

The effect of material optical properties and illuminance on the performance of an AI system in a bin-picking task involving 3D-printed parts using a collaborative robot

Jakub Dąbrowski¹, Szymon Borys¹ , Wojciech Kaczmarek^{1*}, Paweł Sobolewski¹

¹ Faculty of Mechatronics, Armament, and Aerospace, Military University of Technology,
Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland

* Corresponding author's email: wojciech.kaczmarek@wat.edu.pl

ABSTRACT

This article discusses the impact of the optical properties of materials and illuminance on the performance of an AI-based vision system in the bin-picking process. The process is carried out by an ABB GoFa CRB 15000-10/1.52 collaborative robot equipped with a Cambrian vision system. As part of the research, a dedicated test stand was developed. The study was conducted using parts produced via 3D printing from PLA and PETG materials (representing matte and reflective surface properties), analyzing the impact of illuminance, surface properties, partial occlusion of objects, and changes in their geometry on detection performance. The results showed that increasing the illuminance from 100 lx to 350–500 lx reduced the prediction generation time from approximately 0.9 s to approximately 0.3 s. It was noted that higher detection efficiency was achieved for workpieces with matte surfaces compared to reflective surfaces. The developed system correctly detected potential collisions and recognized the objects it had been trained to identify, even when they were partially obscured. Research has shown that a change in the part's geometry prevented its proper identification.

Keywords: bin picking, AI, vision system, collaborative robot, machine learning.

INTRODUCTION

Robotic bin-picking is an automated process involving the extraction of individual, randomly oriented parts from a container. The purpose of this process is to determine the position and orientation of the part in space, pick it up, and then transfer it to the next stage of the production process. In addition to the random arrangement of the parts, another challenge is that the objects obscure each other, which means they may be only partially visible to the vision system. When retrieving parts, there is a risk of collision with other items in the container, with the container itself, and with the robot. Another potential problem is the variety of part geometries – small components, irregular shapes, and reflective surfaces can make it difficult for the vision system to detect them correctly.

Vision systems are used to automate the bin-picking process; they play a key role in modern automation. These are systems consisting of cameras, lighting, computers, and specialized software. The recorded camera footage is analyzed by software that uses advanced image processing algorithms and, increasingly, artificial intelligence (AI), which enables the software to recognize shapes, colors, and textures, as well as compare images to reference patterns. Computers are used to process data from cameras and software and to make decisions based on the analyzed images. They enable the automation of industrial processes, primarily by increasing productivity and efficiency, reducing operating costs, and minimizing the need for human resources.

AI-powered vision systems enable the detection, localization, and orientation of randomly arranged items in containers, even when they

are partially obscured or under varying lighting conditions. Neural network models enable the automatic generation of robot motion trajectories in real time based on data derived from recorded camera images.

A distinctive feature of this study is the evaluation of artificial intelligence algorithms in relation to additively manufactured parts made of polymers (PLA and PETG), whose surfaces exhibit a specific visual anisotropy resulting from their layered structure. In addition, the integration of workspace design and collision analysis within the RobotStudio environment lays the foundation for building a digital twin of an automated workstation. Unlike studies focused on developing novel artificial intelligence algorithms, the primary contribution of this paper is the comprehensive practical evaluation and integration of an existing commercial AI-assisted vision system (Cam-brian) within a collaborative robotic workstation (ABB GoFa). This study aims to systematically assess the workstation's performance limitations by investigating the influence of varying illuminance levels, surface optical properties (matte vs. reflective), partial occlusion, and geometric modifications on the system's operational efficiency.

BACKGROUND

In recent years, advances in deep learning have significantly changed the way bin-picking tasks are performed, leading to a shift from traditional methods based on deterministic geometric matching to solutions that use data-driven models [1]. Currently, research focuses primarily on three areas: the use of artificial intelligence for grasp perception and planning, the estimation of the position and orientation of objects in 6D space, and the integration of these methods into real-world industrial settings.

The use of deep neural networks (DNNs) in robotic manipulation has evolved toward grasp planning methods. The solutions used can be divided into gripper-oriented methods and object-oriented methods, which differ in terms of how the grasp is represented, the scope of the information used, and the process of preparing training data [1]. Regardless of the approach taken, modern systems are increasingly integrating object detection, grip parameter estimation, and manipulator motion planning within a single deep learning architecture.

Initially, these solutions focused primarily on detecting 2D objects and determining oriented bounding boxes, which are used to define grips on flat surfaces [2]. The development of convolutional neural networks (CNNs) has made it possible to significantly improve the effectiveness of object detection while maintaining high computational efficiency. Currently, CNN architectures form the basis of most systems that perform object classification, detection, and segmentation in industrial applications, providing greater robustness against variations in object orientation, partial occlusions, and changing lighting conditions [3, 4]. Significant progress has been made through the use of architectures belonging to the YOLO (you only look once) family, which enable near-real-time object detection. These solutions have been successfully used to identify and locate randomly distributed parts inside industrial containers [3]. The implementation of detection and classification methods based on the YOLO architecture is not limited solely to stationary robotic arms, but is also successfully used in mobile robotics. Research conducted using low-cost computing devices (e.g., Raspberry Pi) has shown that deep learning algorithms provide an optimal balance between mean average precision (mAP) and real-time CPU computational power consumption [5]. An extension of this approach is the integration of stereo cameras with real-time object detectors, which enables the simultaneous determination of objects' spatial coordinates and their use during pick-and-place operations [4]. In addition, the use of deep learning methods to improve the stereo matching process makes it possible to increase the accuracy of depth reconstruction, especially for small features with non-distinctive geometry [6].

The growing complexity of manufacturing environments, particularly in low-volume and high-variety production, has led to increased interest in reinforcement learning (RL). Unlike traditional supervised methods, they enable adaptive refinement of grasping strategies based on the robot's successive interactions with its environment, reducing the need for manual system configuration for each new object [7, 8]. One of the major areas of development is the integration of RL algorithms with computer vision methods. The combination of feature extraction using the YOLO architecture with the Soft Actor-Critic (SAC) algorithm enables simultaneous object identification and optimization of the grasping process in an unstructured working environment [8].

As the level of autonomy of robotic systems increases, issues related to trajectory planning and robotic arm control are becoming increasingly important. Modern solutions use deep neural networks to solve inverse kinematics, path planning, and collision avoidance problems, gradually reducing the use of classical numerical methods [9]. At the same time, neural network models are being developed to identify the manipulators' workspace and compensate for positioning errors, which helps increase the accuracy of manipulation operations and improve process repeatability [10, 11].

As the complexity of industrial bin-picking processes increases, relying solely on object detection based on two-dimensional images and oriented bounding boxes is no longer sufficient. Implementing such applications requires a complete estimation of an object's position and orientation in six-dimensional space (6D pose estimation), which includes three degrees of freedom for translation and three degrees of freedom for rotation. This information is necessary to determine a collision-free trajectory for the manipulator and to ensure the correct orientation of the gripper when retrieving parts from densely packed containers [12].

In traditional approaches, 6D position estimation is based on fitting measurement data to a CAD geometric model. In industrial practice, methods that use geometric feature matching are employed alongside solutions based on deep neural networks for the direct estimation of an object's position. Methods that use point pair features (PPF) and their extensions are particularly important, as they enable the effective localization of objects even in cases of partial occlusions and incomplete point clouds [13]. Hybrid solutions are also being used more and more frequently, in which deep learning models are responsible for detecting or initially localizing an object, while the final position estimation is performed using model-based matching against a CAD model [14, 15]. CAD models also play a significant role in the training process for 6D position estimation systems. They can be used to generate synthetic data, which involves artificially increasing the amount of training data through transformations of existing samples (data augmentation) and random modification of the simulation environment (domain randomization), which reduces the need to collect large sets of real-world data [16].

Despite the high effectiveness of modern methods, most deep learning models still require the preparation of dedicated training datasets for each new object, which has spurred the development of zero-shot 6D pose estimation methods (zero-shot 6D pose estimation), which enable the localization of objects not present during the training process [17]. One of the most advanced solutions is the ZeroBP architecture, which uses position-aware correspondences between real-world observations and CAD models. Combining local geometric features with global spatial information helps reduce the misalignment errors typical of industrial components that lack texture and exhibit significant geometric similarity [17]. Similar assumptions were also adopted in solutions using generalized, recursive flow correction networks, which enable the estimation of the positions of previously unseen objects using predefined geometric elements of CAD models, even under conditions of significant occlusion [18].

The physical properties of the manipulated components remain a significant challenge. The accuracy of 6D position estimation depends to a large extent on the quality of the spatial data provided by the vision system. Industrial environments are dominated by metal components that lack texture and often have highly reflective surfaces, resulting in incomplete or noisy point clouds, which makes it difficult to reconstruct geometry and locate objects, especially in cases of partial occlusion [19, 20]. For this reason, deep learning models are being developed for the direct estimation of 6D positions under conditions of limited texture, monochromatic imaging, and significant occlusions, as well as methods for filtering and reconstructing ordered point clouds, which improve the accuracy of geometric model fitting [15, 19]. Specular reflections are an additional source of errors that degrade the quality of data acquired by stereo vision systems. Modern solutions use deep learning models to mitigate the impact of specular reflections on the input data of stereo vision systems, which leads to increased stability in object position estimation and improved reliability of industrial bin-picking systems [20].

Industrial bin-picking systems must meet requirements for reliability, response time, and integration with advanced automation components. Issues related to real-time communication, collaboration with PLCs, and minimizing total cycle time are becoming increasingly important [21].

A key aspect of such implementations is ensuring the stable integration of image-analysis software with the actuator system. A commonly used solution is data exchange via the TCP/IP protocol, which enables the dynamic transmission of coordinates to a manipulator in industrial sorting (picking) processes based on camera images [22]. The use of external 3D stereo vision systems equipped with AI components is becoming essential to overcome the problem of the limited field of view (FOV) of standard solutions. This allows for a significant expansion of the entire system's digital situational awareness and maintains high positioning accuracy for the industrial robot despite the closed architecture of the control systems [23].

One area of development is the implementation of deep learning algorithms directly within the industrial controller environment. The integration of capture networks, such as Dex-Net, with AI hardware acceleration modules that work in conjunction with PLCs enables real-time inference while maintaining low power consumption. Solutions of this type have demonstrated the ability to achieve a throughput exceeding 350 retrieval operations per hour, confirming the practical applicability of Edge AI architecture in industrial applications [23].

In industrial practice, however, the efficiency of a workstation is determined not only by the time it takes for the artificial intelligence model to perform calculations, but also by the total cycle time, which includes data acquisition, image processing, motion planning, and the execution of the manipulation operation. For this reason, current research focuses on carrying out the individual stages of the process in parallel. The use of synthetic training data generators eliminates the time-consuming process of manually collecting and labeling data, while the use of a parallel workflow allows calculations related to the next grasping operation to be performed while the manipulator is in its return motion. This makes it possible to virtually eliminate downtime caused by the robot waiting for the artificial intelligence model to finish its calculations [24].

An important aspect of implementation remains the integration of artificial intelligence algorithms with the control systems of industrial robots. The gripper position determined by the vision system must be transformed into the robot's coordinate system and used during motion path planning, taking into account kinematic

and collision constraints. In industrial practice, solutions that utilize deep learning algorithms – including at the motion control level – are increasingly being used to directly generate trajectories for industrial robots [9].

The effectiveness of this approach has been confirmed, among other things, in solutions that use the Vision-type Deep Deterministic Policy Gradient (Vision-DDPG) algorithm, which has been validated both in a simulation environment and when controlling actual ABB industrial robots. Research results indicate that combining visual information with deep learning algorithms enables effective planning of a manipulator's motion trajectory while maintaining the required positioning accuracy [25]. A similar direction of development is represented by solutions that integrate deep-learning-based object detectors with classical robot arm control algorithms, ensuring the stability of the grasping process and high accuracy in performing assembly and manipulation tasks [26].

A separate, yet increasingly important, area of research is virtual commissioning of robotic handling processes. The use of environments such as RobotStudio allows for the modeling and verification of the functionality of specialized grippers even before they are physically manufactured and subjected to experimental vibration or inertial tests [27, 28]. The use of virtual controller technology enables multi-criteria simulations and the optimization of test bench operating parameters based on randomly generated data. Research shows that the proper selection of control algorithms not only allows for determining the optimal speed of robots and conveyors for a given throughput, but also ensures that the production process continues automatically in the event of an emergency.

MATERIALS AND METHODS

Design of a robotic workstation for the bin-picking process

The test setup for conducting research on a robotic process for retrieving parts from a container using a vision system and artificial intelligence models is shown in Figure 1. The test stand is an integrated system consisting of several layers: control, communication, data processing, and execution. Communication between devices is handled directly by the ABB OmniCore C30

robot controller, which serves as the central control unit for the entire workstation. Communication between the robot controller, the Cambrian system's computing unit, and the computer running the RobotStudio environment is carried out using an Ethernet connection and the TCP/IP protocol. The vision system is based on a Cambrian stereo camera mounted on the manipulator arm. The camera transmits image data to the Cambrian computing unit, which performs image processing using artificial intelligence models. The vision system communicates with the robot controller using sockets to send and receive data based on the TCP/IP protocol. All devices operate on the same subnet.

A general structural and functional diagram of the workstation is shown in Figure 2.

Two types of objects were designed to conduct research on a robotic process for retrieving parts from a container. This made it possible to verify the accuracy of distinguishing between parts and the methods used to pick them up, given different orientations and positions of the gripper. To investigate the impact of the physical properties of the manipulated parts on the accuracy of their detection, the designed models were printed on a BambuLab X1 Carbon 3D printer using the FDM/FFF (Fused Deposition Modeling/Fused Filament Fabrication) method using various materials, in different colors, and with varying degrees of surface gloss. The matte elements were printed using Polymaker PolyLite filament PLA (polylactic acid), and the reflective elements were printed using Devil Design PET-G filament (polyethylene terephthalate

glycol-modified) (Figure 3). In both cases, filaments were used without additives or admixtures that modify the material's optical properties.

To demonstrate the basic principle of the bin-picking process – which involves retrieving unordered parts from a container and transferring them to a location where they are sorted and placed – a receiving tray designed to position the picked objects was also developed and printed on a 3D printer (Figure 4). The tray has slots designed to accommodate two types of parts: those shaped like a rectangular prism and those shaped like a cylinder. It was assumed that the part could be positioned in two equivalent orientations differing by a rotation of 180°, which eliminated the need to design positioning recesses for the bottle necks as well. This approach made it possible to limit the required rotation of the gripper to a maximum of 90°, which helped reduce the process cycle time.

The first step in preparing CAD models for training is to import them into the Cambrian Training software (Figure 5).

During training, the neural network model learns to recognize and distinguish the type of detail it sees, and to determine its position and orientation in space based on its geometric features – this approach can be used to detect multiple objects in a single scene, particularly in sorting processes. To obtain a properly trained AI model, it is necessary to configure the training parameters appropriately so that the object is clearly visible in the workspace and so that the depth range corresponds to the system's actual operating conditions.

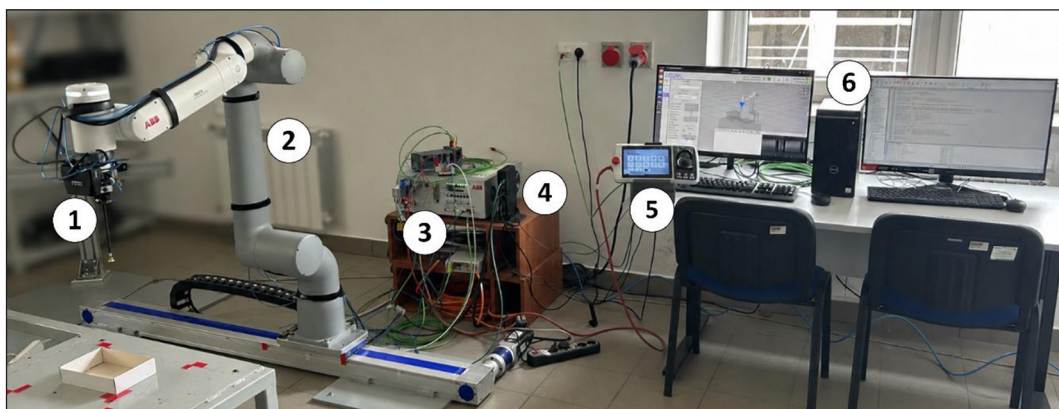


Figure 1. Robotic workstation with a collaborative robot: 1) Cambrian stereo camera, 2) ABB GoFa CRB 15000-10/1.52 collaborative robot, 3) ABB OmniCore C30 robot controller, 4) Cambrian system computing unit, 5) ABB FlexPendant operator panel, 6) computer with RobotStudio software

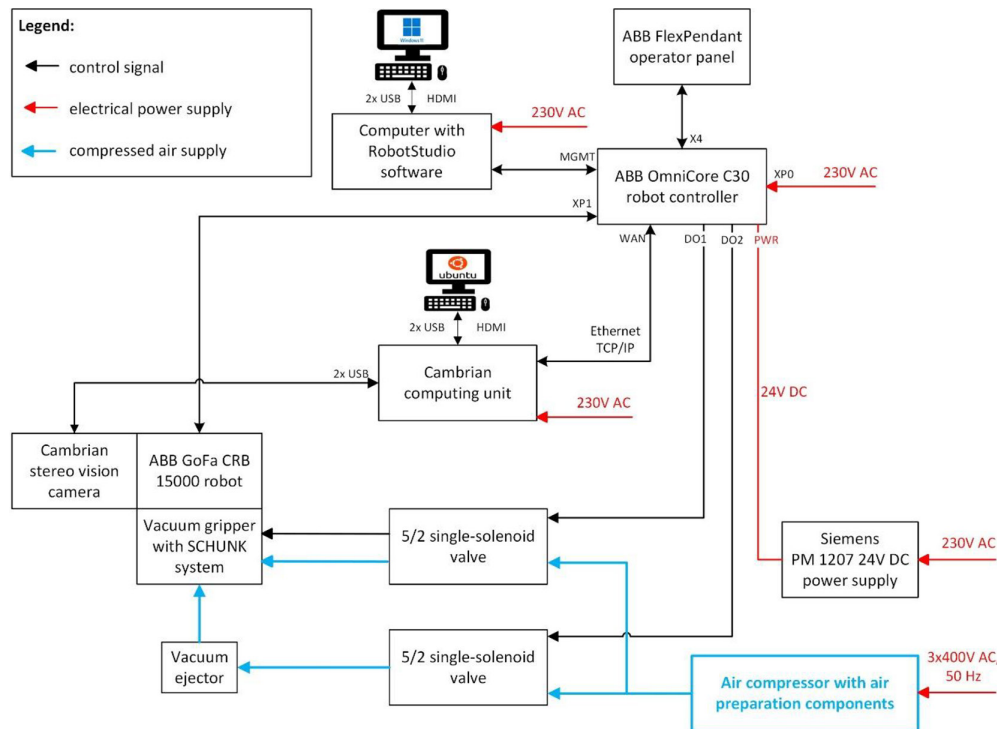


Figure 2. General structural and functional diagram of a bin-picking workstation

An important parameter that requires configuration is the “Z Center,” which is the distance between the camera position and the parts to be captured, as well as the “Z range” for the captured parts (Figure 6). The combination of these two values defines the distance range within which the AI model will detect the object. Workpieces outside this working range will be filtered out. Once the model has been prepared in the Cambrian Training program, the next step is to train it – this is done using the Cambrian Training Portal, an online platform. It generates a personalized AI model based on the submitted object configuration.



Figure 3. Components prepared for conducting bin-picking process tests

The grip points added in the Cambrian VISION environment (indicated by yellow arrows in Figure 7) made it possible to define the position and orientation of the manipulator’s TCP relative to the object, as well as the collision fields, ensuring correct and safe grasping of the object. These were defined taking into account the maximum diameter of the working part of the vacuum gripper tip used in the tests.

In the grip point editor, the following were defined, among other things: the position and orientation of the grip point, the grip type, and



Figure 4. A collection tray prepared for the bin-picking process

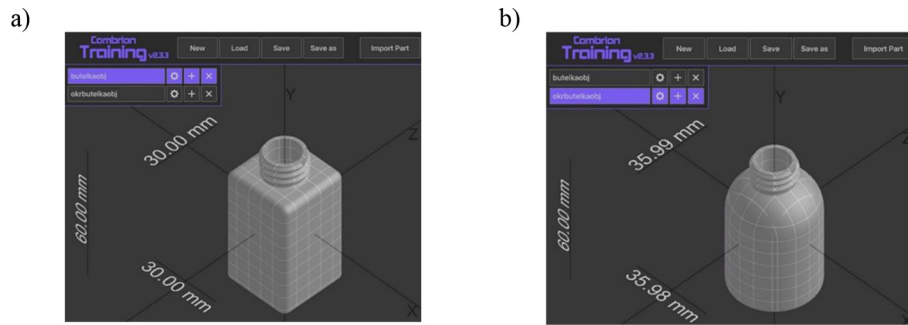


Figure 5. CAD parts imported into the Cambrian Training software: a) a cuboid-shaped bottle model, b) a bottle model with a circular cross-section

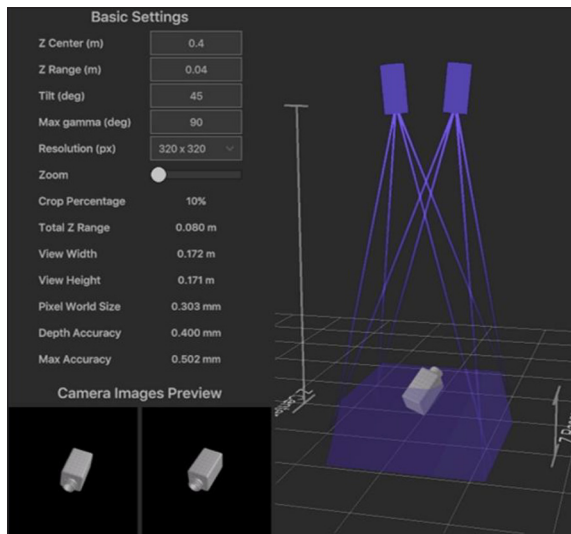


Figure 6. View of training settings and model preview areas in the Cambrian Training software

the symmetry configuration. The gripper type is responsible for assigning an ID to a specific gripper point, which is passed to the robot's API (Application Programming Interface) and influences how the robot approaches grasping objects.

Symmetry types affect how grip orientations are defined and the number of grip points required. For the cuboid-shaped part, the "None" symmetry configuration was selected because the object does not have continuous rotational symmetry. Due to the presence of edges that prevent the workpieces from being gripped with a suction cup, it is necessary to clearly specify the grip orientation. For the bottle-shaped feature with a circular cross-section, a cylindrical symmetry configuration was used because the object has continuous rotational symmetry about a defined vertical axis. Because the grip point is defined on the side surface, rotating the workpiece around a defined axis does not affect the accuracy of the grip.

Defining the workspace – which includes the gripper and the table on which the container is placed (Figure 8) – ensured collision-free part retrieval.

In the AI models panel, the following settings were configured and parameterized: the Confidence Threshold (predictions with a confidence level lower than the selected threshold are not returned by the model) and the Confidence Filter – an additional filter that eliminates predictions below a specified value (such predictions are not sent to the robot). The threshold value and the confidence filter divide the predictions into three groups:

- below the threshold value – completely rejected, not sent to the bot, and not visible in the interface,
- between the confidence threshold and the filter values – displayed in the interface (in orange) but not used by the robot; additionally taken into account in the collision analysis,
- above the filter value – sent to the robot, displayed in the interface (in green or blue), and included in the collision analysis.

The control program was developed in RAPID as a module integrated with the communication module. The main function of the API is to send the robot's current status to the vision system and receive the results of the image analysis. The data sent includes the robot's current TCP position, the current manipulator axis values, the definition of the active tool, and additional parameters related to the gripper configuration. The response from the vision system, received by the robot controller, takes the form of raw data, which is then converted into working coordinates used by the control program. They contain predictions – that is, individual grip proposals for a part generated by the AI model – which include the following information: detection confidence score ("CONF"), part

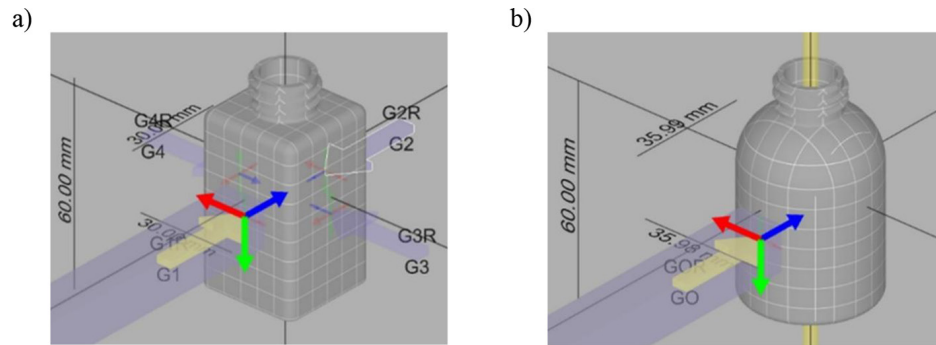


Figure 7. Defined grip points and collision zones in the Cambrian VISION software: a) “None” symmetry, b) cylindrical symmetry

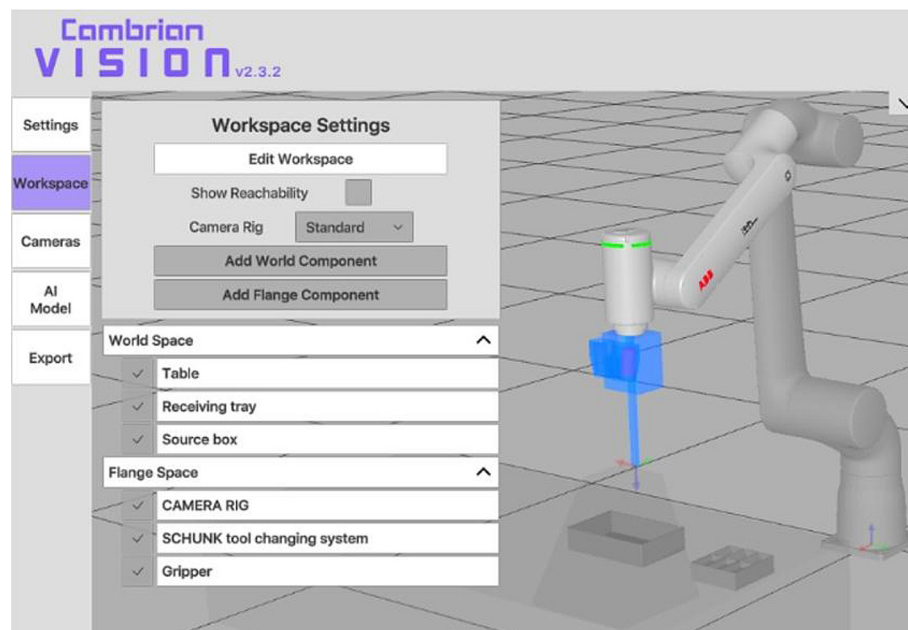


Figure 8. Configuring the workspace in the Cambrian VISION software

visibility index (“PVIS”), the object type number (“OBJ”), the grip type identifier (“GRP”), the orientation (“RX”, “RY”, “RZ”), its distance from the TCP (“DIST”), and any rejection criteria (“Rejection Criteria”). Only those predictions whose confidence score is greater than the confidence threshold specified in the AI model settings are generated and returned.

A general diagram of the control program’s operation is shown in Figure 9.

The flowchart for the workstation consists of four main process steps:

- system initialization, which is responsible for connecting the robot controller to the vision system’s computing unit,
- searching the workspace grid until the correct part predictions are found,

- a collision-free approach to the part and its grip,
- Collision-free placement of the workpiece in the designated location.

To ensure the reliability and reproducibility of the study, the experimental conditions were strictly controlled. The illuminance of the workspace was measured using a CEM DT-8809A digital lux meter. To ensure uniform light distribution analysis, measurements were taken at five distinct points: the four corners of the working table and the center of the picking area. The light source consisted of two halogen lamps located on opposite sides of the working table. During the evaluation of different illuminance levels (100, 350, and 500 lx), the Cambrian vision system’s camera settings were kept constant to evaluate the system’s robustness

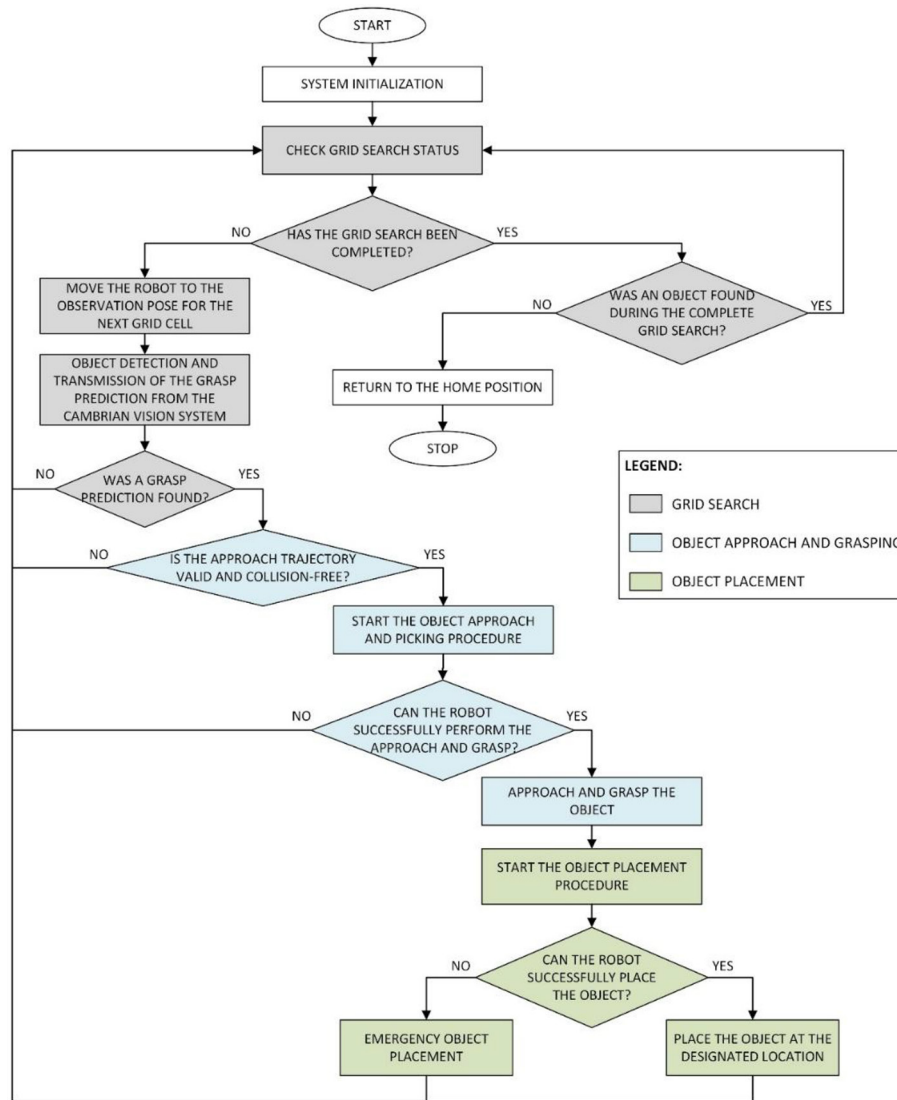


Figure 9. General flowchart of the control program

to varying lightning conditions. Furthermore, it is crucial to emphasize that the exact same pre-trained AI model was applied in every test iteration. To isolate the variables under investigation, comparable pseudo-random object arrangements and identical system parameters were maintained across all lighting and material conditions.

RESEARCH AND RESULTS

As part of the work, studies were conducted on the accuracy of the workspace grid search, the accuracy of collision detection, the impact of workspace lighting on prediction generation time, the effectiveness of detecting partially occluded objects, and the recognition of objects with geometries similar to the trained model.

Testing the correctness of the workspace grid search

To ensure complete visual coverage of the scene being analyzed, the area from which workpieces are extracted (the container space) was divided into a virtual working grid. This division is due to the limited field of view (FOV) of the stereo camera mounted on the robot's arm (eye-in-hand configuration) and the need to maintain an optimal scanning distance. This helps avoid perspective distortion, reduces blind spots at the edges of the container, and maintains a consistently high image resolution, which is essential for artificial intelligence algorithms.

During the grid search procedure, the manipulator sequentially moves the camera over successive defined zones. For each of these positions,

the AI model makes a prediction aimed at locating objects and estimating their orientation in space. Typically, immediately after the control program is launched, the first predictions generated are rejected by the system because they do not meet the rigorous criteria for a collision-free and stable grasp. The situation changes when the camera reaches a position within the AI model's optimal, acceptable operating range – defined in the Cambrian Training environment's training parameters using the “Z Center” parameter. In that case, the system correctly identifies the position and orientation of the parts. Such predictions are classified as correct, which is indicated in the visualization interface by the color green or blue, and simultaneously serves as a trigger to initiate the robot's approach trajectory and the retrieval of the part (Figure 10).

The effectiveness and stability of the implemented grid search algorithm were validated by running the process ten times. In all tested cases, the system executed the specified procedure without any recorded failures in the conducted tests, consistently locating objects within the workspace and generating correct grasping predictions for them.

Testing the accuracy of collision detection

To validate the implemented security mechanisms, tests were conducted on the collision detection algorithms. In the first stage, the accuracy of detection was verified for predefined collision zones within the manipulator's workspace. Thanks to the accurate mapping of the positions

of the source container and the receiving tray in the Cambrian VISION environment (in accordance with their actual layout), the system correctly identified workpieces and generated grasping predictions without registering false collisions with environmental elements (Figure 10). In addition, as part of the boundary condition tests, the scenario was deliberately set up in such a way that a collision-free grasp was physically impossible. These objects were correctly detected by the vision system (as indicated by the pink color); however, as intended, the system prevented the generation of working predictions for them (Figure 11). This analysis of collision zones a priori makes it possible to eliminate risky trajectories of the kinematic system at an early stage of data processing, even before the robot begins to move.

The second phase of the research involved verifying the virtual collision zones defined for the tool mounted on the robot's flange. It has been shown that even when a very high detection confidence coefficient (CONF equal to 0.99) is achieved, the overarching supervision mechanism correctly rejected a prediction if it entailed the risk of the gripper colliding with an obstacle. This situation triggered a “GP Collision” message (Figure 12), and the rejected grip proposals were clearly marked in red both in the visualization window and in the results table.

To confirm the stability and reliability of the safety mechanisms discussed, a series of ten experimental tests was conducted, with varying degrees of scene complexity. In each test, the position, orientation, and mutual occlusion of the workpieces within the container were randomly modified. In

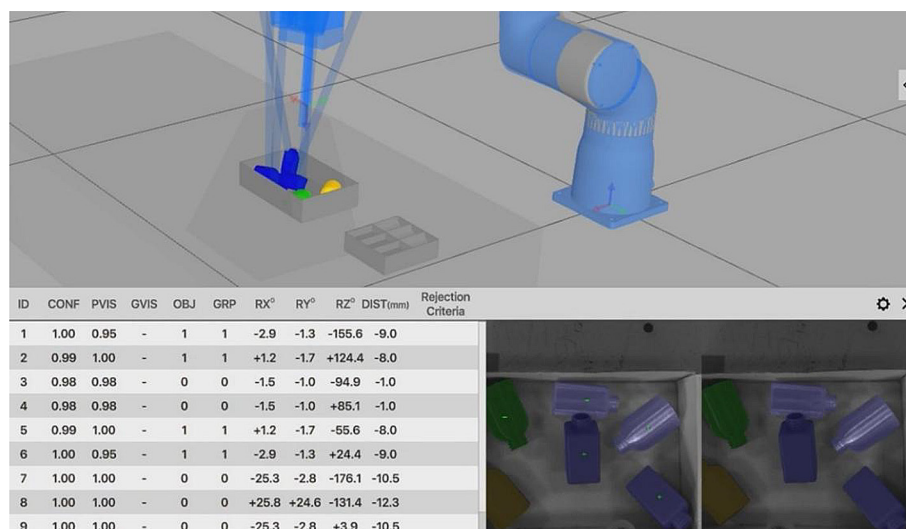


Figure 10. System in action – correct grid search

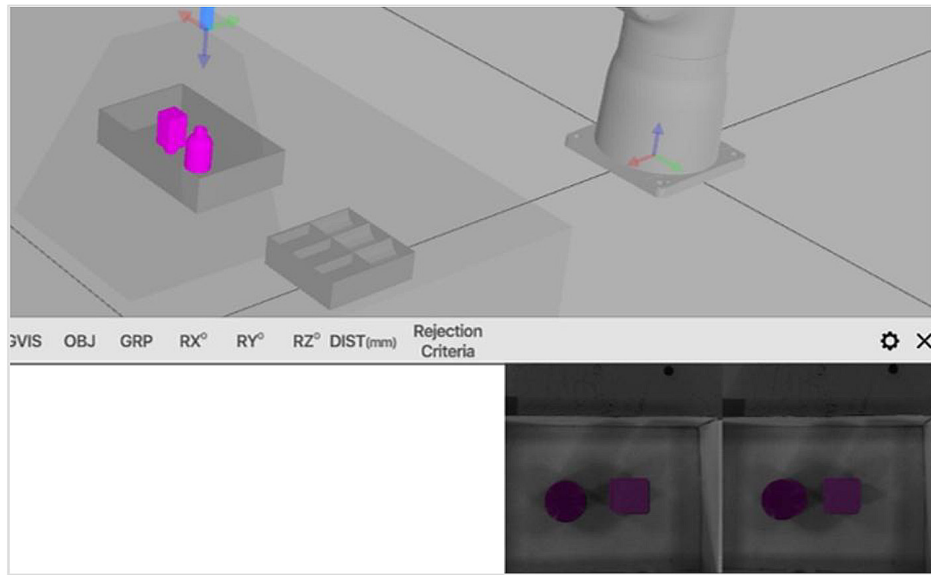


Figure 11. System in action – collision detection in the workspace

all the tests conducted, the integrated anti-collision system functioned without any recorded failures – the system consistently identified safe paths and blocked trajectories that could lead to a collision with the surroundings or with itself.

A study of the effect of workspace lighting on the time required to generate predictions

To assess the impact of lighting conditions in the workspace on the performance of the vision system, a series of experiments was conducted for three different illuminance levels: 100 lx, 350 lx, and 500 lx. The selection of the above levels was based on the guidelines of the PN-EN 12464-1 standard regarding the lighting of indoor workplaces and also reflects the typical photometric conditions found at actual automated industrial workstations. The purpose of the analysis was to evaluate the correlation between the light intensity used and key temporal parameters of the process, including, above all, the time taken for the AI model to generate a prediction and the total time required to complete the manipulation cycle. The collected measurement data, recorded in real time by the control system, were aggregated and presented in detail in Table 1.

For each illuminance level (100 lx, 350 lx, 500 lx), a series of ten ($n=10$) trials was conducted, during which the time taken by the neural network to generate a prediction and the total cycle time were recorded. For the collected samples, an

unbiased estimator of the standard deviation (SD) was calculated, defined by the formula:

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (1)$$

where: x_i denotes the next measurement result, $\bar{x}$ is the arithmetic mean of a given series, and n is the number of measurement trials.

To ensure that the assessment of dispersion is independent of the order of magnitude of the values under study, the coefficient of variation (CV) was also calculated. This parameter expresses the standard deviation as a percentage relative to the mean:

$$CV = \frac{SD}{\bar{x}} \times 100\% \quad (2)$$

By using this parameter, the results indicate which architectural elements (the neural network or the manipulator's kinematics) constitute the actual bottleneck in the automated process (Table 2).

Analysis of the measurement data revealed a clear correlation between illuminance and the performance of the vision system. It was observed that increasing the light intensity results in a three-fold reduction in the time required to generate a prediction. Under conditions of insufficient lighting (100 lx), the average time to initiate a grasp was approximately 0.9 s, whereas at 350 lx and 500 lx, this time was reduced to approximately 0.3 s (Table 1).

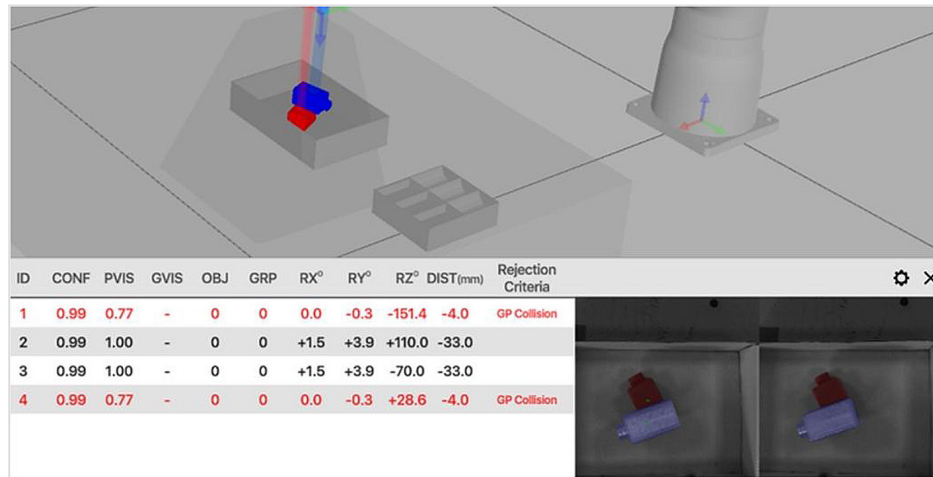


Figure 12. System operation view – tool collision detection

Table 1. Comparison of the time it takes to receive predictions from the vision system under different light intensities: 100 lx, 350 lx, 500 lx

Measurement	100 lx		350 lx		500 lx	
	Prediction time [s]	Cycle time [s]	Prediction time [s]	Cycle time [s]	Prediction time [s]	Cycle time [s]
1	0.89	11.40	0.30	12.18	0.26	10.81
2	0.91	14.24	0.29	12.96	0.28	13.85
3	0.89	13.36	0.31	13.36	0.28	13.17
4	0.90	12.58	0.30	12.74	0.27	11.64
5	0.88	11.92	0.29	11.98	0.27	12.42
6	0.91	13.81	0.31	13.22	0.28	12.95
7	0.90	12.97	0.30	12.43	0.26	11.28
8	0.89	13.44	0.30	13.05	0.27	13.44
9	0.92	14.08	0.29	12.61	0.27	12.83
10	0.90	12.35	0.31	13.11	0.28	11.96
Average	0.90	13.02	0.30	12.76	0.27	12.44

Table 2 shows that the artificial intelligence system has a highly stable and deterministic prediction time, with a very low standard deviation and coefficient of variation (less than 3%). In contrast, the total cycle time varies much more widely, which demonstrates that the main bottleneck and source of fluctuations in the analyzed process are the physical movements of the manipulator resulting from the random arrangement of the workpieces.

At the same time, it was shown that the improvement in photometric parameters at the test station did not significantly reduce the total cycle time. The lack of a clear effect in this area stems from the specific nature of the process under study, in which the physical displacements of the manipulator's kinematic system remain

the dominant temporal component. The duration of each trajectory – from approaching the item, through picking it up, to placing it back in the slot – varies and is determined by the random initial positions of the objects in the container, which, overall, negates the time savings achieved by the artificial intelligence model.

A study on the impact of workspace lighting on confidence levels

An analysis of the effect of lighting conditions on the detection confidence factor (CONF) was conducted for corresponding light intensity levels (100 lx, 350 lx, 500 lx). A summary of the predictions obtained, broken down by type of surface studied, is shown in Figure

Table 2. A statistical analysis of bin-picking process cycle times

Illuminance [lx]	Average prediction time $\bar{x}$ [s]	Standard deviation (SD) [s]	Co-... coefficient of variation (CV) [%]	Average cycle time $\bar{x}$ [s]	Standard deviation (SD) [s]	Co-... coefficient of variation (CV) [%]
100	0.9	± 0.012	1.33	13.02	± 0.943	7.24
350	0.3	± 0.008	2.66	12.76	± 0.458	3.59
500	0.27	± 0.008	2.96	12.44	± 0.988	7.94

13. The results of the experiment showed that, for parts with a matte finish, no change was observed within the tested range of light intensity, consistently achieving a maximum CONF coefficient of 1.0.

Different phenomena were observed in the case of elements with a reflective surface. At a baseline illuminance of 100 lx, a detection reliability of 0.99 was recorded (Figure 13a). Despite the high numerical score, the image acquisition was characterized by significant underexposure of the scene and low contrast between the workpieces and the background, which directly correlated with the increased inference time of the AI model. Although the system functioned properly under these conditions, the acquired image showed increased susceptibility to interference resulting from possible dynamic changes in ambient lighting.

Increasing the illuminance to 350 lx resulted in a decrease in the CONF coefficient to 0.95 (Figure 13b). The workspace was illuminated more evenly, and the image itself became brighter. Nevertheless, the specific properties of the reflective material led to the formation of directional reflections. This phenomenon caused localized overexposure in the image and blurred the edges of objects, which made it difficult for the AI model to unambiguously extract geometric features and reduced the confidence of the predictions.

At the highest illuminance level tested, 500 lx, the confidence coefficient increased again, reaching a value of 0.98 (Figure 13c). Achieving such a high level of brightness significantly improved the overall image sharpness and the visibility of the contours of the objects under examination. Nevertheless, light reflections on reflective surfaces continued to be a source of interference, causing a slight decrease in recognition confidence compared to the ideal results obtained for matte objects.

It should be emphasised that the CONF coefficient reflects the level of confidence of

the artificial intelligence model in recognising objects and estimating their spatial position, which should be distinguished from the actual physical performance of the robotic system. A CONF value of 1.0 indicates the model's maximum mathematical confidence, but this does not automatically imply a 100 per cent success rate in the physical execution of the task. In this study, however, the physical retrieval of objects was carried out following the generation of correct predictions, and no grasping failures were recorded in the trials conducted, indicating the high performance of the workstation.

Analysis of the effectiveness of detecting partially obscured features

To evaluate the artificial intelligence model's resilience to the phenomenon of partial occlusion (obstruction) of the manipulated elements, a series of ten test runs was conducted. As part of the experiment, both types of parts under analysis were intentionally obscured to varying degrees by an opaque barrier (a white sheet of paper), with the degree of occlusion gradually increased. In cases of partial occlusion, the vision system correctly located and classified objects, which was indicated in the software's graphical interface by an orange mask superimposed on the camera image (Figure 14). Despite correctly detecting the presence of workpieces, the master control algorithm deliberately did not generate a grasp prediction for any of the trials conducted. This phenomenon is the intended behavior of the safety system—the rejection of a potential sampling point resulted from the detection confidence factor falling below the specified filtering threshold or from the part visibility index failing to meet the minimum required value. In two of the ten tests, the degree of occlusion exceeded approximately 70% of the part's surface area, which prevented the vision system from recognizing it (no color mask was superimposed on the camera images).

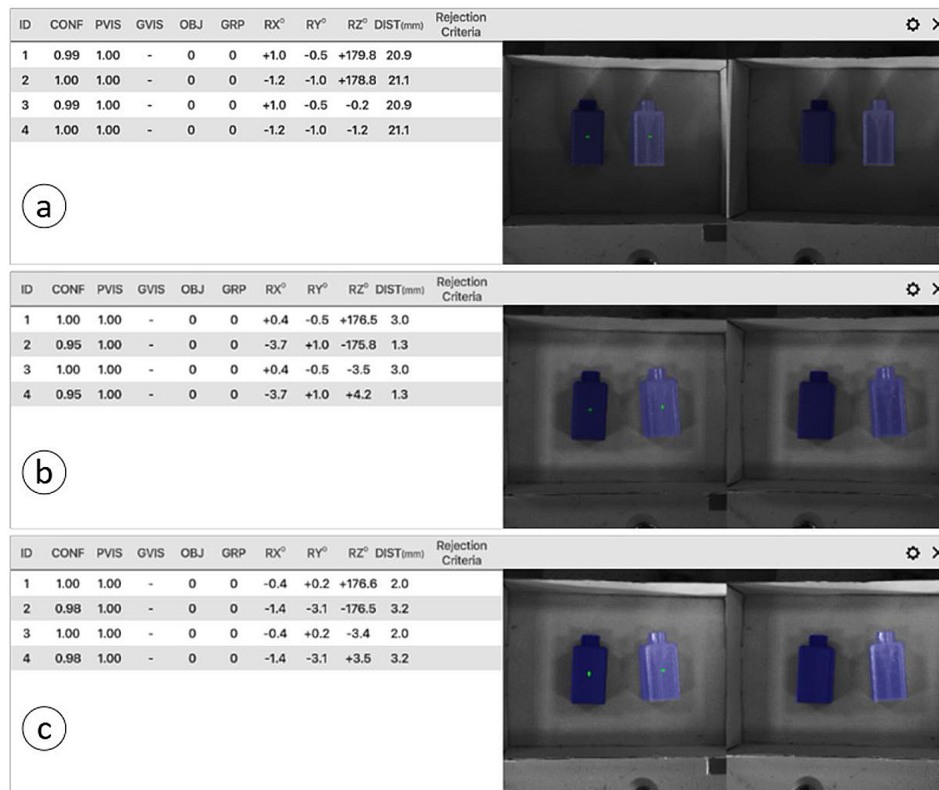


Figure 13. Comparison of detail recognition accuracy results at different light intensities in the work area:
a) 100 lx, b) 350 lx, c) 500 lx

Analysis of the recognition of features with geometry similar to that of the trained model

To evaluate the artificial intelligence model's ability to verify objects with geometries that deviate from the reference model, a bottle model with a circular cross-section was designed and fabricated, with a height reduced by one-third compared to the nominal CAD model (which had been used earlier in the training process). The study was conducted with a confidence filter (CONF) threshold set at 0.9. The analysis focused not only on the detection of the modified element, but also on the impact of occlusion in various spatial configurations on classification accuracy (Figure 15).

In the baseline scenario, the object with reduced height was not identified by the vision system; the algorithm did not generate a grasp prediction for it, which was reflected in the interface by the absence of any mask on the camera image (Figure 15a). This demonstrates the high selectivity of the AI model used, which correctly verifies that the observed features conform to the nominal geometry stored in the reference

CAD model. Importantly, even when a trained detail was intentionally obscured (which visually made its proportions resemble those of a foreshortened detail), the system demonstrated adequate resistance to this type of interference and correctly classified the object (Figure 15a).

The introduction of an additional occlusion covering the upper part of the modified (shortened) feature caused the system to attempt classification and generate a preliminary prediction. However, it had a confidence coefficient below the established threshold of 0.9. In accordance with safety protocols, the system rejected this grasping proposal, marking the object in question only with an orange warning mask (Figure 15b). The situation did not change until the incomplete portion of the truncated workpiece was completely obscured. The forced elimination of the height difference through complete occlusion caused the remnant visible to the cameras to fit into the learned geometric pattern. As a result of the reduced availability of spatial features, the vision system interpreted the visible fragment as a reference detail and generated a full-fledged grasp prediction for it (Figure 15c).



Figure 14. Analysis of the effectiveness of detecting partially obscured features

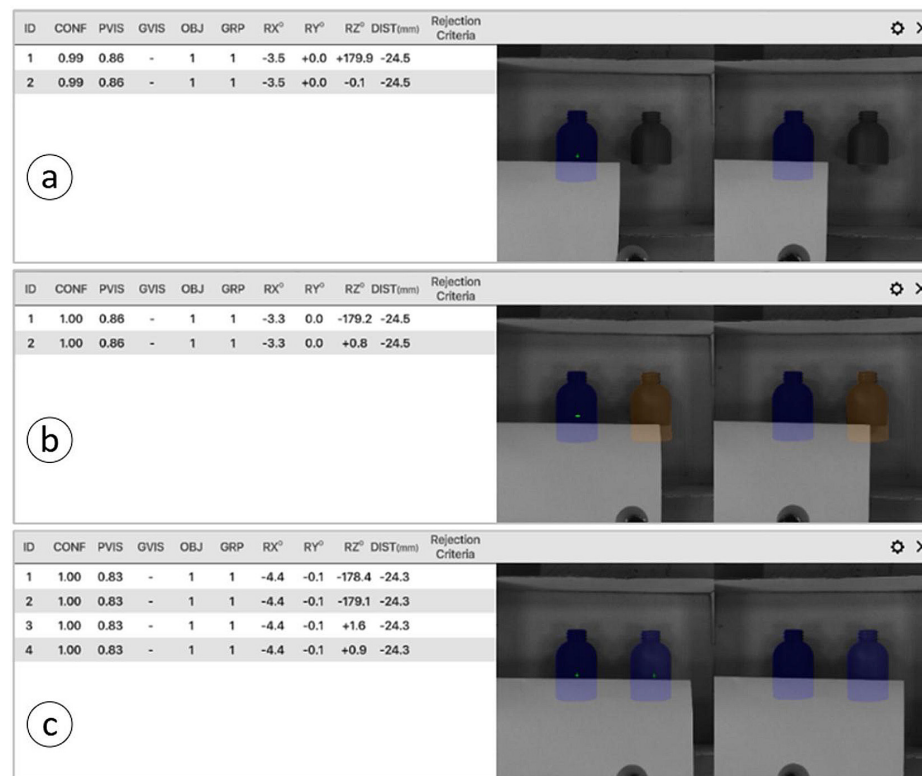


Figure 15. Analysis of the recognition of a feature with reduced height under various conditions of partial occlusion: a) partial occlusion of the trained feature, b) partial occlusion of the trained feature and a portion of the truncated feature, c) partial occlusion of the trained feature and complete occlusion of a portion of the truncated feature

DISCUSSION

The results of the experimental studies confirm that the optical properties of the manipulated objects determine the stability and response time of deep learning-based vision systems. The

different behavior of the algorithms for matte and reflective surfaces (represented in this study by PLA and PETG, respectively) is reflected in the issues discussed in the relevant literature [18, 19]. Therefore, the conclusions drawn from this comparison refer primarily to the objects' optical

surface characteristics rather than broadly generalizing the properties of the two polymer groups. While matte surfaces ensure uniform light diffusion, reflective materials are prone to specular reflections. This results in localized overexposure that masks important geometric features, which explains the minimal decreases in the detection confidence factor (CONF) at higher illuminance levels. Nevertheless, it has been shown that optimizing the illuminance to 500 lx reduces the AI model's inference time by a factor of three (from 0.9 s to 0.3 s). It is worth noting, however, that the acceleration of the detection process itself did not result in a dramatic reduction in the total cycle time. This observation is consistent with analyses by other researchers, which indicate that in classical robotic systems, the physical time required for the kinematic system to move remains the bottleneck. This points to the need to implement parallel workflow, where the calculation of the next move takes place while the manipulator is in motion [21].

Another important aspect addressed in the studies was the system's susceptibility to geometric modifications and occlusions. The model demonstrated high selectivity, effectively rejecting workpieces with inconsistent proportions and blocking predictions with low confidence. This is a key safety mechanism in industrial applications and bin picking, designed to prevent the picking of an incorrect object. A case in which the system incorrectly generated a prediction for a truncated feature after it was completely obscured (in a way that made it resemble the reference feature) illustrates the inherent limitations of architectures based solely on matching to a CAD model.

The behaviour of the vision system when modified geometric features are obscured deserves particular attention, as it reveals a significant limitation of the system. In an experiment in which a modified (defective) part of an object was obscured, the system successfully, albeit incorrectly, recognised it as a complete and correct feature. This indicates that the artificial intelligence model actively infers and extrapolates hidden geometry solely on the basis of visible fragments. It demonstrates that, in industrial applications, advanced AI models may exhibit overconfidence by fitting fragmentary spatial features to a known whole. This extrapolation constitutes a critical shortcoming in industrial applications: if key distinguishing geometric features are obscured, the system is unable to

distinguish between a good and a defective part, generating false positives with a high confidence score. This finding justifies the search for hybrid methods [13, 14] or estimation models and zero-shot models, which are inherently more robust to severe occlusions and less prone to unverified geometric extrapolation [16, 17].

CONCLUSIONS

As part of the work carried out, a robotic test station equipped with an ABB collaborative robot and a Cambrian vision system was designed and integrated. A proprietary control program written in RAPID was implemented, enabling the automation of the bin-picking process, which involves the detection and safe retrieval of randomly arranged parts from a container and their precise placement into a dedicated receiving tray. The main research focused on evaluating the impact of lighting conditions and the optical properties of materials on the effectiveness of artificial intelligence algorithms.

The findings of this study are inherently limited to the specific configuration of the tested robotic workstation, the analyzed geometries of the objects, and the three selected illuminance levels. The analysis showed a significant correlation between light intensity in the workspace and the performance of the vision system. Increasing the illuminance from the baseline of 100 lx to 350–500 lx resulted in a threefold reduction in the time required to generate a grasp prediction (from approximately 0.9 s to 0.3 s). It has also been proven that the optical surface properties of the objects, rather than the specific polymer type, primarily determine the behavior of detection algorithms. For matte workpieces (represented by PLA), the system demonstrated complete stability, consistently generating the maximum confidence coefficient (CONF = 1.0). In the case of reflective surfaces, due to the presence of directional reflections and localized overexposure in the image, this coefficient fluctuated slightly within the range of 0.95–0.99.

Within these defined parameters, an illuminance of approximately 500 lx yielded the best results among the tested conditions, ensuring the algorithm's highest effectiveness with the shortest inference time. However, it must be explicitly stated that this reduction in prediction time did not translate to a proportionally significant decrease

in the total cycle time, as the physical movement of the manipulator remained the dominant temporal component of the entire bin-picking process.

Furthermore, the developed artificial intelligence model exhibited high spatial selectivity, correctly recognizing reference workpieces and maintaining satisfactory robustness to minor partial occlusions. The algorithm successfully refrained from generating grip proposals for objects with intentionally modified dimensions, provided the modified sections were clearly visible to the camera. However, the study also revealed a critical operational limitation: when the modified or defective features of an object were heavily obscured, the system falsely identified it as a complete reference part by extrapolating the hidden geometry. This confirms that while classification accuracy strongly depends on the correspondence between the visible real-world image and the reference CAD model, the AI's tendency to infer missing geometric data based solely on fragmentary features poses a challenge for applications requiring absolute reliability.

REFERENCES

- Cordeiro A, Rocha LF, Costa C, Costa P, Silva MF. Bin picking approaches based on deep learning techniques: a state-of-the-art survey. 2022. <https://doi.org/10.1109/ICARSC55462.2022.9784795>
- Gomes NM, Martins FN, Lima J, Wörtche H. Reinforcement learning for collaborative robots pick-and-place applications: A case study. *Automation*. 2022;3:223–241. <https://doi.org/10.3390/automation3010011>
- Torres P, Arents J, Marques H, Marques P. Bin-picking solution for randomly placed automotive connectors based on machine learning techniques. *Electronics*. 2022;11:476. <https://doi.org/10.3390/electronics11030476>
- Le HP, Phan XT, et al. Development of intelligent vision for robotic object pick-and-place system. *FME Transactions*. 2025;25. <https://doi.org/10.5937/fme2502233P>
- Cherubin S, Kaczmarek W, Siwek M. YOLO object detection and classification using low-cost mobile robot. *Przegląd Elektrotechniczny*. 2024;(9):29–33. <https://doi.org/10.15199/48.2024.09.04>
- Yoshizawa M, Motegi K, Shiraishi Y. A deep learning-enhanced stereo matching method and its application to bin picking problems involving tiny cubic workpieces. *Electronics*. 2023;12(18):3978. <https://doi.org/10.3390/electronics12183978>
- Chen YL, Cai YR, Cheng MY. Vision-based robotic object grasping—a deep reinforcement learning approach. *Machines*. 2023;11:275. <https://doi.org/10.3390/machines11020275>
- Nwakeze OM, Okeke OC, Mgbemfulike IJ. Intelligent robotic object grasping system using computer vision and deep reinforcement learning techniques. *International Journal of Science and Research Archive*. 2025;14(3):511–521. <https://doi.org/10.30574/ijrsra.2025.14.3.0693>
- Calzada-Garcia A, Victores JG, Naranjo-Campos FJ, Balaguer C. A review on inverse kinematics, control and planning for robotic manipulators with and without obstacles via deep neural networks. *Algorithms*. 2025;18:23. <https://doi.org/10.3390/a18010023>
- Alebooyeh M, Urbanic RJ. Neural network model for identifying workspace, forward and inverse kinematics of the 7-DOF YuMi 14000 ABB collaborative robot. *IFAC-PapersOnLine*. 2019;52(10):176–181. <https://doi.org/10.1016/j.ifacol.2019.10.019>
- Li B, Tian W, Zhang C, Hua F, Cui G, Li Y. Positioning error compensation of an industrial robot using neural networks and experimental study. *Chinese Journal of Aeronautics*. 2022;35(2):346–360. <https://doi.org/10.1016/j.cja.2021.03.027>
- Govi E, Sapienza D, Toscani S, Cotti I, Franchini G, Bertogna M. Addressing challenges in industrial pick and place: a deep learning-based 6 degrees-of-freedom pose estimation solution. *Computers in Industry*. 2024;161:104130. <https://doi.org/10.1016/j.compind.2024.104130>
- Cui X, Yu M, Wu L, Wu S. A 6D pose estimation for robotic bin-picking using point-pair features with curvature (Cur-PPF). *Sensors*. 2022;22(5):1805. <https://doi.org/10.3390/s22051805>
- He Z, Feng W, Zhao X, Lv Y. 6D pose estimation of objects: recent technologies and challenges. *Applied Sciences*. 2021;11(1):228. <https://doi.org/10.3390/app11010228>
- Franaszek M, Rachakonda P, Saidi KS. Filtering organized 3D point clouds for bin picking applications. *Applied Sciences*. 2024;14(3):961. <https://doi.org/10.3390/app14030961>
- Borrego J, Dehban A, Figueiredo R, Moreno P, Bernardino A, Santos-Victor J. Applying domain randomization to synthetic data for object category detection. 2018. <https://doi.org/10.48550/arXiv.1807.09834>
- Chen J, Zhou Z, Li X, Zheng Y, Bao T, He Z. ZeroBP: learning position-aware correspondence for zero-shot 6D pose estimation in bin-picking. 2025. <https://doi.org/10.48550/arXiv.2502.01004>
- Zhang Z, Lu H, Cai J, Kamiya T, Serikawa S. Generalizable zero-shot object pose estimation

- for bin-picking. 2025 IEEE International Conference on Robotics and Automation (ICRA). 2025;12394:12400. <https://doi.org/10.1109/ICRA55743.2025.11128200>
19. Kozák V, Sushkov R, Kulich M, Přeučil L. Data-driven object pose estimation in a practical bin-picking application. *Sensors*. 2021;21(18):6093. <https://doi.org/10.3390/s21186093>
20. Jayasinghe D, Abeysinghe C, Opanayaka R, Dinalankara R, Silva BN, Wijesinghe RE, et al. Minimizing the effect of specular reflection on object detection and pose estimation of bin picking systems using deep learning. *Machines*. 2023;11:91. <https://doi.org/10.3390/machines11010091>
21. Panasiuk J. Controlling an industrial robot using stereo 3D vision systems with AI elements. *Sensors*. 2025;25:6402. <https://doi.org/10.3390/s25206402>
22. Prusaczyk P, Kaczmarek W, Panasiuk J, Besseghieur K. Integration of robotic arm and vision system with processing software using TCP/IP protocol in industrial sorting application. *AIP Conference Proceedings*. 2019;2078:020032. <https://doi.org/10.1063/1.5092035>
23. Solowjow E, Ugalde I, Shahapurkar Y, Aparicio J, Mahler J, Satish V, et al. Industrial robot grasping with deep learning using a programmable logic controller (PLC). 2020. <https://doi.org/10.48550/arXiv.2004.10251>
24. Ojer M, Alvarez H, Lin X, Lazkano E. Advances in industrial bin picking through sim-to-real transfer and cycle optimization. 2025. <https://doi.org/10.1177/17298806251374367>
25. Dikole R, Baloyi A, Sekopa T. Vision-type deep deterministic policy gradient for robotic manipulation with applications to ABB robots. 2025 7th International Conference on Control and Robotics (ICCR); 2025; Kyoto, Japan. p. 272-277. <https://doi.org/10.1109/ICCR67607.2025.11372069>
26. Wei BY, Zhang QH, Ma T, Wu XY, Li YP, Hao LD, et al. Research on robotic object grasping and detection technology based on deep learning and computer vision. *Journal of Computers*. 2025;36(6):241–258. <https://doi.org/10.63367/199115992025123606016>
27. Kaczmarek W, Borys S, Panasiuk J, Siwek M, Prusaczyk P. Experimental study of the vibrations of a roller shutter gripper. *Applied Sciences (Switzerland)*. 2022;12(19). <https://doi.org/10.3390/app12199996>
28. Borys S, Kaczmarek W, Laskowski D. Selection and optimization of the parameters of the robotized packaging process of one type of product. *Sensors*. 2020;20:5378. <https://doi.org/10.3390/s20185378>