UNDERWATER DETECTION OF ANCIENT POTTERY SHERDS USING DEEP LEARNING

Konstantinos Paraskevas, Ioannis Mariolis, Georgios Giouvanis, Georgia Peleka, Georgios Zampokas and Dimitrios Tzovaras

Centre for Research & Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

ABSTRACT

This paper outlines the creation of a machine learning model designed to identify ancient pottery fragments near a submerged shipwreck off Modi Island, Greece. We trained multiple iterations of the YOLOv8 model using a custom dataset comprised of underwater videos taken during diving expeditions at the wreck site. The primary goal of this research is to integrate the resulting object detection system into a remotely operated vehicle (ROV) for automated pottery shard recognition, thereby aiding archaeological excavations. The paper elaborates on the model's development methodology and presents comprehensive experimental and evaluative results. These findings underscore the model's potential to significantly enhance the efficiency and accuracy of underwater archaeological exploration and analysis.

KEYWORDS

Ancient pottery shreds detection, underwater archaeological excavations, machine learning, object detection, remotely operated vehicle (ROV), underwater shipwrecks, YOLOv8 model.

1. INTRODUCTION

Underwater archaeology is a costly, dangerous, and often misappropriated field, making it crucial to develop methodologies that enable accurate and efficient mapping of potential shipwrecks across vast geographical areas. Such an approach would assist archaeologists in prioritizing site selection and planning excavations. Pottery plays a fundamental role in comprehending archaeological contexts, offering insights into production, trade networks, and social interactions. However, pottery characterization and classification remain primarily manual processes, reliant on analog catalogs curated by specialists and stored in archives and libraries.

Recent studies have made progress in this field. In [1], a system for automated detection and mapping of shipwrecks was developed, utilizing open-source topo-bathymetric data and employing machine learning techniques. Specifically, the YOLOv3 architecture was used, achieving promising results with high F1 and precision scores.

In [2] the ArchAIDE project is introduced driving a large collaborative effort aiming to automate the classification of archaeological pottery. It focuses on digitizing printed catalogues and utilizing Convolutional Neural Networks (CNN) for pottery recognition. Although the current system can identify a limited number of pottery types, the expansion of the database with more records and photographs will enable recognition of a wider variety of pottery types. The project also developed an innovative app for tablets and smartphones to support classification and interpretation work during fieldwork and post-excavation analysis. This work resulted in a streamlined pipeline, where pottery shreds were photographed, compared against a trained neural

Dhinaharan Nagamalai (Eds): NLAII, VISCOM, CCSITA, AIDD, BCYIoT -2023
pp. 01-08, 2023. IJCI – 2023 DOI:10.5121/ijci.2023.120601

network, and matched with suggested types from a comparative collection. The relevant information of the identified pottery type was linked to the new sherd and stored in a shareable online database.

In [3], the aim was to develop a reliable classification model for ancient Chinese celadon using EDXRF, machine learning algorithms, and Mahalanobis distance. Among the four machine learning models considered, the Random Forest algorithm was identified as the most suitable for celadon classification. The chemical compositions within the predicted categories aligned with their corresponding composition ranges. Key chemical compositions that influenced category identification were selected as characteristic parameters. Mahalanobis distance was employed to describe the similarity between samples and predicted categories, resulting in a more specific and accurate prediction. The developed celadon classification model successfully predicted categories for samples from various sites, validated by comparison with background information.

The work in [4] presents the development of two complementary machine learning tools for proposing identifications of pottery sherds based on on-site captured images. The first tool utilizes the shape of fracture outlines of sherds, incorporating a novel deep learning architecture that integrates shape information from points along the inner and outer surfaces. The second tool focuses on decorative features, employing standard image recognition architectures. Training these classifiers posed challenges encountered when working with real-world archaeological data, including limited labeled data, imbalanced class instances, and the need to capture distinguishing features of rare classes. To overcome these challenges, synthetically produced virtual potsherds and various data augmentation techniques were employed. Additionally, a novel training loss function was devised to handle under-populated classes and non-homogeneous distribution of discriminative features.

Regarding object detection in underwater environments, recent studies such as [5] and [6] utilized modern Machine Learning models, specifically the YOLOv5 model, to detect defects in fish farming nets. These studies showed promising detection results using Machine Learning models. However, it is evident that the widely varying underwater conditions pose challenges in achieving high-performance results.

The work in [7] presents a survey of various methods for underwater marine object detection, categorizing these methods based on the type of object they detect. They reviewed both deep learning and feature detection architectures, resulting in a comprehensive analysis of the current related work. The studied categories included fish, plankton, and corals. However, the paper notes the absence of research on seagrass detection, which is equally crucial for the oceanic ecosystem. The conclusion drawn is that improved detection and classification results can be achieved by combining color and texture features.

Upon reviewing related work, it becomes evident that while there are some instances of research concerning the classification of archaeological findings and numerous examples of object detection in underwater conditions, there is a limited amount of research that combines the detection of archaeological findings in underwater environments. This paper introduces a proposed system intended for use in a remotely operated vehicle (ROV) as the initial identification system for pottery shards. These discoveries can then be verified as genuine by archaeologists. Given the mobility of the vehicle, the chosen method must be fast and computationally efficient, while maximizing the number of detections to minimize undetected pottery shards. According to the literature, the You Only Look Once (YOLO) model is a very fast and accurate detection model. Given its successful application in underwater environments in several papers such as in [1], [5], and [6], YOLO has been selected for the presented system.

2. PROPOSED SYSTEM

2.1. Dataset

The proposed system for potsherd detection encompasses an object detection model trained on a custom dataset, yielding bounding boxes around the identified pottery shreds. For this study, the YOLOv8 model was selected, necessitating the dataset to conform to the required format. Specifically, YOLOv8 mandates the labels to be stored in text (.txt) files, wherein the object classes are denoted by an index, and the bounding box coordinates and dimensions are specified in a format normalized relative to the image size. Each bounding box comprises the relative x and y coordinates of its center, as well as its width and height.

To construct the custom dataset, a total of 3495 images were extracted from multiple videos, out of which 2531 underwent manual annotation, designating them as instances of the "amphora-frag" class. Among these annotated images, 2186 were allocated for training, 229 for validation, and a further 116 images were set aside for subsequent testing purposes. Fig. 1 shows a sample of an annotated image, depicting three instances of “amphora-frag” class. The final images were resized to 640x640 pixels for faster training.

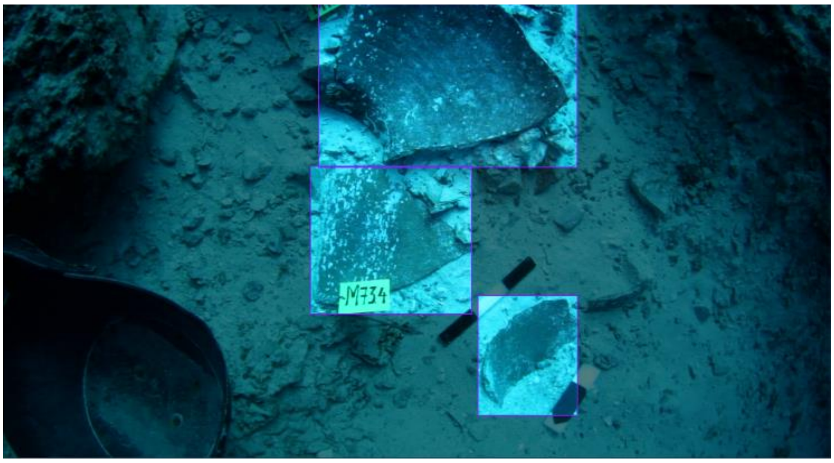

Figure 1. Sample of an annotated image from the custom dataset

2.2. Model

The work by Joseph Redmon et al. [8], introduced YOLO (You Only Look Once), an end-to-end deep learning approach for real-time object detection. YOLO revolutionized the field by achieving detection in a single pass of the network, unlike previous methods that used sliding windows or two-step processes. It simplified detection by utilizing regression-based prediction for both classification probabilities and bounding box coordinates. The YOLO framework has since evolved through multiple versions. The latest version, YOLOv8, demonstrates impressive performance, as reported in the literature [9].

The YOLOv8 model offers the advantage of flexible size configurations, offering five options ranging from "nano" to "extra-large." Each size variant represents a trade-off between precision and speed, where larger models achieve higher precision but at the cost of reduced speed. In this paper, all available models were trained using consistent settings and evaluated to compare their performance. The primary objective was to identify the most suitable model for the specific application based on the obtained results. Specifically, all models were trained using the same dataset and the maximum batch size allowed by the hardware. The training was conducted for 150 epochs or until the losses ceased to decrease. Indicatively, in Fig. 2 the training metrics of the "small" model are shown with respect to the training epochs.

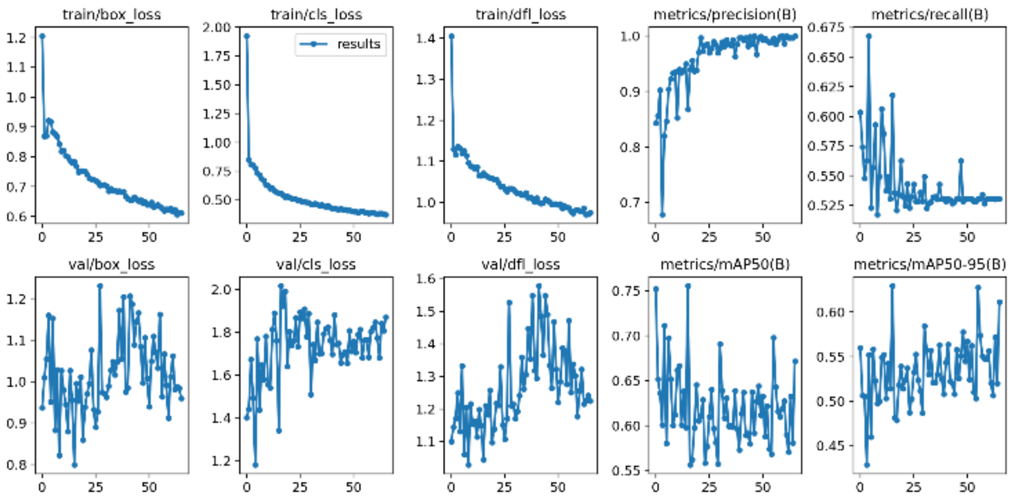

Figure 2. Exported metrics during the training of yolov8s.

2.3. Evaluation

After the training procedure, the evaluation of the trained models was done using the evaluation set of images. In this procedure, each image is passed through the trained model, and the predicted results are compared to the ground-truths from the given labels, thus the metrics such as mean Average Precision (mAP) and Average Recall (AR) are calculated. Specifically, mAP is the measurement of the true-positive detections relative to the total detections, whereas AR measures the true-positives relative to all the instances that should have been detected. Equations (1) and (2) show the calculation of each metric respectively.

$$mAP = \frac{True\ positives}{True\ positives + False\ positives} \quad (1)$$

$$AR = \frac{True\ positives}{True\ positives + False\ negatives} \quad (2)$$

To ensure dissimilarity between the training and evaluation sets, a specific video containing pottery sherds captured from various angles and under different lighting conditions was excluded from the training set. These images were exclusively reserved for evaluating the models. By adopting this approach, more realistic performance metrics were obtained, as shown in Table 1.

Table 1. Performance metrics

Model	**mAP50**	**AR**	**Speed (ms)**
Nano	0.724	0.587	3.6
Small	**0.755**	**0.6**	**6.5**
Medium	0.727	0.53	14.7
Large	0.747	0.53	22.5

3. RESULTS

Based on Table 1, the "small" version of the YOLOv8 model demonstrated the best performance metrics, albeit slightly slower than the "nano" version. Conversely, the larger models exhibited inferior performance due to overfitting on this specific dataset. Consequently, for this research, the selected model was the "small" version. Additional performance results for the "small" model include Precision over Confidence (Fig. 3), revealing a sharp increase in precision at low confidence values, reaching a maximum precision of 0.649. The Recall over Confidence (Fig. 4) demonstrates relatively stable recall values until a confidence value of 0.8. When deploying the model, it is crucial to select an appropriate confidence threshold to effectively balance precision and recall. Additionally, the confidence threshold can be tuned based on the detection task and the specific use-case. For example, if the objective is to verify archaeologists' findings, a higher confidence threshold should be set to increase precision while potentially producing some false negatives. Conversely, if the goal is to alert archaeologists of possible findings, a lower confidence threshold should be set to detect most suspected items, even at the expense of more false positives. Figures 5 and 6 depict Precision over Recall and F1 over Confidence, respectively, assisting in the selection of the most suitable confidence value for this model.

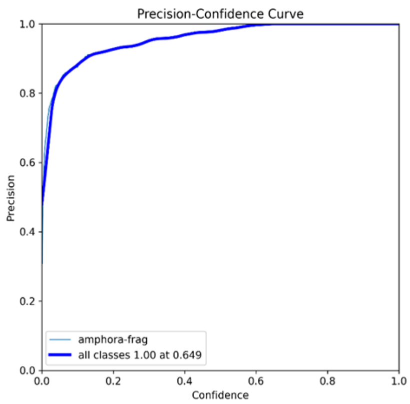

Figure 3. Precision-Confidence curve

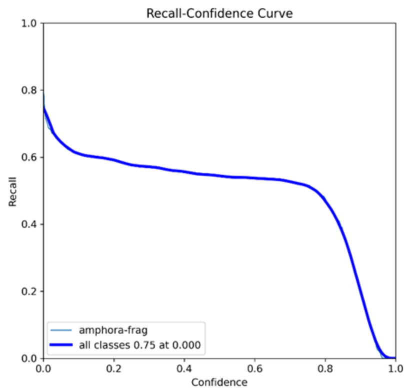

Figure 4. Recall-Confidence curve

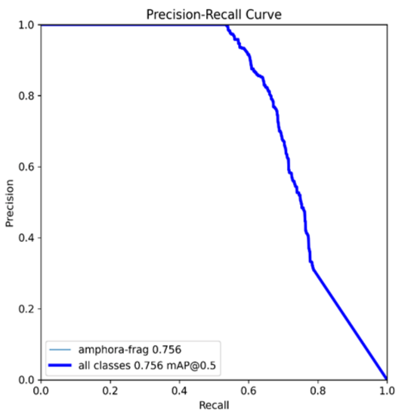

Figure 5. Precision-Recall curve

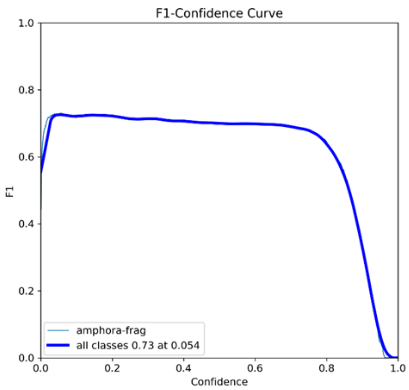

Figure 6. F1-Confidence curve

When conducting an inference on the selected model using the evaluation video, numerous detections are generated, as depicted in Fig 7. It can be observed that the model successfully detects the majority of the visible pottery sherds. However, there are cases where not all visible sherds are detected. For example, in Fig. 8, some instances are not detected, despite being present in similar frames.

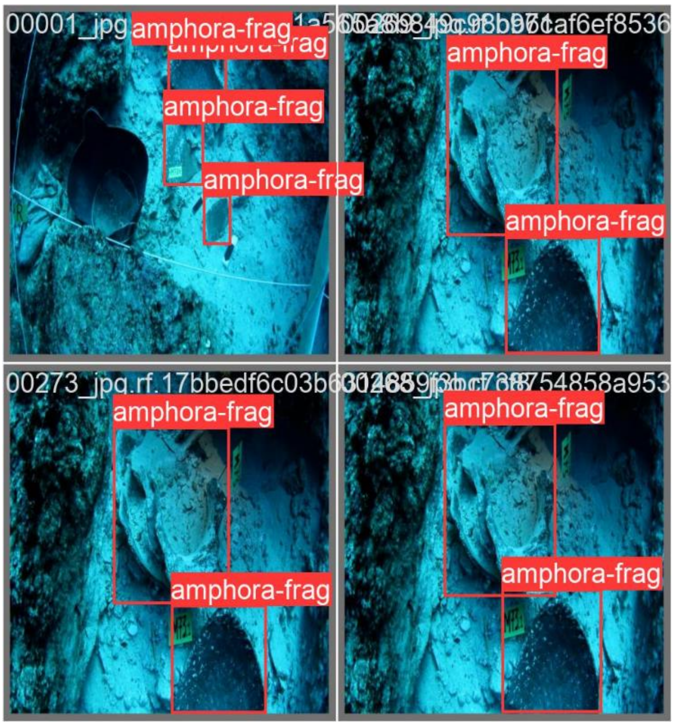

Figure 7. Sample detections of the "small" model on the evaluation video, shown in four separate screenshots.

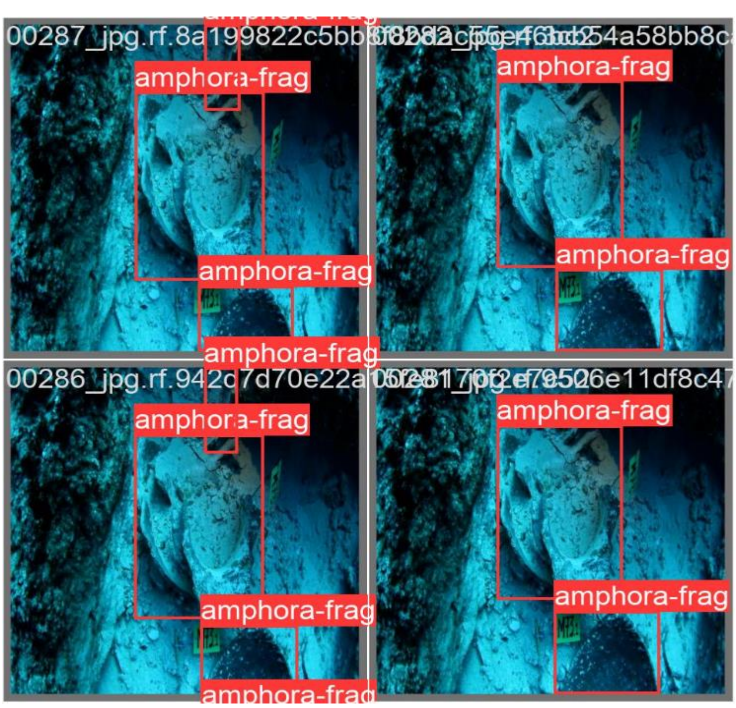

Figure 8. Sample detections including false negative detections.

4. CONCLUSION

In general, the utilization of Deep Learning techniques, as presented in this paper, demonstrates promising results in underwater object detection, particularly in pottery sherds. Additionally, it is crucial to employ a large dataset with meticulously annotated images to prevent model overfitting, as observed in the results of our larger models. Moreover, the specific purpose of this system is to facilitate archaeological excavations through the utilization of an underwater Remotely Operated Vehicle (ROV), emphasizing the significance of inference speed. However, despite being slower than the "nano" version, the "small" version of the trained model was selected due to its superior performance results. The inference speed of the "small" model, at 6.5 ms, remains satisfactory for the intended application. Given that our system will operate within an ROV, in collaboration with professional divers who will manually search for pottery shards, we can set an extremely low confidence threshold for detection. This will result in numerous false positives, which the divers can subsequently discard, while minimizing the number of undetected sherds.

Future research endeavors could enhance these outcomes by incorporating Online Learning, enabling the model to adapt and learn new classes and instances of pottery sherds.

ACKNOWLEDGEMENTS

The authors would like to thank the Hellenic Institute of Marine Archaeology (H.I.M.A.) for providing the video recordings of their underwater excavations at Modi, Greece, which were used for the creation of this work's dataset.

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project name u-ArchaeoRoV, code: T2EDK- 03656).

REFERENCES

[1] Character, L., Ortiz Jr, A., Beach, T., & Luzzadder-Beach, S. (2021). Archaeologic machine learning for shipwreck detection using lidar and sonar. Remote Sensing, 13(9), 1759.

[2] Anichini, F., Banterle, F., Buxeda i Garrigós, J., Callieri, M., Dershowitz, N., Dubbini, N., ... & Zallocco, M. (2020). Developing the ArchAIDE application: a digital workflow for identifying, organising and sharing archaeological pottery using automated image recognition. Internet Archaeology, 2020, num. 52.

[3] Sun, H., Liu, M., Li, L., Yan, L., Zhou, Y., & Feng, X. (2020). A new classification method of ancient Chinese ceramics based on machine learning and component analysis. Ceramics International, 46(6), 8104-8110.

[4] Itkin, B., Wolf, L., & Dershowitz, N. (2019). Computational ceramicology. arXiv preprint arXiv:1911.09960.

[5] Paraskevas, K., & Kavallieratou, E. (2022, November). Detecting Holes in Fishery Nets using an ROV. In 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-5). IEEE.

[6] Paraskevas, K., & Kavallieratou, E. (2023, May). Detecting Holes in Fish Farming nets: A Two–Method approach. In 2023 International Conference on Control, Automation and Diagnosis (ICCAD) (pp. 1-7). IEEE.

[7] Moniruzzaman, M., Islam, S. M. S., Bennamoun, M., & Lavery, P. (2017). Deep learning on underwater marine object detection: A survey. In Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18 (pp. 150-160). Springer International Publishing.

[8] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[9] Terven, J., & Cordova-Esparza, D. (2023). A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. arXiv preprint arXiv:2304.00501.