International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mar-)

http://dx.doi.org/10.12785/ijcds/150194

Automatic Deep-Sea Amphorae Detection Using Optimal 2D Ultralytics Deep Learning

Maad kamal Al-anni[1,2] **and Pierre DRAP**[2]

[1]*Computer Engineering Department, College of Engineering, Al-Iraqia University, 7366 Baghdad, Iraq.*
[2]*Aix Marseille University, CNRS, ENSAM, Universit´e De Toulon, LIS UMR 7020, 13397 Marseille, France.*

Received . 2024, Revised . 2024, Accepted . 2024, Published . 2024

Abstract: Despite the challenges in modern digital documentation, current research prioritizes computer-aided semantic segmentation in underwater environments and temporal monitoring, particularly for the digital documentation of deep-sea sites. Using cutting-edge technologies, exemplified by our automated archetype of archaeological sites (e.g., the Xlendi shipwreck), we present research on an archaeological shipwreck known as Xlendi, located off the coast of Malta, aiming to facilitate digital model acquisition for professionals and amateurs. This enhances archaeological insights and yields promising results across challenging sites, promoting virtual exploration, awareness, and advocacy for underwater cultural heritage(UCH).
Indubitably, current 3D instance segmentation methods enhance archaeological site comprehension, but, they encounter challenges such as computational complexity and labor-intensive annotation. This article addresses these issues by utilizing automated 2D object detection extended to 3D through photogrammetry, minimizing human effort by focusing on ad-hoc 2D annotation methods seen in previous research, and facilitating 3D segmentation through 2D 3D projection via photogrammetry.
intriguingly, the construction of this proposed model relies heavily on achieving precise 3D detection and identification. Its success is contingent upon the performance of the 2D object detection and its projections in an end-to-end scene. In this study, we evaluate the performance of YOLOv8 for object detection, focusing on underwater archaeological sites. Previous research using YOLOv4 reported an accuracy range of 78%-88% (mAP). Building on this, we assessed YOLOv8 using sensitivity, specificity, and mean average precision (mAP), achieving mAP values ranging from 98.2% to 99.2%. Specifically, we measured mAP@0.50 and mAP@0.50:0.95 to comprehensively evaluate model performance. Our findings demonstrate significant improvements over previous methods, highlighting the efficacy of YOLOv8 in archaeological contexts. We have also included a future workflow to inspect further enhancements.

Keywords: underwater cultural heritage (UCH), AI, Deep Learning(DL), 3D Instance Segmentation, 2D Object Detection, Deep Sea Photogrammetry.

1. Introduction

The discovery of the Xlendi shipwreck in Malta near the coast of Gozo [1], dating back to the 7th century BC, provides valuable insights into the maritime trade practices of that era and sheds light on the trading relationships between the western Phoenician and Tyrrhenian regions. The excavation at such a depth of approximately 100 m is a remarkable archaeological undertaking, offering a glimpse into ancient seafaring practices and trade routes. The mixed cargo of amphorae found at the site hints at the diverse goods that were being transported across the Mediterranean during that period, as illustrated in Figure 1.

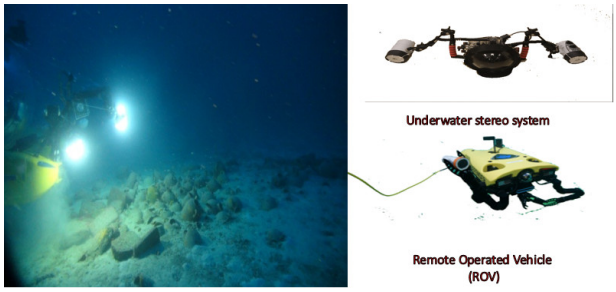

Figure 1. A remote operated vehicle (ROV) and stereo system were deployed to discover a shipwreck dating back to the 7th century BC, located approximately 100m deep.

The University of Malta's initiative to enhance research methods on the exceptional wreck involves long-standing cumulative high-technology employments in 3D surveying of archaeological excavations at greater depths. The entire dataset was pooled simultaneously with timelines of

E-mail address: muad.rashed@aliraqia.edu.iq, pierre.drap@lis-lab.fr

excavations through collaborative efforts from COMEX[1] and CNRS[2] in 2014. The project, funded by the French Agence Nationale de la Recherche under the GROPLAN project, marks a significant advancement in understanding and preserving submerged archaeological sites [2]. The University of Malta's initiative to deploy a team of skilled divers for excavating the site, triaging strata, as illustrated in Figure 2, focuses on 3D photogrammetric documentation. This effort has resulted in a substantial collection of 30,000 images that capture the archaeological site's evolution over a decade of distinct excavations. This meticulous approach underscores the significance of comprehensive data integration and sharing for an in-depth analysis and understanding of the site's history. The extensive image collection obtained through this audacious excavation serves as a valuable resource for further research and exploration, demonstrating the university's commitment to advancing archaeological knowledge.

The full paper will detail the entire pipeline and enhancement of previous research. Using YOLOv4, an accuracy range of 78%-88% (mAP) was reported [3], focusing on selecting the best 2D object detection algorithm for consistent 3D instance segmentation through the 2D_3D bidirectional relationship offered by photogrammetry. The study will highlight recent methods of semi-supervised 2D object detection, discussing the advantages of the latest versions for higher accuracy and precision, and addressing the limitations of older methods along with strategies to overcome them. The dataset for this research was specifically collected for this task, forming the foundation for linking knowledge bases with VR SDKs (virtual reality software development kits). Furthermore, the core of the photogrammetric acquisition design includes a stereo system mounted on an underwater scooter, which captures around 30,000 images from various timelines. This setup enables the use of the camera and various sensors present in devices like smartphones or tablets, along with SLAM (simultaneous localization and mapping). This technology allows us to evaluate the user's movements and adjust the device's point of view accordingly. These advancements facilitate capturing isolated amphorae in 3D scenes through the precomputed photogrammetry pipeline, an approach already being implemented at the LIS Lab.

Figure 2. Records showing outermost layer of excavated cargo as progress continues

At present, cyber-archaeology[4] employs digital tools such as virtual reality and pre-computed modelling to explore archaeological sites, extending beyond basic object identification or metadata retrieval to enable deeper analysis and interpretation of represented entities[5], illustrated in Figure 3.

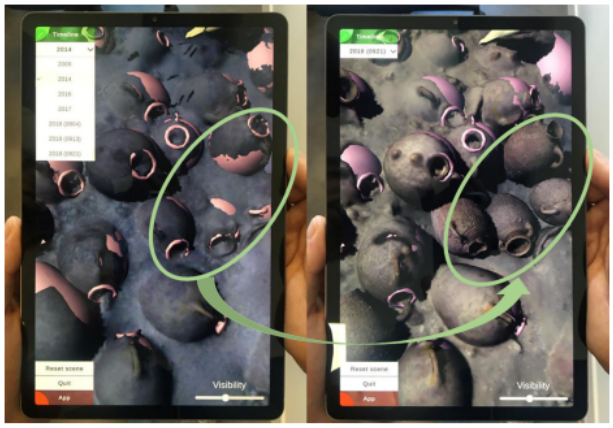

Figure 3. This figure, referencing our published article[5], illustrates that four amphorae were retrieved from the wreck before 2018, facilitating a deeper analysis and interpretation of the represented entities and their associated metadata.

In recent years, artificial intelligence, deep learning, and semantic network approaches have facilitated the visualization of complex data, supported educational initiatives, and fostered interdisciplinary collaboration in archaeology. This is particularly evident when these applications function as dashboards, assisting in decision-making, guiding site evolution, and revising existing knowledge [3]. The images used for 2D_3D reconstruction and 2D object detection models are captured synchronously, and the system is calibrated to scale the survey. The utilization of the scooter[3], along with a powerful and unique continuous lighting system, ensures fast acquisition speed to prevent motion blur and consistent lighting for uniform photographic coverage. Photogrammetric image processing is automated through scripting[4], and the resulting textured models are imported into virtual and augmented reality[5] tools developed for further analysis.

The stereo system developed at Septentrion Environment is based on extensive experience in underwater photography and photogrammetric surveying. It consists of two full-frame SLR cameras and a powerful continuous light system.

[1] https://comex.fr/
[2] https://www.cnrs.fr/
[3] https://www.septentrion-env.com/en/projet/photogrammetry-submarine-and-subaquatic-caves/
[4] https://www.agisoft.com/
[5] https://www.unrealengine.com/

The setup includes two Nauticam waterproof housings, each equipped with a Nikon D700 camera and a 14mm lens, fixed on a 60 cm long aluminium profile (40mm in diameter) (Figure 4). The two cameras are synchronized using a specific external cable, with the housings spaced 7 cm apart. A spherical joint in the middle of the aluminium profile allows the stereo set to be attached to an underwater scooter (DPV). Each housing is equipped with a float arm and an external Ikelite DS160 flash.

This stereo system enables precise and synchronized image capture for underwater photogrammetry. It has recorded a total of 30,000 images across all excavation campaigns since the first one. These categorized datasets form the cornerstone of our research, and we have selected a subset of 3,307 images to conduct this experiment.

Figure 4. Two Nauticam housings with Nikon D700 cameras on 60cm profile mounted on scooter.

This system is assembled underwater with a second device for continuous remote lighting. It consists of two LED projectors with a unitary power of 66,000 lumens mounted on a 150 cm long aluminium profile (40mm in diameter). The buoyancy of this second system is balanced with specific incompressible foams to achieve slightly negative buoyancy. Finally, the two systems are coupled in the water on a Suex xj37 underwater thruster using a specific support.

This article is organized as follows: Section II delves into object detection approaches, highlighting the shortcomings of traditional methods in fulfilling the task. Our discussion extends to the decision-making process of excluding YOLOv4 from consideration and shifting our focus to YOLOv8. Section III provides comprehensive details about YOLOv8, starting from elucidating the fundamental principles of its functionality. It then proceeds with the dataset labelling, training process, validation of the results, and concludes by comparing the results using matrix evaluation parameters. Section IV discusses the author's perspective on visualization of underwater cultural heritage (UCH). Finally, Section V consists of the discussion and conclusion of the research conducted.

2. Literature Survey

An essential step in processing the surveys is to localize the artifacts in the captured images. The main problem is that the artifacts, which are amphorae, are partially occluded and covered with sediment, and some objects are damaged. Traditional object detection approaches based purely on color or spatial information may fail [2]. Recent advances in machine learning, and more specifically in deep learning, have resulted in robust end-to-end object detection methods, also known as one-stage detectors, as illustrated in Figure 5. In particular, object detection algorithms [6] [7] have shown significant improvements in this area. Deep learning

Figure 5. Example images of the underwater site displaying artifacts (amphorae), with one image juxtaposed that requires automatic detection. The detection of amphorae in the adjacent image is indispensable.

has been widely applied to image classification for over a decade, aiming to find the label of an image among several learned labels. In object detection, multi-scale sliding windows have been used to identify the position and bounding box of objects within an image. Since 2015, several versions based on object of interest (OOI) have evolved in state-of-the-art (SOTA) methods[8][9][10][11]. The goal is to detect and localize objects within an image accurately. Numerous metrics are available for evaluating the performance of object-detection algorithms[12][13], as illustrated in Figure 6.

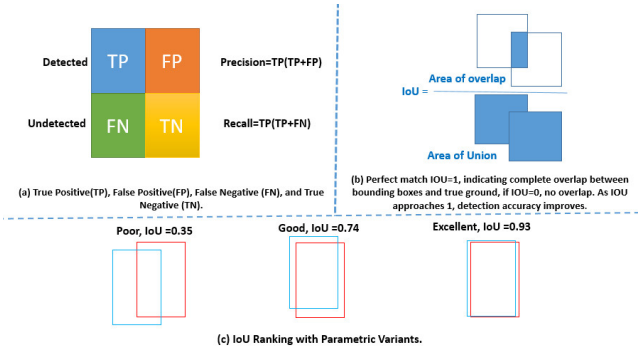

Figure 6. matrix performance for prioritizing object detection algorithms

In the literature, numerous competitive methods exist, but our focus is on earlier and more recent approaches that have demonstrated higher accuracy and significantly faster speeds. These approaches include:

- you only look once (YOLO) V4 [14]: This method discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to handle objects of various sizes naturally.

- you only look once (YOLO) V8 [15]: YOLOv4 features a sophisticated architecture with a CSPDarknet53 backbone, SPP module, and PANet enhancements for improved accuracy and speed. YOLOv8, an evolution of YOLOv4, integrates focal loss, CSP attention, and further PANet enhancements to enhance performance, making it a promising advancement in object detection algorithms.

Figure 7 illustrates the model results in detecting amphorae, whereas $mAP = \frac{\sum_{i=1}^{k} AP_i}{K}$ and variants of IoU, showing competitive outcomes based on Yolov4 + AlexeyAB's Darknet [6] and Yolov8 + Ultralytics [7]. YOLO (single-shot object

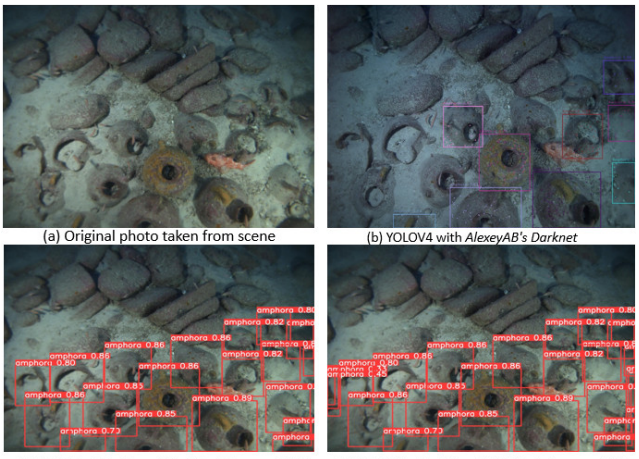

Figure 7. The detection effect of two mainstream detection methods: (a) Original image, (b) YOLOv4+AlexeyAB's Darknet with mAP@0.50:0.95, (c) YOLOv8+Ultralytics with mAP@0.50, and (d) YOLOv8+Ultralytics with mAP@0.50:0.95.

detection) prioritizes real-time performance and efficiency. Designed to achieve high accuracy while maintaining fast inference speeds, YOLO models are compatible with a wide range of applications. They are particularly suitable for deployment in resource-constrained environments and applications requiring rapid processing of visual data.

YOLOv4 with AlexeyAB's Darknet achieved 88% accuracy, 92% precision, and 95% recall, while YOLOv8 with Ultralytics attained 93% accuracy, 95% precision, and 97% recall. These results were obtained using a dataset of 3,307 labeled images for ad-hoc 2D annotation methods and an on-demand evaluation process covering an overarching 3D scene, as illustrated in Figure 8.

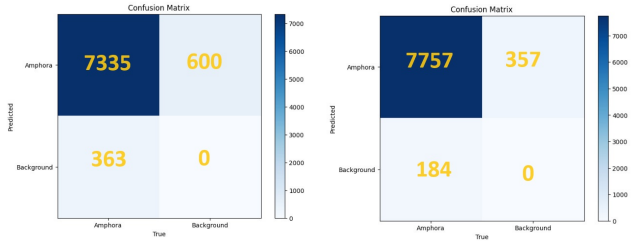

Figure 8. The first figure on the left shows the confusion matrix for YOLOv4, while the second figure on the right shows the confusion matrix for YOLOv8.

YOLO detection network has 24 convolutional layers followed by two fully connected layers. Alternating convolutional layers reduce the features space from preceding layers. Both methods have undergone several improvements since the time they appeared. For instance, now we have Yolo V8 that is much more precise and many times faster than original. The overarching architecture of YOLO is illustrated in Figure 9 . In this paper, we explore the

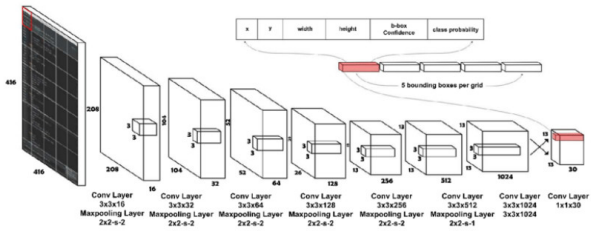

Figure 9. Overreaching architecture of YOLO. The network has 24 convolutional layers followed by two fully connected layers.

performance of two approaches for detecting OOI, specifically amphorae in our dataset. We exclude YOLOv4 + AlexeyAB's Darknet from our analysis and concentrate on improving the evaluation performance of YOLOv8 + Ultralytics by increasing the number of epochs, as discussed in the next section. Additionally, the selected best model aids in generating a comprehensive 3D model of the site using photogrammetry. We establish a bidirectional 2D to 3D association between the detected OOI in the images and their representation in the 3D model. This process of isolating amphorae through a 2D_3D relationship is invaluable for archaeologists attempting to correlate a 3D model of an amphora with its corresponding images for topological

[6] https://github.com/AlexeyAB
[7] https://github.com/ultralytics/ultralytics

investigation. Subsequently, we construct a dense cloud for mesh and texture modeling.

3. Materials and Methods

The primary challenge in our work is managing the size of the dataset, which comprises over 30,000 images from surveys conducted over the last decade. Supervised machine learning requires a large number of manually labelled images, entailing substantial human effort. To mitigate this, we propose adapting a semi-supervised learning approach, leveraging the sequential nature of the images to reduce manual labelling.

Our approach involves measuring similarity between consecutive images and selecting one image per group of similar images. Only the selected images need to be labelled. We then conduct an initial training phase using this subset of labelled images. The resulting model is used to detect OOI in the remaining unlabelled images. Detection is refined using results from a sparse feature point matching approach applied to labelled and unlabelled images within each image group. Finally, a second training phase is conducted using all images to obtain a robust final model.

Figure 10 illustrates the comprehensive approach to sketching models for enhancing the knowledge base in photogrammetry. The process involves implementing 2D object detection techniques for precise 3D model generation and 3D instance segmentation. This facilitates archaeological enrichment by enabling the creation of accurate 3D models turning to good use of 2D to 3D reconstruction. The subsequent sections will detail the practical workflows required to achieve the goals of the proposed model.

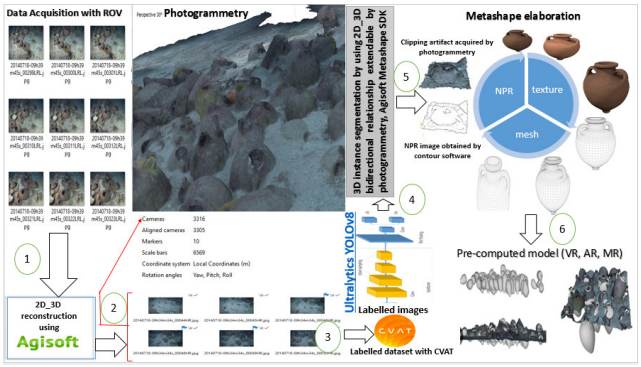

Figure 10. Comprehensive approach to enhancing photogrammetry knowledge base using 2D object detection and 3D instance segmentation.

A. Labelling Dataset with YOLO Format

The model uses CVAT for image labelling and a CNN for automated artifact localization and labelling. CVAT[8] is an open-source AI-powered tool for annotating 2D images, compatible with PyQt4 and PyQt5, supporting various formats, and extendable to semi-automatic annotation. Ultralytics' YOLO format compatibility suits our CNN model for object detection tasks, appreciated for its user-friendly interface (Figure 11).

Overall, image annotation plays a crucial role in object

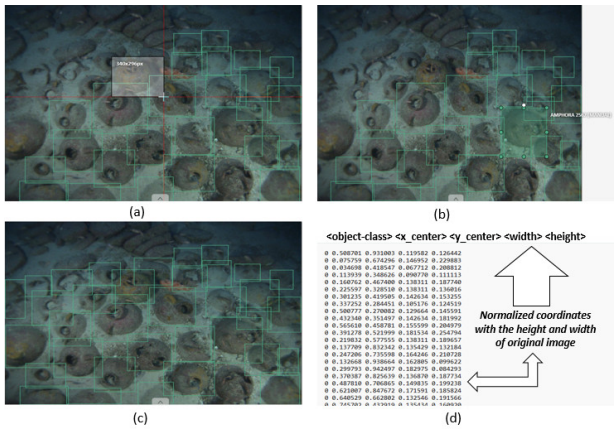

Figure 11. (a) The top-left figure provides guides to help users align and adjust the bounding box accurately around the object of interest in the image. (b) The top-right figure displays the sequential number of the bounding box within the overall annotation process. (c) The bottom-left figure illustrates the entire annotation process once completed, while the bottom-right figure shows the corresponding text file for the annotated image in YOLO format.

recognition approaches by providing the necessary labelled data for model training, evaluation, and dataset creation. Adhikari et al. [16] introduce a semi-automatic method for bounding box annotation, reducing manual effort by up to 75% across three datasets. Yoon et al. [17] address the challenge of sparsely annotated datasets in anchor-based object detection, proposing anchor-less and single-object tracker approaches. Their results demonstrate competitive performance on the EPIC-KITCHENS 2020 dataset. Mundher et al. [18] review automatic image annotation (AIA) methods, focusing on deep learning models, and categorize them into five types, emphasizing the importance of continued research in this area. Russell et al. [19] developed a web-based annotation tool to build a large dataset for object detection research, enhancing labels with WordNet.

B. Decoding the Core Principles of YOLO: Modern Perspectives on its Workflow

In recent years, the development of various AI architectures, notably the YOLO algorithm, has revolutionized object detection via CNN-based regression. YOLO efficiently determines object coordinates and classes in images. Thanks to technological and scientific advancements, end-to-end object detection approaches have become increasingly faster and more accurate. The neural network processes entire images at once, extracting object coordinates and classes by dividing them into S∗S grids[20]. The YOLO algorithm excels in real-time applications, also known as single-shot detection, where input images undergo convolutional layers to predict bounding boxes and class probabilities. Each box includes coordinates (x, y), width (w), height (h), non-maximum suppression (NMS), and confidence

[8]https://github.com/cvat-ai/cvat

(IoU). Thus, YOLO detects objects and their coordinates simultaneously[21], as illustrated in Figure 12.

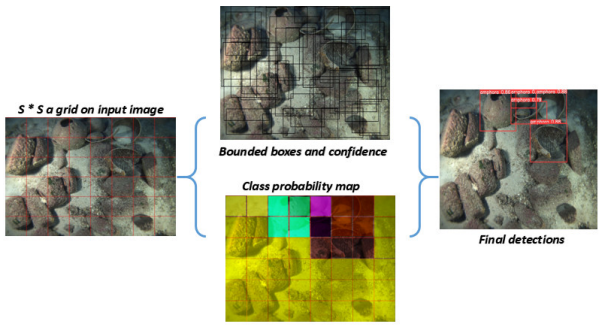

Figure 12. YOLO:single network principle predicts bounding boxes and class probabilities. [20].

Various versions of the YOLO algorithm have demonstrated remarkable success in object detection applications within the realm of artificial intelligence in recent years[22][23][24]. In this study, training and testing of the YOLOv8 Ultralytics algorithm were conducted using Python 3.10.13 within a virtual environment. The setup included dependencies and an NVIDIA GeForce Quadro RTX 6000 graphics processor, operating as a multi-GPU system, alongside a computer featuring an Intel Core i7-9750H 2.60 GHz processor, 16 GB DDR5 RAM, and 512 GB SSD. The system ran on a Linux-based operating system (Ubuntu) and utilized CUDA and cuDNN to accelerate the workflow.

C. *Training a Labelling Dataset with YOLOv4 and Ultralytics*

In this study, we utilized a labelled dataset focusing on a single class, previously employed in generating a 3D orthomosaic through photogrammetry. Our objective was to assess the efficacy of YOLO algorithms, particularly YOLOv8, on the same set of 2D images. The dataset comprises a total of 3,307 images, including high-resolution (HR) and low-resolution (LR) images. To facilitate 3D instance segmentation using YOLOv8, we developed specific Python scripts to accurately project 2D labelled instances. This process aimed to establish a probability area indicating the presence of amphorae in 3D space. By aligning labels depicted in photos with the corresponding 3D amphorae instances, we achieved a comprehensive 3D model capturing visible portions of amphorae and surrounding sediment.

Our future efforts will focus on proposing methodologies for determining the typology of isolated amphorae. To achieve this goal, we first need to train the network on the labelled dataset and achieve high detection accuracy. We propose dividing the dataset into 70% for training, 20% for testing, and 10% for validation, as illustrated in Algorithm 1.

Algorithm 1 Dataset Splitting

$N \leftarrow$ total number of samples
$N_{\text{train}} \leftarrow 0.7 \times N$ ▷ 70% for training
$N_{\text{test}} \leftarrow 0.2 \times N$ ▷ 20% for testing
$N_{\text{val}} \leftarrow 0.1 \times N$ ▷ 10% for validation
Shuffle the dataset randomly
for $i \leftarrow 1$ ***to*** N **do**
$i \leq N_{\text{train}}$
Add sample i to training set
$i \leq N_{\text{train}} + N_{\text{test}}$
Add sample i to testing set **else**
Add sample i to validation set

The study utilized YOLOv8 algorithms with a dataset comprising 3,309 images to detect objects, using 7,941 bounding boxes in the 2D survey. Seventy percent of the dataset's images were allocated for training, and 20% for testing. Each YOLO algorithm underwent training and testing across 100, 200, and 300 epochs, as illustrated in Algorithm 2. Figure 13 shows the snippet used for training YOLOv8.

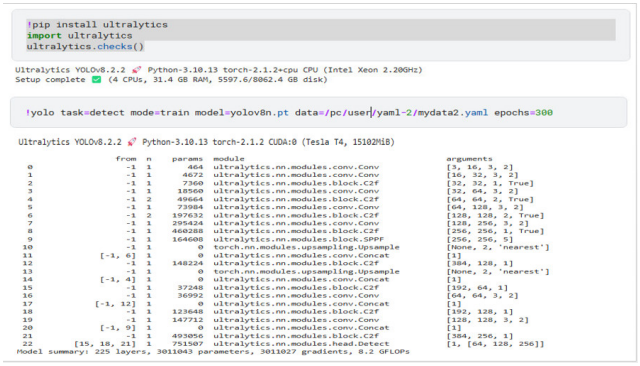

Figure 13. The installation of YOLOv8 version 2.2 is demonstrated for Python 3.10.13, using a CUDA Tesla T4. The model configuration consists of 225 layers and 3,011,043 parameters required to commence network training.

The YOLOv8 algorithm was trained and tested using a dataset containing labelling for a single class. The algorithm was executed for 100 and 300 epochs, with results recorded for each epoch with mAP@0.50:0.95 to comprehensively evaluate model performance. Numerous experimental analyses were conducted to evaluate the training and testing of the algorithm, yielding successful findings.

Comparative assessments, including confusion matrices illustrated in Figure 14, were made. For 100 epochs, the results indicate TP=7757, FP=357, and FN=184, with precision at 95% and recall at 97%. For 300 epochs, the results reveal TP=7897, FP=228, and FN=44, with precision at 97% and recall at 99%. Figures 15 and 17 illustrate that the YOLOv8 algorithm demonstrates the most effective learn-

Algorithm 2 Install Ultralytics and Train

```
Install Ultralytics:
!pip install ultralytics

Import Necessary Libraries:
import os
import shutil

Define Directories:
DATA_DIR = "path/to/data"
MODEL_DIR = "path/to/model"

Prepare Dataset:
...                          ▷ Code to prepare your dataset

Train the Model:
!python train.py --data $DATA_DIR --cfg
   yolov5s.yaml --weights '' --batch-size 16
   --epochs 100, 200, and 300
# Note: Adjust parameters as needed
```

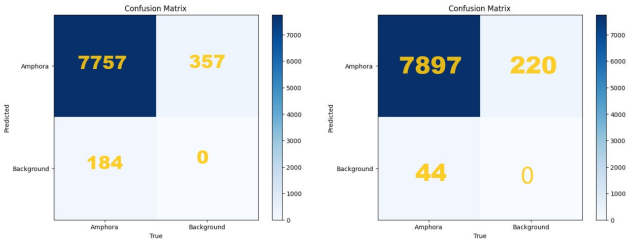

Figure 14. Comparative success graphs obtained as a result of training and testing YOLOV8 (a) graph on the left side shows the confusion matrix at 100 epochs (b) graph on the right side shows the confusion matrix at 300 epochs

ing overall, despite showing the least favourable outcomes after certain epochs. A review of Figure 16 and 18, indicates that after 300 epochs, the YOLOv8 algorithm achieves the lowest average loss value, while the earlier epochs register a significantly higher average loss value. Additionally, all numerical results from training and testing YOLOv8 are presented comparatively in Table I.

TABLE I. A resizable table with 5 columns and 6 rows.

epoch	Gpu_mem	box_loss	cls_loss	mAP@0.50:0.95
50	2.78G	1.165	0.5285	0.985
62	2.51G	1.153	0.5209	0.986
100	2.68G	1.123	0.4816	0.99
200	3.18G	1.052	0.4333	0.991
228	3.01G	1.009	0.4159	0.991
300	2.33G	0.9476	0.3588	0.992

According to the experimental analyses for the training and testing of the YOLOv8 algorithm, the best results were observed at epoch 240, with the algorithm successfully detecting amphorae at a mean average precision (mAP@0.50:0.95) rate of 99.2%. Consequently, training was halted prematurely due to no improvement was observed in the last 70 epochs.

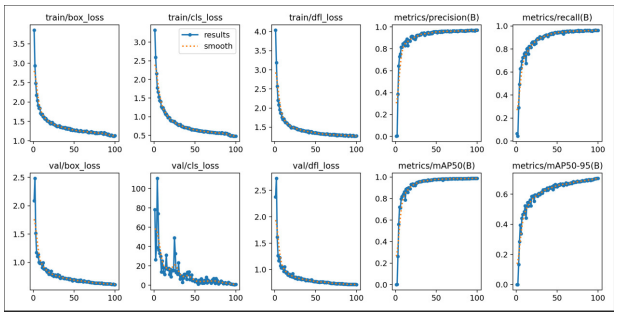

Figure 15. Adjacent figures sequentially display the results of training for 100 epochs, including the loss function for both training and validation, as well as the mAP@0.50:0.95 curve, precision curve, and recall curve.

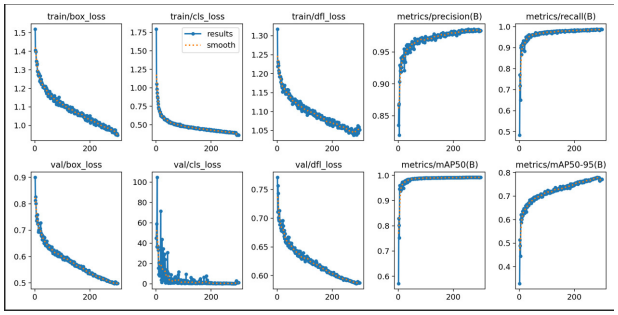

Figure 16. Adjacent Figures sequentially display the results of training for 300 epochs, including the loss function for both training and validation, as well as the mAP@0.50:0.95 curve, precision curve, and recall curve.

To validate the success rates obtained from experimental analyses with the YOLOv8 algorithm trained with different epochs, a performance test was conducted separately for each epoch. Each type of training was tested using real-life images. Detailed examples of the success rates of each epoch in real-image performance are illustrated in Figures 16 and 18.

D. Assessing Trained Network Performance with Real-Life Images

To assess the validity of the success rates obtained from experimental analyses using YOLOV8 trained with 100 epochs and 300 epochs, a performance evaluation was conducted for each training phase independently. Each phase was tested using real images. Examples of the success rates achieved by each algorithm in real-world scenarios are depicted in Figure 19, and Algorithm **??** shows the steps

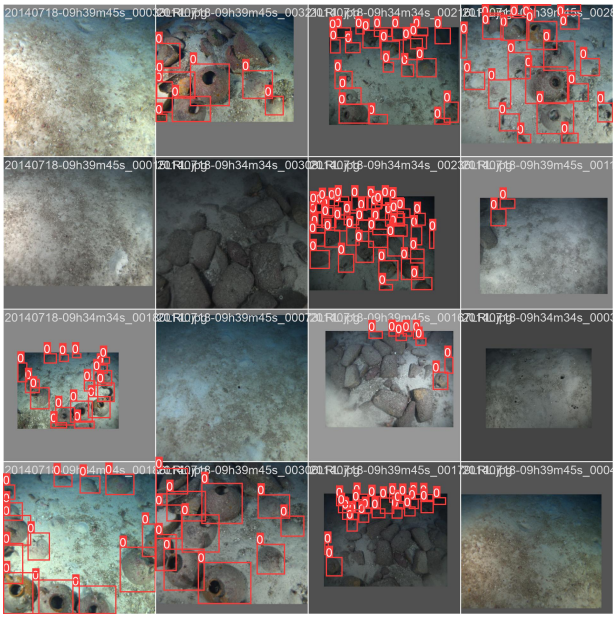

Figure 17. Photo segments of performance testing when the training dataset has epochs = 100

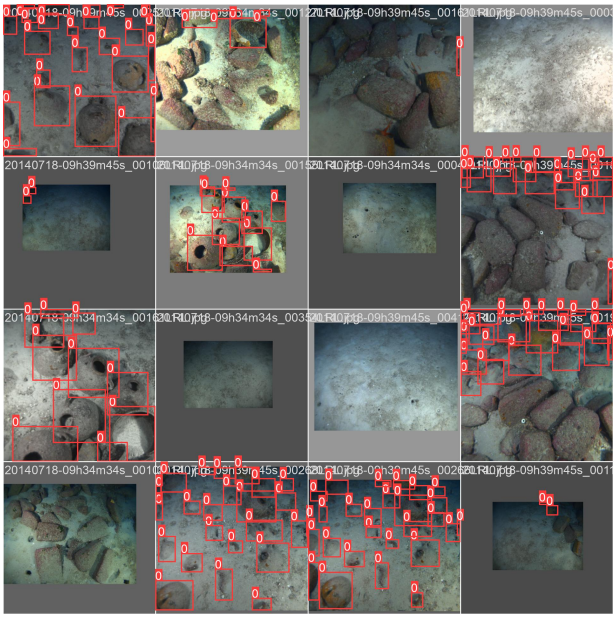

Figure 18. Photo segments of performance testing when the training dataset has epochs = 300

to test the real-world scenarios. Additionally, Figures 20 and 21 illustrate the real predictions made by the networks trained with 100 epochs and 300 epochs, respectively, along with the corresponding label correlation graphs, revealing their spatial dependencies and relative distribution patterns among object detection.

Upon careful examination of Figures 20 and 21, which represent the results of network training with varying numbers of epochs (100 and 300 epochs of mAP@0.50:0.95), the final training achieves tangible results in distinguishing amphorae more accurately and precisely. This is observed in the network's enhanced capability to identify the necessary features and correlation coefficients of instances.

Figure 19. Real-world scenarios featuring photos captured from deep-sea expeditions.

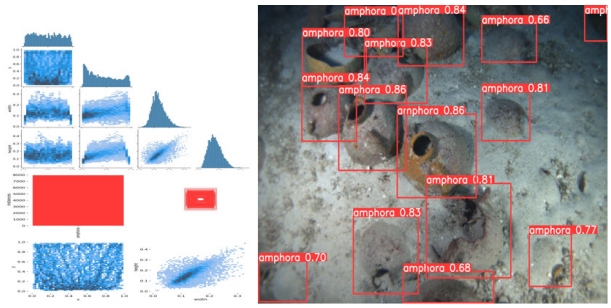

Figure 20. Detected amphorae along with correlogram of instances (height, width) at 100 epochs.

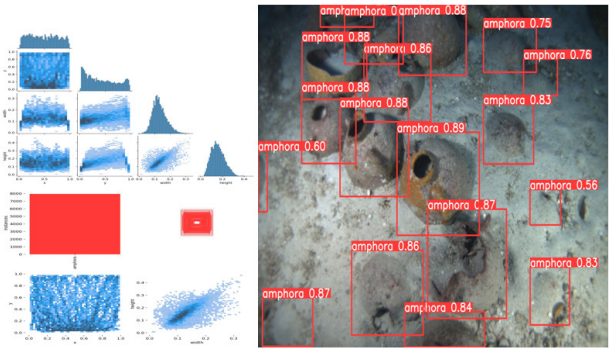

Figure 21. Detected amphorae along with correlogram of instances (height, width) at 300 epochs.

Algorithm 3 Testing YOLO on an Image

1: **Input:** Image to be tested, Trained YOLO model
2: **Output:** Detected objects with bounding boxes
3: Load the trained YOLO model
4: Load the image to be tested
5: Preprocess the image (resize, normalization, etc.)
6: Pass the preprocessed image through the YOLO model
7: Obtain predictions for detected objects
8: Post-process predictions (filtering, non-max suppression, etc.)
9: Display the original image with detected objects and bounding boxes

4. Discussion and Future work (underwater cultural heritage (UCH))

The photogrammetric technique is well-established for virtualizing cultural assets. Digital twins support archaeological analyses, scientific dissemination, monitoring, and preservation. These models are used in various forms, including web viewers, physical replicas, and virtual/augmented/mixed reality. Technological advancements and increased heritage fragility have boosted the popularity of underwater site virtualization.

A test pilot has demonstrated the efficiency of using 3,307 2D images projected into a 3D orthomosaic. The autocoder successfully isolated vertices of the OOI, specifically amphorae, by leveraging the bidirectional relationship between 2D and 3D data. This approach enabled the establishment of 3D instance segmentation using YOLOv8. Figure 22 Plots isolated vertices from camera references 165 to 170, captured in PLY file format.

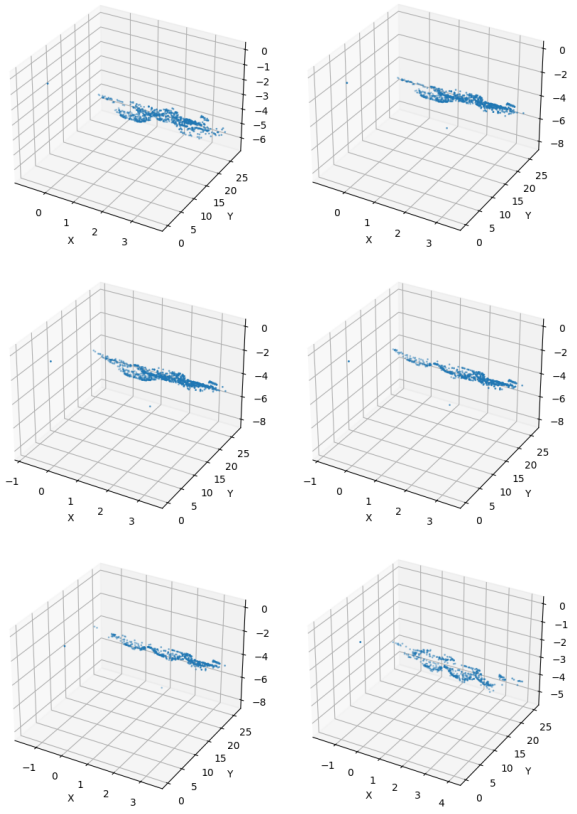

Figure 22. Plotting isolated vertices from camera references 165 to 170, captured in PLY file format

Surveying and digitizing through photogrammetry is now standard for aerial and underwater applications. While improvements in automation, acquisition, and processing speed continue, challenges like underwater color correction and temporal monitoring in dynamic environments remain active research areas. Obtaining digital models for dissemination is straightforward with photography and basic surveying skills. Our research aims to enhance 3D scenes by applying an extendable version of photogrammetry from 2D to 3D reconstruction. This allows for automated coding to isolate vertices of OOI, such as amphorae, within the 3D orthomosaic deposit. Ultimately, rebuilding the model using texture and mesh enables the visualization of 3D instance segmentation while exploring non-invasive underwater sites.

Similarly, the role and importance of virtual reality for archaeology and cultural heritage has been well established for over 30 years [25][26]. The development of these techniques has enabled the possibility of being immersed in the archaeological site, not only reproducing the current state of the heritage, but also allowing the simulation of the past, a process defined as cyberarcheology by Forte et al. [27]. Virtual and augmented reality techniques for underwater sites have been explored in several projects aimed at allowing virtual visits to non-divers [28], increasing awareness, and promoting underwater cultural heritage (UCH) through serious games [29], as well as studying and analyzing complex excavations and their evolution over time [30].

We plan to create a seamless virtualization solution from photogrammetric survey to virtual and augmented reality tools for underwater sites, visualizing amphorae and other site components. Our automated and optimized steps ensure a comprehensive and robust photogrammetric survey, safe for divers at great depths, and easily integrated into developed virtual and augmented reality tools.

At the heart of the 3D instance segmentation is a pre-computed model integrated with photogrammetry. Photogrammetric image processing is performed automatically through scripting, and the generated textured models are imported into virtual and augmented reality tools.

We anticipate providing more technical details of the developed solution from survey to virtual and augmented reality applications, which will be expanded in upcoming work.

5. Conclusion

The workflow is part of a series of mainstream efforts conducted under the auspices of the LIS UMR. This project, along with others, is based on a long-standing cooperation since 2009 between the University of Malta and Aix-Marseille University. In recent years, the University of Malta, under the direction of Prof. T. Gambin, conducted excavations of the Xlendi wreck at a depth of over 100 meters. Utilizing semantic web technology and 3D tools, this article showcases a decade-long excavation monitoring at Xlendi, employing ontology and AR/VR [5].
Additionally, Mohamed et al. [31] propose integrating cultural heritage (CH) data with domain-specific knowledge for intelligent visualization, fostering semantic interoperability and user-friendly querying. Ben et al. [32] propose using semantic web technology to enhance access and visualization of diverse CH resources, demonstrated with the Xlendi shipwreck dataset, available on the Google Play Store app. Our work aims to enhance 3D scenes for VR/AR exploration, building on previous efforts [3] to seamlessly integrate exploration with the latest hardware, leveraging extendable photogrammetry and advancements in 3D documentary development to enable robust 2D object detection algorithms to interact with 3D scenes. Subsequently, reconstructing 3D scenes with dense clouds, mesh, and texture vividly virtualizes immersive, lifelike environments, enriching archaeological insights and exploration of the Xlendi site. Accurate 3D instance segmentation using 2D object detection approaches is crucial for maintaining consistency and precision in projecting 3D objects like amphorae. Evaluating YOLOv8 and YOLOv4 on a dataset of 3,307 labeled images, YOLOv8 achieved a recognition rate of 99.2%, while YOLOv4 achieved 87.94%. Successful tests to recognize and detect amphorae in 3D scenes using YOLOv8 validate its effectiveness in 3D instance segmentation-based applications and its importance in isolating vertices within targeted 3D objects.

Acknowledgment

The authors would like to express sincere gratitude to the University of Malta, Aix-Marseille University, and Al-Iraqia University, whose insightful guidance and unwavering support throughout this research have been invaluable. We are also thankful to the reviewers for their constructive feedback and invaluable comments. This work was made possible by the generous support of Prof. T. Gambin. Special thanks to the researchers at the LIS UMR. Finally, we acknowledge the efforts of all participants and contributors who made this research possible.

References

[1] T. Gambin, "A phoenician shipwreck off gozo, malta," *Malta Archaeological Review*, pp. 69–71, 01 2015.

[2] P. Drap, D. Merad, B. Hijazi, L. Gaoua, M. M. Nawaf, M. Saccone, B. Chemisky, J. Seinturier, J.-C. Sourisseau, T. Gambin, and F. Castro, "Underwater photogrammetry and object modeling: A case study of xlendi wreck in malta," *Sensors*, vol. 15, no. 12, pp. 30 351–30 384, 2015. [Online]. Available: https://www.mdpi.com/1424-8220/15/12/29802

[3] M. Al-anni and P. DRAP, "Efficient 3d instance segmentation for archaeological sites using 2d object detection and tracking (pp. 133–1342)," 10 Mar. 2024.

[4] M. Forte, "Virtual reality, cyberarchaeology, teleimmersive archaeology 3d recording and modelling in archaeology and cultural heritage: theory and best practices (pp. 113–127)," 2014.

[5] M. Nawaf, P. Drap, M. Ben-Ellefi, E. Nocerino, B. Chemisky, T. Chassaing, A. Colpani, V. Noumossie, K. Hyttinen, J. Wood, T. Gambin, and J. C. Sourisseau, "Using virtual or augmented reality for the time-based study of complex underwater archaeological excavations," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. VIII-M-1-2021, pp. 117–124, 2021. [Online]. Available: https://isprs-annals.copernicus.org/articles/VIII-M-1-2021/117/2021/

[6] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. PP, pp. 1–20, 03 2023.

[7] E. Arkin, N. Yadikar, X. Xu, A. Aysa, and K. Ubul, "A survey: object detection methods from CNN to transformer," *Multimedia Tools and Applications*, vol. 82, no. 14, pp. 21 353–21 383, Jun. 2023.

[8] A. Vijayakumar and S. Vairavasundaram, "YOLO-based object detection models: A review and its applications," *Multimedia Tools and Applications*, Mar. 2024.

[9] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.

[10] V. Kaya and Akgül, *OBJECT DETECTION WITH ARTIFICIAL INTELLIGENCE: YOLO APPLICATION*, 12 2022, pp. 109–126.

[11] A. Alhardi and M. Afeef, *Object Detection Algorithms & Techniques*. Springer, 03 2024, pp. 391–399.

[12] R. Padilla, S. Netto, and E. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proceedings of the 27th International Conference on Systems, Signals, and Image Processing (IWSSIP)*, 07 2020.

[13] J. Tian, Q. Jin, Y. Wang, J. Yang, S. Zhang, and D. Sun, "Performance analysis of deep learning-based object detection algorithms on COCO benchmark: a comparative study," *Journal of Engineering and Applied Science*, vol. 71, no. 1, p. 76, Mar. 2024.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[15] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolo," Jan. 2023, version 8.0.0, AGPL-3.0 license. [Online]. Available: https://github.com/ultralytics/ultralytics

[16] B. Adhikari and H. Huttunen, "Iterative bounding box annotation for object detection," in *2021 16th International Conference on Pattern Recognition (ICPR)*, 01 2021, pp. 4040–4046.

[17] J. Yoon, S. Hong, and M.-K. Choi, "Semi-supervised object detection with sparsely annotated dataset," in *Proceedings of the IEEE*

International Conference on Image Processing (ICIP), 09 2021, pp. 719–723.

[18] M. Mundher, M. Rahim, A. Rehman, Z. Mehmood, T. Saba, and R. Naqvi, "Automatic image annotation based on deep learning models: A systematic review and future challenges," *IEEE Access*, vol. PP, 03 2021.

[19] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and Web-Based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 157–173, May 2008.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 06 2016, pp. 779–788.

[21] J. Sang, Z. Wu, P. Guo, H. Hu, H. Xiang, Q. Zhang, and B. Cai, "An improved yolov2 for vehicle detection," *Sensors*, vol. 18, no. 12, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/12/4272

[22] E. Shreyas, M. H. Sheth, and Mohana, "3d object detection and tracking methods using deep learning for computer vision applications," in *2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*, 2021, pp. 735–738.

[23] G. Lavanya and S. Pande, "Enhancing real-time object detection with yolo algorithm," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 12 2023.

[24] N. M. Krishna, R. Y. Reddy, M. S. C. Reddy, K. P. Madhav, and G. Sudham, "Object detection and tracking using yolo," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021, pp. 1–7.

[25] J. A. Barceló, M. Forte, and D. H. Sanders, *Virtual reality in archaeology*. ArchaeoPress Oxford, 2000.

[26] P. Reilly, "Towards a virtual archaeology," in *Computer Applications in Archaeology*. Oxford: British Archaeological Reports, 1990, pp. 133–139.

[27] M. Forte, "Virtual reality, cyberarchaeology, teleimmersive archaeology 3d recording and modelling in archaeology and cultural heritage: theory and best practices (pp. 113–127)," 2014.

[28] F. Liarokapis, P. Kouřil, P. Agrafiotis, S. Demesticha, J. Chmelik, and D. Skarlatos, "3d modelling and mapping for virtual exploration of underwater archaeology assets," in *Proceedings of the ISPRS Conference*. ISPRS, 2017.

[29] M. Cozza, S. Isabella, P. Di Cuia, A. Cozza, R. Peluso, V. Cosentino, L. Barbieri, M. Muzzupappa, and F. Bruno, "Dive in the past: A serious game to promote the underwater cultural heritage of the mediterranean sea," *Heritage*, vol. 4, no. 4, pp. 4001–4016, 2021.

[30] M. Nawaf, P. Drap, M. Ben-Ellefi, E. Nocerino, B. Chemisky, T. Chassaing, A. Colpani, V. Noumossie, K. Hyttinen, J. Wood *et al.*, "Using virtual or augmented reality for the time-based study of complex underwater archaeological excavations," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 8, pp. 117–124, 2021.

[31] M. Ben Ellefi, P. Drap, O. Papini, D. Merad, J. Royer, M. Nawaf, E. Nocerino, K. Hyttinen, J.-C. Sourisseau, T. Gambin, and F. Castro, "Ontology-based web tools for retrieving photogrammetric cultural heritage models," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W10, pp. 31–38, 04 2019.

[32] M. Ben Ellefi, M. Nawaf, J.-C. Sourisseau, T. Gambin, F. Castro, and P. Drap, "Clustering over the cultural heritage linked open dataset: Xlendi shipwreck," in *Proceedings of the International Workshop on Semantic Big Data (SBD)*, 05 2018.