Asia-Pacific Journal of Information Technology and Multimedia 13(02) 2024: 204 – 220
https://doi.org/10.17576/apjitm-2024-1302-04

Comparison of YOLOv7, YOLOv8, and YOLOv9 for Underwater Coral Reef Fish Detection

Perbandingan YOLOv7, YOLOv8 dan YOLOv9 untuk Pengesanan Ikan Terumbu Karang Bawah Air

Mohammad Amyruddin Shamsuddin[1], Wan Nural Jawahir Hj Wan Yussof[1], Muhammad Suzuri Hitam[1], Ezmahamrul Afreen Awalludin[2], Muhammad Afiq-Firdaus Aminudin[3], Zainudin Bachok[3]*

[1]Artificial Intelligent Group, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

[2]Faculty of Fisheries and Food Sciences, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

[3]Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

**Corresponding author: wannurwy@umt.edu.my*

Received 29 May 2024
Accepted 9 September 2024, Available online 7 October 2024

ABSTRACT

Automated underwater fish detection offers an appealing solution to improve efficiency and cost-effectiveness compared to labor-intensive manual detection methods. This study conducted a thorough assessment of three state-of-the-art single-stage detectors belonging to the You Only Look Once (YOLO) series – namely, YOLOv7, YOLOv8, and YOLOv9 – focusing on the detection and classification of four dominant coral reef fish species. These YOLO models were trained using a customized dataset comprised of underwater images showcasing the fish species, sourced from Pulau Bidong and neighboring islands in Terengganu, Malaysia. Data collection was facilitated using the Stereo-Diver Operated Underwater Video System (Stereo-DOVs). The main objective of this study is to determine the top-performing model for precisely detecting and classifying the fish in the images. Notably, each of the YOLO models achieved high mean Average Precision (mAP)@0.5 scores, with percentages of 96.6%, 97.9%, and 94.3% respectively. Further visual examination showcased the models' adeptness in accurately detecting the majority of fish instances within the test dataset and dataset images from the internet, confirming their robust performance. Taking into account both the evaluation metrics and visual results, YOLOv7 and YOLOv8 stand out as appealing choices to be used as the base models for our future study.

Keywords**:** Artificial intelligence; Computer vision; Deep learning; Fish detection; YOLO

ABSTRAK

Pengesanan ikan bawah air secara automatik menawarkan penyelesaian yang menarik untuk meningkatkan kecekapan dan keberkesanan dari segi kos berbanding kaedah pengesanan secara manual. Kajian ini menjalankan penilaian menyeluruh terhadap tiga pengesan peringkat tunggal tercanggih milik siri You Only Look Once (YOLO) - iaitu, YOLOv7, YOLOv8 dan YOLOv9 - memfokuskan pada pengesanan dan pengelasan empat dominan spesies ikan karang. Model YOLO ini dilatih menggunakan set data tersuai yang terdiri daripada imej bawah air yang mempamerkan spesies ikan, yang diperolehi dari Pulau Bidong dan pulau berhampiran di Terengganu, Malaysia. Pengumpulan data telah dikumpul menggunakan Sistem Video Bawah Air (Stereo-DOVs). Objektif utama kajian ini adalah untuk menentukan model yang berprestasi tinggi untuk mengesan dan mengklasifikasikan ikan dalam imej dengan tepat. Secara keseluruhannya, setiap model YOLO mencapai skor min Purata Ketepatan (mAP)@0.5 yang tinggi, dengan peratusan masing-masing 96.6%, 97.9% dan 94.3%. Pemeriksaan visual selanjutnya mempamerkan kebolehan model dalam mengesan dengan tepat kebanyakan contoh ikan dalam set data ujian dan imej set data daripada internet, mengesahkan prestasi mantap mereka. Dengan mengambil kira kedua-dua metrik penilaian dan hasil visual, YOLOv7 dan YOLOv8 menonjol sebagai pilihan yang menarik untuk digunakan sebagai model asas untuk kajian masa depan kami.

Kata kunci**:** Kecerdasan buatan; Penglihatan computer; Pembelajaran mendalam; Pengesanan ikan; YOLO

INTRODUCTION

Coral reefs are often called the "tropical rainforests of the sea" for their remarkable diversity, provide a wide array of ecosystem services and advantages to humanity. Numerous studies have demonstrated the significance of coral reefs in various aspects including fisheries (Moberg & Folke 1999; Teh et al. 2013), coastal protection (Costanza et al. 1997), supporting tourism (Cesar et al. 2003), recreation (Adey 2000), and contributing to the exploration of potential medicinal resources (Bruckner 2023). However, as noted by Eddy et al. (2021), the ongoing degradation of coral reef ecosystems is primarily attributed to the persistent effects of global environmental changes and human activities. Therefore, it is imperative to take action to continuously monitor the condition of coral reef environments. One such action is fish population monitoring, as this method can help evaluate the relative importance of environmental threats impacting local reefs and prioritize management efforts.

Therefore, accurately detecting underwater fish in coral reef environments is critically important. It allows for the estimation of fish species' relative abundance in their natural habitats and facilitates the monitoring of their populations. However, as highlighted by Weinstein (2018), manual data processing bears resemblances to physical data collection in terms of being labor-intensive and time-consuming. Currently, some researchers, such as Afiq-Firdaus et al. (2023), continue to utilize manual fish counting alongside software tools for analyzing fish abundance. While this method can effectively identify and measure fish in

images manually, it may not be the most efficient or accurate approach, especially for large-scale fish monitoring projects.

Recent advancements in machine learning technologies have propelled deep learning to the forefront as an invaluable tool for addressing this challenge. As outlined by LeCun et al. (2015), deep learning is a subset of machine learning that employs multiple computational layers to process complex data, such as raw images and video recordings, which are difficult to analyze through traditional methods. Yet, the hurdles of acquiring usable footage in marine settings to attain satisfactory computational performance differ significantly from those encountered in terrestrial environments. The intricate characteristics of the underwater environment present the most significant challenge in detecting and recognizing objects within underwater images (Awan et al. 2019; Nair et al. 2021; Shen et al. 2021). The primary difficulty in underwater imaging stems from the restricted presence of light, resulting in significant fluctuations in light intensity that lead to inadequate luminosity, distortion, and light attenuation (Kong et al. 2018; Marshall 2017; Rizzini et al. 2015; Xu & Matzner 2018). Additional obstacles include scale variations, complex cluttered backgrounds, arbitrary orientations, and degradation in image quality (Li et al. 2022; Yeh et al. 2021). Although these factors may affect the quality of images and videos, deep-learning methods have shown remarkable effectiveness in classifying tropical reef fish. Besides, Siddiqui et al. (2018) and Xu & Matzner (2018) have presented compelling evidence indicating that these methods have outperformed human capabilities in species recognition.

Furthermore, employing deep-learning methods for fish detection can alleviate researchers from the laborious process of analyzing fish abundance, thereby saving time. Moreover, studies conducted by researchers such as Li et al. (2020), have shown that utilizing deep-learning methods for underwater fish detection offers a non-intrusive method, facilitating accurate and effective monitoring of fish populations and behaviors while safeguarding their natural habitat. As asserted by Pagire & Phadke (2022), the integration of Convolutional Neural Networks (CNN) in deep learning facilitates the extraction of features from underwater images, enabling the classification of these images into different fish species. Numerous researchers have also applied the CNN models to develop a smart underwater vision system (UVS), resulting in satisfactory performance (Han et al. 2020; Huang et al. 2019; Moniruzzaman et al. 2019; Zhao et al. 2019).

The advancement of deep learning and the evolution of more advanced algorithms have led to the rapid growth of You Only Look Once (YOLO) architectures. YOLO, a member of the family of CNN models, is a leading algorithm for object detection due to its remarkable speed and precision. Moreover, it has found applications in numerous research endeavors aimed at identifying fish in underwater habitats. As shown in Figure 1, YOLO has evolved through multiple iterations and enhancements, resulting in improved performance in object detection, faster inference times, and the development of more resilient algorithms applicable across diverse domains. Hence, in this study, we have trained various iterations of YOLO models – namely, YOLOv7, YOLOv8, and YOLOv9 – using a customized coral reef fish dataset. We

then conducted a comparative analysis of their accuracy to determine the most fitting YOLO model version for our dataset's requirements.

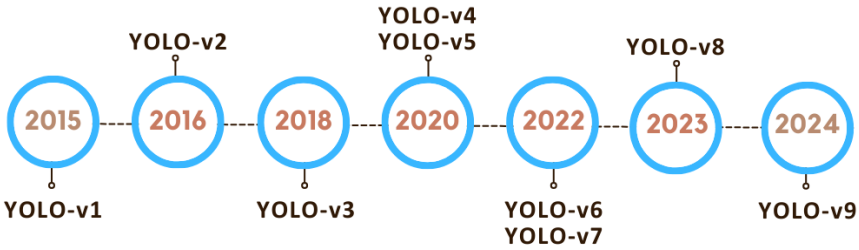


FIGURE 1. YOLO evolution timeline

RELATED WORKS

Various methods have been employed to detect fish and estimate their populations using image and video processing algorithms. Initially, Stratchan (1993), identified fish based on their shape and color, while Storbeck & Daan (2001), developed 3D models of fish to capture dimensions such as height, width, and thickness. However, these methods were applied in controlled sampling environments. Detecting and classifying fish underwater without constraints and without assuming specific environmental conditions is challenging due to significant variations in factors such as water clarity, lighting, and the presence of other objects. To address this, previous research by Spampinato et al. (2008), proposed an image-processing method that captures the textural patterns of fish in underwater environments. This method enables the detection and counting of fish in low-quality, unrestricted underwater footage, demonstrating remarkable overall accuracy of up to 85% across a set of 20 underwater videos.

Subsequently, to mitigate the effects of environmental changes, Sheikh & Shah (2005), integrated color information into their background pixel modeling in images by using Kernel Descriptors within Kernel Density Estimation (KDE). Meanwhile, Yao & Odobez (2007), introduced a background modeling method based on texture-specific features calculated using local binary patterns. Despite employing these traditional machine learning algorithms and image processing techniques, accurately capturing the complex characteristics unique to fish in highly dynamic and diverse environments remains challenging. Consequently, Siddiqui et al. (2018), have noted that fish detection methods relying on video or image data often fall short in real-world settings.

Recently, researchers such as Petrellis et al. (2023) and Salman et al. (2020), have demonstrated that deep learning approaches can achieve outstanding accuracy in detecting and classifying fish in unrestricted underwater environments, achieving accuracies of approximately 95% and 87.44%, respectively. In the domain of deep learning, the Convolutional Neural Network (CNN) is a specialized algorithm designed for tasks like image recognition and the processing of pixel data. It can extract features and offer a more sophisticated approach to addressing challenges associated with object detection. Currently, two primary classifications that exist for deep learning approaches in object detection. One is the two-stage object detectors

exemplified by architectures like Region-based Convolutional Neural Network (RCNN), Fast R-CNN, and Faster R-CNN. These types of algorithms typically involve two steps. The initial step involves employing a selective search or Region Proposal Net (RPN) to produce potential target regions, followed by conducting classification and regression on the proposed regions. Although these detectors boast high accuracy, they also exhibit slower detection speeds. Another algorithm is the one-stage object detectors exemplified by RetinaNet, Single Shot Multibox Detector (SSD), and You Only Look Once (YOLO). The one-stage algorithms employ a single network to predict object bounding boxes and class probability scores directly from images. While they are less precise compared to the two-stage detectors, they exhibit faster detection speeds and are generally easier to train and implement.

As indicated by Redmon et al. (2016), the YOLO algorithm introduces a novel approach to target detection by conceptualizing detection as a regression problem. By framing detection as a regression problem, the YOLO algorithm obviates the necessity for a complex pipeline. The YOLO algorithm utilizes a straightforward CNN architecture to directly manage regression for target detection, predicting both the position of the bounding box and the class of the candidate box. It has also been applied for detecting fish in underwater settings, yielding promising outcomes in terms of both accuracy and speed. Over time, the YOLO architecture has undergone evolution, with each iteration introducing new features and enhancements.

As shown in Figure 2, the YOLO model comprises three primary components: the backbone, neck, and head, which form its fundamental architecture. The backbone's primary role involves feature extraction from input images, commonly achieved through the utilization of a CNN. The neck layer improves feature maps through the integration of varied scales, while the head layer utilizes these enhanced feature maps to generate predictions, encompassing bounding box coordinates and class probabilities. The backbone component is pivotal in determining the overall efficacy of the object detection model. Various iterations of YOLO have implemented diverse backbone architectures, including Darknet-19, Darknet-53, CSPDarknet53, EfficientNet-B3, and more. Over time, these backbone architectures have undergone evolution aimed at enhancing the efficiency and performance of YOLO models.

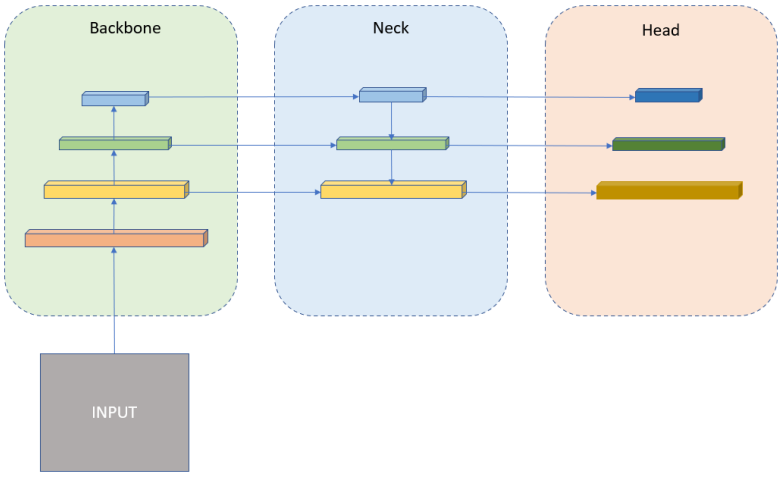


FIGURE 2. General overview of YOLO architecture

YOLOv7, proposed by Wang et al. (2023), features the Efficient Layer Aggregation Network (ELAN) as its backbone. This architecture is crucial for extracting significant features from the input data. Another enhancement in YOLOv7 is the use of anchor boxes, which consist of a set of pre-defined boxes with varying aspect ratios tailored for detecting objects of different shapes. Notably, some researchers such as Liu et al. (2023) and Yu et al. (2023), have employed YOLOv7 as a base model for underwater object detection, achieving promising results with mAP values of 89.6% and 73.5%, respectively.

Following this, YOLOv8 was introduced by Ultralytics, the creators of the influential YOLOv5 model. This iteration is an anchor-free model, meaning predictions are made directly regarding the center of an object rather than calculating the offset from predetermined anchor boxes. This anchor-free detection reduces the number of box predictions, resulting in faster Non-Maximum Suppression (NMS). Researchers have built upon the YOLOv8 model for underwater object detection by using a combination of the Pascal VOC dataset and a custom dataset, achieving a notable mAP of 86.6% (Liu et al 2023).

The most recent advancement, YOLOv9, developed by an independent open-source team, introduces groundbreaking methods such as Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN) (Wang et al. 2024). YOLOv9 aims to achieve top-tier real-time object detection through innovative strategies that address information loss challenges inherent in deep neural networks. By integrating PGI and the versatile GELAN framework, YOLOv9 enhances the model's learning capabilities while ensuring the preservation of crucial information during the detection process, resulting in outstanding accuracy and performance. Given its recent launch, the application of YOLOv9 for underwater object detection is still somewhat limited.

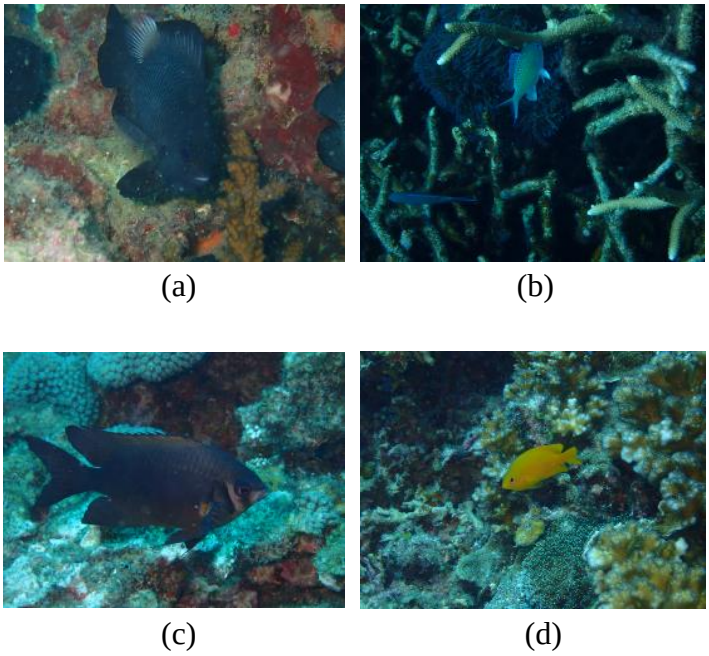
METHODOLOGY

DATASET PREPARATION

To attain reliable parameters and models in deep learning, a substantial volume of data samples is typically required throughout the training phase. The underwater image dataset utilized in this study was sourced from the Institute of Oceanography and Environment (INOS) at the University of Malaysia Terengganu. The footage was captured using the Stereo-Diver Operated Underwater Video System (Stereo-DOVs) and covers the region of Pulau Bidong and its neighboring islands. The operator of the Stereo-DOVs maintained a hovering distance of approximately 0.7 meters above the surface, ensuring a horizontal viewpoint with minimal water visibility during video capture. All footage was subsequently extracted and transformed into image frames using Video Images Master Pro V1.2.8.

For this study, we selected 100 images for training purposes, featuring four distinct fish species belonging to the Pomacentridae family: *Dascyllus trimaculatus*, *Chromis viridis*, *Neoglyphidodon melas*, and *Pomacentrus moluccensis*. These species were chosen based on the findings of Afiq-Firdaus et al. (2023), who identified them as the predominant reef fish

species with some of the highest density values (ind. m^{-3}). To ensure the model's efficiency and accuracy, the images were resized from 4608 × 3456 to fit the optimal input size during the training phase, which in this case was 640 × 640. The optimal input size was chosen based on the standard practices recommended for training YOLO models and empirical testing to balance computational efficiency and model accuracy. It is acknowledged that different input sizes can affect the models' performance, and this size was selected to achieve the best results for our specific dataset and objectives. Sample images from the dataset, showcasing the four fish species, are depicted in Figure 3.



(a) (b)

(c) (d)

FIGURE 3**.** Sample images. (a) *Dascyllus trimaculatus*, (b) *Chromis viridis*, (c) *Neoglyphidodon melas* and (d) *Pomacentrus moluccensis*

After resizing the images, data augmentation methods were employed to expand the training set artificially by generating modified versions of the existing images. In this study, the images underwent rotations at angles of 45, 135, 225, and 315 degrees, along with a horizontal flip. This process resulted in a dataset comprising 1000 images. The dataset was then split into three sets: 80% for the training set, 10% for the testing set, and the remaining 10% for the validation set. Figure 4 showcases examples of the augmented images.

(a) (b) (c)

(d) (e) (f)

FIGURE 4. Sample augmentation of the images. (a) Original image, (b) Horizontal flip, (c) 45-degree rotation, (d) 135-degree rotation, (e) 225-degree rotation, and (f) 315-degree rotation

OBJECT IMAGE ANNOTATION

The purpose of object image annotation is to utilize text or annotation tools to label or classify an image, highlighting the data features that a machine learning model needs to identify. In the context of fish detection, it is essential to acquire both the fish species information and the fish border location information for the image. For this study, we selected LabelImg, a Python-based software tool, for its compatibility with the YOLO labeling format and its intuitive graphical interface to annotate the dataset images. Figure 5 illustrates an instance of the image annotation process.

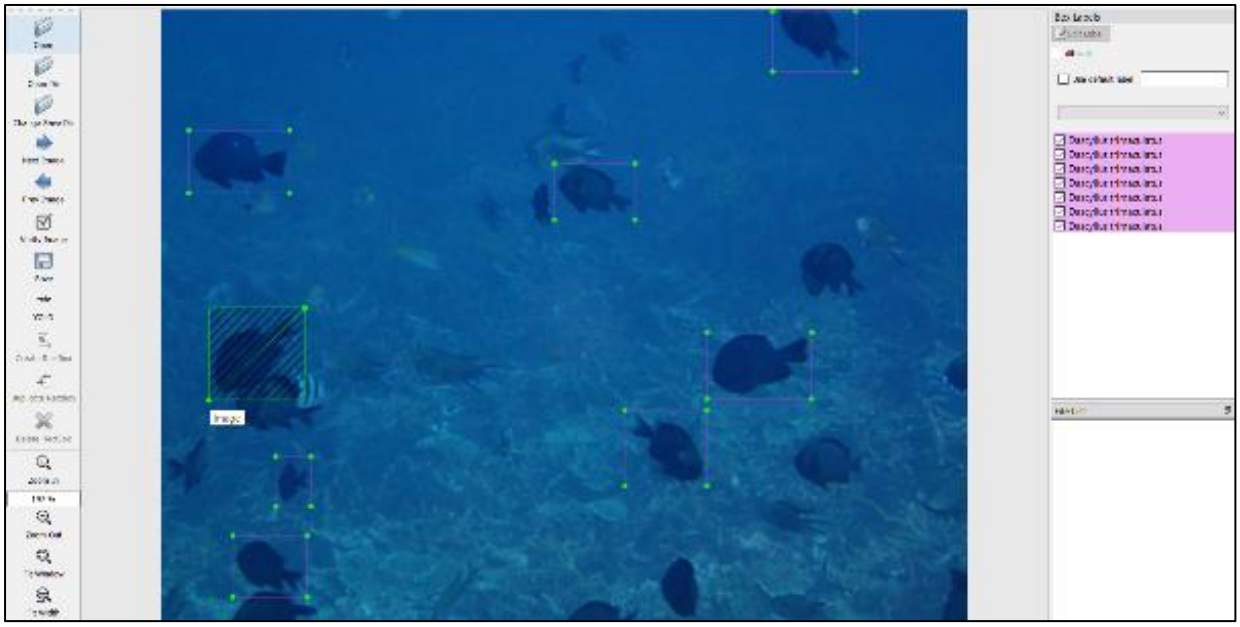


FIGURE 5. LabelImg interface during the labeling process

After finishing the labeling process, the relevant image data is systematically stored in associated XML files. The XML files contain extensive data necessary for network training, encompassing vital information such as the object class and its exact spatial coordinates within the images.

TRAINING MODELS

In this study, we opted to use Google Colab to train all the YOLO models due to its provision of a free GPU, which significantly speeds up the training time and enhances the models' accuracy. The training set was used to train the models, while the validation set served to monitor the models' performance during training and to check for overfitting. After completing the training process, we assessed the models' ability to generalize to new, unseen data by evaluating them using the test set. Each YOLO model was trained for 100 epochs with a batch size of 8. The images were resized to an input size of 640 × 640 to ensure compatibility with all models.

EVALUATION METRICS

To assess the performance of the YOLO models in detecting underwater coral reef fish, we employed a range of standard metrics commonly used to evaluate such models. These include precision, recall, F1 score (F1), and mean average precision (mAP). The equations for computing these evaluation metrics are provided as Equations (1), (2), (3), and (4) below:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1\ score = 2 * \frac{(precision * recall)}{precision + recall} \tag{3}$$

$$Mean\ average\ precision\ (mAP) = \frac{1}{n}\sum_{k=1}^{k=n} AP\ k \tag{4}$$

where:
TP signifies the true positive.
FP signifies the false positive.
FN signifies the false negative.
AP k is the average precision of class k.
N is the number of classes.
In object detection, precision indicates the ratio of accurately identified objects to all detected objects, while recall signifies the proportion of correctly identified objects to all objects within the sample set. The F1 score is a machine learning evaluation metric that measures a model's accuracy by combining precision and recall through a weighted harmonic average. Average precision (AP) serves as a metric for assessing the performance of object detection models, computing the precision for all elements associated with a specific class or fish species. Conversely, the mAP is determined as the numerical average of the aggregated AP values across all species, providing an evaluation of the model's overall performance.

EXPERIMENTAL RESULTS

Upon completing the training phase, we implemented the YOLO models to detect fish within the test dataset. More specifically, our focus was on detecting four coral reef fish species as mentioned in the preceding section. The total training durations for the YOLOv7, YOLOv8, and YOLOv9 are approximately 1.793, 1.385, and 2.708 hours, respectively. Table 1 provides a summary of the evaluation metrics results of the three YOLO models. The evaluation metrics demonstrate that all YOLO models yield outstanding performance, effectively detecting and classifying fish species within the test dataset.

TABLE 1**.** Overall evaluation metrics of trained YOLO models

YOLO Model	Precision (%)	Recall (%)	F1 Score (%)	mAP@0.5 (%)
YOLOv7	96.0	93.1	94.5	96.6
YOLOv8	95.7	93.4	94.5	97.9
YOLOv9	93.7	89.5	91.6	94.3

YOLOv8 outperforms both YOLOv7 and YOLOv9 in mAP@0.5, achieving a 1.3% and 3.6% higher score, respectively. The mAP@0.5 denotes the mAP computed at an Intersection over Union (IoU) threshold of 50%. IoU measures the overlap between the predicted and ground truth bounding boxes, distinguishing between true positives and false positives in detections. YOLOv8 stands out for its rapid training time, finishing in approximately 1.358 hours. The swift training process and impressive mAP@0.5 score make YOLOv8 an attractive choice as the base model for our future study.

Figure 6 below illustrates the progression of this metric throughout the training of all three YOLO models over a span of 100 epochs. YOLOv7 and YOLOv8 demonstrate rapid convergence, maintaining stability throughout the epochs without notable performance degradation, indicating consistent performance across both models. In contrast, YOLOv9 exhibits a slower convergence rate but it maintains its performance without any significant degradation. Although YOLOv8 ultimately achieved the highest value at the end of training, YOLOv9's slower convergence suggests that with additional epochs, it could potentially surpass both YOLOv7 and YOLOv8.

Figure 7 below illustrates the effectiveness of the three YOLO models in detecting the four fish species within images from the test set. The test set images are utilized to assess the models' performance in generalizing to new, unseen data. In the first row, it shows that both YOLOv7 and YOLOv8 accurately identify the fish species and precisely place the bounding boxes. YOLOv9, on the other hand, shows a couple of false positive detections where it mistakenly identifies unknown objects as one of the classes it was trained on.

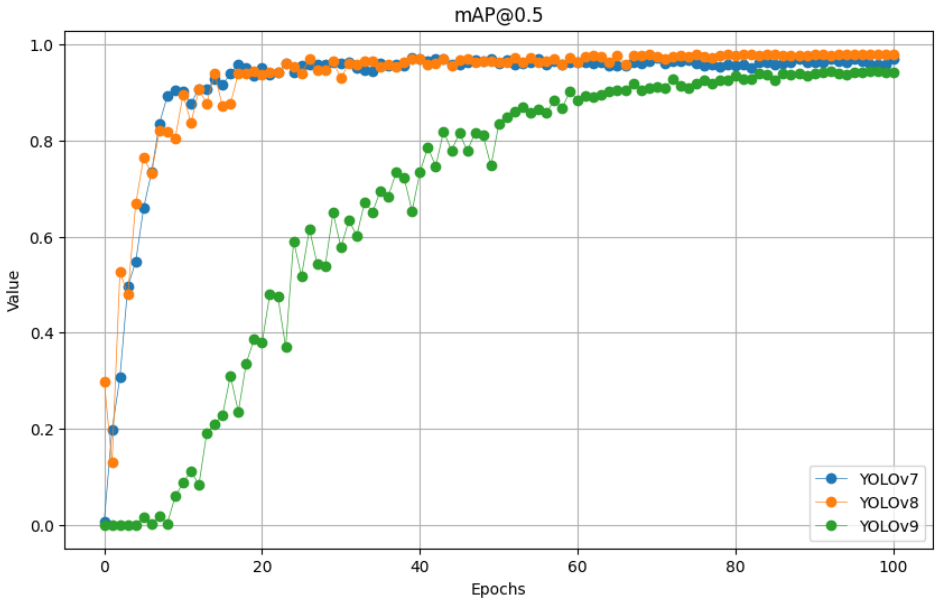
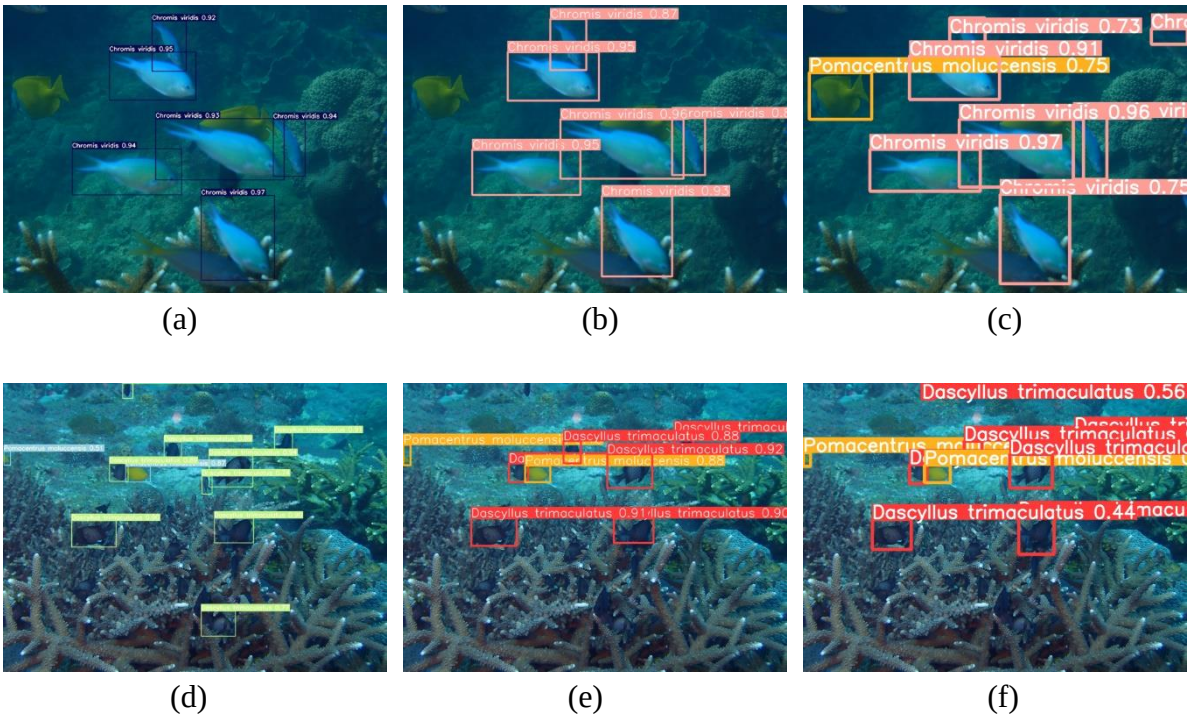


FIGURE 6. Graph of the mAP@0.5 of the three YOLO models throughout 100 epochs during the training phase



(a) (b) (c)

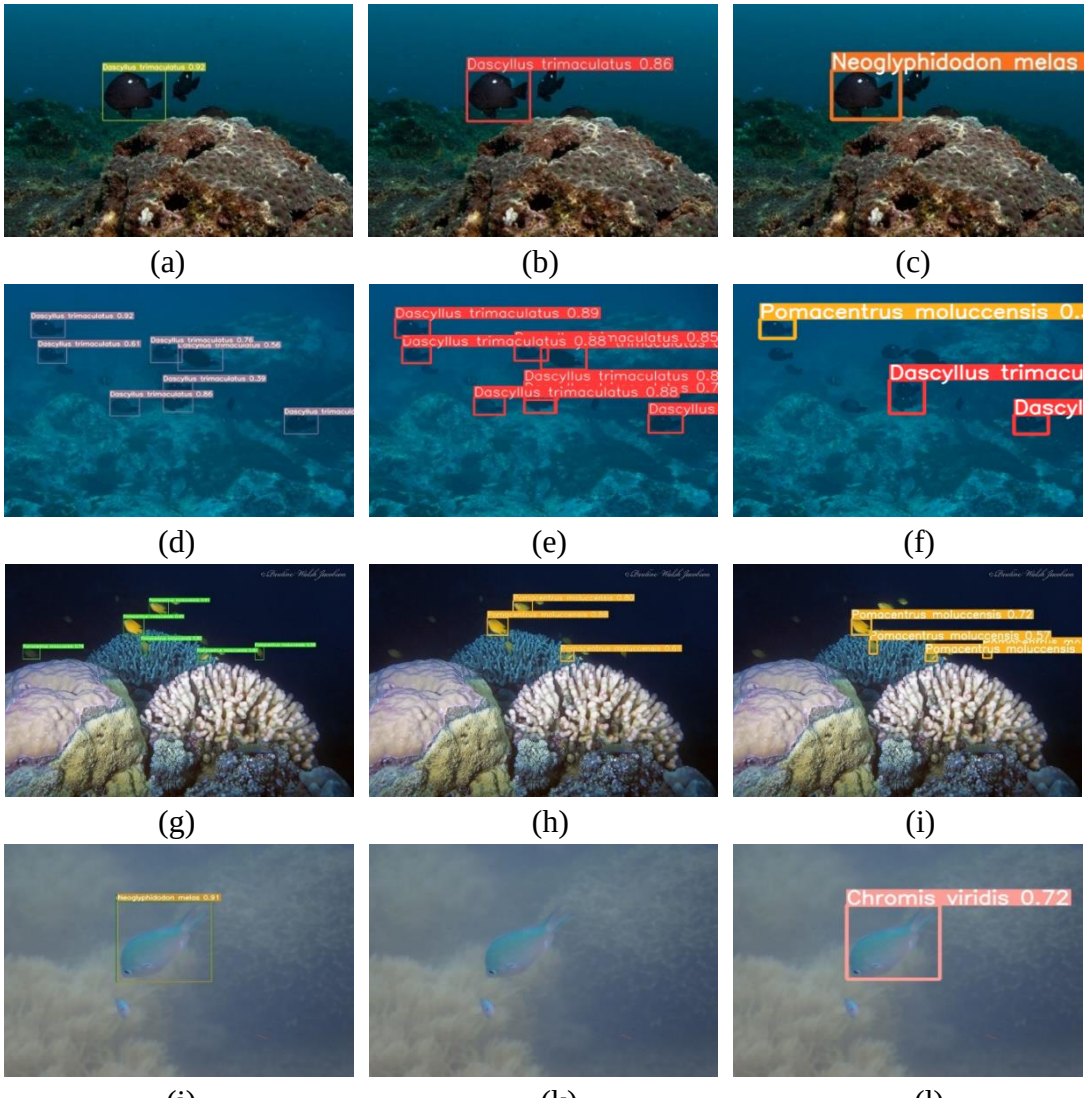
(d) (e) (f)

FIGURE 7. Detection results of the YOLO models on the test set. First column: (a) and (d) YOLOv7. Second column: (b) and (e) YOLOv8. Third column: (c) and (f) YOLOv9

Moving on to the images detected in the second row, we examined those containing multiple fish species. While it is clear that all the YOLO models successfully detect and identify the majority of fish species present in the images, there are a few instances of detection failure. This might be due to insufficient diversity within the training image dataset. Significantly, the YOLOv7 model excels in detecting the majority of fish instances within these images. This is

noteworthy considering that during training and validation, YOLOv7 demonstrated lower recall levels and mAP@0.5 values as compared to YOLOv8.

Previously, we examined images from the test set that might display similarities in terms of underwater environment with those used in training and validation. To further the assessment of the models, we conducted supplementary tests utilizing freely available images from the internet, as illustrated in Figure 8. These images were chosen from the Global Biodiversity Information Facility (GBIF) database and were sourced from the citizen science platform, known as iNaturalist. These underwater images constitute a completely new dataset for the models, featuring diverse lighting conditions, varying distances, and other unique underwater settings.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

FIGURE 8. Detection results on images from the GBIF database. First column: (a), (d), (g) and (j) YOLOv6. Second column: (b), (e), (h) and (k) YOLOv7. Third column: (c), (f), (i) and (l) YOLOv8

In the first row, it is evident that none of the three YOLO models successfully detected every instance of *Dascyllus trimaculatus* within the image. Both YOLOv7 and YOLOv8 precisely detected just a single instance within the images with high confidence levels. The most notable discovery is with YOLOv9, which incorrectly classified the detected fish as another class. This misclassification could be attributed to YOLOv9's constraints in distinguishing objects with similar appearances, particularly when the training data is insufficient or when object features lack distinctiveness.

In the second row, which showcases numerous instances of *Dascyllus trimaculatus* amid significantly hazier underwater conditions, YOLOv7 and YOLOv8 accurately detected seven and eight instances, respectively. However, YOLOv7 detected two fish within a single bounding box. In contrast, YOLOv9 correctly identified only two instances but also made an additional misclassification.

The third row displays multiple instances of *Pomacentrus moluccensis* under considerably dimmer underwater lighting conditions. Here, YOLOv7 exhibited the highest performance, detecting six fish. YOLOv9 successfully detected four fish, whereas YOLOv8 detected only three. Although YOLOv8 achieved the highest mean Average Precision (mAP) value, this result highlights its limitations in detecting smaller objects or objects under diverse lighting conditions that were not adequately represented in the training dataset.

Finally, in the fourth row, which contains a couple of *Chromis viridis* fish in a slightly hazier underwater environment, all YOLO models fail to accurately detect and identify all instances of fish. YOLOv7 makes a false positive detection by misidentifying the fish as another species within the trained class, while YOLOv8 makes a false negative by not detecting any fish in the image. Remarkably, the YOLOv9 model successfully detects one fish present in the image. This observation is significant, especially given that YOLOv9 exhibited lower recall and mAP@0.5 values during training and validation compared to the other two models.

CONCLUSION

This study comprehensively assessed YOLO models (YOLOv7, YOLOv8, and YOLOv9) for detecting coral reef fish underwater, with a particular emphasis on four dominant fish species. All the YOLO models were trained for 100 epochs, and the assessment of various performance metrics showed a balanced performance across all models, particularly highlighting YOLOv7 and YOLOv8. YOLOv7 demonstrated superior precision, whereas YOLOv8 showcased outstanding recall. Both YOLOv7 and YOLOv8 attained an impressive F1 score of 94.5%. YOLOv8 achieved the highest mAP@0.5 value at 97.9%. Even though YOLOv9 is the newest version of the YOLO family, it demonstrated multiple occurrences of false positives, indicating misclassification of species multiple times. The visual results demonstrated effective detection of relevant instances, although certain limitations emerged, such as the smaller size of fish, and challenging lighting underwater conditions.

Considering the evaluation metrics and visual results, YOLOv7 and YOLOv8 emerge as appealing options as the foundational models for our forthcoming research. Future research should incorporate more extensive image datasets and enhanced variability in image data to improve the model's generalization capabilities. We also intend to explore various optimization strategies aimed at enhancing the performance of the chosen YOLO models. This could entail fine-tuning hyperparameters or altering the YOLO model's architecture by adjusting parameters such as the number of layers, feature maps, or filters.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the Ministry of Higher Education Malaysia for providing funding through the Fundamental Research Grant Scheme (FRGS) under grant number FRGS/1/2020/ICT02/UMT/02/1 (Vote No. 59621) for this research work.

REFERENCES

Adey, Walter H. "Coral reef ecosystems and human health: Biodiversity counts!" *Ecosystem health* 6, No. 4 (2000): 227-236. https://doi.org/10.1046/j.1526-0992.2000.006004227.x.

Afiq-Firdaus, Aminudin Muhammad, Che Din Mohd Safuan, Suhaidi Shafie, Lila Iznita Izhar, Ezmahamrul Afreen Awalludin, Muhammad Faiz Ahmad, Nur Arbaeen Mohd Johari, and Zainudin Bachok. "Current Status of Coral Reef Fish Abundances at Pulau Bidong and Nearby Islands, South China Sea Using Stereo-Diver Operated Video System." *Ocean Science Journal* 58, no. 2 (2023): 16. https://doi.org/10.1007/s12601-023-00110-5.

Awan, Khalid Mahmood, Peer Azmat Shah, Khalid Iqbal, Saira Gillani, Waqas Ahmad, and Yunyoung Nam. "Underwater wireless sensor networks: A review of recent issues and challenges." *Wireless Communications and Mobile Computing* 2019, no. 1 (2019): 6470359. https://doi.org/10.1155/2019/6470359.

Bruckner, Andrew W. "Life-saving products from coral reefs." *Issues in Science and Technology* 18, no. 3 (2002): 39-44.

Cesar, Herman, Lauretta Burke, and Lida Pet-Soede. "The economics of worldwide coral reef degradation." (2003).

Costanza, Robert, Ralph d'Arge, Rudolf De Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg et al. "The value of the world's ecosystem services and natural capital." *nature* 387, no. 6630 (1997): 253-260. https://doi.org/10.1038/387253a0.

Eddy, Tyler D., Vicky WY Lam, Gabriel Reygondeau, Andrés M. Cisneros-Montemayor, Krista Greer, Maria Lourdes D. Palomares, John F. Bruno, Yoshitaka Ota, and William WL Cheung. "Global decline in capacity of coral reefs to provide ecosystem services." *One Earth* 4, no. 9 (2021): 1278-1285. https://doi.org/10.1016/j.oneear.2021.08.016.

GBIF. "GBIF Home Page." 2024. https://www.gbif.org/

Han, Fenglei, Jingzheng Yao, Haitao Zhu, and Chunhui Wang. "Marine organism detection and classification from underwater vision based on the deep CNN method."

Mathematical Problems in Engineering 2020, no. 1 (2020): 3937580. https://doi.org/10.1155/2020/3937580.

Huang, Hai, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi, and Ai-Yun Zang. "Faster R-CNN for marine organisms detection and recognition using data augmentation." *Neurocomputing* 337 (2019): 372-384. https://doi.org/10.1016/j.neucom.2019.01.084.

Kong, Shihan, Xi Fang, Xingyu Chen, Zhengxing Wu, and Junzhi Yu. "A real-time underwater robotic visual tracking strategy based on image restoration and kernelized correlation filters." In *2018 Chinese Control and Decision Conference (CCDC)*, pp. 6436-6441. IEEE, 2018. https://doi.org/10.1109/CCDC.2018.8408261.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444. https://doi.org/10.1038/nature14539.

Li, Daoliang, Yinfeng Hao, and Yanqing Duan. "Nonintrusive methods for biomass estimation in aquaculture with emphasis on fish: a review." *Reviews in Aquaculture* 12, no. 3 (2020): 1390-1411. https://doi.org/10.1111/raq.12388.

Li, Zheng, Yongcheng Wang, Ning Zhang, Yuxi Zhang, Zhikang Zhao, Dongdong Xu, Guangli Ben, and Yunxiao Gao. "Deep learning-based object detection techniques for remote sensing images: A survey." *Remote Sensing* 14, no. 10 (2022): 2385. https://doi.org/10.3390/rs14102385.

Liu, Kaiyue, Qi Sun, Daming Sun, Lin Peng, Mengduo Yang, and Nizhuan Wang. "Underwater target detection based on improved YOLOv7." *Journal of Marine Science and Engineering* 11, no. 3 (2023): 677. https://doi.org/10.3390/jmse11030677.

Liu, Qiang, Wei Huang, Xiaoqiu Duan, Jianghao Wei, Tao Hu, Jie Yu, and Jiahuan Huang. "DSW-YOLOv8n: A new underwater target detection algorithm based on improved YOLOv8n." *Electronics* 12, no. 18 (2023): 3892. https://doi.org/10.3390/electronics12183892.

Marshall, Justin. "Vision and lack of vision in the ocean." *Current biology* 27, no. 11 (2017): R494-R502. https://doi.org/10.1016/j.cub.2017.03.012

Moberg, Fredrik, and Carl Folke. "Ecological goods and services of coral reef ecosystems." *Ecological economics* 29, no. 2 (1999): 215-233. https://doi.org/10.1016/S0921-8009(99)00009-9.

Moniruzzaman, Md, Syed Mohammed Shamsul Islam, Paul Lavery, and Mohammed Bennamoun. "Faster R-CNN based deep learning for seagrass detection from underwater digital images." In *2019 digital image computing: techniques and applications (DICTA)*, pp. 1-7. IEEE, 2019. https://doi.org/10.1109/DICTA47822.2019.8946048.

Nair, Rashmi S., Rohit Agrawal, S. Domnic, and Anil Kumar. "Image mining applications for underwater environment management-A review and research agenda." *International Journal of Information Management Data Insights* 1, no. 2 (2021): 100023. https://doi.org/10.1016/j.jjimei.2021.100023.

Pagire, Vrushali, and Anuradha Phadke. "Underwater fish detection and classification using deep learning." In *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, pp. 1-4. IEEE, 2022. https://doi.org/10.1109/ICICCSP53532.2022.9862410.

Petrellis, Nikos, Georgios Keramidas, Christos P. Antonopoulos, and Nikolaos Voros. "Fish monitoring from low-contrast underwater images." *Electronics* 12, no. 15 (2023): 3338. https://doi.org/10.3390/electronics12153338.

Redmon, J. "You only look once: Unified, real-time object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. https://doi.org/10.1109/CVPR.2016.91.

Rizzini, Dario Lodi, Fabjan Kallasi, Fabio Oleari, and Stefano Caselli. "Investigation of vision-based underwater object detection with multiple datasets." *International Journal of Advanced Robotic Systems* 12, no. 6 (2015): 77. https://doi.org/10.5772/60526.

Roboflow. "What is YOLOv8? The Ultimate Guide." 2023. https://blog.roboflow.com/whats-new-in-yolov8/#what-is-yolov8/.

Salman, Ahmad, Shoaib Ahmad Siddiqui, Faisal Shafait, Ajmal Mian, Mark R. Shortis, Khawar Khurshid, Adrian Ulges, and Ulrich Schwanecke. "Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system." *ICES Journal of Marine Science* 77, no. 4 (2020): 1295-1307. https://doi.org/10.1093/icesjms/fsz025.

Sheikh, Yaser, and Mubarak Shah. "Bayesian modeling of dynamic scenes for object detection." *IEEE transactions on pattern analysis and machine intelligence* 27, no. 11 (2005): 1778-1792. https://doi.org/10.1109/TPAMI.2005.213.

Shen, Ying, Chuanjiang Zhao, Yu Liu, Shu Wang, and Feng Huang. "Underwater optical imaging: Key technologies and applications review." *IEEE Access* 9 (2021): 85500-85514. https://doi.org/10.1109/ACCESS.2021.3086820.

Siddiqui, Shoaib Ahmed, Ahmad Salman, Muhammad Imran Malik, Faisal Shafait, Ajmal Mian, Mark R. Shortis, and Euan S. Harvey. "Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data." *ICES Journal of Marine Science* 75, no. 1 (2018): 374-389. https://doi.org/10.1093/icesjms/fsx109.

Spampinato, Concetto, Yun-Heh Chen-Burger, Gayathri Nadarajan, and Robert B. Fisher. "Detecting, tracking and counting fish in low quality unconstrained underwater videos." In *International Conference on Computer Vision Theory and Applications*, vol. 2, pp. 514-519. SciTePress, 2008.

Storbeck, Frank, and Berent Daan. "Fish species recognition using computer vision and a neural network." *Fisheries Research* 51, no. 1 (2001): 11-15. https://doi.org/10.1016/S0165-7836(00)00254-X.

Strachan, N. J. C. "Recognition of fish species by colour and shape." *Image and vision computing* 11, no. 1 (1993): 2-10. https://doi.org/10.1016/0262-8856(93)90027-E

Teh, Louise SL, Lydia CL Teh, and U. Rashid Sumaila. "A global estimate of the number of coral reef fishers." *PLoS One* 8, no. 6 (2013): e65397.https://doi.org/10.1371/journal.pone.0065397.

Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7464-7475. 2023. https://doi.org/10.1109/CVPR52729.2023.00721.

Wang, Chien-Yao, I-Hau Yeh, and Hong-Yuan Mark Liao. "Yolov9: Learning what you want

to learn using programmable gradient information." *arXiv preprint arXiv:2402.13616* (2024).

Weinstein, Ben G. "A computer vision for animal ecology." *Journal of Animal Ecology* 87, no. 3 (2018): 533-545. https://doi.org/10.1111/1365-2656.12780.

Xu, Wenwei, and Shari Matzner. "Underwater fish detection using deep learning for water power applications." In *2018 International conference on computational science and computational intelligence (CSCI)*, pp. 313-318. IEEE, 2018. https://doi.org/10.1109/CSCI46756.2018.00067.

Yao, Jian, and Jean-Marc Odobez. "Multi-layer background subtraction based on color and texture." In *2007 IEEE conference on computer vision and pattern recognition*, pp. 1-8. IEEE, 2007. https://doi.org/10.1109/CVPR.2007.383497.

Yeh, Chia-Hung, Chu-Han Lin, Li-Wei Kang, Chih-Hsiang Huang, Min-Hui Lin, Chuan-Yu Chang, and Chua-Chin Wang. "Lightweight deep neural network for joint learning of underwater object detection and color conversion." *IEEE Transactions on Neural Networks and Learning Systems* 33, no. 11 (2021): 6129-6143. https://doi.org/10.1109/TNNLS.2021.3072414.

Yu, Guoyan, Ruilin Cai, Jinping Su, Mingxin Hou, and Ruoling Deng. "U-YOLOv7: a network or underwater organism detection." *Ecological Informatics* 75 (2023): 102108. https://doi.org/10.1016/j.ecoinf.2023.102108.

Zhao, Minghao, Chengquan Hu, Fenglin Wei, Kai Wang, Chong Wang, and Yu Jiang. "Real-time underwater image recognition with FPGA embedded system for convolutional neural network." *Sensors* 19, no. 2 (2019): 350. https://doi.org/10.3390/s19020350.