwater photos. Li et al. introduced deep convolutional networks for the first time. Fish identification and recognition To detect fish, it employed a Fast Region-based Convolutional Neural Network (Fast-RCNN). We also built a clean fish dataset of 24272 pictures divided into 12 classes, which is a subset of the ImageCLIEF training and testing datasets. They updated AlexNet to train Fast R-CNN parameters using stochastic gradient descent (SGD). Their experimental results demonstrated that larger maximum posterior estimates resulted in better performance (mAP). These are 9.4% more accurate on average than the deformable part model (DPM).

In the Fish Knowledge project, Villon et al. evaluated deep learning performance on his GroundTruth dataset. They also compared deep learning's performance for fish identification to that of a typical system that used SVM classification and HOG feature extraction. Their deep network architecture is modeled by GoogleNet, which has 9 initial layers and 27 layers with softmax classifiers.

2.2. Deep Learning in Plankton Classification

Plankton are commonly used as markers of ecological health since they form the foundation of aquatic food webs. For large-scale surveys, traditional plankton monitoring and measurement approaches are insufficient. In 2015, organized the National Data Science Bowl in partnership with Oregon State University's Hatfield Marine Science Center to classify photographs of plankton. The professor Joni Dambre from the University of Ghent, Belgium, led the winning team, which used convolutional neural networks. Deep learning algorithms are widely assumed to require massive data sets, yet in this case, the classification accuracy is 81.52% with approximately 30,000 examples in 121 classes.The victorious team's output feature map was similar to the input map, pooling and overlapping with a window size of 3 and an increment of 2. The ultimate structure has 16 layers after starting with a very flat 6-level model and gradually increasing the amount of layers.

To allow the network to assess the input from numerous perspectives while using the same feature extraction pipeline, a cyclic pooling method was used. The same stack of convolutional layers was used, which was then fed into a stack of dense layers and feature maps that were merged on top.Finally, stacks of cyclic pooling output feature maps from separate directions are combined into a single huge stack, and this combined input is used to train the next level, which includes four times the number of filters as previously. The technique of integrating feature maps from multiple directions is known as "rolling." It planned and built an initial module with convolutional layers to avoid distortion and optimize visual information extraction using the same dataset from the 2015 National Dataset Bowl. GoogleNet served as inspiration. Network architecture has been defined by increased utilization of computational resources within a network. To alter rotational and translational invariants, data augmentation was used, and rotational affinity was used to enhance the data.

The deep convolutional neural network was partitioned into two parts: classifiers and features. However, if the dataset is too tiny, this form of classifier subdesign overfits, thus it substitute the last two fully connected layers with small kernels. It turned out to be better than expected. The model outperformed state-of-the-art approaches at particular image sizes. Lee and colleagues created a deep network technique for categorizing plankton using very large datasets. They made advantage of his WHOI plankton dataset (produced by the Woods Hole Oceanographic Institution). This dataset included 3.4 million expert-labeled images of him from 103 distinct classes. They primarily concentrated their methods on addressing the issue of class imbalance in large datasets.

To eliminate the bias induced by class imbalance, they employed the CIFAR 10 CNN model as a classifier. Three levels of convolutions were followed by two completely connected layers in the suggested design. Their classifier was pre trained using class-normalized data before being restrained using the original data. We were able to remove the bias from the class imbalance as a result of this. Dai and colleagues We developed a deep folding network specifically to classify zooplankton. For the data set, the ZooScan system acquired 9460 photomicrographs and grayscale pictures of zooplankton from 13 distinct classes. They proposed ZooplanktoN, a new deep learning architecture for classifying zooplankton. After experimenting with various convolution sizes, he determined that ZooplanktoN net provided the best performance to date at 11 layers.To buttress up their claims, they ran comparative trials with other deep learning architectures, including his AlexNet, CaffeNet, VGGNet, and GoogleNet, and discovered that ZooplanktoN outperformed with 93.7% accuracy.

2.3. Deep Learning in Coral Classification

Coral color, size, shape, and texture can also differ according to the class. Furthermore, the border distinctions are hazy and organic. In addition, currents, algae blooms, and plankton abundance can change water turbidity and mild availability, impacting image color. Traditional annotation solutions such as bounding boxes, picture labels, and entire segmentation are ineffective as a result of these issues, Schechner et al. (2005). used the texture Local Binary Pattern (LBP) and the color Normalized Chromaticity Coordinate (NCC). They employed a three-layer returned propagation neural network for categorization. Beijbom et al., on the other hand, were the first to address automatic annotation on a large scale for coral reef survey photos by expanding the Moorea Labeled Corals (MLC) collection.

They developed a texture classifier that is completely based on color and texture descriptors spanning several scales and surpassed existing approaches. Elawady et al. classified corals by employing supervised Convolutional Neural Networks (CNNs). They computed Phase Congruency (PC), Zero Component Analysis (ZCA), and Weber Local Descriptor using Moorea Labeled Corals and the Atlantic Deep Sea Dataset from Heriot-Watt University (WLD). They investigated the form and texture characteristics of entry images while using spatial color channels, Rout et al. (2018). Mahmood et al. proposed a function extraction technique based entirely on Spatial Pyramid Pooling to make traditional point-annotated marine data well matched with CNNs' entry barriers (SPP). They used deep capabilities obtained from the VGGNet for coral class, Kumari et al. (2019). They also combined textual content and color-primarily based completely handmade features to improve categorization capability.

2.4. Deep Learning Opportunities for Seagrass Detection and Classification

Seagrasses are critical for sediment stabilization, carbon sequestration, and supplying food and habitat for large marine animals. It is critical to monitor seagrasses in various areas to acquire a better understanding of the temporal and spatial patterns in species composition, reproductive phenology and frequency, and the consequences of commercialization and human contact. Ten et al. (2013) used hyperspectral imaging of seagrass ecosystems to identify tubeworms from the rest of the seagrass surface and performed binary classification of seagrass. Vasamsetti et al. (2018) undertook a more extensive analysis to quantify the presence of the seagrass Posidonia oceanica in the Gulf of Palma.They made use of analogue RGB data.

The classifier Law's Logistic Model Tree (LMT) and energy measures were chosen. A grayscale co-occurrence matrix was used to identify texture changes. Using sparse coding and morphological filters, Spampinato et al. (2016) detected seagrass scars on the ocean floor using panchromatic data taken by orbiting the WorldView2 satellite. This technique was only useful for detecting coastline scars in flat coastal locations. A common technique to digital imaging, presently recommended by Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO) and Health Safety and Environment Policies (HSE), is to snap an image from a digital camera every three seconds.

The camera is normally mounted on a frame and towed behind a boat cruising at 1.5-3 knots, ensuring that the photo is around 2-3 meters away. The pictures are then evaluated using the Photo Grid or TranscetMeasure (®SeaGIS) programmes. A standard 20-point grid is overlaid, and a human operator identifies the presence and species of seagrass. Technicians frequently spend several hours analyzing a single transect of 50 m and 25-50 pictures. Because most surveys span hundreds of meters of seafloor, the analysis can take several days. Furthermore, specialists' ability to identify seagrass in pictures varies.

Deep learning algorithms have the potential to increase efficiency while reducing observer bias from analysis. To the best of our knowledge, there is no method for detecting seagrass in digital photographs that uses deep learning. As a result, there is a significant chance to study the deep ocean floor using deep neural networks to detect and classify seagrass species.

2.5. Methods for Addressing the Issue of Insufficient Underwater Images

Useful training data is required for a deep learning-based UOD model. Unfortunately, gathering enough picture data in an underwater setting is difficult. Many solutions have been proposed to address these issues, particularly data augmentation and image synthesis. Data Augmentation increases the size of the data collection by changing the labeled data set. Data can be extended in a variety of ways. Add noise to the image or reshape it. Recent work has employed generative adversarial neural networks to enhance data for domain matching. To generate distorted underwater views from high-quality aerial RGB or RGB-D shots, image synthesis methods are given. The image synthesis method is divided into two types: those based on physical models and those based on deep learning. To make underwater photographs, physical model-based technologies employ an underwater image model.

Generative Adversarial Networks have lately been researched in the field of underwater picture synthesis due to its success in image-to-image conversion difficulties. It treated underwater image synthesis as an image-to-image conversion, generating underwater images from RGB-D images collected in the air with a single GAN. To avoid the necessity for image pair training, it used an unpaired snapshot to train a two-sided cycle consistent adversarial network to learn interconversions between the air and undersea domains. Underwater image synthesis algorithms based on both physical models and deep learning, on the other hand, are unable to accurately predict the deterioration history of underwater images, resulting in subpar generated images.The commonly used underwater physical image model can only synthesis 10 Jerlov water types and takes only two characteristics into account during the degradation process, leading in considerable flaws in the resultant image. Additionally, generative adversarial networks are prone to model collapse, which results in images with monotonous tones and frequent artifacts.

3. Proposed Methodology

Object recognition is the ability to properly recognize objects, calculate their position and dimensions, and conduct semantic or instance segmentation. Previous studies relied on algorithms based on shape, color, and contour matching to recognize things, which are unsuitable for real-world object detection. Deep learning frameworks are categorized as object-regression-based, classification-based, or both. Methods for recommending regions include Region-based CNN (RCNN), Fast RCNN, Faster RCNN, and Mask RCNN, for example, indicate regions of interest and attempt to identify objects within them, but are incapable of doing so. For direct object detection, classification-based algorithms make use of an integrated framework.

Figure 2 depicts a flowchart for anticipating underwater images. A convolutional neural network-based underwater object prediction is designed using deep learning. Using additional number arrays in the model, this network detects images after encoding them into numeric arrays. Images are recorded and sent to the internet software as the user adds various characteristics of the forecast into web form and the model with TensorFlow and scaled it down from its original enormous size. The numerical values of the various objects are entered into the data collection by this model.To identify object prediction from photographs found online, the proposed technique leverages a CNN object recognition model. Image acquisition, image preprocessing, segmentation, feature extraction, and grading are five primary stages of object identification. Scanners are used for image processing tasks such as picture enhancement, segmenting a photograph into separate sections, locating infection foci, and extracting features that aid in image classification.

3.1. Dataset

The fundamental data gathering strategy employed in this investigation is depicted in Figure 3. Kaggle was used to collect data. Image acquisition refers to the process of capturing images from underwater. A digital camera or scanner is used to capture images of objects or to collect data.The type and location of the digital camera have an impact on the quality of the

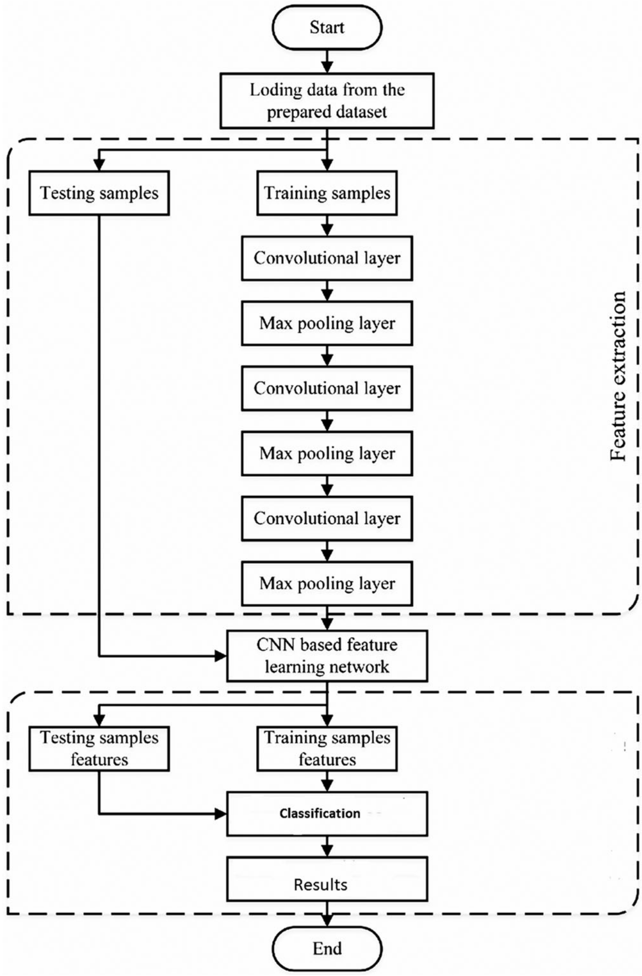

Figure 2. Flow diagram of object detection in under water images.

images. The initial step in using image data as computational input is to collect image data.

3.2. Data Preprocessing

As demonstrated in Figure 4, preprocessing methods such as resizing, scaling, zooming, and mirroring were used. Preprocessing occurs after the image is captured. Enhancing, resizing, enlarging, cropping, changing color space, smoothing, and eliminating noise from photos are all examples of preprocessing. Some of the photographs exhibit disorientation, but they appear to have all been denoised. A distorted image can be enhanced by removing the distortion with a noise reduction filter. If the image lacks contrast, actions should be done to improve it.

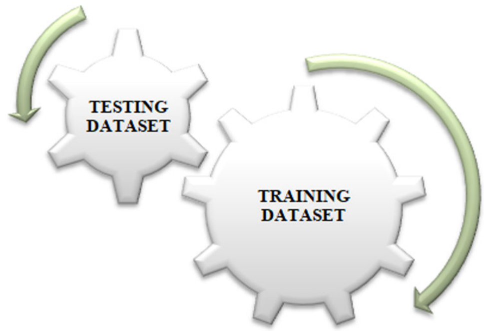

Figure 3. Dataset representation.

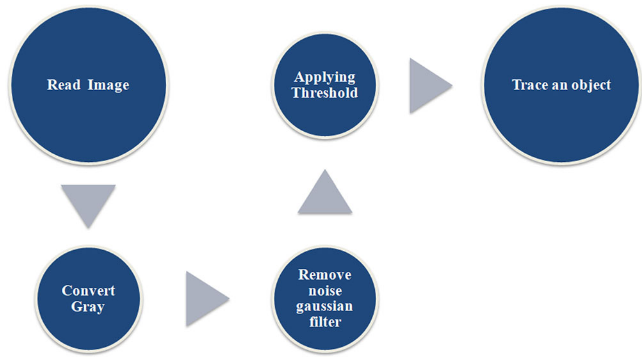

Figure 4. Preprocessing of input image.

A Gaussian filter was used to soften the image. A histogram graphically depicts the image intensity distribution. Image contrast with histograms is improved by anti-aliasing techniques in image processing. This is accomplished by dividing the image's broadest intensity value. Histograms represent a collection of photos from the preprocessing stage. Each dataset comprises two image classifications. Datasets for testing and training As a result, the training and test datasets account for 80% and 20% of the total dataset, respectively. CNN models are tested on a training set. The validation set allows to objectively examine the model while fine-tuning the hyperparameters. The test set analyses the model's training success by determining whether the technique used was correct.

3.3. Image Segmentation

After collecting the preprocessed images from the region of interest, image segmentation is required for object detection. The underwater photographs in the proposal are broken into portions. Here's an example of edge detection in modified segmentation, which separates an image into sections based on intensity values. In order to detect color changes, it uses the k-means clustering technique to partition the image and differentiate between clear and unclear object categories.

3.4. Feature Extraction

Image recognition descriptive features are sought for and extracted as part of the extraction operation. Color, texture, and shape are all widely reported characteristics. The primary color components of histograms and moments are the color qualities that color disease. A texture is created that depicts the variance of the image texture utilized for disease categorization. Entropy, uniformity, and contrast are structural characteristics.

3.5. CNN

Because of its ability to perceive and comprehend patterns, CNNs have evolved substantially. The output accuracy is quite high, making it the most efficient design for image categorization, retrieval, and recognition applications (Figure 5). Because it accepts any photo as input, this quality makes it the best prediction method.A CNN must be able to satisfy crucial properties such as spatial invariance in order to learn to recognize and extract visual information from random points within an image. CNNs automatically extract pictures and learn features from photos and data. As a result, CNNs can deliver accurate deep learning outcomes.

3.6. Object Prediction

As demonstrated in Figure 2, images are categorized based on the qualities gathered. Classification is a monitoring technique that divides element predictions into various categories. Figure 2 depicts how the classifier approach

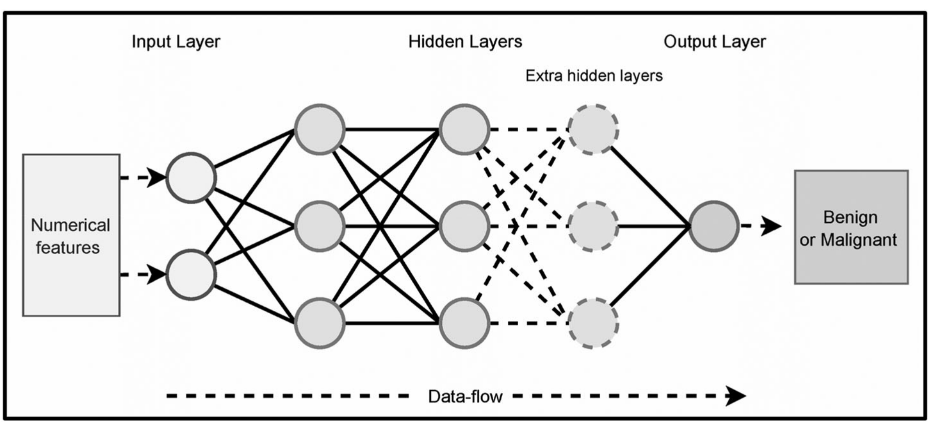

Figure 5. Convolutional neural network.

learns to characterize a predefined set of underwater images from a photograph. The training phase is the name given to this stage of learning. The trained classifier used to assess the images has an impact on accuracy.

4. Simulation Results

The technology recommended will be used in a remotely operated underwater vehicle (ROV). A total of 33,000 pictures have been manually and artificially labeled. Deep learning employs 9720 photos for training, 8910 images for validation, and 14370 images for testing. Accuracy, recall, and mean are often used to assess object detecting accuracy. Using a split training and testing strategy, divide the data into a 70% training model and a 30% testing model.

- Accuracy: Use the best-fit model to find patterns in the data set.
- Prediction: Predictions with positive outcomes split by total positive predictions
- Recall: It should be noted that measures are used to establish the type of TP
- F1 score: Weighted average of precision and recall measurements.

Confusion matrices are used to evaluate and summarize the efficacy of classifiers.

a. True Positive (TP): Correctly forecast whether the class is positive or negative.
b. False Positive (FP): An incorrect forecast of a positive class.
c. False Negative (FN): An incorrect categorization prediction.
d. True Negative (TN): Correctly anticipate false classes.

Predictions are used to classify each image in the trained set. The output of each image is predicted based on its precision.

As illustrated in Figure 6, this technique uses deep learning to produce results such as object recognition and categorization. So the overall accuracy is 98.48%. This can be done with accurate proportions.The training and validation accuracy are represented on the y-axis while the epoch is plotted on the x-axis as the model trains the dataset. The link between training accuracy and epochs is depicted in Figure 7.

The training and validation losses are displayed on the Y-axis and the epochs are exhibited on the X-axis as the model trains the dataset, as illustrated in the graph (Figures 8–10).

```
Epoch 1/15
9/9 [==============================] - 26s 3s/step - loss: 0.6372 - accuracy: 0.5437 - val_loss: 0.6720 - val_accuracy: 0.3824
Epoch 2/15
9/9 [==============================] - 26s 3s/step - loss: 0.6131 - accuracy: 0.5856 - val_loss: 0.6703 - val_accuracy: 0.3824
Epoch 3/15
9/9 [==============================] - 26s 3s/step - loss: 0.5764 - accuracy: 0.6236 - val_loss: 0.6883 - val_accuracy: 0.3824
Epoch 4/15
9/9 [==============================] - 26s 3s/step - loss: 0.4704 - accuracy: 0.7338 - val_loss: 0.5165 - val_accuracy: 0.4706
Epoch 5/15
9/9 [==============================] - 25s 3s/step - loss: 0.3656 - accuracy: 0.8897 - val_loss: 0.2308 - val_accuracy: 1.0000
Epoch 6/15
9/9 [==============================] - 25s 3s/step - loss: 0.2674 - accuracy: 0.9544 - val_loss: 0.0684 - val_accuracy: 1.0000
Epoch 7/15
9/9 [==============================] - 25s 3s/step - loss: 0.1423 - accuracy: 0.9506 - val_loss: 0.0766 - val_accuracy: 1.0000
Epoch 8/15
9/9 [==============================] - 26s 3s/step - loss: 0.1224 - accuracy: 0.9734 - val_loss: 0.0921 - val_accuracy: 1.0000
Epoch 9/15
9/9 [==============================] - 17s 2s/step - loss: 0.0649 - accuracy: 0.9772 - val_loss: 0.0104 - val_accuracy: 1.0000
Epoch 10/15
9/9 [==============================] - 15s 2s/step - loss: 0.0400 - accuracy: 0.9848 - val_loss: 0.0135 - val_accuracy: 1.0000
Epoch 11/15
9/9 [==============================] - 17s 2s/step - loss: 0.0949 - accuracy: 0.9696 - val_loss: 0.0368 - val_accuracy: 1.0000
Epoch 12/15
9/9 [==============================] - 17s 2s/step - loss: 0.0770 - accuracy: 0.9848 - val_loss: 0.3870 - val_accuracy: 0.7353
Epoch 13/15
9/9 [==============================] - 17s 2s/step - loss: 0.1163 - accuracy: 0.9620 - val_loss: 0.1581 - val_accuracy: 1.0000
Epoch 14/15
9/9 [==============================] - 18s 2s/step - loss: 0.0772 - accuracy: 0.9810 - val_loss: 0.0351 - val_accuracy: 0.9706
Epoch 15/15
9/9 [==============================] - 17s 2s/step - loss: 0.0345 - accuracy: 0.9924 - val_loss: 0.0104 - val_accuracy: 1.0000
```

Figure 6. Accuracy and loss metrics.

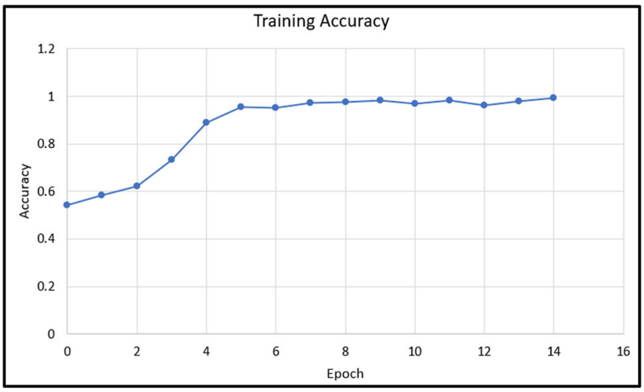

Figure 7. Training accuracy vs epoch.

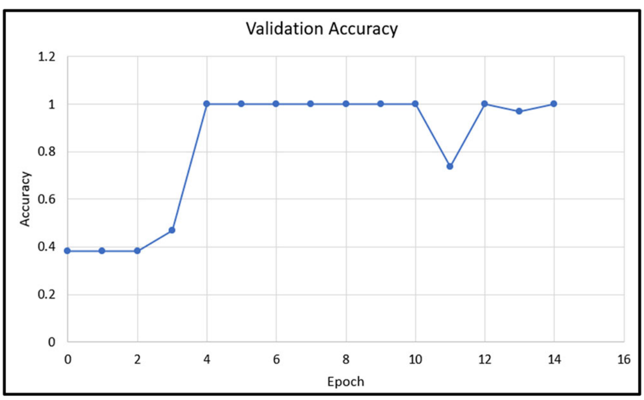

Figure 8. Validation accuracy vs epoch.

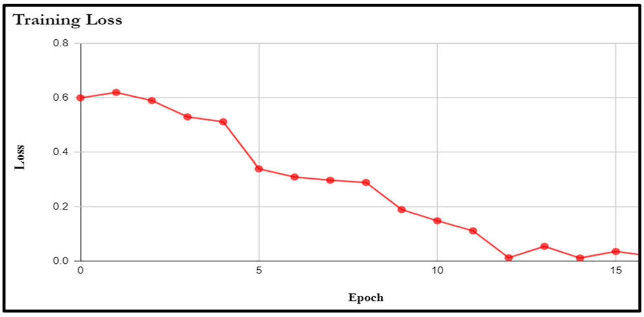

Figure 9. Training loss vs epoch.

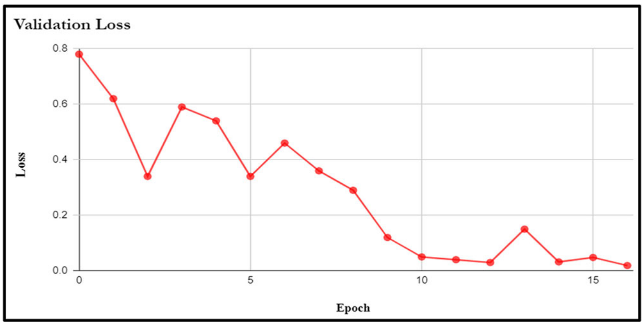

Figure 10. Validation loss vs epoch.

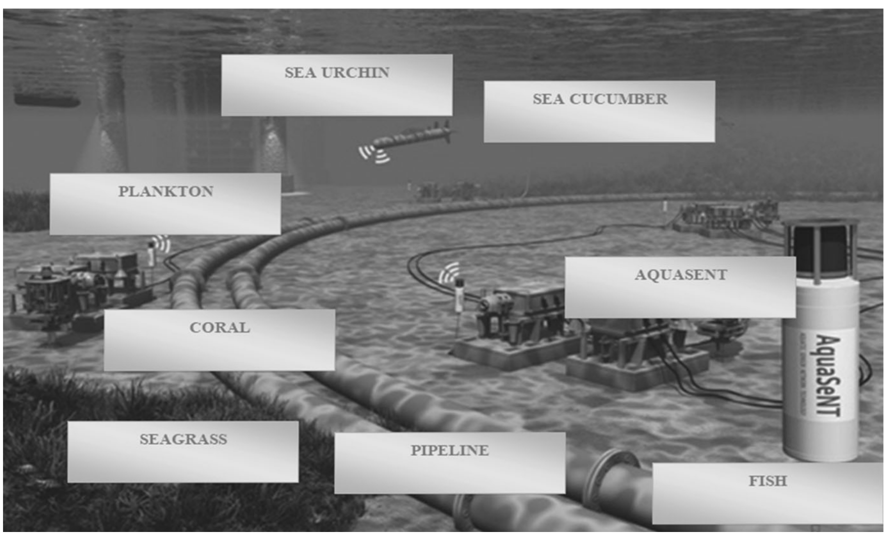

Figure 11. Under water objection detection.

Because more representational information is saved, this approach is suited for detection in water.In the ROV, a training specific model is utilized to assess the detection effectiveness. The sky is hazy, and the sea is exceedingly cloudy. Figure 11 depicts the results of the real-time detection.

As illustrated in Figure 11, certain objects are missed because the dataset is too small, especially if the images in the dataset are so similar. The lighting and surroundings are straightforward. As a result, if the trained model is employed for detection in various sea areas or under different climatic conditions, the detection accuracy will be diminished. As a result, the proposal intends to capture additional underwater images in diverse marine locations and climatic conditions.

5. Conclusion

The primary goal of underwater object detection technology is to detect objects as quickly as possible. It built and tested an autonomous underwater object detection system that can detect objects in challenging underwater images in the proposal. The suggested automatic underwater object detection output is evaluated for accuracy in terms of reduced tracking error compared to earlier detection approaches. The proposed detection system could be used as an automated module for an ocean explorer's high-end computer-equipped underwater object detection. The proposed method is designed to discover hidden or camouflaged visualization in underwater situations. The application of this method to object detection is widespread, despite its superior detection accuracy compared to existing systems.

ORCID

Manoj Prabu M http://orcid.org/0000-0002-7316-4810

References

Chiang, J. Y., and Y.-C. Chen. 2012. Underwater image enhancement by wavelength compensation and dehazing. *IEEE Transactions on Image Processing : a Publication of the IEEE Signal Processing Society* 21 (4):1756–69. doi:10.1109/TIP.2011.2179666.

Galdran, A., D. Pardo, A. Picón, and A. Alvarez-Gila. 2015. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation* 26:132–45. doi:10.1016/j.jvcir.2014.11.006.

Jalal, A., A. Salman, A. Mian, M. Shortis, and F. Shafait. 2020. Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics* 57:101088. doi:10.1016/j.ecoinf.2020.101088.

Kumari, C. U., D. Samiappan, R. Kumar, and T. Sudhakar. 2019. Fiber optic sensors in ocean observation: A comprehensive review. *Optik* 179:351–60. doi:10.1016/j.ijleo.2018.10.186.

Lee, H., M. Park, and J. Kim. 2016. Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In 2016 IEEE International Conference on Image Processing (ICIP), 3713–7. IEEE.

Li, C.-Y., J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang. 2016. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing : a Publication of the IEEE Signal Processing Society* 25 (12):5664–77. doi:10.1109/TIP.2016.2612882.

Li, X., M. Shang, H. Qin, and L. Chen. 2015. Fast accurate fish detection and recognition of underwater images with fast r-cnn. In *OCEANS 2015-MTS/IEEE Washington*, pp. 1–5. IEEE.

Marini, S., E. Fanelli, V. Sbragaglia, E. Azzurro, J. Del Rio Fernandez, and J. Aguzzi. 2018. Tracking fish abundance by underwater image recognition. *Scientific Reports* 8 (1):1–12. doi:10.1038/s41598-018-32089-8.

Mathias, A., and D. Samiappan. 2019. Underwater image restoration based on diffraction bounded optimization algorithm with dark channel prior. *Optik* 192:162925. doi:10.1016/j.ijleo.2019.06.025.

Mathias, A., S. Dhanalakshmi, R. Kumar, and R. Narayanamoorthi. 2021. Underwater object detection based on bi-dimensional empirical mode decomposition and Gaussian Mixture Model approach. *Ecological Informatics* 66:101469. doi:10.1016/j.ecoinf.2021.101469.

Nunes, J. C., S. Guyot, and E. DeléChelle. 2005. Texture analysis based on local analysis of the bidimensional empirical mode decomposition. *Machine Vision and Applications* 16 (3):177–88. doi:10.1007/s00138-004-0170-5.

Rout, D. K., B. N. Subudhi, T. Veerakumar, and S. Chaudhury. 2018. Spatio-contextual Gaussian mixture model for local change detection in underwater video. *Expert Systems with Applications* 97:117–36. doi:10.1016/j.eswa.2017.12.009.

Saleh, A., I. H. Laradji, D. A. Konovalov, M. Bradley, D. Vazquez, and M. Sheaves. 2020. A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports* 10 (1):1–10. doi:10.1038/s41598-020-71639-x.

Savant, R. R., J. V. Nasriwala, and P. P. Bhatt. 2023. Different skin tone segmentation from an image using KNN for sign language recognition. In *Proceedings of Emerging Trends and Technologies on Intelligent Systems*, 109–18. Singapore: Springer.

Schechner, Y. Y., and N. Karpel. 2005. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering* 30 (3):570–87. doi:10.1109/JOE.2005.850871.

Simonyan, K., and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Spampinato, C., S. Palazzo, P.-H. Joalland, S. Paris, H. Glotin, K. Blanc, D. Lingrand, and F. Precioso. 2016. Fine-grained object recognition in underwater visual data. *Multimedia Tools and Applications* 75 (3):1701–20. doi:10.1007/s11042-015-2601-x.

Vasamsetti, S., S. Setia, N. Mittal, H. K. Sardana, and G. Babbar. 2018. Automatic underwater moving object detection using multi-feature integration framework in complex backgrounds. *IET Computer Vision* 12 (6):770–8. doi:10.1049/iet-cvi.2017.0013.

Veiga, R. J., I. E. Ochoa, A. Belackova, L. Bentes, J. P. Silva, J. Semião, and J. M. Rodrigues. 2022. Autonomous Temporal Pseudo-Labeling for Fish Detection. *Applied Sciences* 12 (12):5910. doi:10.3390/app12125910.

Yan, Z., J. Ma, J. Tian, H. Liu, J. Yu, and Y. Zhang. 2014. A gravity gradient differential ratio method for underwater object detection. *IEEE Geoscience and Remote Sensing Letters* 11 (4):833–7. doi:10.1109/LGRS.2013.2279485.

Yang, H., P. Liu, Y. Hu, and J. Fu. 2021. Research on underwater object recognition based on YOLOv3. *Microsystem Technologies* 27 (4):1837–44. doi:10.1007/s00542-019-04694-8.

Yang, H.-Y., P.-Y. Chen, C.-C. Huang, Y.-Z. Zhuang, and Y.-H. Shiau. 2011. Low complexity underwater image enhancement based on dark channel prior. In 2011 Second International Conference on Innovations in Bio-Inspired Computing and Applications, 17–20. IEEE.