GA – PROJECT NUMBER:	**101158046**
PROJECT ACRONYM:	AUTOMATA
PROJECT TITLE:	AUTOmated enriched digitisation of Archaeological liThics and cerAmics
CALL/TOPIC:	HORIZON-CL2-2023-HERITAGE-ECCCH-01-02
TYPE OF ACTION	HORIZON RIA
PRINCIPAL INVESTIGATOR	Prof Gabriele Gattiglia, UNIPI
TEL:	+39 050 2215228
E-MAIL:	gabriele.gattiglia@unipi.it

This project has received funding from the European Union's HORIZON RIA research and innovation programme under grant agreement N. 101158046

D 3.2 Grasp and planning Strategies

Version: 1.0

Revision: first release

Work Package: 3 - Robotic automation system development - Hardware and Grasp Planning

Lead Author (Org): Do Won Park (IIT)

Contributing Author(s) (Org): Antonio Bicchi (IIT), Manuel G. Catalano (IIT), Arianna Traviglia (IIT), Fabio Bonomo (QB), Riccardo Persichini (QB), Giuliano Dami (QB), Nevio Dubbini (MIN), Gabriele Gattiglia (UNIPI), Martina Naso (Unipi)

Due Date: M13 - 30/09/2025

Date: 30/09/2025

Dissemination Level		
P	**Public**	**X**
C	**Confidential, only for members of the consortium and the Commission Services**	

Revision History

Revision	Date	Author	Description
0.1	01/09/2025	Do Won Park	Creation
0.2	11/09/2025	Do Won Park	Draft content added
0.3	22/09/2025	Do Won Park	Draft content added
0.4	24/09/2025	Riccardo Persichini	Draft content added
0.5	25/09/2025	Arianna Traviglia	Revision of the document
0.6	26/09/2025	Do Won Park	Revision of the document
0.7	29/09/2025	Gabriele Gattiglia, Nevio Dubbini, Martina Naso	Revision of the document
0.8	01/10/2025	Gabriele Gattiglia	Content added
0.9	02/10/2025	Do Won Park, Gabriele Gattiglia	Final revision

Disclaimer

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Abbreviations

WP: Work package

M: Month

UNIPI: Università di Pisa

UBM: Université Bordeaux Montaigne

UoY: University of York

INRAP Institut National de Recherches Archéologiques

AMZ: Arheoloski Muzej u Zagrebu

QB: QBrobotics Srl

HUJ: The Hebrew University of Jerusalem

MIN: Miningful srls

KCL: King's College London

IIT: Fondazione Istituto Italiano di Tecnologia

UB: Universitat de Barcelona

CL: Culture Lab

Executive summary

Deliverable D3.2 outlines the creation and validation of algorithms developed for the autonomous manipulation and detection of archaeological artefacts. This work addresses the critical challenge of handling fragile and non-uniform objects, requiring a system that combines precision and soft grasping for cultural heritage applications.

The core of this deliverable is the study of a planning and control framework. This system also integrates the capabilities of the soft end-effectors, largely tested in Deliverable D3.1. This integration enables a robotic arm to gently grasp artefacts, manipulate them, and securely place them into protective boxes.

The framework is tested through the real implementation of algorithms, which demonstrates the successful manipulation of artefacts. This validation serves as a crucial proof of concept, laying the foundation for the final robotic workcell.

This document is organised as follows. *Section 1* provides a brief overview of the state of the art in object detection, pose estimation for grasping, and motion planning. *Section 2* presents the methodology and the experiments conducted on archaeological artefacts. *Section 3* discusses the results of the simulations. *Section 4* outlines the design concepts of the Graphical User Interface (GUI) for intuitive programming and management by non-expert roboticists. Finally, *Section 5* draws conclusions and suggests directions for future work.

Introduction

The AUTOMATA project focuses on the development and implementation of an autonomous robotic workcell for the documentation, digitisation, and analysis of archaeological artefacts, with a specific focus on ceramics and lithics. The Deliverable D3.2 integrates state-of-the-art soft robotic technologies with innovative grasping solutions specifically designed to manipulate the uncertainties and delicacy of the artefacts. From the software perspective, a novel set of algorithms for manipulation and planning will be developed, explicitly leveraging the unique characteristics of the artefacts.

Within the AUTOMATA robotic system, a key requirement is the safe and effective manipulation of artefacts. Considering the broad variability of cultural heritage objects, the system will ensure adaptability within the range of shapes and dimensions defined by the consortium. This is achieved by means of adaptive and soft end-effectors. This report outlines the work done in Task T3.2. The core challenge is to develop control algorithms and planning solutions for grasping and manipulating objects for 3D digitisation, as well as for grasping and positioning them in storage boxes. Key technical aspects addressed include the generation of robust grasping strategies for delicate and non-uniform objects (artefacts) and the integration of soft end-effectors with the robotic arm controls in constrained environments.

1 State of the Art

1.1 Object Detection

Object detection, the task of identifying and localising objects within an environment, represents a critical and challenging step for any robotic system interacting with the physical world. In the context of hand pose estimation and manipulation, object detection is not limited to classifying an object; it must also provide its position in space, often through a “bounding box” (Palleschi et al., 2023) and a segmentation mask (Dong & Zhang, 2023). This constitutes an essential step for subsequent tasks such as object pose estimation, grasp planning, and the generation of grasping trajectories for a robotic manipulator (Xie et al., 2023).

The evolution of object detection techniques can be described as a transition between two approaches: analytical methods, based on searching contact points on the object (typically known) for grasping (Bicchi, 1995), and modern methods, based on deep learning, which learn features automatically from data and have redefined the field in terms of accuracy and robustness (Tsirtsakis et al., 2025).

Traditional, or analytical, methods for object detection are based on the manual extraction of object features. For example, the Scale-Invariant Feature Transform identifies stable keypoints based on local image gradients, creating a description that is robust to changes in object scale, rotation, and lighting, making it ideal for recognising known objects from different viewpoints (Lowe, 2004). Another method is the Histogram of Oriented Gradients, which describes an object’s shape through the distribution of gradient directions and has become a benchmark approach for general object detection (Dalal & Triggs, 2005). A further analytical method is template-based matching. An example is LINEMOD, which uses colour gradients and surface normals from RGB-D data to detect objects and their 3D pose, even in heavily cluttered scenes (Hinterstoisser et al., 2012). In addition to analytical and RGB-D methods, there are also approaches that exploit point clouds to decompose objects and obtain pose estimations (Palleschi et al., 2023).

A large number of modern methods rely on deep learning techniques, particularly Convolutional Neural Networks (CNNs) (Zhiqiang & Jun, 2017; Dhillon & Verma 2020; Cai et al., 2016). Unlike manually designed feature extractors, CNNs automatically learn features directly from training data. Two main branches of detectors have emerged: two-stage detectors, such as R-CNN (Girshick et al., 2014) and its successors Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2015); and one-stage detectors, such as YOLO (You Only Look Once) (Redmon et al., 2016), which formulate object detection as a single regression problem and are well-suited for real-time applications.

1.2 Motion Planning

Once an object has been detected and its pose estimated, the next step for an autonomous robotic system is motion planning, that is, generating a motion to interact with the object. It is important to distinguish between path planning and trajectory planning. The former aims to find a collision-free geometric path from a start to a goal state, while ignoring time and dynamics. The latter is a more complex problem that takes a geometric path and assigns it a temporal profile, specifying the robot's position, velocity, and acceleration, while respecting its kinematic and dynamic constraints. The ultimate goal is to generate a trajectory that is not only collision-free, but also smooth, efficient, and safe for the robot, for humans in the workspace, and for the manipulated object, such as an archaeological artefact in our use case.

There are approaches (Ding et al., 2023; Meng et al., 2022) that include sampling-based planners, such as Rapidly-exploring Random Tree, which remains a benchmark for its ability to find feasible paths in complex configuration spaces (Karaman & Frazzoli, 2011). Optimisation-based methods formulate planning as the minimisation of a cost function to generate smooth and safe trajectories (Yu et al., 2022). A further methodology is represented by data-driven approaches, which learn motion strategies directly from data. Reinforcement Learning, particularly Deep Reinforcement Learning (Chen et al., 2022), has enabled robots to acquire complex control policies directly through interaction with the environment (Akkaya et al., 2019).

These algorithms are typically used through practical software frameworks. MoveIt!, built on Robotic Operative System (ROS), is the standard in the research community, providing an interface to core planners in the Open Motion Planning Library (Sucan et al., 2012) alongside tools for kinematics and collision avoidance (Sucan & Chitta, n.d.).

2 System Implementation and Methodology

Building on the methods described in Section 1, this section presents the approach adopted for our archaeological use case. The methodology is designed to address the specific challenges posed by the substantial and diverse inventory of artefacts. Consequently, a generic object detection strategy was selected as an initial solution. The system is structured into two main modules:

1. **Object Detection and Soft End-Effector Pose Estimation**, responsible for accurately segmenting each artefact, estimating its centroid, and computing an appropriate grasping pose for the soft robotic gripper.
2. **Motion Planning**, responsible for generating collision-free, smooth, and precise pick-and-place trajectories that move the artefacts between predefined locations for digitisation and analysis.

The following subsections provide a detailed description of each module, including the algorithms and software frameworks employed, the calibration procedure, and the considerations taken to ensure real-time performance and high accuracy in the manipulation of irregularly shaped archaeological objects.

2.1 Object Detection and Soft End-Effector Pose Estimation Module

2.1.1 Object Detection

Object detection and localisation is a particularly challenging problem, especially when approaching unique and irregularly shaped objects like archaeological artefacts. To address this challenge, the Bilateral Reference Network (BiRefNet) (Zheng et al., 2024) is employed for high-resolution image segmentation. This approach is selected for its ability to operate at high resolution, a critical requirement for capturing the fine details of the artefacts' irregular contours, while maintaining high computational efficacy since the whole system should be real-time.

Unlike traditional methods that often compromise detail by downsampling images to manage the computational load, BiRefNet applies an innovative bilateral reference strategy. It leverages a low-resolution version of the image to generate a coarse reference map, which guides a refinement process on the original high-resolution image (Fig. 1). This allows the network to focus on critical boundary details without leading to unfeasible computational costs, making the system suitable for real-time applications.

Fig. 1 - BiRefNet high-resolution segmentation of a video frame showing a ceramic fragment used in the tests described in Section 3.

Once the object's segmentation mask is obtained, the next step is to determine a suitable grasping pose for a soft end-effector. Accordingly, an efficient methodology is developed to detect the shortest line segment that passes through the artefact's centroid and connects two points on its contour.

The algorithm is structured into the following steps.

1. **Contour Extraction and Centroid Calculation**: using the OpenCV library (OpenCV, n.d.-a), the artefact's contour, a list of N points $\{p_1, p_2, \ldots, p_N\}$ defining its boundary, is extracted (Fig.2a).
2. **Coordinate System Transformation**: the contour is normalised by translating the coordinate system so that its geometric centroid matches the origin $(0, 0)$ (Fig.2b).
3. **Polar Coordinate Representation and Sorting**: each translate point $p_i = (x_i, y_i)$ is converted to polar coordinates by calculating its angle θ_i using the $atan2(y_i, x_i)$ function to obtain an unique angular representation within the range [-π, +π] (Fig.2c). To enable an efficient search for diametrically opposite points, the contour points are ordered according to their polar angle θ_i (OpenCV does not generate contour points with a specific order). This gives a sequence of points ordered counter-clockwise along the contour perimeter (Fig.2d).
4. **Opposite Point Search via Binary Search**: for each point p_i (with angle θ_i), its diametrically opposite point is determined. The target angle for this point is defined as $\theta_{target} = \theta_i + \pi$. To optimise the search, a *binary search algorithm* is used on the sorted list of angles to find the index closest to θ_{target} in logarithmic time ($O \log(N)$). The optimal candidate is

selected by comparing the angle at the found index and its adjacent one, choosing the point that minimizes the angular deviation (Fig.2e).

5. **Euclidean Distance Calculation**: finally, for each candidate point (p_i, p_j), the Euclidean distance is calculated and the pair with the minimum distance is iteratively saved.

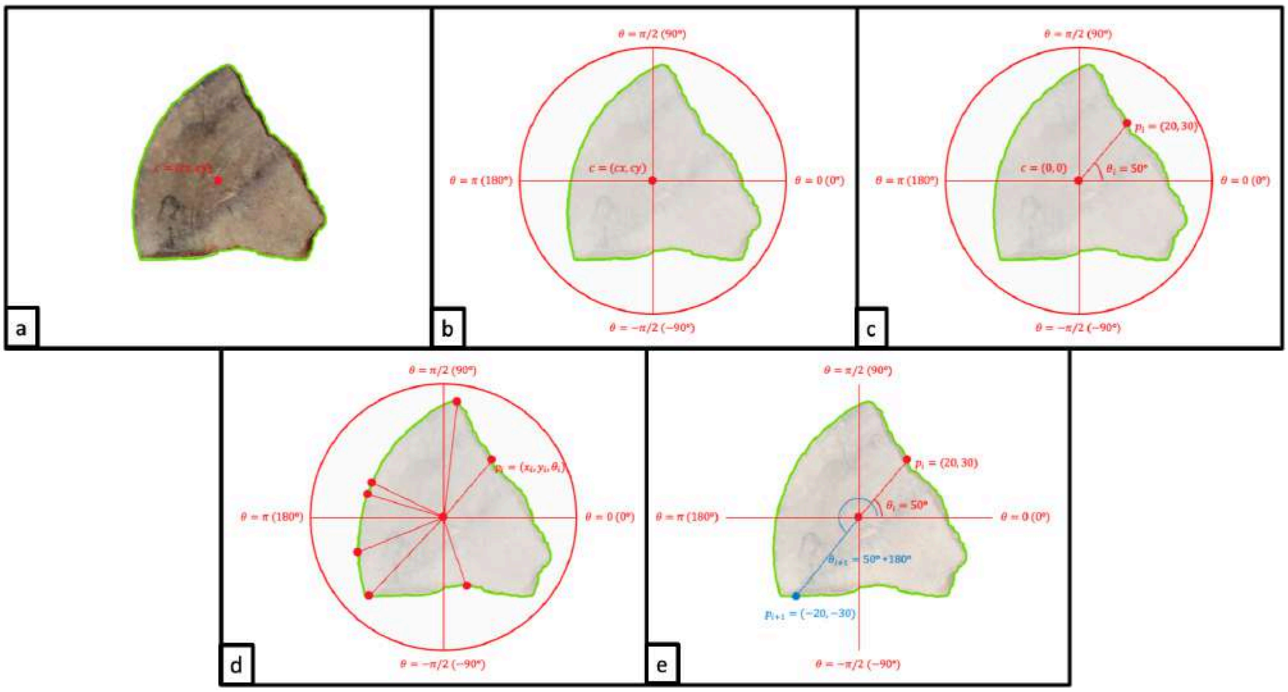

Fig. 2 - Visual description of the algorithm after the BiRefNet segmentation: (a) Contour Extraction and Centroid Calculation, (b) Coordinate System Transformation, (c-d) Polar Coordinate Representation and Sorting and (e) Opposite Point Search via Binary Search.

The algorithm thus returns the pair of points defining the shortest segment, with total computational complexity, mainly due to the initial sorting step, $(O \log(N))$, making this approach suitable for real-time analysis and robot manipulation.

2.1.2 Pose Estimation

The purpose of computing the shortest segment through the centroid and identifying the coordinates of its endpoints on the contour is to estimate the target position of the end-effector. This segment also provides its orientation, specifically its opening direction from the current position toward the centroid along the chosen segment.

At this stage, only the pixel coordinates, ${}^{pixel}p_i = (x_i, y_i)$ are provided. However, by using an RGB-D camera, the depth value of a given point can be obtained, thereby establishing a 3D coordinate system, ${}^{camera}p_i = (x_i, y_i, z_i)$, with respect to the camera system frame (Fig. 3).

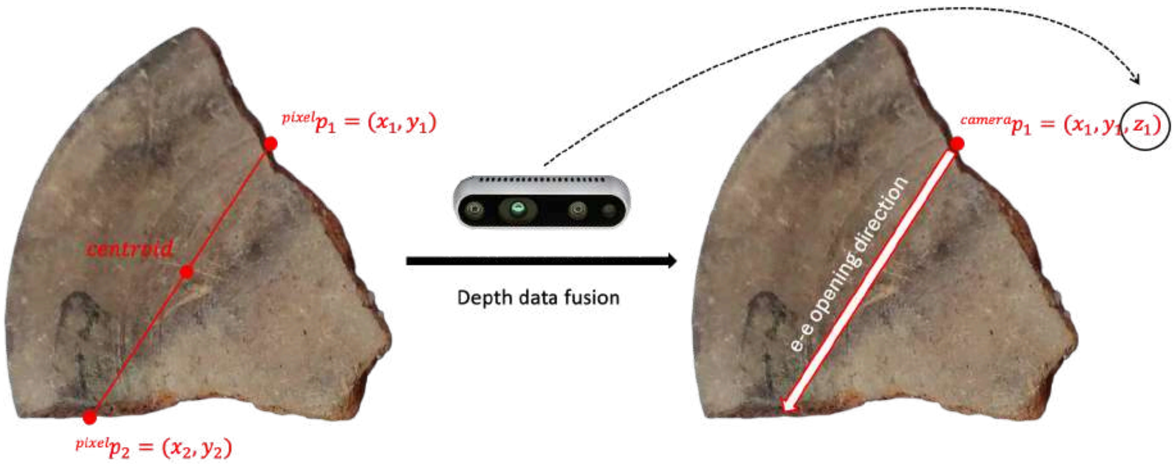

Fig. 3 - from the output of the object detection algorithm to the soft end-effector grasping pose (position+orientation).

The final step of this module is to transform the estimated pose into the robot base frame so that it can be sent to the robot. This is achieved by calibrating the RGB-D camera using the OpenCV ArUco calibration tool (OpenCV, n.d.-b). A QR code is attached as a fixed joint to the end-effector, allowing the camera to determine its pose relative to the robot base frame (Fig.4).

Fig. 4 - Screenshot from Rviz during the calibration process.

2.2 Motion Planning Module

In the AUTOMATA robotic cell, the robot must be able to perform pick-and-place tasks in a robust way, including obstacle avoidance. Starting from the initial location, a box containing all the artefacts to be digitised, the robot must pick and place each artefact on the platform in front of the sensors for archaeological analysis, and finally deposit it into the output box where the processed artefacts are placed.

Preliminary experiments were performed before integrating the estimated pose from object detection, and so decoupling that module from the robot's motion planning module. The robot was moved to specific points on the workspace (plain table) by manually defining waypoints, in order to grasp the artefact and place it across different spots. These trials were mainly aimed at tuning the MoveIt! parameters and validating the robot's motion execution.

Motion planning is implemented using MoveIt! within the ROS2 framework, adopting Cartesian path planning for end-effector trajectories. Cartesian motion planning allows the end-effector to follow a trajectory defined directly in Cartesian coordinates. This method interpolates a sequence of waypoints at a specified spatial resolution and incrementally computes the robot's joint configurations via inverse kinematics.

This approach is suitable for digitisation tasks, where predictable and linear motion of the end-effector is required. It enables the robot to manipulate artefacts within a constrained workspace with high spatial accuracy and repeatability.

The computed Cartesian paths were validated both in simulation and on a real robot. Trajectory planning is visualised in Rviz (Fig. 5) and executed from the robot (Fig. 6).

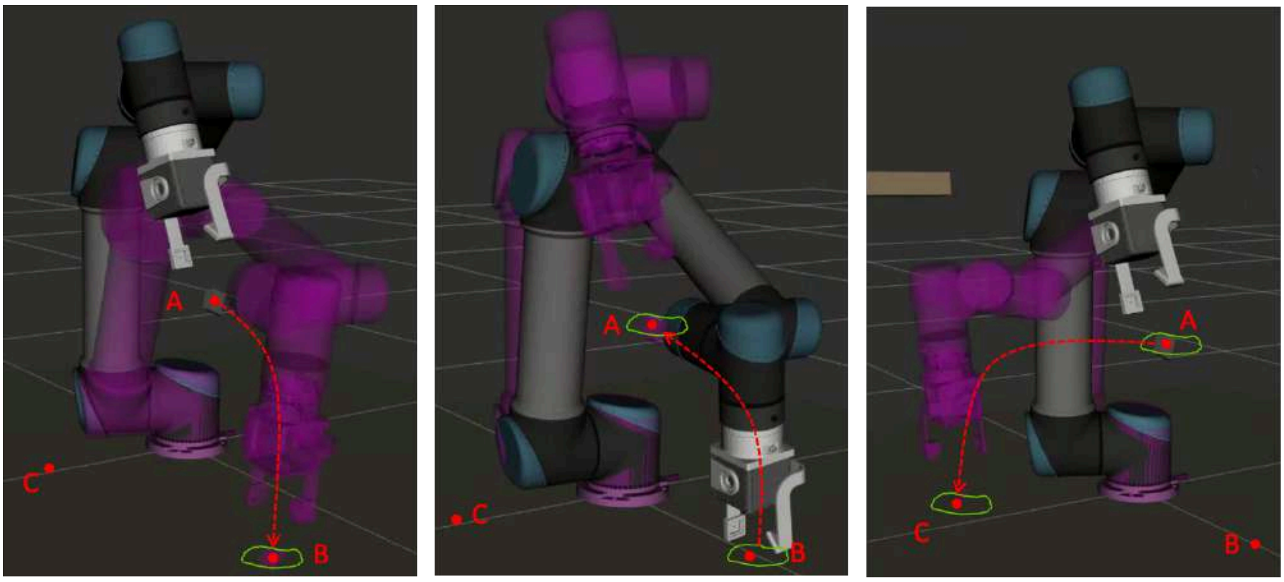

Fig. 5 - Trajectory planning visualised in Rviz, with waypoints A, B, and C indicated in execution order.

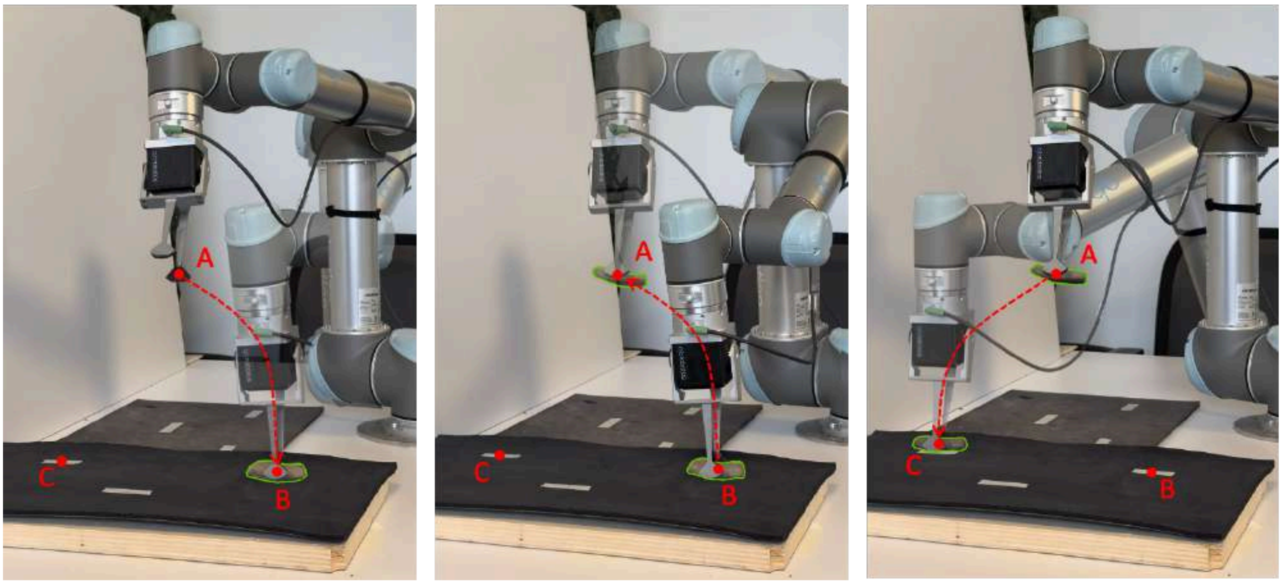

Fig. 6 - Robot movements during execution, stopping at waypoints A, B, and C as shown in Fig. 5.

3 Integrated System Testing and Evaluation

The integration of the object detection and motion planning modules is implemented through the pipeline shown in Fig. 7. The RGB-D camera first detects the artefact, and then the algorithm extracts the end-effector pose, position, and orientation in 2D pixel coordinates. Then, the depth information of the camera is used to convert the pose in 3D with respect to the manipulator's reference frame. Finally, the trajectory is calculated and executed from the real robot.

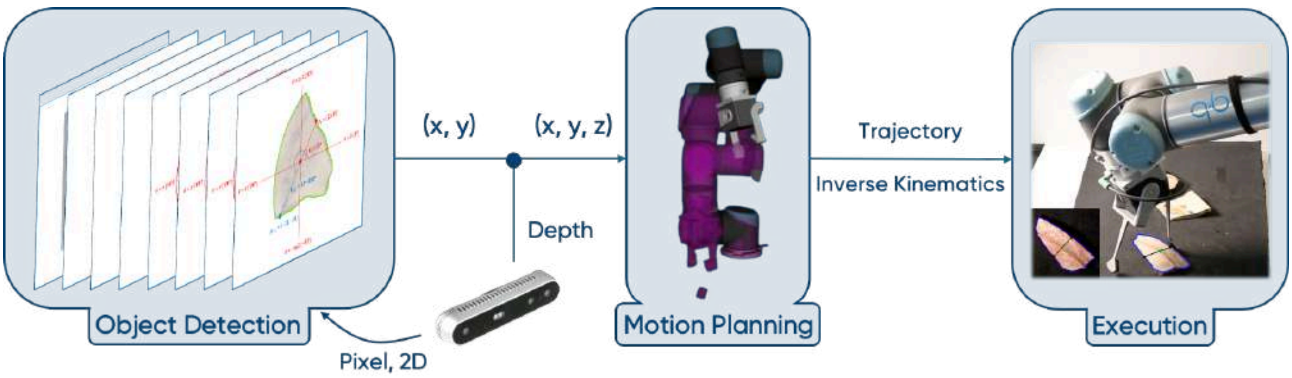

Fig. 7 - Manipulation workflow.

3.1 Experimental Setup

As stated in Deliverable D2.3, the robotic cell includes a robotic arm equipped with a dual-tool gripper, a 3D model/HSI acquisition system (A), and two additional data acquisition stations: pXRF (B) and a handheld Raman spectrometer (C) (Fig. 8). However, this configuration represents a conceptual design, which will be realised and refined in the upcoming WPs once the specific sensors to be integrated into the robot cell are finally decided by the consortium.

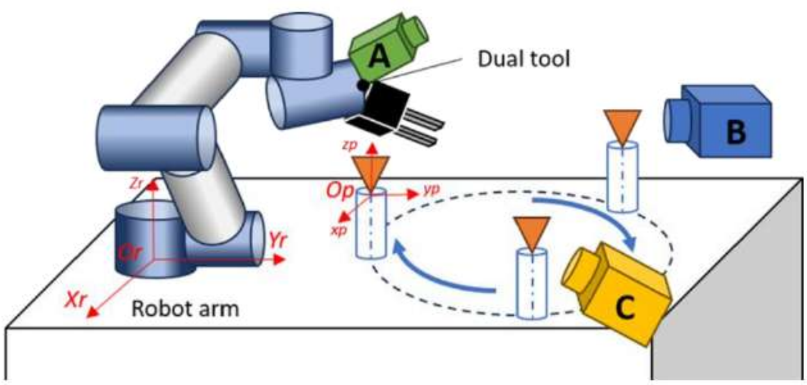

Fig. 8 - Work cell layout with robotic arm equipped with gripper and data acquisition systems: A, B and C.

In the meantime, preliminary tests of integration of the object detection module, pose estimation and motion planning module are carried out in a simplified workspace, where the robot performed pick-and-place tasks with different artefacts, handling them one at a time.

The modules integration test setup (Fig. 9) was as follows:

- **Robot**: Universal Robots UR5 (a);
- **End-effector**: SoftClaw (b);
- **RGB-D Camera**: Intel RealSense D415 (c);
- **Light**: Neewer NL 660, bi-color 3200-5600K;
- **PC**: Dell G16;
- **Control Framework**: ROS2 Humble

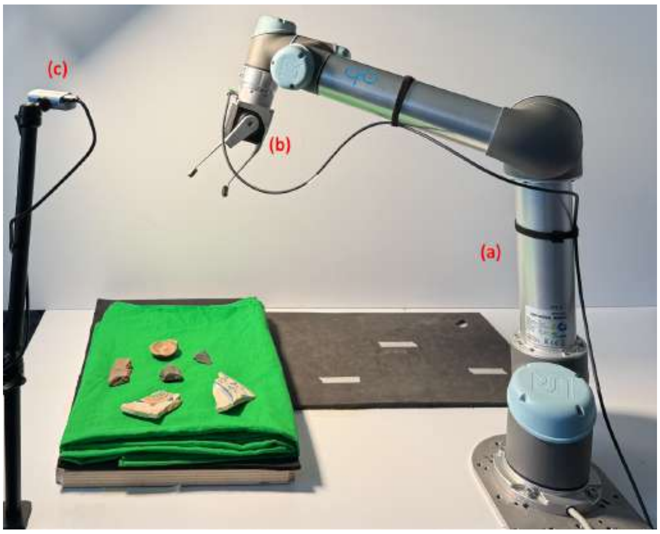

Fig. 9 - Experiment setup: (a) UR5, (b) SoftClaw and (c) Realsense D415 Camera.

For the tests, some of the samples provided by UNIPI (Fig. 10) with the dimensions indicated in Tab.1 were used.

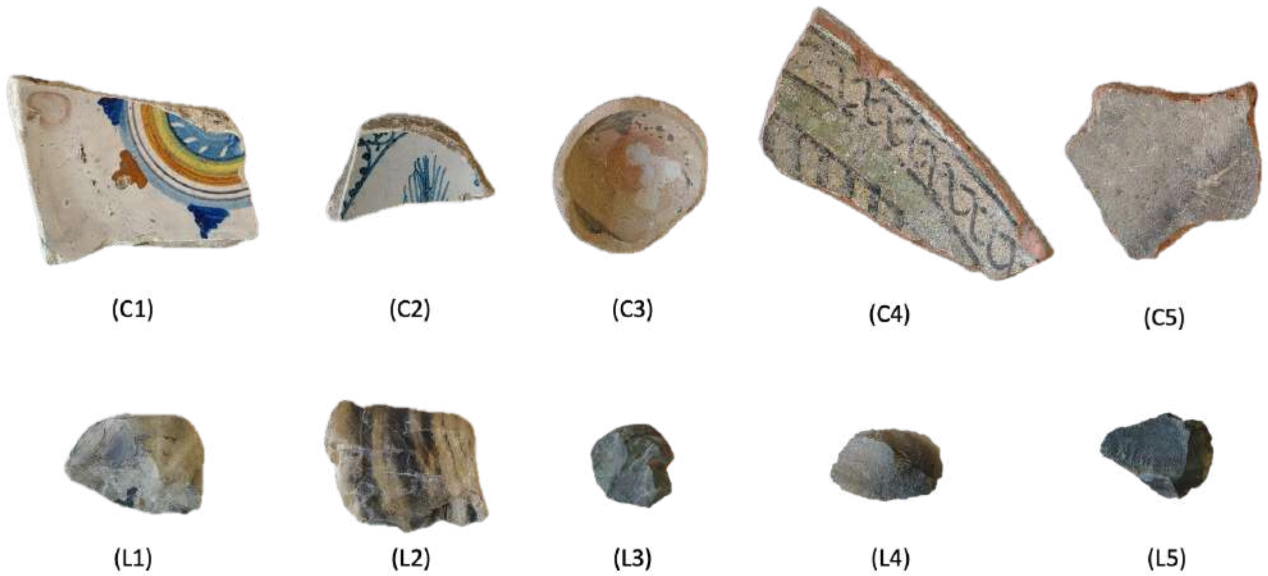

Fig. 10 - Artefacts provided by UNIPI used for the experiments, dimensions in Tab. 1.

Reference	Artefact Type	Lenght (mm)	Height (mm)	Thickness (mm)
C1	Ceramic	113	55	14
C2	Ceramic	65	40	9
C3	Ceramic	60	52	16
C4	Ceramic	125	45	16
C5	Ceramic	70	55	13
L1	Lithic	55	40	20
L2	Lithic	58	47	22
L3	Lithic	32	24	20
L4	Lithic	42	25	11
L5	Lithic	45	30	7

Tab. 1 - Morphometric characteristics of the artefacts in Fig. 9.

3.2 Analysis and Results

3.2.1 Analysis

An experimental trial with ceramic artefacts is illustrated in Fig. 11. These experiments were conducted using the robotic setup described in Section 3.1, ensuring consistency in the configuration of the RGB-D camera, robotic arm, and workspace. In each trial, the operator randomly places the artefacts within the camera's field of view (Fig. 11a). The pose of the end-effector is then estimated in real time, allowing the robot to reach and grasp the object (Fig. 11b–c) and subsequently complete the pick-and-place task (Fig. 11d).

The complete sequence of these experiments is documented in the attached video (https://youtu.be/d_7redqRWZo), providing a comprehensive view of the integration between detection, pose estimation, and motion planning modules.

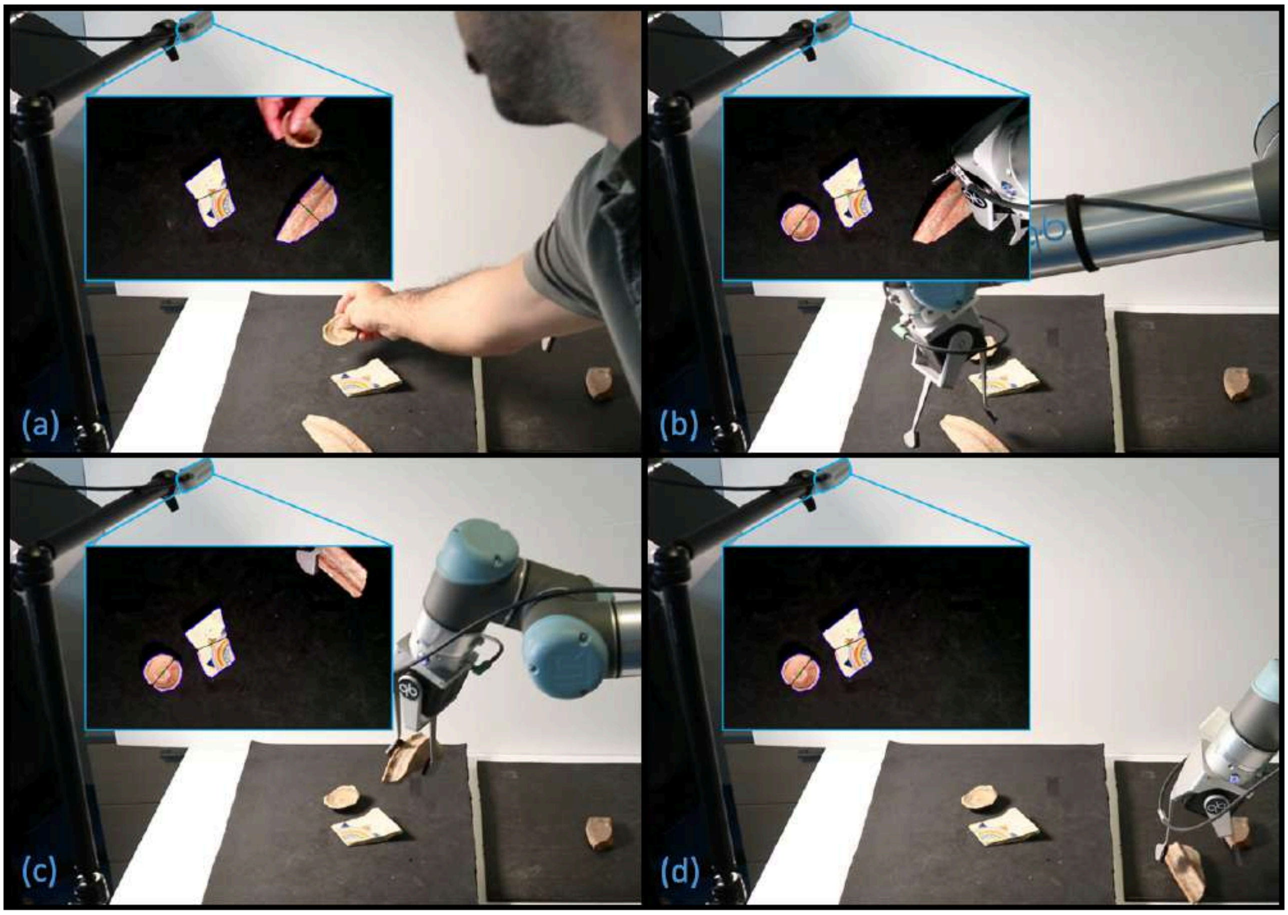

Fig. 11 - Sequential illustration of the integration between artefact detection and motion planning.

3.2.2 Results

In terms of grasping success, 100% of ceramic samples were successfully grasped, a result that can be partly explained by their relatively uniform and sufficient thickness. For lithic samples, grasping performance is closed to 60% of success rate, mainly depending on their thickness. This indicates that the observed limitations are not due to pose estimation errors, but rather to the physical constraints imposed by the tested geometry of the customised fingers, which is less effective on thinner objects. This will involve further evaluations and tests with different finger shapes to increase the efficiency. To overcome this issue and ensure consistent grasping of both ceramics and lithics across different thickness ranges, the development of a dedicated tool will be required.

Regarding execution times, it is not yet possible to provide quantitative results, as this aspect is currently out of the scope of this Deliverable. At this stage of the project, object detection and pose estimation are performed in real-time, whereas for motion planning the execution time is manually defined. In the experiments, this parameter was set to 3 seconds for the robot to move from point A to point B.

These observations confirm that the overall strategy for object detection and pose estimation is effective under the tested conditions. Each artefact placed under the RGB-D camera is successfully segmented (Fig. 12), and the pose estimation provides precise orientation information, which guides the opening and closing direction of the SoftClaw gripper (Fig. 13). Such performance ensures that the robot can approach, grasp, and manipulate objects reliably, meeting the operational objectives defined for the automated pick-and-place pipeline.

In conclusion, the integrated pipeline demonstrates robust performance and confirms the objectives outlined in Deliverable D.3.2, showing that the combined object detection and motion planning modules can reliably perform pick-and-place operations on archaeological artefacts.

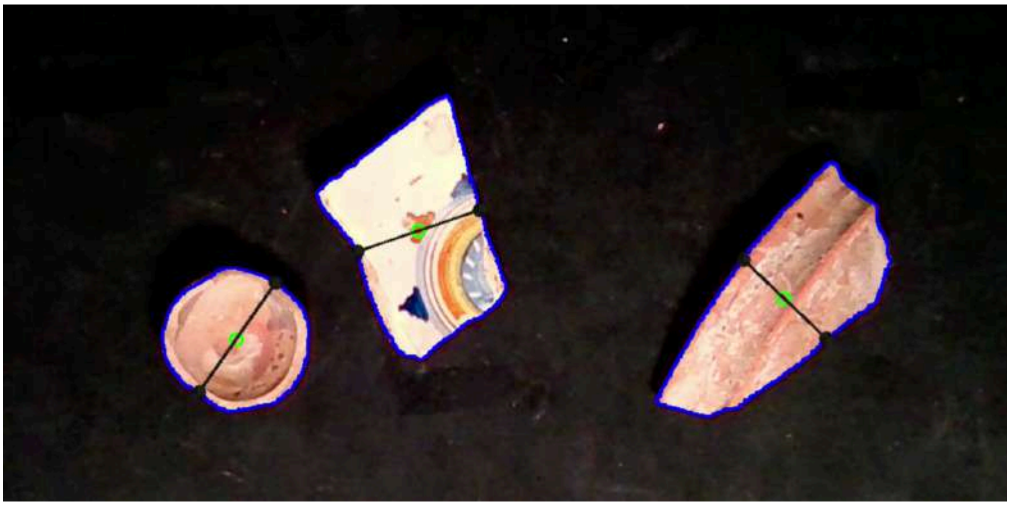

Fig. 12 - Real-time object detection, segmentation and pose estimation.

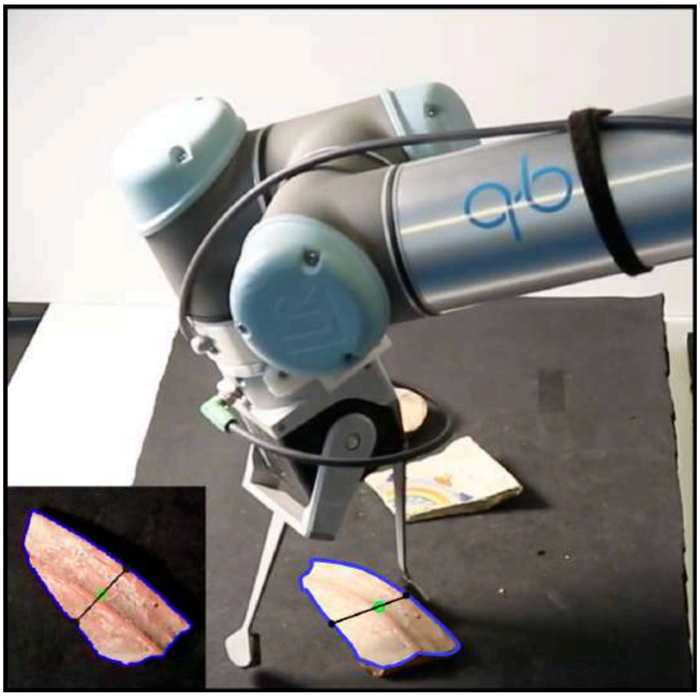

Fig. 13 - SoftClaw opening the gripper in the direction of the segment passing the centroid.

4 Graphical User Interface Design Concepts

4.1 User Requirements and Objective

The robotic cell interface is intended for users with expertise in archaeology but without specific knowledge of robotics. It should provide access to general process parameters, including real-time process status, type of analysis being performed, and progress on the samples under examination.
The interface adopts a simple and intuitive layout, comparable to that of a mobile application, with clearly identifiable buttons and a logical navigation flow. This structure will guide users in correctly entering process parameters and ensure that safety procedures are consistently followed.

4.2 Artefact Arrangement within the robotic cell

The system must be able to operate in various contexts, both where a standardised inventory system based on shelves and respective containers positions is in place, and where no such system exists. In fact, existing inventory systems may vary in terms of sophistication and level of detail. The digitisation of the finds must take these inventory codes into account and associate them with all the information acquired, then upload it to the cloud. To achieve this, the proposed solution is a container equipped with multiple slots, designed to host artefacts of different sizes. The container must therefore be modular and adaptable, so as to accommodate the variability of archaeological finds. Each slot has coordinates managed by the robotic system itself, which creates the reference framework for pick-and-place operations, as in the following figure. The operator places the artefacts in the slots and enters the corresponding inventory code through the user interface. In this way, the system establishes the link between the physical position, the inventory code, and the digital data acquired.

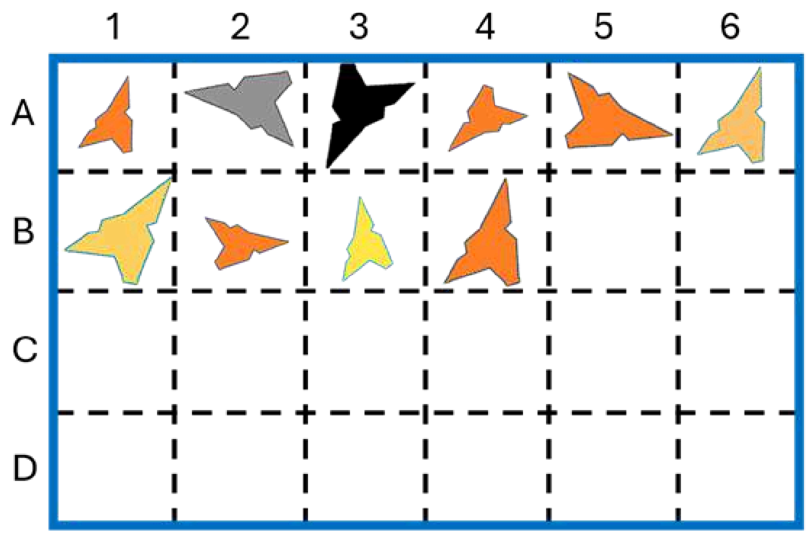

Fig. 14 - Container layout for find placement.

The system loading procedure will be:

1. Place the individual item in a slot in the container.
2. Enter the inventory code corresponding to the slot using the user interface.
3. Repeat the procedure until the container is full. If one or more slots are empty at startup, the system will still perform the task, bypassing the empty slots.
4. At the end of the process, if the system detects anomalies for a specific item. This information will be indicated in the report, and the item will be identified by code and its location in the container. At this point, the operator can select it and choose whether to perform only certain data acquisitions on that item.

4.3 Proposed Design and Mockups

As shown in Fig. 15, the concept design of the HOME page of the user interface follows a clear structure, arranged as follows:

- system status;
- general process parameters;
- quick buttons for process interaction (start and reset need confirmation from users);
- selected analysis for the process.

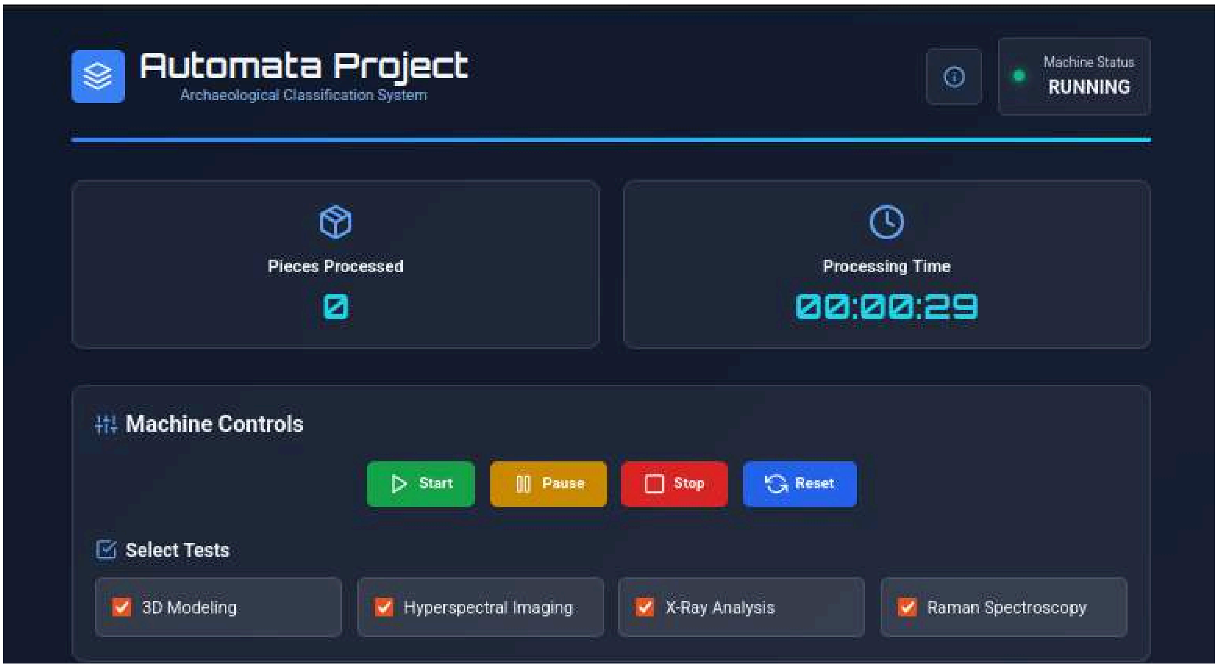

Fig. 15 - Concept of the GUI.

From the *HOME* page, users can access the other functional pages, like:

- *INFO*: general information about system release, software, contacts, etc…

- *ARTEFACTS LOADING*: When preparing the items in the reference box, equipped with a coded grid and reference to the robotic system, the operator will write the inventory code corresponding to the position chosen in the box. At this point, the system will have the codes and positions as input, so as to create a correspondence between the real object and the database.

- *PROCESS RESULTS AND WARNINGS*: list of digitised items, sorted by location, code and a picture, indicating the processing status for each type of analysis and any warnings. Once the process has been completed or interrupted by the operator, it will be possible to select the single finding or multiple findings that have encountered anomalies and perform the analysis again only on them.

5 Conclusion and Future Integrations

In Deliverable D.3.2, the development of a pipeline was presented, structured into two main modules: object detection and motion planning. The object detection module (segmentation and pose estimation) was validated for real-time use, while the motion planning module was executed through the implementation of MoveIt! framework. Finally, the complete pipeline was integrated and tested on real hardware, achieving successful results and effectively carrying out the tasks defined for the deliverable.

These achievements represent a crucial development, confirming that the proposed approach can support reliable pick-and-place operations on archaeological artefacts. At the same time, the experiments highlighted specific challenges that provide important guidance for future work. The SoftHand2 proved effective in grasping large artefacts and achieved a high success rate with medium-sized objects (around 90%). However, smaller and thinner items, particularly flat lithics, remain consistently difficult to grasp. Moreover, even when grasping is successful, the orientation of medium-sized artefacts sometimes changes once lifted, as both the hand and the object adapt to each other, leading to unintended rotations or flips. These results are consistent with the observations reported in Deliverable D3.1 and confirm that the SoftHand2 remains a suitable solution, while also requiring critical evaluation and the development of complementary strategies.

These findings highlight the need to refine gripper design and grasping strategies, while keeping in mind the trade-off with computational efficiency. In parallel, the modular container system and GUI workflow will be fundamental in ensuring that artefacts of different shapes and sizes can be safely managed and consistently linked with their corresponding inventory information.

Building on the results of D3.2, , the next steps will focus on:

- designing customised SoftClaw tools to improve the grasping of thin and flat artefacts;
- optimising the grasping strategies to efficiently handle the full range of sizes and morphologies (both thinner objects and larger ones);
- testing of the workflow also with other soft end-effectors;
- finalising the robotic cell layout and testing the pipeline developed with the definitive set of sensors and input box;;
- analysing the results obtained within the defined robotic cell layout and, if necessary, further fine-tuning the algorithm.

References

Akkaya, I., Andrychowicz, M., Chociej, M., Litwin, M., McGrew, B., Petron, A., ... & Zhang, L. (2019). Solving rubik's cube with a robot hand. *arXiv preprint* arXiv:1910.07113.

Bicchi, A. (1995). On the closure properties of robotic grasping. The International Journal of Robotics Research, 14(4), 319-334.

Cai, Z., Fan, Q., Feris, R. S., & Vasconcelos, N. (2016, September). A unified multi-scale deep convolutional neural network for fast object detection. In *European conference on computer vision* (pp. 354-370). Cham: Springer International Publishing.

Chen, P., Pei, J., Lu, W., & Li, M. (2022). A deep reinforcement learning based method for real-time path planning and dynamic obstacle avoidance. *Neurocomputing*, *497*, 64-75.

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence, 9*(2), 85-112.

Ding, J., Zhou, Y., Huang, X., Song, K., Lu, S., & Wang, L. (2023). An improved RRT* algorithm for robot path planning based on path expansion heuristic sampling. *Journal of Computational Science*, *67*, 101937.

Dong, M., & Zhang, J. (2023). A review of robotic grasp detection technology. Robotica, 41(12), 3846-3885.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., & Navab, N. (2012, November). Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In *Asian conference on computer vision* (pp. 548-562). Berlin, Heidelberg: Springer Berlin Heidelberg.

Karaman, S., & Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. *The international journal of robotics research*, *30*(7), 846-894.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Meng, B. H., Godage, I. S., & Kanj, I. (2022). RRT*-based path planning for continuum arms. *IEEE Robotics and Automation Letters*, *7*(3), 6830-6837.

OpenCV. (n.d.-a). *OpenCV*. Retrieved from https://opencv.org/

OpenCV. (n.d.-b). *Camera calibration with ArUco markers*. OpenCV Documentation. Retrieved from https://docs.opencv.org/4.x/da/d13/tutorial_aruco_calibration.html

Palleschi, A., Angelini, F., Gabellieri, C., Park, D. W., Pallottino, L., Bicchi, A., & Garabini, M. (2023). Grasp it like a pro 2.0: A data-driven approach exploiting basic shape decomposition and human data for grasping unknown objects. IEEE Transactions on Robotics, 39(5), 4016-4036.

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).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Sucan, I. A., & Chitta, S. (n.d.). "MoveIt", [Online] Available at moveit.ros.org.

Sucan, I. A., Moll, M., & Kavraki, L. E. (2012). The open motion planning library. *IEEE Robotics & Automation Magazine*, *19*(4), 72-82.

Tsirtsakis, P., Zacharis, G., Maraslidis, G. S., & Fragulis, G. F. (2025). Deep learning for object recognition: A comprehensive review of models and algorithms. International Journal of Cognitive Computing in Engineering.

Xie, Z., Liang, X., & Roberto, C. (2023). Learning-based robotic grasping: A review. Frontiers in Robotics and AI, 10, 1038658.

Yu, X., Dong, M., & Yin, W. (2022). Time-optimal trajectory planning of manipulator with simultaneously searching the optimal path. *Computer Communications*, *181*, 446-453.

Zheng, P., Gao, D., Fan, D. P., Liu, L., Laaksonen, J., Ouyang, W., & Sebe, N. (2024). Bilateral reference for high-resolution dichotomous image segmentation. *arXiv preprint arXiv:2401.03407*.

Zhiqiang, W., & Jun, L. (2017, July). A review of object detection based on convolutional neural network. In 2017 36th Chinese control conference (CCC) (pp. 11104-11109). IEEE.