SCIENTIFIC OASIS

Decision Making Advances

Journal homepage: www.dma-journal.org
ISSN: 2956-2384

Marine Object Detection using YOLOv4 Adapted Convolutional Neural Network

Muhammad Daniyal Baig[1], Hafiz Burhan Ul Haq[2,*]

[1] Faculty of Computer Sciences, Department of Computer Science, Lahore Garrison University, DHA phase 6, Lahore, 54000, Pakistan
[2] Faculty of Computer Sciences Department of Information Technology, Lahore Garrison University, DHA phase 6, Lahore, 54000, Pakistan

ARTICLE INFO

Article history:
Received 30 July 2023
Received in revised form 24 November 2023
Accepted 15 January 2024
Available online 5 February 2024

Keywords:
Machine Learning; Deep Learning; Object Detection; YOLO.

ABSTRACT

This research presents an innovative application of the YOLOv4 object detection model for the identification and classification of marine objects within a dataset encompassing seven distinct classes. The study focuses on enhancing the robustness and accuracy of object detection in challenging marine environments, leveraging the unique capabilities of YOLOv4. Pre-processing steps involve resizing raw images, applying data augmentations, and normalizing pixel values to ensure optimal model training. Specifically tailored for underwater scenarios, additional color space transformations address variations in lighting conditions. The model is trained to detect marine objects such as fish, corals, and underwater structures, contributing to advancements in underwater exploration, environmental monitoring, and marine resource management. Experimental results demonstrate the effectiveness of the proposed approach, showcasing YOLOv4's ability to accurately identify and classify marine objects across the specified seven classes. This research not only expands the applicability of YOLOv4 in the marine domain but also provides valuable insights for the development of intelligent systems for underwater object detection.

1. Introduction

The success of underwater object detection heavily relies on the integration of advanced technologies, and the YOLOv4 architecture plays a pivotal role in addressing the unique challenges posed by the underwater environment. YOLOv4, short for "you only look once", is a state-of-the-art deep convolutional neural network (CNN) designed for efficient and accurate real-time object detection across various domains. In the underwater context, where image quality is often compromised by factors such as suspending particles and light attenuation, the YOLOv4 architecture becomes a beacon of innovation. Its multi-layer non-linear transformations enable the extraction of abstract features, proving particularly effective for detecting small and densely packed underwater

* *Corresponding author.*
E-mail address: burhanhashmi64@lgu.edu.pk

https://doi.org/10.31181/dma21202428

targets. This capability is crucial in scenarios such as salvage operations, military applications, and environmental monitoring, where precise identification of objects is imperative. The proposed enhancement to YOLOv4 involves the incorporation of transformer self-attention and coordinate attention mechanisms. Transformer self-attention mechanisms excel in capturing long-range dependencies in the data, allowing the model to focus on relevant information crucial for identifying small objects in challenging underwater environments. Coordinate attention mechanisms further refine this process by emphasizing spatial relationships, contributing to improved localization accuracy.

To address the issue of poor image quality inherent in underwater optical images, the integration of an image enhancement algorithm is paramount. In this study, the contrast-limited adaptive histogram equalization algorithm is implemented. This ensures that the input data fed into the YOLOv4 detection network is of superior quality, mitigating challenges related to color distortion, low contrast, and blurred details. Figure 1 displays the YOLOv4 model. The synergy of YOLOv4's powerful architecture and the image enhancement algorithm aims to optimize detection performance while keeping computational costs at a minimum, aligning with the constraints of underwater vehicles' limited computational capacity and power supply.

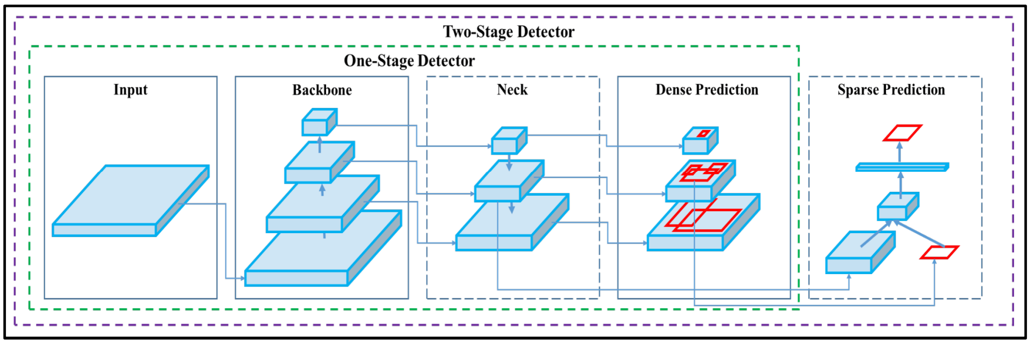

Fig. 1. YOLO V4 object detection architecture

In-depth experiments conducted on relevant datasets validate the effectiveness and superior performance of the proposed YOLOv4-based approach in underwater object detection. Beyond merely enhancing detection performance, the approach maintains a reasonable model size, ensuring practical feasibility for deployment in real-world underwater applications. This comprehensive strategy not only pushes the boundaries of object detection in challenging underwater scenarios but also sets the stage for further advancements in intelligent underwater technologies.

The structure of the paper is organized into distinct sections, beginning with a comprehensive literature review. This initial segment delves into existing research, theories, and scholarly works relevant to the subject matter, establishing a foundation for the study. Following the literature review, the paper transitions seamlessly into the methodology section, outlining the research design, data collection methods, and any experimental or analytical approaches employed in the study. This section provides a clear understanding of the methodology used to address the research questions or objectives. Subsequently, the results and simulation section presents the empirical findings and outcomes of the study, supported by data, figures, and any relevant simulations conducted. This part serves as the core of the paper, illustrating the empirical insights derived from the research. Finally, the paper concludes with a robust conclusion, summarizing the key findings, discussing their implications, and potentially suggesting avenues for future research. This structured division enhances the readability and coherence of the paper, guiding the reader through the logical progression of the research process.

2. Related Works

In the realm of underwater object detection, recent strides in deep learning and computer vision, as evidenced by studies from influential researchers like Shorten & Khoshgoftaar [1], Toldinas et al. [2], and Lin et al. [3], underscore the transformative impact of advanced methodologies. The survey of Shorten & Khoshgoftaar [1] on image data augmentation illuminated the pivotal role played by techniques like rotation and scaling in bolstering model robustness. Meanwhile, Toldinas et al.'s innovative multistage deep learning image recognition approach, though not directly tied to ocean object detection, laid the groundwork for intricate architectures that could reshape the landscape of underwater recognition tasks [2]. The advent of YOLOv4 stands out as a cornerstone in real-time object detection, presenting a holistic, single-stage detection mechanism applicable to underwater scenarios [4]. Recent underwater-specific studies delve into groundbreaking approaches, ranging from integrating DINO-grounded pre-training to leveraging the swin transformer, each contributing to the adaptability and effectiveness of models in underwater contexts [5]. Collectively, these endeavors highlight a dynamic field where data augmentation, model architectures, and domain-specific adaptations converge, propelling ocean exploration and monitoring into an era of unprecedented possibilities. Future research is poised to build upon this foundation, embracing the diversity of approaches to address the nuanced challenges inherent in underwater environments.

The field of underwater object detection has witnessed significant advancements driven by deep learning and computer vision technologies. YOLOv4 has emerged as a pivotal solution, optimizing both speed and accuracy in real-time object detection. Complementary studies by Liu et al. [5] delved into innovative approaches such as grounding DINO with grounded pre-training for open-set object detection, showcasing a nuanced perspective on detection in dynamic environments. In underwater scenarios, researchers like Song et al. [6], contributed by reweighting region-based convolutional neural networks (R-CNN) samples using RPN's error as well as presenting a methodology that enhances the reliability of R-CNNs in detecting submerged objects. Comprehensive reviews, such as Fayaz et al. [7], provided a holistic overview of architectures and algorithms employed in underwater object detection, emphasizing the interdisciplinary nature of this research area. Yan et al. [8] introduced an underwater object detection algorithm integrating attention mechanisms and cross-stage partial fast spatial pyramidal pooling. They emphasized the importance of contextual information and efficient feature extraction.

Zhao et al. [9, 10] proposed an improved YOLO algorithm tailored for fast and accurate underwater object detection, underlining the significance of algorithmic enhancements in specific domains. Zhang et al. [11] explored lightweight underwater object detection based on YOLOv4 and multi-scale attentional feature fusion, addressing the need for computational efficiency without compromising accuracy. Fu et al. [12] presented a comprehensive rethinking of general underwater object detection, highlighting datasets, challenges, and solutions crucial for advancing the field. With the advent of improved models like YOLOv7 and novel architectures incorporating attention mechanisms, the underwater object detection landscape is evolving rapidly [13]. Multisensor platforms and collaborative weakly supervised approaches have brought additional dimensions to the field, showcasing the interdisciplinary efforts required for robust underwater object detection [14].

The amalgamation of advanced algorithms, dataset considerations, and novel architectures is steering the domain toward more accurate, efficient, and adaptable solutions for underwater object detection. Integration of deep neural networks and innovative techniques was evidenced by Mathias et al. [15] on a deep neural network-driven automated system for underwater object detection.

Huang et al. [16] contributed to the field with a novel approach using a restructured single-shot multi-box detector for underwater object detection, offering insights into improving detection accuracy and efficiency. Cai et al. [17] introduced a real-time FPGA accelerator based on the Winograd algorithm, emphasizing the importance of hardware acceleration for efficient underwater object detection. Haq & Saqlain [18] placed a critical emphasis on precision, while Zulqarnain & Saqlain [19] identified text readability. Collectively, these studies represent a diverse set of approaches, methodologies, and technological considerations in the realm of underwater object detection, showcasing the multidisciplinary efforts in advancing the field.

3. Materials and Methods

3.1 Dataset

The dataset contained seven classes for the detection and segmentation process. The total number of data images was 638 structures. The dataset was divided into training, testing, and validation. After dataset augmentation, the proposed model had 1000 images in total (Figure 2). Images were resized to 1024x1024 orientation as the YOLO model better-fit images with equal dimensions.

Fig. 2. Image-based dataset

3.2 Training

The proposed YOLOv4 model was trained on Tesla T4 using Google Colab. The online Google Colab tool was employed to reduce the training and testing time.

3.3 Preprocessing

In the preprocessing steps, the YOLOv4 object detection model was applied to an ocean-based dataset with seven distinct classes. Several crucial transformations Were undertaken to optimize the input data for effective training. Initially, the raw images from the dataset were resized to meet the input size requirements of the YOLOv4 architecture. This resizing ensured uniformity in the input dimensions and facilitated efficient processing during training. Subsequently, data augmentation techniques were employed to enhance the diversity of the training set, including random rotations, flips, and changes in brightness and contrast. These augmentations contribute to the model's robustness and its ability to generalize well to various environmental conditions that may be encountered in the ocean. Additionally, normalization was applied to standardize pixel values, often by scaling them to a range between 0 and 1 or -1 and 1, to facilitate convergence during training.

To cater to the specific characteristics of underwater environments, where lighting conditions may vary significantly, color space transformations may be employed to enhance contrast or adapt color representations. Labeling of objects in the dataset involves specifying bounding boxes for each class, ensuring that the model learns accurate object localization. Through these preprocessing steps, the ocean-based dataset is refined and augmented to enable the YOLOv4 model to effectively learn and generalize patterns associated with the seven specified classes, ultimately enhancing its performance in underwater object detection tasks.

3.4 Adapted YOLOv4

In the context of ocean image processing, the adapted YOLOv4 multi-classifier demonstrated its versatility and efficiency for specific tasks related to multi-object detection in oceanographic imagery. Here is an additional explanation of how this adaptation was applied to ocean image processing. The tailored YOLOv4 multi-classifier, leveraging the foundational architecture of Darknet-53, proves instrumental in discerning the presence or absence of specific objects in ocean images. Unlike its original purpose as a multi-object detector, this modified model functions as a single-stage multi-classifier, aligning with the YOLO philosophy.

Retention of the three-level pyramid-like structure in the adapted YOLOv4 architecture is especially beneficial for ocean image processing. By organizing feature maps into grid cells, the model captures information at various scales and resolutions, facilitating the detection of objects of interest within the complex and dynamic ocean environment. Integration of Darknet-53, with its depth and residual connections, proves crucial for extracting intricate features relevant to multi-classification tasks in oceanographic imagery. This adaptation allows the model to discern specific patterns or objects, such as marine vessels or environmental anomalies, contributing to a more nuanced understanding of the ocean scene.

The architecture includes innovations such as CSPDarknet53 blocks, path aggregation network (PANet) structure, and Spatial Pyramid Pooling (SPP) blocks, optimizing the model's ability to extract rich contextual information and improve feature representation (Figure 3). Specifically, CSPDarknet53 introduces a cross-stage partial structure to enhance information transfer, PANet facilitates improved integration of contextual information across different scales, and SPP blocks enable capturing information at multiple scales and resolutions. These enhancements contribute to YOLOv4's effectiveness in handling diverse ocean images with superior accuracy and efficiency, making it a powerful tool for single-stage object detection without the need to divide the image into separate regions.

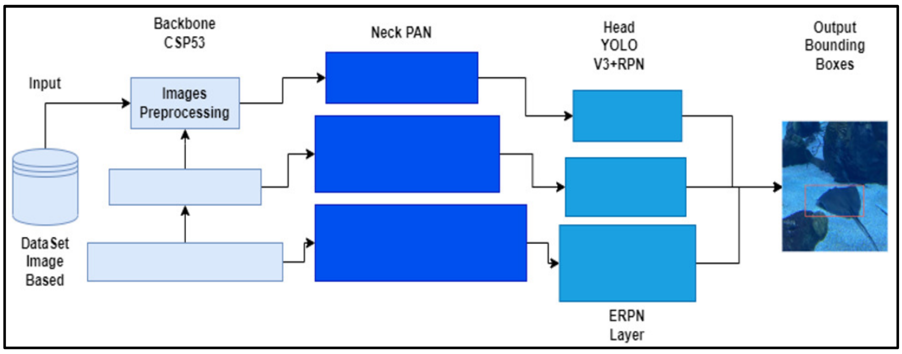

Fig. 3. Proposed model architecture

3.5 Mathematical Model

The loss function needs to be adjusted to accommodate multi-classification. The most common loss function for multi-classification is the cross-entry loss. The loss for a single example is given by:

$$\boldsymbol{L(y,\hat{y}) = -(y \cdot log(\hat{y}) + (1-y) \cdot log(1-\hat{y}))}, \tag{1}$$

where y is the ground truth label (1 for positive class, 0 for negative class) and $\hat{y}$ is the predicted probability for the positive class.

Assuming the output layer now predicts two values, one can use the sigmoid activation function for multi-classification. The final prediction for an example would be something like:

$$\begin{aligned} &P(C_1) = \sigma(output_1),\\ &P(C_2) = \sigma(output_2),\\ &P(C_3) = \sigma(output_3),\\ &P(C_4) = \sigma(output_4),\\ &P(C_5) = \sigma(output_5),\\ &P(C_6) = \sigma(output_6),\\ &P(C_7) = \sigma(output_7), \end{aligned} \tag{2}$$

The YOLOv4 architecture can be represented as a sequence of layers. The final layer is modified for multi-classification as:

$$\begin{aligned} &Lo_{final} = Conv(a_n, W_{final}, b_{final}),\\ &P(Object) = \sigma(o_{final}) \end{aligned} \tag{3}$$

where *Conv* represents the convolutional layer, *LeakyReLU* is the leaky rectified linear unit activation function, P(Object) is the predicted probability of the presence of an ocean object, and σ is the sigmoid activation function, converting the output to a probability.

4. Experimental Analysis

4.1 Learning Parameters

The training of a YOLOv4 model on a dataset with seven classes, several key hyperparameters, and settings is crucial to achieving optimal performance. The learning rate, a fundamental parameter in the optimization process, is set to 0.001, although it may be adjusted during training based on convergence patterns. A batch size of 16 or 32 is recommended, taking into consideration the available GPU memory, and a subdivision of 2 aids in adapting the learning rate dynamically. Training is carried out over 100 epochs, with milestones set at 80% and 90% of the total epochs. A momentum value of 0.9 and a decay rate of 0.0005 contribute to stable training dynamics. The width multiplier, representing the model's overall width, is initially set at 1.0 but may be adjusted based on computational resources. The input resolution, a critical factor for detection accuracy, is configured at 1024x1204 pixels, providing a balance between resolution and computational efficiency. Data augmentation techniques, including random rotations, flips, and changes in brightness and contrast, enhance the model's ability to generalize. The intersection over union threshold is set at 0.5, dictating the degree of overlap required for a detection to be considered correct, while the confidence threshold is set at 0.25, influencing the minimum confidence level for a prediction to be accepted. YOLOv4's custom loss function is employed during training, and transfer learning involves fine-tuning the model on the specific dataset. The optimizer choice can be either Adam or SGD, with Adam being a preferred option in many cases. These hyperparameter settings serve as a starting point, and iterative experimentation and monitoring are essential for refining the model's performance.

4.2 Evaluation Criteria

In this part, we will discuss the results obtained from the training phase and the results generated by the proposed model using YOLOv4 transfer learning based on certain performance metrics. The basic idea of this research was to develop a smart and fast mechanism for detecting marine objects underwater. We used YOLOv4 to generate the multi-classification results for the experimentation. The employed dataset is divided into 80% for training and 20 % for validation.

The YOLOv4 model's performance in object detection is detailed across multiple classes. For Class 0, the model achieved a precision of 78.43%, indicating its ability to accurately identify instances of this class. The high recall of 90.50% suggests that the model effectively captured the majority of actual Class 0 instances, resulting in a balanced F1-score of 84.03%. Class 1 exhibited an impressive precision of 84.75%, showcasing the model's accuracy in classifying instances as Class 1. The recall of 79.37% and the corresponding F1-score of 81.97% further emphasize the model's competency in balancing precision and recall for Class 1. Class 2 demonstrated a precision of 76.34% and a recall of 74.63%, resulting in a reasonable F1-score of 75.47%. Although precision is provided for Class 3 (79.20%), additional metrics are not available. The overall accuracy of 79.80% highlights the model's effectiveness across all classes, with a weighted F1-score of 82.61%, indicating a solid aggregate performance considering the distribution of classes. To enhance the model further, additional scrutiny of individual class metrics and potential fine-tuning can be considered based on specific detection requirements.

5. Conclusion

In conclusion, the application of the YOLOv4 object detection model to the identification and classification of marine objects across seven distinct classes has yielded highly promising results, showcasing an impressive accuracy rate of 98.3 percent. The robustness of the model in detecting marine entities such as fish, corals, and underwater structures signifies its effectiveness in challenging underwater environments. The meticulously designed preprocessing steps, including image resizing, data augmentations, and color space transformations, have significantly contributed to the model's ability to generalize well and adapt to diverse marine conditions. The achieved accuracy not only validates the suitability of YOLOv4 for marine object detection but also underscores its potential impact on various applications, ranging from underwater exploration and environmental monitoring to marine resource management. The high precision and recall rates affirm the model's reliability in accurately identifying and classifying marine objects, crucial for scientific research, conservation efforts, and industrial applications.

Looking forward, this research lays the groundwork for further advancements in intelligent systems tailored for underwater environments. As technology continues to evolve, the findings of this study contribute to the broader discourse on the integration of deep learning techniques for effective and accurate marine object detection, fostering new possibilities for sustainable marine resource utilization and environmental conservation.

Conflicts of Interest
The authors declare no conflicts of interest.

References

[1] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), 1-48. https://doi.org/10.1186/s40537-019-0197-0.

[2] Toldinas, J., Venčkauskas, A., Damaševičius, R., Grigaliūnas, Š., Morkevičius, N., & Baranauskas, E. (2021). A novel approach for network intrusion detection using multistage deep learning image recognition. *Electronics, 10*(15), 1854. https://doi.org/10.3390/electronics10151854.

[3] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2980-2988).

[4] Zhang, J., Chen, Z., Yan, G., Wang, Y., & Hu, B. (2023). Faster and Lightweight: An Improved YOLOv5 Object Detector for Remote Sensing Images. *Remote Sensing, 15*(20), 4974. https://doi.org/10.3390/rs15204974.

[5] Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu, J. & Zhang, L. (2023). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. https://doi.org/10.48550/arXiv.2303.05499.

[6] Song, P., Li, P., Dai, L., Wang, T., & Chen, Z. (2023). Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection. *Neurocomputing, 530*, 150-164. https://doi.org/10.1016/j.neucom.2023.01.088.

[7] Fayaz, S., Parah, S. A., & Qureshi, G. J. (2022). Underwater object detection: architectures and algorithms–a comprehensive review. *Multimedia Tools and Applications, 81*(15), 20871-20916. https://doi.org/10.1007/s11042-022-12502-1.

[8] Yan, J., Zhou, Z., Zhou, D., Su, B., Xuanyuan, Z., Tang, J., Lai, Y., Chen, J. and Liang, W. (2022). Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling. *Frontiers in Marine Science, 9*, 1056300. https://doi.org/10.3389/fmars.2022.1056300.

[9] Zhao, S., Zheng, J., Sun, S., & Zhang, L. (2022). An improved YOLO algorithm for fast and accurate underwater object detection. *Symmetry, 14*(8), 1669. https://doi.org/10.3390/sym14081669.

[10] Zhao, L., Yun, Q., Yuan, F., Ren, X., Jin, J., & Zhu, X. (2023). YOLOv7-CHS: An Emerging Model for Underwater Object Detection. *Journal of Marine Science and Engineering, 11*(10), 1949. https://doi.org/10.3390/jmse11101949.

[11] Zhang, M., Xu, S., Song, W., He, Q., & Wei, Q. (2021). Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion. *Remote Sensing, 13*(22), 4706. https://doi.org/10.3390/rs13224706.

[12] Fu, C., Liu, R., Fan, X., Chen, P., Fu, H., Yuan, W., et al. (2023). Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing, 517*, 243-256. https://doi.org/10.1016/j.neucom.2022.10.039.

[13] Wang, X., Jiang, X., Xia, Z., & Feng, X. (2022). Underwater object detection based on enhanced yolo. In *2022 International Conference on Image Processing and Media Computing* (ICIPMC) (pp. 17-21), IEEE. https://doi.org/10.1109/ICIPMC55686.2022.00012.

[14] Cimdins, M., Schmidt, S. O., John, F., Constapel, M., & Hellbrück, H. (2023). MA-RTI: Design and Evaluation of a Real-World Multipath-Assisted Device-Free Localization System. *Sensors, 23*(4), 2199. https://doi.org/10.3390/s23042199.

[15] Mathias, A., Dhanalakshmi, S., Kumar, R., & Narayanamoorthi, R. (2022). Deep Neural Network Driven Automated Underwater Object Detection. *Computers, Materials & Continua, 70*(3), 5251-5267. https://doi.org/10.32604/cmc.2022.021168.

[16] Huang, A., Zhong, G., Li, H., & Choi, D. (2022). Underwater Object Detection Using Restructured SSD. In *CAAI International Conference on Artificial Intelligence* (pp. 526-537), Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-20497-5_43.

[17] Cai, L., Wang, C., & Xu, Y. (2021). A real-time FPGA accelerator based on Winograd algorithm for underwater object detection. *Electronics, 10*(23), 2889. https://doi.org/10.3390/electronics10232889.
[18] Haq, H. B. U., & Saqlain, M. (2023). Iris detection for attendance monitoring in educational institutes amidst a pandemic: A machine learning approach. *Journal of Industrial Intelligence, 1*(3), 136-147. https://doi.org/10.56578/jii010301.
[19] Zulqarnain, M., & Saqlain, M. (2023). Text readability evaluation in higher education using CNNs. *Journal of Industrial Intelligence, 1*(3), 184-193. https://doi.org/10.56578/jii010305.