

Object Detection Model for Marine Organisms Based on Faster R-CNN

JunHan Hu

University Of New South Wales, Sydney, 2032, Australia

Abstract. With the development of marine resources, image-based biological target detection technology has gradually become the core method of marine ecological monitoring. This paper adopts Faster R-CNN technology, combined with two deep learning models, VGG and ResNet50, to improve the efficiency of target detection and recognition of underwater organisms. By combining large-scale annotated seabed image datasets for training, accurate localization and recognition of biological targets in images can be achieved. Compared to ResNet50, VGG performs better in complex seabed environments, with its mAP 1.75% higher than ResNet50, indicating higher detection accuracy and robustness. Besides, this study provides a practical and feasible solution for underwater ecological monitoring, verifying the excellent performance of ResNet50 in marine biological target detection, and providing an important and reliable support tool for deep-sea scientific research and ecological protection.

Keywords: Object Detection Model, Machine learning, Faster R-CNN, VGG16, ResNet50.

1. Introduction

With the continuous development of deep-sea scientific research and marine resources, efficient investigation of benthic biological species has gradually become a key challenge. Traditional fishing methods are difficult to sort and study deep-sea fishery resources, thus their efficiency is low. However, image processing technology provides new solutions for improving the marine ecological monitoring level and resource development efficiency. This article takes the Faster Region-based Convolutional Neural Network (Faster R-CNN) algorithm in machine learning technology as the core framework and uses deep learning models such as Visual Geometry Group (VGG16) and Residual Network (ResNet50) to train object detection models for marine organisms on the seabed. Through this innovative method, the aim is to achieve automatic monitoring and analysis of marine ecosystems through image processing of benthic organisms, to improve work efficiency and data processing capabilities. This comprehensive application injects modern intelligent elements into traditional fishing resources and environmental monitoring, providing more advanced technological support for marine scientific research, resource development, and ecological protection.

Image processing techniques include image enhancement, image restoration, image recognition, and so on. These algorithms are concerned with the imaging effect and image processing of images, which are widely used in daily life. The combination of image recognition and artificial intelligence algorithms can reduce labor costs in many fields. In the medical field, artificial neural networks can help doctors analyze the area where the lesion is located and make preliminary judgments [1]. In a comprehensive plant disease and pest detection system, image processing technology can help farms detect vegetable diseases, achieve planting automation, and reduce manual inspection costs [2]. If image processing is applied to the study of drone flight trajectories, the obstacle avoidance ability of drones can be achieved by analyzing the collected image streams [3].

The application of image processing technology in marine biological target detection provides an efficient and advanced method for marine ecological monitoring and resource investigation. The complexity and variability of the marine environment make traditional fishing face many challenges, which often include rapid and accurate identification and positioning of marine biological targets. Traditional fishing methods often rely on manual experience, but in the complex and diverse terrain and ecological environment at the bottom of the ocean, this method is difficult to meet the needs of efficient fishing. Taking the Norwegian deep-sea sea cucumber collection as an example, the

traditional method of using remotely operated vehicles (ROVs) requires strict manual operation, which places extremely high professional requirements on operators and requires them to maintain a high level of mental concentration at all times. This method that relies on manual operation often comes with issues such as operator fatigue, limited subjective judgment, and difficulty adapting to the complexity of deep-sea environments. These issues may lead to low collection efficiency, missed detections, or false detections. Therefore, the introduction of intelligent image processing technology can greatly reduce workload, which not only improves acquisition efficiency and reduces excessive reliance on operator professional skills, but also increases the system's robustness to complex seabed environmental changes [4].

This study used the Faster R-CNN algorithm, combined with VGG16 and ResNet50 as the backbone networks of deep learning models, to train the seabed biological image dataset. The purpose of this is to extract image features during the model training process, and through deep learning, enable the model to automatically locate and recognize marine biological targets. This process effectively reduces the complexity of marine species investigation and improves the accuracy and speed of biological target detection.

Considering the complexity of the marine environment, in model training, not only should challenges such as insufficient lighting, texture distortion, and uneven lighting be considered, but also factors such as the transparency of underwater plankton, water quality turbidity, and seawater temperature difference should be taken into account. These factors may lead to a decrease in image contrast, blurred edges of target objects, and increased background interference, thereby affecting the accuracy of object detection. Considering the impact of the dataset on training effectiveness, this paper adopts Kaggle's FathomNet 2023 dataset, which provides a solid foundation for research due to its wide applicability and accuracy, and provides strong support for exploring technological breakthroughs in the field of marine biological target detection [5].

The method proposed in this article can effectively process underwater biological images, and improve the accuracy and speed of biological target detection. This comprehensive application provides more efficient and cost-effective solutions for marine scientific research, resource development, and ecological protection. Through training on the seabed biological image dataset, all models used can automatically locate and recognize marine biological targets. This training process effectively reduces the complexity of marine species investigation and the reliance on manual operations.

2. Model methods

The Faster R-CNN model has been an important algorithm in the field of object detection since its inception. Proposed by Ross Girshick in 2015, it has now become the preferred framework widely used in various object detection tasks [6]. The unique feature of this model is the introduction of the Regional Proposal Network (RPN), which achieves more efficient detection speed by combining region generation and target detection in a single model. Under the same benchmark face dataset in the field of facial recognition, Faster R-CNN has faster and more accurate training results than R-CNN [7]. However, Fan and his team found that Faster R-CNN performs poorly in direct vehicle detection applications, but the adjusted network has better performance [8].

2.1 The structure of Faster R-CNN

In the prediction process of Faster R-CNN, the model can be combined with various backbone networks to extract features and obtain feature maps. After obtaining the feature map, it has two main applications: one is integration with ROI pools, and the other is applied to RPN (Regional Proposal Network).

RPN consists of two convolutional layers, 9x4 and 9x2, respectively. Among them, the 9x4 convolutional layer is used to predict the changes of each previous box on each grid point in the shared feature layer, and the convolution result will adjust these previous boxes to obtain new

bounding boxes. The 9x2 convolutional layer is used to predict whether objects exist within each proposal box on each grid point in the shared feature layer. At this stage, a rough suggestion box will be decoded to find the position to be cropped, and then the shared feature layer will be cropped. After adjusting the size and performing the next convolution, the system obtains the final prediction result.

During the training procedure of Faster R-CNN, the first step is to perform network training on the suggestion boxes. In this process, the model calculates the loss function relative to the predicted results of the Faster R-CNN recommendation network. The next step is to train the ROI network, which requires calculating the degree of overlap between all proposal boxes and the actual bounding boxes, and then proceeding with the selection process.

Specifically, the selection process between positive and negative samples is crucial. If the overlap between the actual box and the proposed box is greater than 0.5, the proposed box is considered a positive sample; On the contrary, if the overlap is less than 0.5, it is recommended that the box be considered a negative sample. This strategy helps the model to accurately locate and classify learning objectives. The calculation of the loss function includes classification loss and bounding box regression loss. The classification loss measures the accuracy of the model for the target category, while the bounding box regression loss is used to adjust the suggestion box to better fit the actual bounding box.

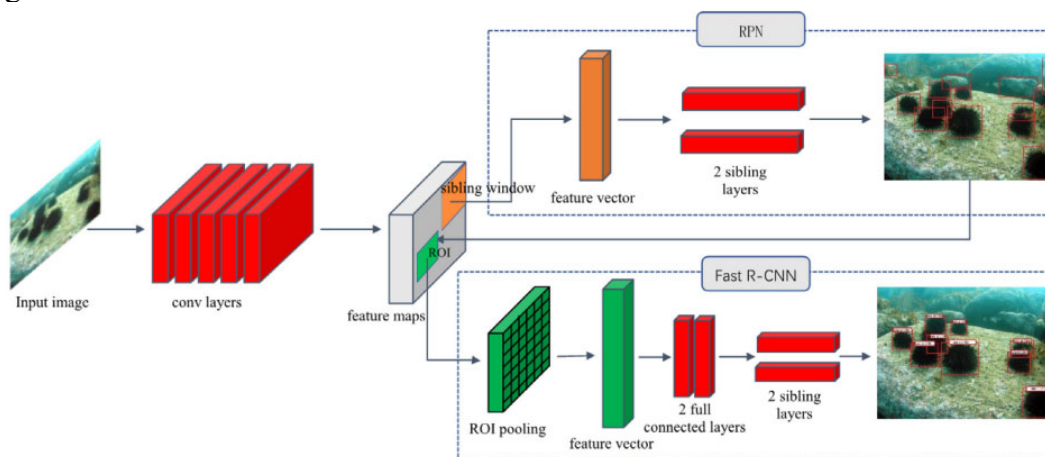


Figure 1 Structural diagram of Faster R-CNN

As shown in Figure 1, Faster RCNN consists of a convolutional network and two sub-networks, a Region Proposal Network (RPN) and a Fast Region-Based Convolutional Neural Network (Fast R-CNN). The stacking and combination of convolutional layers result in certain CNN structures, also known as backbone networks, including VGG, ResNet, Inception, etc.

When the input image passes through a convolutional layer, the generated feature map serves as the input for a 3x3 sliding window, followed by the application of a predefined set of anchors. Subsequently, two 1x1 convolutional layers are used to generate object scores and bounding box offsets, respectively. This constitutes the basic principle of the Regional Proposal Network (RPN) in Faster RCNN, which generates suggestions.

On the other hand, the Faster R-CNN network is used to classify proposals detected by RPN. The candidate regions are cropped from the feature map, and processed through an ROI pooling layer and two convolutions, ultimately producing outputs for softmax classification and bounding box regression.

The main advantage of Fast R-CNN lies in its two-stage structure: first, high-quality candidate regions are generated through RPN, and then Fast R-CNN is used for accurate classification and localization. This structure enables Faster R-CNN to achieve higher accuracy in processing complex and diverse images.

The loss function of Faster R-CNN is used to adjust the internal weights of the model, which is usually combined with backpropagation because a correct learning rate setting can help the model converge.

2.2 Backbone network

Currently, there are two main methods for processing ocean images, both based on convolutional neural networks (CNNs) [9].

The first type consists of a one-stage object detection algorithm, which directly obtains the class probability and position coordinates through regression. The noteworthy methods in this category include SSD, RFB, YOLO, and EfficientDet. These methods provide fast detection but typically have lower accuracy. For example, Zhang Youbo et al. [10] applied a multi-granularity pruning strategy to YOLOv4 for bidirectional compression of channels and convolutional layers. Chen's team [11] proposed a multi-class inverse Adaboost method based on SSD networks to improve the detection accuracy of small targets.

The second type includes two-stage object detection algorithms, which are known to have high detection accuracy but relatively slow computational speed. Common methods in this category include R-CNN, Fast R-CNN, and Faster R-CNN. In addition, Zhang introduced an underwater fish detection method based on Faster RCNN in his article [12], which achieved the automation of fish data collection. Han [13] used grayscale shadow technology to achieve underwater visual enhancement and adjusted the CNN structure to improve mean Average Precision (mAP).

Kaggle's challenge focuses on identifying animals, with a focus on the accuracy of species identification. Therefore, it is more inclined to use a two-stage detection method. Meanwhile, the training dataset consists of images captured from the upper ocean (less than 800 meters) (see appendix), characterized by clear images and rich species diversity, while the target dataset comes from deep sea areas. That is to say, the training dataset can include situations where multiple species of different sizes appear in a single image, thus a flexible model that can handle complex images is essential. Compared with traditional machine learning techniques, deep learning has better performance. Faster R-CNN ensures processing accuracy, while VGG16 and RESnet50 serve as backbone networks for deep convolutional neural network structures used for complex and high-precision feature extraction. Therefore, this article will use a more efficient RCNN, combined with VGG16 and RESnet50, to address this challenge.

2.3 VGG network

The performance of two backbone feature extraction networks was compared. Compared to traditional machine learning techniques, the VGG network is known for its concise and consistent structure. One of its notable features is the use of 3x3 small convolution kernels, which allows the network to capture image features at finer granularity while reducing the number of parameters. In addition, VGG includes 2x2 pooling layers for downsampling, which helps capture larger spatial structures [14]. Through small convolution kernels and deep stacking, it performs well in image recognition tasks, providing convenience for transfer learning. However, due to the large depth of the network, VGG has a high number of parameters and computational costs, which limits its application in resource-constrained environments. In addition, relatively simple structures may not perform as well on some complex tasks as networks with deeper and more complex design elements. Nevertheless, the design approach of the VGG network still provides valuable training results in the deep learning model presented in this paper.

In the training procedure, VGG16 is employed. Compared to VGG19, VGG16 has fewer convolutional layers, improving computational efficiency while avoiding potential issues with smaller datasets.

To further improve the performance of the VGG network, we have introduced BatchNormalization. This technology can normalize the input of each layer of neurons to a standard normal distribution with a mean of 0 and a variance of 1, which prevents the problem of gradient vanishing, ensuring that small changes in input will lead to significant changes in the loss function and accelerate training speed.

2.4 ResNet network

ResNet has introduced bottleneck structures and skip connections, and through these innovative designs, it has successfully improved computational efficiency and allowed residuals to flow between layers further in the network, effectively alleviating the problem of gradient vanishing [15]. Its advantage lies in better training of deep networks while maintaining relatively light computational requirements. Even as the network architecture continues to deepen, the performance of ResNet50 can be maintained.

In the structure of ResNet50, the initial convolutional layer is used to capture the basic features of the input image, followed by the max pooling layer. The following four stages include different numbers and configurations of residual blocks, each containing multiple convolutional layers. Finally, the network is classified through a global average pooling layer, followed by a fully connected layer.

The advantage of ResNet50 lies in its ability to maintain good training performance even with increased depth and complexity. However, due to its relatively deep network structure, ResNet50 may face some challenges in computing and storage resources, especially in resource-constrained environments. Therefore, when applying ResNet50, it is necessary to balance its performance advantages and resource costs to ensure optimal performance in different scenarios.

3. Experiments and results

Before training machine learning models, dataset preprocessing is necessary to ensure the effectiveness of subsequent inputs. An excellent model can achieve fast convergence, prevent overfitting effectively, make the training process efficient and not time-consuming, and perform well when facing the test set.

3.1 Dataset preprocessing

The dataset for this article comes from the Kaggle Challenge (Fathom 2023 Ocean Change, Species Change: Out of Sea Sample Detection). This dataset contains 290 unidentified species, each recorded in a file named "category_key.csv". The training data stored through "train.json" and "evaluate.json" provides detailed object information, including bounding boxes, areas, etc. To meet the training needs of machine learning models, the research team manually converted the dataset into VOC format using Python scripts to provide standardized input. A total of 15000 training images are included, each of which records annotation information in the form of an XML file.

The purpose of this dataset is to provide effective training tools for the automatic detection and classification of deep-sea organisms. By utilizing large-scale and diverse data, models can learn and recognize the characteristics of different species, improving their generalization ability. The task of external detection of deep-sea samples is of great significance for ecological research, resource management, and environmental protection. Therefore, by adopting advanced machine learning techniques, this study aims to improve the efficiency and accuracy of deep-sea biological species investigation and provide stronger support for deep-sea scientific research and protection.

3.2 Model evaluation indicators

In machine learning, the loss function is a key indicator for evaluating the difference between model predictions and actual labels, used to quantify model performance and serve as an optimization objective for the training process. The loss functions used in this article mainly include Mean Squared Error (MSE) and Cross-Entropy. Among them, MSE is suitable for regression problems, measuring the average square difference between the predicted values of the model and the true labels, while cross-entropy is commonly used for classification problems, measuring the distance between the predicted probability distribution of the model for the true category and the actual distribution. When calculating the Loss value, the predicted value p_i generated by the model and the target value t_i

jointly calculate the regularization loss function. The difference is measured by L_{cls} and normalized by λ / N_{reg} .

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, t_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

To comprehensively evaluate the performance of the object detection model, Mean Average Precision (mAP) was used as a commonly used evaluation metric. MAP comprehensively considers the precision and recall of the model under different thresholds. Among them, precision measures the ratio of the number of correctly detected positive categories to the total number of detections, while Recall measures the ratio of the number of correctly detected positive categories to the actual positive categories. By calculating the average accuracy under the Precision-Recall curve, the mAP metric can provide a comprehensive and comprehensive performance evaluation of the model.

$$AP = \int_0^1 p(r) dr$$

The calculation method of the loss function varies depending on the task but usually involves the difference between the predicted value of the function calculation and the actual label. In deep learning, optimization algorithms such as gradient descent are used to adjust the model parameters to minimize the loss function, thus enabling more accurate learning of tasks on the training set. This method helps the model better adapt to training data and improve generalization performance.

In the Faster R-CNN model, regression loss is primarily implemented through the `fast_rcnn_loc_loss` function. The calculation of this loss adopts a smooth L1 loss function, whose core idea is to measure the difference between the predicted box position and the true box position. Firstly, filter out non-zero elements from all positive samples, and then calculate the absolute difference between the predicted position and the true position. For each difference term, apply square loss or linear loss, depending on its relative size to the smoothing parameter, sigma. This design can smoothly reduce the loss when the difference is small while using linear loss when the difference is large improves the robustness. The confusion matrix comprises four key indicators: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

$$MissRate = \frac{FN}{TP + FN}$$

$$LAMR = \frac{1}{n} \sum_{i=1}^n \log(MissRate)$$

The entire loss calculation process involves summing up the losses of all differential terms and normalizing them based on the number of positive samples. This ensures that the calculation of losses not only effectively guides the model in learning the accurate position of targets, but also adapts to targets of different sizes and shapes. By minimizing regression loss, the model can better adjust the bounding boxes to improve the accuracy of object detection.

4. Training results

In this study, experiments of two different models will be compared on the same dataset. Specifically, Model 1 uses VGG16 as the backbone, while Model 2 uses RESnet50 as the backbone architecture.

4.1 VGG16

A notable feature of VGGNet is its utilization of 3x3 small convolutional kernels. This design enables the network to capture image features at a finer granularity while minimizing the number of parameters. Additionally, VGG incorporates 2x2 pooling layers, which play a crucial role in downsampling and aiding the capture of larger spatial structures. This architectural choice contributes to VGGNet's effectiveness in image recognition tasks, as it allows the model to learn intricate patterns

and retain important spatial information. The combination of small kernels and pooling layers has been proven successful in creating a deeper network without excessive computational burden.

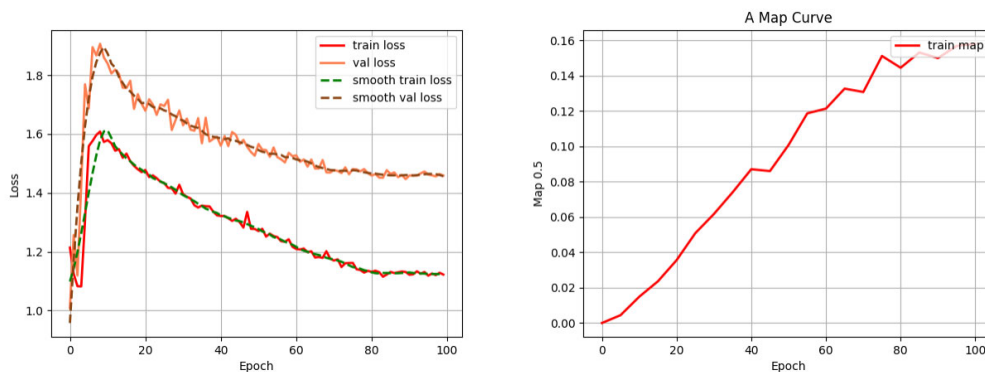


Figure 2 Loss and mAP of VGG

The observed trend in the training loss and val loss suggests that the model encounters some difficulty in effectively capturing the underlying patterns and complexities within the data. The initial increase in both losses may be attributed to the model's struggle to converge to an optimal solution. However, as the training progresses, the model starts to learn better representations, leading to a decrease in the losses.

Despite the improvements in train loss, the consistently higher val loss indicates that the model's performance on unseen data is subpar. This suggests that the model may not be able to generalize well beyond the training dataset. To address the issue of underfitting, several strategies could be explored, such as increasing the model's capacity, adjusting hyperparameters, or acquiring more diverse training data.

Regarding the mAP curve in Figure 2, the initial value of 0 indicates a lack of meaningful predictions in the early stages of training. However, as the model learns from the data, the mAP curve shows progress, reaching 0.16 at the end of the 100 epochs. While this improvement is positive.

It can be observed that when the color of the marine organisms is similar to their environment, the mAP value tends to be smaller, possibly approaching 0. However, when the color of the organisms differs from their environment, the mAP value becomes significantly higher. In such cases, the model achieves accuracy and recall rates of over 95%, demonstrating excellent recognition performance. That means the current model performs poorly for marine organisms with protective coloration, as it struggles to accurately identify them. However, for marine organisms with distinct color differences from the environment, the model can precisely detect and classify them.

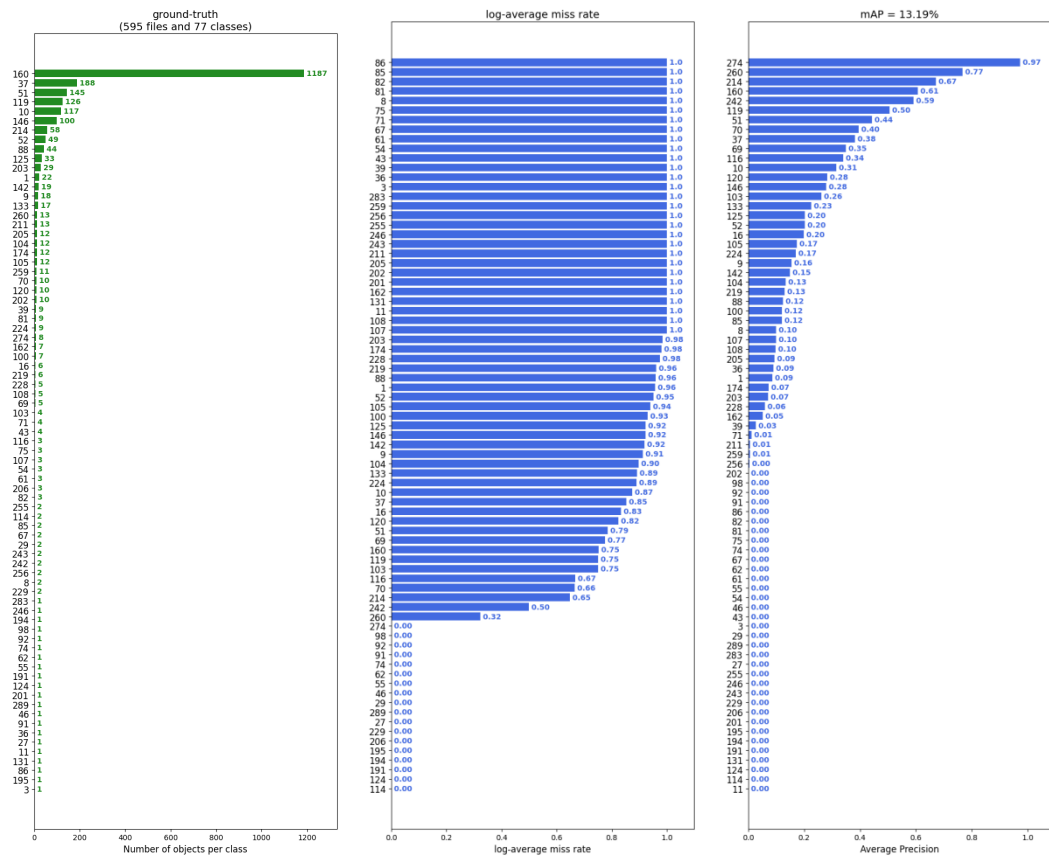


Figure 3 Confusion matrix of VGG

The following images are the training results regarding VGG16, predicted using the VGG16 model.

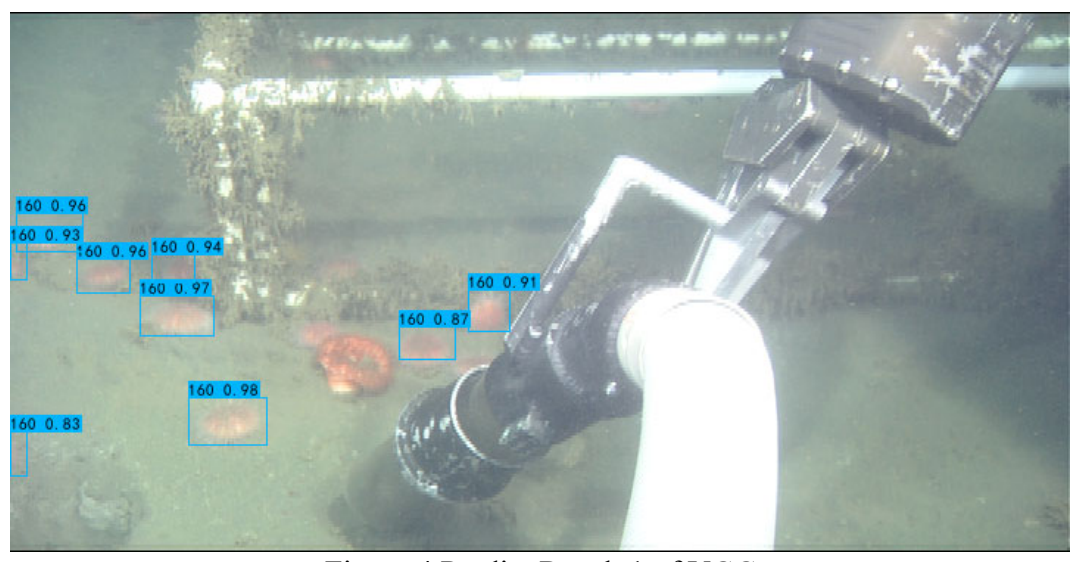


Figure 4 Predict Result 1 of VGG

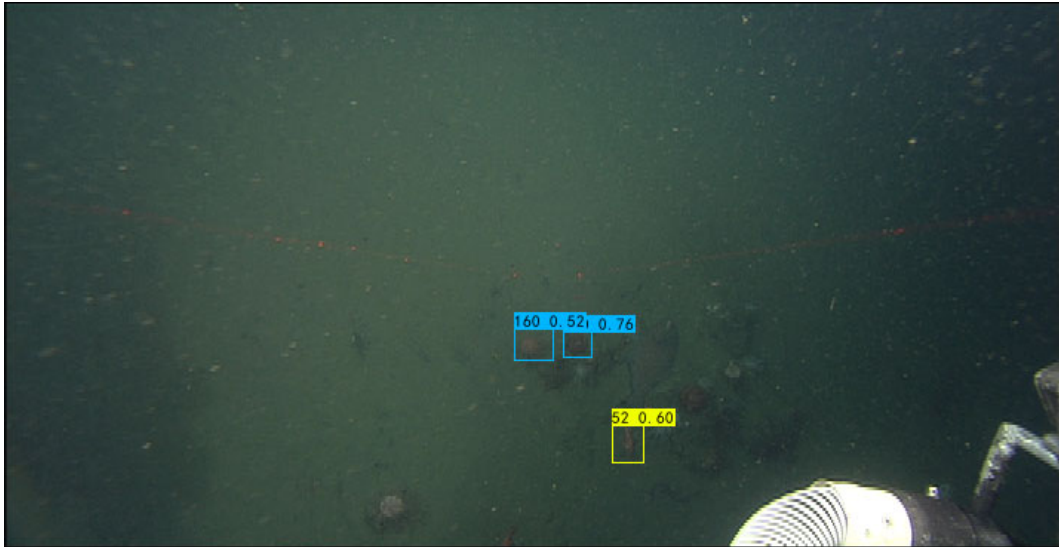


Figure 5 Predict Result 2 of VGG

4.2 ResNet50

For ResNet50, the architecture starts with a convolutional layer to capture the basic features of the input image, followed by a max pooling layer. It consists of four stages, each containing a different number and configuration of residual blocks, where each residual block includes multiple convolutional layers. Then, the network uses a global average pooling layer to perform classification, followed by a fully connected layer. The design choices in ResNet50 can achieve effective training even with increasing depth and complexity, making it the preferred backbone network in this report, as it can maintain good performance while maintaining controllable computational requirements.

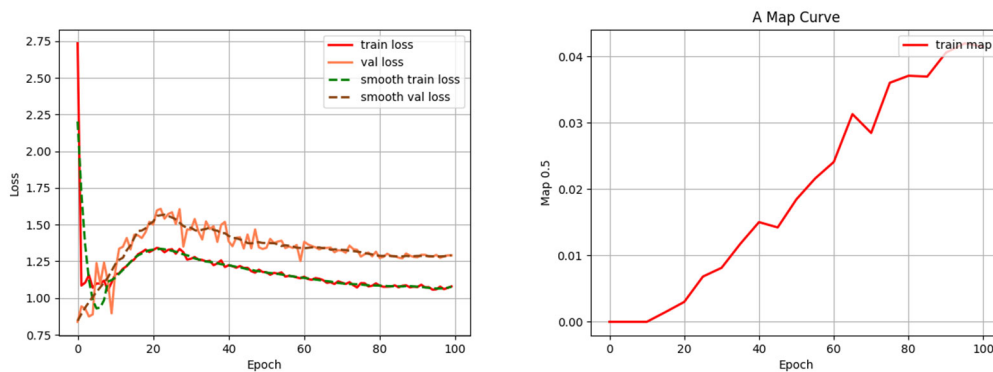


Figure 6 Loss and mAP of ResNet50

ResNet50, as a powerful deep neural network, performs well in many image recognition tasks. However, when applied to image recognition of marine benthic organisms, its training effectiveness may be affected by various factors.

Firstly, the characteristics of the dataset may not match the design of ResNet50. The characteristics of underwater biological images may differ from image datasets that have achieved success in other fields, leading to performance degradation. In addition, the size and diversity of the dataset are also key factors. If the training set is too small or not diverse enough, the model may not be able to fully learn the complex features of marine biological images.

Label quality is also an important consideration factor. If the labels in the training data are inaccurate or inconsistent, the model may be misled, thereby reducing performance. In addition, the pre-training of the model in other fields may not be sufficient to adapt to specific features of marine seabed biological images, and more emphasis needs to be placed on transfer learning and domain adaptation.

The methods address these issues include data augmentation, which enhances the diversity of the dataset by performing various transformations on the images. Domain adaptation is also an important strategy, which can be pre-trained in fields related to marine benthic organisms to better adapt to target tasks. At the same time, it is necessary to ensure high-quality labels and improve their accuracy through professional annotation, or the use of semi-supervised learning methods.

The models have been trained for 100 iterations. Within this period, the precision reached an impressive 80%, while the recall rate hit 83%. These figures illustrate that the selected model has been successful in identifying a particular category of marine organisms with a high degree of accuracy, ensuring that the true positives are well recognized. Furthermore, the increased recall rate underlines the model’s effectiveness in finding relevant instances within the data set, making VGG16 a more reliable choice for this specific task.

The slight edge that VGG16 has over ResNet50 in this comparison could be attributed to differences in architecture and may lead to insights that could help in further refining models for specific applications like marine organism recognition.

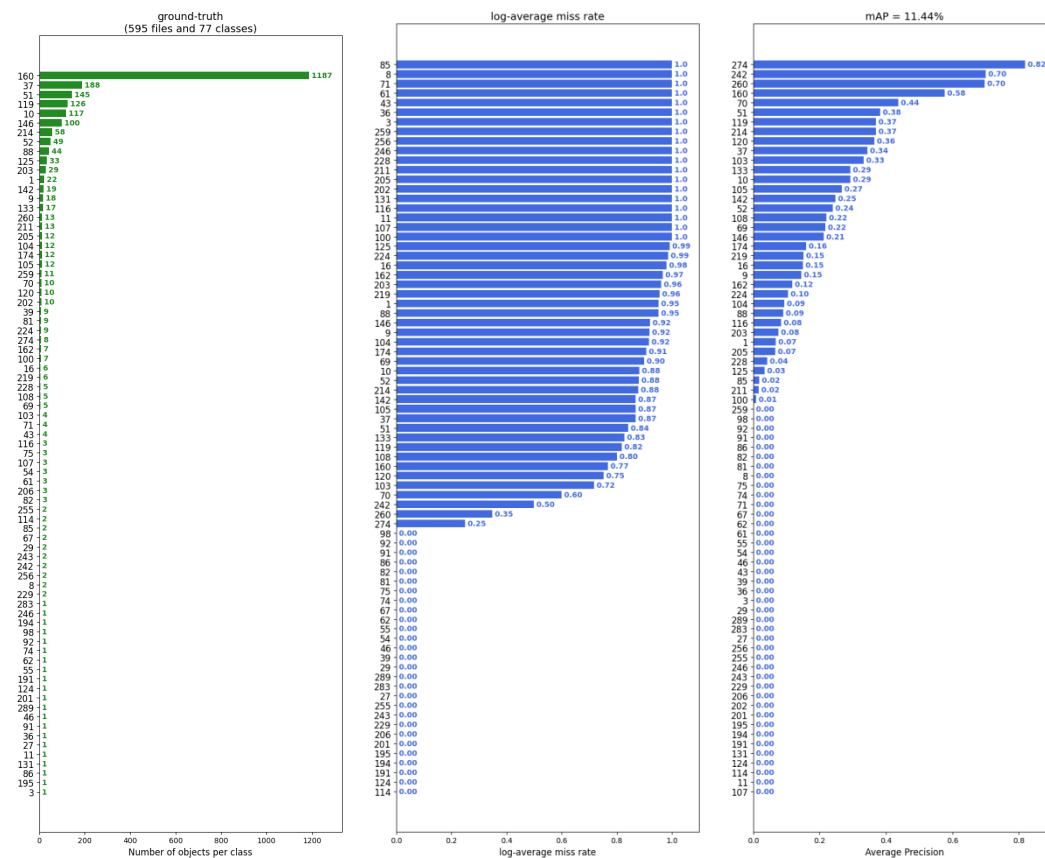


Figure 7 Confusion matrix of ResNet50

The following images are the training results regarding ResNet50, predicted using the ResNet50 model.

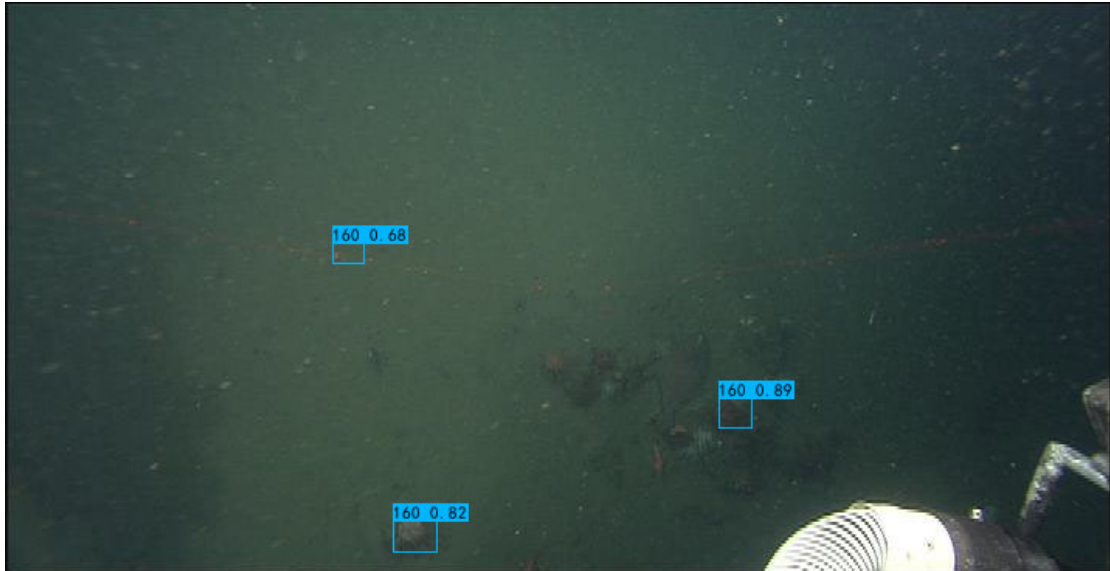


Figure 8 Predict Result 1 of ResNet50

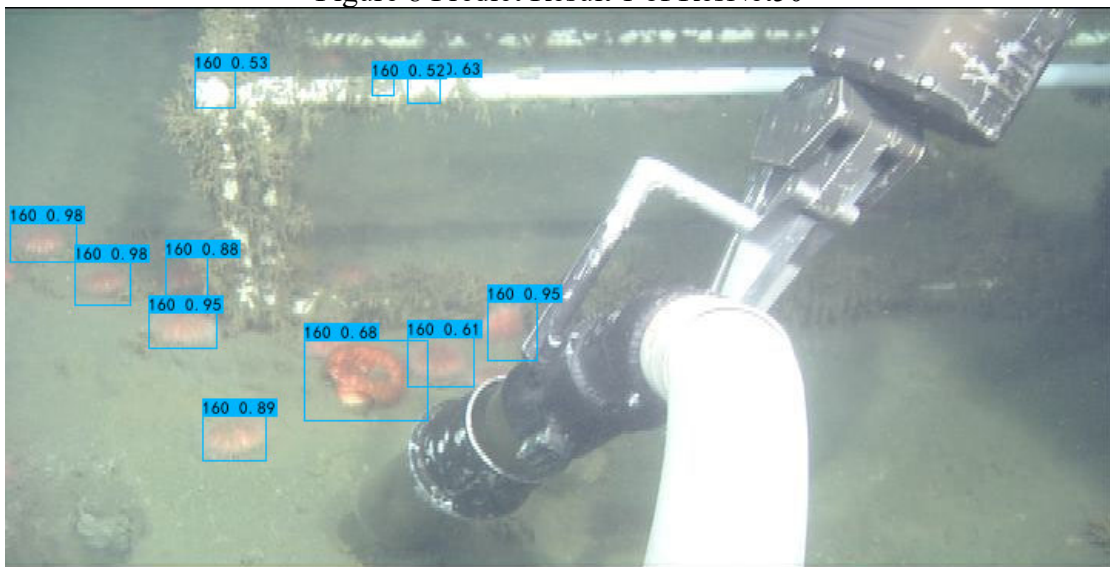


Figure 9 Predict Result 2 of ResNet50

In this comparison, VGG16 has a slight advantage over ResNet50, which may be attributed to differences in architecture and may bring insights that can help further improve specific application models such as marine biometrics.

5. Conclusion

This study aims to use the Fast R-CNN model and two different backbone networks (RESnet50 and VGG16) to accurately detect marine life images in the Kaggle dataset. The experimental results indicate that although the detection performance of both backbone networks did not reach the ideal level, VGG16 performs better than ResNet-50, possibly due to its relatively simple structure that is more suitable for the specific context of the current dataset. However, ResNet-50 may not achieve better performance in all cases, which may be related to the characteristics of the training data.

Another notable observation is that although the dataset contains labels related to certain categories, these categories are not included in all retrieved images. To overcome the problem of inaccurate category labels, more detailed design and consideration should be carried out in the data preprocessing stage. Firstly, semi-supervised learning methods can be considered to improve the generalization performance of the model by utilizing unlabeled data. Secondly, with the participation

of domain knowledge and professionals, an in-depth review and validation of the category labels of the dataset can be carried out. This helps to reduce training bias caused by label errors and improve the credibility of the model in practical application scenarios. In addition, data augmentation techniques can be used to generate more diverse samples through random rotation, flipping, and other methods, thereby improving the robustness of the model.

Overall, the subjective analysis of model performance and dataset labels has given us a deeper understanding of the subjective nature of existing problems and has provided clearer directions and improvement paths for future research. With the continuous progress of technology, in-depth research on underwater biodiversity will help to better understand the complexity of marine ecosystems, thereby promoting the development of marine protection and sustainable management. By combining advanced computer vision technology and deep learning models, it is expected to achieve more accurate and efficient detection and monitoring of underwater biological communities in the future, providing strong support for protecting the marine ecological environment and promoting biological research. The in-depth exploration in this field will have a profound and positive impact on global environmental protection and biodiversity conservation.

References

- [1] Miller, A. S., B. H. Blott, and T. K. Hames. "Review of neural network applications in medical imaging and signal processing." *Medical and Biological Engineering and Computing* 30 (1992): 449-464.
- [2] Mugithe, Pradeep Kumar, et al. "Image processing technique for automatic detection of plant diseases and alerting system in agricultural farms." 2020 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2020.
- [3] Carnie, Ryan, Rodney Walker, and Peter Corke. "Image processing algorithms for UAV" sense and avoid". Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006. IEEE, 2006.
- [4] Han, Fenglei, et al. "Underwater image processing and object detection based on deep CNN method." *Journal of Sensors* 2020 (2020).
- [5] Ben Woodward, eor123, GenevievePatterson, Lilli Carlsen. (2023). FathomNet 2023. Kaggle. <https://kaggle.com/competitions/fathomnet-out-of-sample-detectio>
- [6] Faster, R. C. N. N. "Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 9199.10.5555 (2015): 2969239-2969250.
- [7] Jiang, Huaizu, and Erik Learned-Miller. "Face detection with the faster R-CNN." 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017). IEEE, 2017.
- [8] Fan, Quanfu, Lisa Brown, and John Smith. "A closer look at Faster R-CNN for vehicle detection." 2016 IEEE intelligent vehicles symposium (IV). IEEE, 2016.
- [9] Wang, Hao, and Nanfeng Xiao. "Underwater Object Detection Method Based on Improved Faster RCNN." *Applied Sciences* 13.4 (2023): 2746.
- [10] Zhang, Y. B., W. Guo, and Y. Zhou. "Real-time target detection of underwater relics based on multigranularity pruning." *Laser & Optoelectronics Progress* 58.14 (2021): 1410019.
- [11] Chen, Long, et al. "SWIPENET: Object detection in noisy underwater images." arXiv preprint arXiv:2010.10006 (2020).
- [12] Zhang, David, et al. "Unsupervised underwater fish detection fusing flow and objectiveness." 2016 IEEE Winter Applications of Computer Vision Workshops (WACVW). IEEE, 2016.
- [13] Yao, Jingzheng, Fenglei Han, and Haitao Zhu. "Underwater image processing and object detection based on modified YOLO method." ISOPE International Ocean and Polar Engineering Conference. ISOPE, 2019.
- [14] Jiang, Zhi-Peng, et al. "An improved VGG16 model for pneumonia image classification." *Applied Sciences* 11.23 (2021): 11185.
- [15] Sarwinda, Devvi, et al. "Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer." *Procedia Computer Science* 179 (2021): 423-431.