

Article

Neural Network for Robotic Control and Security in Resistant Settings

Kubra Kose , Nuri Alperen Kose  and Fan Liang * 

Department of Computer Science, Sam Houston State University, Huntsville, TX 77341, USA;
kxg067@shsu.edu (K.K.); nxk022@shsu.edu (N.A.K.)

* Correspondence: fxl027@shsu.edu

Abstract

As the industrial automation landscape advances, the integration of sophisticated perception and manipulation technologies into robotic systems has become crucial for enhancing operational efficiency and precision. This paper presents a significant enhancement to a robotic system by incorporating the Mask R-CNN deep learning algorithm and the Intel® RealSense™ D435 camera with the UFactory xArm 5 robotic arm. The Mask R-CNN algorithm, known for its powerful object detection and segmentation capabilities, combined with the depth-sensing features of the D435, enables the robotic system to perform complex tasks with high accuracy. This integration facilitates the detection, manipulation, and precise placement of single objects, achieving 98% detection accuracy, 98% gripping accuracy, and 100% transport accuracy, resulting in a peak manipulation accuracy of 99%. Experimental evaluations demonstrate a 20% improvement in manipulation success rates with the incorporation of depth data, reflecting significant enhancements in operational flexibility and efficiency. Additionally, the system was evaluated under adversarial conditions where structured noise was introduced to test its stability, leading to only a minor reduction in performance. Furthermore, this study delves into cybersecurity concerns pertinent to robotic systems, addressing vulnerabilities such as physical attacks, network breaches, and operating system exploits. The study also addresses specific threats, including sabotage and service disruptions, and emphasizes the importance of implementing comprehensive cybersecurity measures to protect advanced robotic systems in manufacturing environments. To ensure truly robust, secure, and reliable robotic operations in industrial environments, this paper highlights the critical role of international cybersecurity standards and safety standards for the physical protection of industrial robot applications and their human operators.

Keywords: smart manufacturing; adversary learning; industrial automation



Academic Editors: Dazhao Cheng,
Chuang Hu and Fang He

Received: 22 July 2025

Revised: 3 September 2025

Accepted: 10 September 2025

Published: 12 September 2025

Citation: Kose, K.; Kose, N.A.; Liang, F. Neural Network for Robotic Control and Security in Resistant Settings.

Electronics **2025**, *14*, 3618. <https://doi.org/10.3390/electronics14183618>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Smart manufacturing is emerging as an important sector characterized by its dynamic landscape and evolving nature, particularly in the areas of automation and artificial intelligence (AI) [1]. With the incorporation of robotics, computer vision technology, and cybersecurity measures, the advancement of smart robotic arms is ushering in a new era of manufacturing. To meet these challenges, traditional industrial robotic arms must exhibit rapid adaptability and perform complex tasks. Traditionally, these arms have been used for repetitive tasks in controlled environments; however, their inflexibility and vulnerability

to cybersecurity threats pose limitations in addressing the dynamic and often demanding requirements of modern manufacturing.

In smart manufacturing, high-tech robotics has signaled a significant shift from conventional manufacturing models, ushering in an era of enhanced flexibility and intelligence for robotic systems [2]. This evolution is crucial as it enables them to perform a wider range of tasks with greater efficiency. The focus of our research is on improving the capabilities of robotic arms in terms of navigation and operation in dynamic environments. Compared to traditional production processes, which are characterized by fixed and predefined tasks, smart manufacturing processes are flexible and quick to respond to changing needs.

Our proposal leverages advancements in robotics and cybersecurity to meet these evolving requirements. We aim to improve the functionality and security of robotic arms within smart manufacturing environments. For example, a flexible robotic pick-and-place work cell based on the Digital Twin concept illustrates strategies for meeting current manufacturing demands. This study aims to improve the functionality, adaptability, and security of robotic arms in such environments by developing highly adaptive and reconfigurable systems. Our study explores the application of a deep learning framework with robotic arms, focusing on improving their object detection and manipulation capabilities. The study also provides a comprehensive analysis of specific cybersecurity threats, such as hijacking, sabotage, and physical harm, supported by targeted cybersecurity recommendations for robotic arms. Our foundational work equips the robotic arms with the precision necessary to interact with individual objects, laying the groundwork for potentially more complex configurations in future developments. Specifically, this foundational phase aims to equip robotic arms with the necessary precision for individual object recognition and interaction while safeguarding against potential cybersecurity risks. The resulting capability forms the basis for further complexity, although our scope in this phase is limited to manipulating single objects rather than multi-object stacks.

By focusing on optimizing single-object interaction, cybersecurity measures, and adversarial resilience, we establish the foundation for future advancements in robotic arm technology, enabling these systems to evolve to meet more complex demands, such as handling multi-object stacks, navigating more complex operational environments, and enhancing resistance to cyber threats. Achieving intelligent, adaptable, efficient, and secure manufacturing processes requires a strategic approach that ensures the progressive advancement of smart manufacturing technologies.

In this paper, we have the following contributions:

- We first optimize object detection algorithms by fusing Mask R-CNN's pixel-wise segmentation with Intel® RealSense™ D435 depth data. This integration enables the robotic arm to recognize and locate individual objects effectively in its environment with enhanced 3D spatial precision and a 10% increase in detection accuracy.
- We then utilize the outputs of the Mask R-CNN deep learning model (segmentation masks and 3D localization) to inform and develop effective gripping strategies based on heuristic rules. This approach enables the robotic arm to adjust its gripping mechanism in real time for secure and efficient object handling, accounting for the object's geometry and orientation.
- We integrate Mask R-CNN with Intel® RealSense™ D435 depth sensing, achieving a 20% improvement in manipulation success rate compared to RGB-only systems and demonstrating the practical benefit of combining depth and segmentation for robotic grasping. This integration constitutes a methodological contribution through a depth-augmented grasp planning pipeline that fuses RGB and depth data for precise 3D gripping point calculation, enhancing spatial reasoning.

- We conduct a pioneering evaluation of the manipulation pipeline's robustness against adversarial conditions by applying the Fast Gradient Sign Method (FGSM). This study quantifies the impact of varying noise intensities on detection, gripping, and overall manipulation accuracy and efficiency, revealing critical insights into the system's resilience and potential vulnerabilities.
- Finally, we address the critical need for comprehensive cybersecurity measures to protect robotic systems against cyber threats. We categorize and discuss specific risks, including physical attacks, network breaches, operating system exploits, hijacking, sabotage, and physical harm, drawing on real-world examples and relevant literature. To mitigate these risks, we propose and detail technical strategies such as securing communication channels, enhancing authentication protocols, implementing intrusion monitoring, integrating adversarial training for resilience, and advocating strict adherence to international standards.

The remainder of this paper is organized as follows: In Section 2, we provide a brief literature review of related studies on smart manufacturing and robotics. In Section 3, we outline the proposed system developed to enhance manipulation capabilities. In Section 4, we present the results of our experiments, demonstrating the efficacy of our approach. In Section 5, we discuss the cybersecurity threats and mitigation strategies. Finally, we conclude the paper in Section 6.

2. Related Work

In the field of smart manufacturing, the integration of technological advances in robotics has resulted in significant innovations in object detection, sorting, and manipulation. This Related Work section reviews studies that have shaped current practices and identifies emerging trends that inform our research. By examining key advancements and applications, from smart manufacturing to adaptive systems, we position our work within the dynamic landscape of robotic automation.

2.1. Smart Manufacturing and Robotic Automation

Smart manufacturing is at the forefront of technological advancement, where innovations are enhancing industrial processes through the application of computer vision and deep learning algorithms to enable system perception and decision-making. For example, Islam et al. [3] introduce a vision-based system that effectively detects and precisely locates objects within 3D environments. This system utilizes SSD MobileNet V2 in conjunction with RealSense cameras, achieving an object detection accuracy of 79.84%. Its applications extend to helping robots navigate unstructured environments, showcasing the potential of vision systems to enhance real-world robotic capabilities. Similarly, Shahria et al. [4] present real-time object localization and tracking through a vision-based robotic manipulation system. The combination of RealSense Depth Cameras and a UFactory xArm-5 robotic arm results in a detection accuracy of 75.2% and an average depth accuracy of 94.5%. This technology finds use in exploration, mobile robotics, and assistive systems, emphasizing the flexibility of vision-based solutions.

2.2. Advanced Sorting Mechanisms and Handling Complexity

For complex tasks such as picking up stacked objects, a reliable strategy is crucial. For instance, Farag et al. [5] delve into the development of an effective sorting algorithm, specifically engineered for object classification based on color, in conjunction with a proficient robotic arm. The system comprises a conveyor belt, a precision-driven robotic arm, and a dedicated camera for RGB color-based object detection. This algorithm effectively identifies and localizes red, green, and blue objects. This capability allows the robotic arm to pre-

cisely pick items of the desired color from the moving conveyor belt and deposit them into corresponding bins. In-depth kinematic analysis of the robotic arm, precise object detection utilizing RGB color ranges, and autonomous object sorting collectively contribute to the system's capabilities, achieving high sorting efficiency in optimized operations. Moreover, Capkan et al. [6] focus on robotic handling within industrial settings, specifically on objects with diverse geometric configurations. The study introduces two contributions: firstly, the configuration of a robotic arm designed for complex pick-and-place tasks, equipped with inverse kinematics and a virtual simulation environment; secondly, a revolutionary edge detection methodology grounded in knowledge-based rules, which generates stable training patterns and precisely identifies object centers. This approach achieved significant results by training a deep learning classifier, including AlexNet, GoogleNet, and ResNet, using edge image datasets. This approach enables autonomous object recognition, coordinate extraction, and seamless guidance of a microcontroller-controlled robotic arm for pinpoint pick-and-place operations more effectively than with raw images.

2.3. Enhancements in Object Detection and Sorting

Further contributions to object detection and sorting mechanisms have been made by exploring color and proximity detection. The novel design presented in [7] employs an Arduino Mega microcontroller to identify and sort objects based on color and distance. In manufacturing and beyond, this smart robotic solution demonstrates an innovative approach to object differentiation. Similarly, the integration of object detection [8] exemplifies the fusion of robotic arms with intelligent software to autonomously manipulate object positions, illustrating the significant potential for increasing efficiency and reducing labor costs in industrial automation. Adding to these advancements, a study by [9] refines the sorting algorithm to classify objects by their colors effectively. The research demonstrated improved sorting capabilities for the robotic arm by utilizing a color-detection system and a conveyor belt setup.

2.4. Adaptive Systems for Diverse Applications

Adaptability in robotic systems is further demonstrated in specialized applications. The design described in [10] leverages a Dobot Magician robotic arm, integrating machine vision to optimize agricultural sorting processes, highlighting the versatility of robotic arms in agricultural domains. Similarly, the system proposed in [11] introduces an autonomous robotic solution specifically designed for pruning sweet pepper leaves, showcasing the application of advanced perception and manipulation in complex, non-industrial agricultural environments. This system employs a semantic segmentation neural network to distinguish different plant parts and utilizes 3D point clouds from an Intel RealSense L515 camera to precisely detect pruning positions and control an articulated manipulator. Additionally, the work by [12] demonstrated the use of low-cost robotic arms equipped with sensors and computer vision to assist individuals with visual challenges, illustrating the significant humanitarian potential of robotic technologies.

2.5. Enhanced Control and Detection Mechanisms

Deep learning and 3D point cloud data have made significant contributions to the advancement of robotic arm control systems. Studies focusing on robotic arm control systems based on deep learning [13] and 3D Object Detection combined with inverse kinematics [14] demonstrate the integration of sophisticated algorithms for enhanced control and object manipulation. A key aspect of these contributions is the use of high-performance computational techniques to achieve precise and efficient robotic arm operations, which align closely with the objectives of smart manufacturing.

Building on these advancements, GraspSAM [15] extends the Segment Anything Model (SAM) for prompt-driven, category-agnostic grasp detection. Unlike two-stage methods, it unifies object segmentation and grasp prediction in a single framework, requiring minimal fine-tuning via adapters, learnable tokens, and a lightweight decoder. GraspSAM supports diverse prompt types (points, boxes, and language) and achieves state-of-the-art results on the Jacquard, Grasp-Anything, and Grasp-Anything++ datasets, outperforming prior two-stage approaches in real-world grasp tests.

2.6. Cybersecurity in Robotic Systems

With the integration of technologies like Mask R-CNN for object detection, comprehensive cybersecurity measures are becoming increasingly important to mitigate potential vulnerabilities, such as physical attacks, network breaches, and operating system vulnerabilities. Research [16] emphasizes the importance of machine learning techniques for improving security protocols for both the hardware and software aspects of robotic systems. Additionally, [17] illustrates how misleading techniques can be used to manipulate industrial robots while masking their activities from SCADA systems, demonstrating the importance of verifying data integrity. Furthermore, ref. [18] discusses the risks of adversarial attacks that can lead to dangerous misoperations in collaborative robotics, underscoring the urgent need for adversarial training and stability testing during the development phase. These studies demonstrate the importance of implementing strong cybersecurity frameworks to ensure the safe operation of robotic systems in smart manufacturing.

Beyond these individual measures, adopting standardized cybersecurity frameworks is crucial for safeguarding robotic systems in industrial environments. The ISA/IEC 62443 series, a globally recognized standard for Industrial Automation and Control Systems (IACS), provides a comprehensive, implementation-independent framework to identify, mitigate, and manage security vulnerabilities [19]. Unlike generic IT security, it specifically addresses industrial risks related to health, safety, and operational continuity. By establishing structured requirements for secure system design, continuous monitoring, and risk management, ISA/IEC 62443 supports the safe and reliable operation of robotic systems against threats such as hijacking, sabotage, and physical harm [20].

Using cutting-edge detection algorithms, control systems, deep learning techniques, and strong cybersecurity measures, robot arms have become more adaptable, precise, and efficient, resulting in smarter, safer manufacturing processes. In addition to improving industrial operations, these advancements also make robotic arms more useful in assistive technologies and agricultural production, thus bringing in a new era of versatility, intelligence, and security.

3. Proposed System

In our study, we focus on implementing object detection and manipulation using a UFactory xArm 5 robotic arm integrated with a D435 camera and employing the Mask R-CNN deep learning algorithm. The setup represents the core of our system, which is designed to enhance the capabilities of robotic arms in smart manufacturing environments by enabling precise object interaction. Our goal is to develop a robotic arm that can detect, grasp, and move objects accurately, which requires the integration of specialized hardware and software components.

Figure 1 presents a system structure that summarizes the complete process from observation to action. In general, the robotic arm always starts from the initial position. The depth camera then captures object location and depth information. Subsequently, our optimized deep learning model analyzes this data to detect and segment the objects before controlling the robotic arm to grasp them. Fundamentally, the collected data is combined

into a robot grasping strategy that controls the arm's manipulation movements to ensure accuracy and secure engagement with the target.

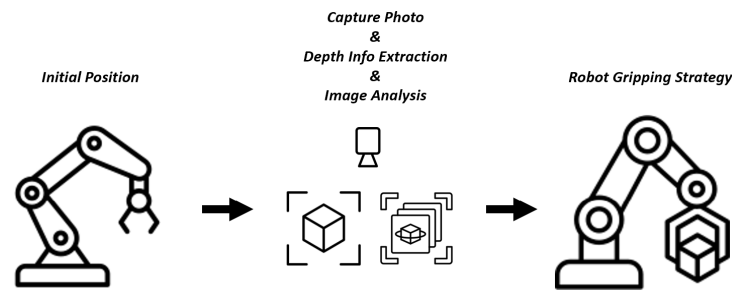


Figure 1. Overall System Structure and Workflow from Observation to Robotic Action.

Figure 2 shows the experimental apparatus, including the UFactory xArm 5 robotic arm, the Intel® RealSense™ D435 depth camera mounted above the workspace, the control laptop, and the test objects. This setup ensures consistent positioning for repeatable trials.



Figure 2. Experimental apparatus showing robotic arm, depth camera, control system, and object placement area.

From a theoretical perspective, the proposed approach leverages the principle that combining semantic segmentation with spatial depth estimation enables more accurate 3D object localization. Mask R-CNN provides pixel-wise segmentation masks and object classification, while the D435 camera supplies depth data for each pixel. By fusing these outputs, the robotic arm can calculate precise gripping points in 3D space, improving manipulation accuracy and reducing collision risk. This integration aligns with robotic perception theory, where depth-augmented vision systems enhance the robot's spatial reasoning capabilities.

3.1. System Design

3.1.1. Hardware Design

This study utilizes the UFactory xArm 5, a highly adaptable and precise robotic arm characterized by its ease of integration and versatility. A combination of its robust mechanical design and responsive control system makes it particularly suitable for complex tasks such as object manipulation. Additionally, the D435 camera is leveraged to capture depth data. This cutting-edge 3D camera provides depth sensing and high-resolution RGB

imaging, critical for accurate environmental perception and object detection. The camera's ability to provide three-dimensional data enables the robotic arm to interact with objects more effectively.

The Denavit–Hartenberg (DH) convention is a standardized method to describe the geometry of a robot arm using four parameters for each joint: joint rotation (θ), link offset (d), link twist (α), and link length (a). This provides a concise formal description of the arm's structure, allowing accurate modeling of its movement and position in space. The corresponding DH parameters are given in Table 1, defining the geometric relationships used in kinematic modeling [21].

Table 1. Denavit–Hartenberg parameters of the UFactory xArm 5 [21].

Joint	θ (rad)	d (mm)	α (rad)	a (mm)	Offset (rad)
J1	0	267	0	0	0
J2	0	0	$-\pi/2$	0	$T2_{offset}$
J3	0	0	0	a_2	$T3_{offset}$
J4	0	0	0	a_3	$T4_{offset}$
J5	0	97	$-\pi/2$	76	0

In our setup, the D435 camera was mounted at the centerline of the UFactory xArm 5 operational range. The camera was calibrated using the RealSense calibration API to produce an extrinsic transformation matrix mapping image pixel coordinates to the robot base frame. The system operates on a workstation running Ubuntu 20.04, Python 3.9, and PyTorch 1.13.0, with an NVIDIA RTX 3060 GPU. The camera streams RGB-D frames at 30 fps, and the Mask R-CNN inference runs at an average of 8 fps. Grasp points are calculated in 3D space and transformed to joint coordinates via the inverse kinematics solver from the UFactory xArm Python SDK, which issues control commands at 125 Hz. Communication latency between perception and actuation averages 42 ms.

3.1.2. Software Implementation

In our study, we employ a Mask R-CNN model as the base model, which is widely known for its segmentation capabilities. We optimize the Mask R-CNN model to process three-channel color videos. Additionally, we integrate UFactory Studio (an integrated development environment designed specifically for the UFactory line of robotic arms) to control the actions of the robotic arm. Based on the analysis results from the Mask R-CNN model, UFactory Studio enables precise and rapid object pickup. To facilitate efficient communication with the robotic arm, we utilize the PySerial library, which supports communication between Mask R-CNN and the robotic arm. We configured the Mask R-CNN with a ResNet-101 backbone and Feature Pyramid Network (FPN), pre-trained on the COCO dataset, and ran it in inference mode within our experimental environment.

3.2. Our Approach

Throughout this subsection, we detail the implementation process of integrating a UFactory xArm 5 robotic arm with a D435 camera and the Mask R-CNN algorithm to achieve precise object detection and manipulation. Through the integration of specialized hardware and software components, robotic arms can significantly improve their capabilities in smart manufacturing environments.

3.2.1. Object Detection Enhanced with RealSense D435 Depth Data and Mask R-CNN

In our study, we utilize the Intel RealSense D435 camera alongside the Mask R-CNN algorithm to provide real-time RGB and depth data. These data streams are processed by

the Mask R-CNN model, producing outputs such as instance segmentation masks, object labels, and median depth values for each detected object. The Mask R-CNN was configured with a ResNet-101 backbone and Feature Pyramid Network (FPN), pre-trained on the COCO dataset, to process the RealSense camera data streams.

High-resolution RGB data from the D435 camera are synchronized with depth data, forming the foundation of our object detection and segmentation system. Accurate 3D object detection and localization rely heavily on this synchronization. This method achieves optimization by combining semantic segmentation with spatial depth estimation, allowing for more accurate 3D object localization. The Mask R-CNN employs a ResNet-101-FPN backbone to process the RealSense camera data streams. The ROI Heads score threshold was empirically set to 0.7 after preliminary testing, minimizing false positives while preserving detection recall. By integrating dynamic depth data with advanced spatial and temporal filtering techniques, the system refines depth measurements, reduces noise, and ensures accurate object representation in the 3D model. This allowed us to link each detected object's contour and centroid to precise spatial location data, enabling accurate 3D localization and grasp planning in real time.

Table 2 summarizes the Mask R-CNN configuration parameters used in our experiments, including backbone, pre-training dataset, and key thresholds. The overall perception-to-actuation process implemented in our system is summarized in Algorithm 1.

Table 2. Mask R-CNN configuration parameters used in the proposed system.

Parameter	Value
Backbone	ResNet-101 with FPN
Pre-trained Dataset	COCO
Batch Size	2
Learning Rate	0.001
Optimizer	Stochastic Gradient Descent (SGD)
Epochs	20
ROI Heads Score Threshold	0.7
Frameworks	Python, OpenCV, pyrealsense2

Algorithm 1 Depth-Augmented Grasp Planning Pipeline

- 1: Capture synchronized RGB and depth frames from the Intel® RealSense™ D435 camera.
 - 2: Apply the Mask R-CNN model to the RGB frame to obtain the segmentation mask and object label.
 - 3: Compute the centroid (u, v) of the segmented object in image coordinates.
 - 4: Retrieve the corresponding depth value z from the aligned depth frame.
 - 5: Transform (u, v, z) from image coordinates to 3D coordinates (X, Y, Z) in the robot base frame using the pre-computed calibration matrix.
 - 6: Solve the inverse kinematics to determine the target joint configuration for the computed grasp point.
 - 7: Execute the grasp sequence: approach the target pose, close the gripper, lift the object, and place it at the predefined location.
-

Once these enhanced data streams are synchronized, the Mask R-CNN performs highly precise instance segmentation. The algorithm identifies and categorizes each object in the scene, generating segmentation masks that accurately depict the objects' shapes and boundaries. These detailed representations enable precise robotic arm manipulation.

The fusion of depth data with visual segmentation enhances the overall accuracy of the system. Based on the generated segmentation masks, the system calculates the physical dimensions of each detected object, combining this information with median depth values

from the depth stream. This enables the system to accurately estimate the size and spatial location of each object within the camera's field of view. Additionally, this information informs motion planning for the robotic arm, including calculating object trajectories, grip strengths, and end-effector orientations.

A fine-tuned set of parameters, such as the Region of Interest (ROI) Heads score threshold for mask visualization, enhances Mask R-CNN's real-time performance. Through transfer learning, the system achieves fast inference times and high average precision (AP) for mask identification by setting a confidence threshold for segmentation mask detection and leveraging pre-trained weights. These optimizations, combined with real-time processing rates, enable the robotic arm to respond to its environment both quickly and accurately.

This combination of depth sensing and Mask R-CNN-based segmentation creates a robust system for object detection that significantly improves the robotic arm's capabilities. By merging high-precision visual segmentation with depth data, the system is capable of detecting, understanding, and controlling objects with high precision and efficiency.

The D435 camera captures RGB and depth data synchronously during the object detection phase. The ability to perceive the environment accurately depends on the dual acquisition of data. Figure 3 illustrates the depth stream output (left) alongside the RGB stream (right). The depth stream is presented in a color-coded depth map format, where varying colors correspond to different distances from the camera, a feature that helps determine the precise location of objects in the workspace. The RGB stream provides detailed color imagery, capturing the object, in this case, a standard water bottle, in high resolution. The integration of these two streams allows the Mask R-CNN algorithm to exploit both textural and depth cues for reliable object detection.

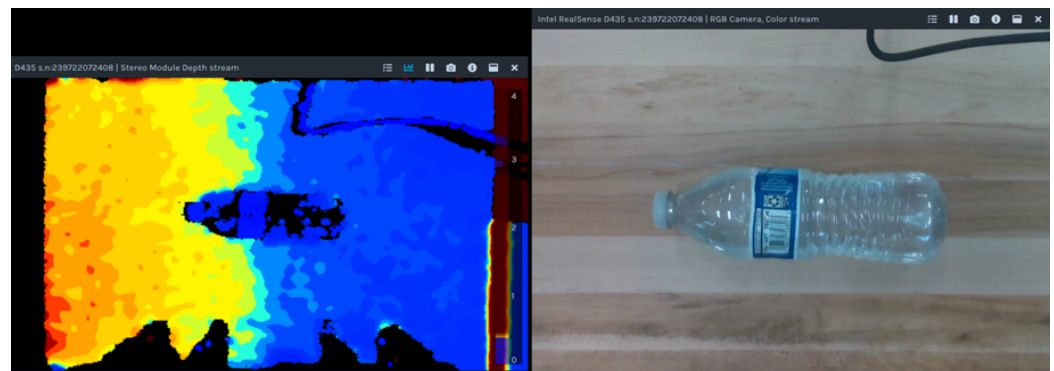


Figure 3. Depth and RGB Streams from Intel RealSense D435.

The Mask R-CNN model accurately segments and defines the object's boundaries, as shown in Figure 4. Segmentation masks accurately identify the object's boundaries, size, and location. The precise definition of the object's dimensions and corners is important for accurate manipulation, as this information is crucial. Mask R-CNN is designed to recognize and outline objects based on shape and size.

Figure 5 illustrates how the camera perceives an object's depth. A 3D point cloud visualization provides detailed depth data, which measures how far different parts of the object are from the camera. The system uses different colors to illustrate how the object is perceived in three dimensions. Based on this depth information, the robotic arm can calculate the object's distance from the robotic arm, enabling precise reach and manipulation.



Figure 4. Instance Segmentation Mask by Mask R-CNN.

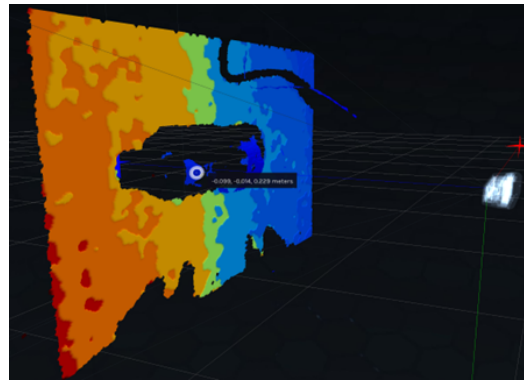


Figure 5. 3D Point Cloud Visualization of the Detected Object.

3.2.2. Gripping Strategy Formulation

Upon successful object detection and segmentation via the Mask R-CNN algorithm, the system begins creating the optimal gripping strategy. The Mask R-CNN algorithm provides detailed segmentation of the object, enabling the system to conduct a comprehensive analysis of the object's geometry and orientation. Employing heuristic rules informed by this deep learning output, the system calculates the potential gripping points that promise maximal stability and safety. The end-effector's approach is planned to ensure a secure grip that minimizes the risk of slippage or damage to the object. This calculation takes into account the physical characteristics of the object, such as shape, size, texture, and estimated weight. The relationship can be represented as follows:

$$\vec{F}_{grip} = f(\vec{P}_{obj}, \vec{O}_{obj}, \vec{T}_{obj}, W_{obj}) \quad (1)$$

Here, $\vec{F}_{grip}$ represents the gripping force vector, $\vec{P}_{obj}$ denotes the gripping points on the object's surface, $\vec{O}_{obj}$ represents the object's orientation, $\vec{T}_{obj}$ the texture, and W_{obj} the weight. The gripping force is dynamically adjusted according to the object's properties, ensuring an optimal balance between grip strength and delicacy.

3.2.3. Object Manipulation

With the gripping strategy defined, command signals are sent to the xArm 5, instructing it to align its end-effector with the predetermined gripping point on the object. The arm then executes a controlled motion to approach, grip, and securely hold the object. This phase leverages the robotic arm's precise movement capabilities and the end-effector's design to adapt to the object's surface and shape, ensuring a firm grasp. Following a successful grip, the arm lifts the object and transports it to a designated location, where it is placed or released, completing the manipulation task. A controlled motion sequence is then used to approach the object, establish a grip, and secure the object. The motion sequence is defined as follows:

$$\vec{M}_{seq} = g(\vec{P}_{grip}, \vec{S}_{path}) \quad (2)$$

where $\vec{M}_{seq}$ is the motion sequence vector, $\vec{P}_{grip}$ represents the gripping point positions, and $\vec{S}_{path}$ the pre-computed safe path to the object. This ensures a smooth approach and a solid grip. Once a solid grip is established, the robotic arm lifts the object and transports it to a predefined location. The task concludes as the object is gently placed or released, signifying the successful completion of the object manipulation task.

The robotic arm demonstrates high capabilities in object manipulation due to its powerful algorithms and precise control mechanisms, thereby providing significant advances in smart manufacturing.

3.2.4. Robotic Arm Operational Workflow

Robotic arms operate to ensure precision and efficiency in intelligent manufacturing environments. The process begins with the arm in its standby configuration, designed for quick activation. Utilizing the Mask R-CNN algorithm, the arm operates within the visual range of the D435 camera to detect and locate the target object within the environment. Based on the physical characteristics of the object, the arm moves to the object's coordinates and executes a precise grip. After securing the object, the arm lifts and relocates it to a predefined location, gently releasing it. After completing the task, the arm returns to its initial position to ensure its readiness for the next cycle. The workflow is designed for continuous operation so that the arm can consistently perform repeated tasks with high precision, simulating the principles of modern industrial automation. Each workflow phase demonstrates the robotic arm's capacity to operate autonomously, ensuring accuracy and reliability.

3.2.5. Adversarial Learning with Fast Gradient Sign Method

Our object detection process is further enhanced by incorporating an adversarial learning approach utilizing the Fast Gradient Sign Method (FGSM). The method simulates adversarial attacks with varying noise intensity levels by introducing controlled changes to the detected object images. We evaluate how such noise affects the Mask R-CNN model's detection accuracy and the robotic arm's operational reliability during object manipulation. FGSM modifies input images by adjusting the gradient of the loss function, which is calculated based on the input data. Different levels of adversarial noise, ranging from mild to severe, can be simulated using varying intensities of perturbations and observed under various conditions. The following are the steps involved in applying adversarial noise:

$$X_{adv} = X + \epsilon \cdot \text{sign}(\nabla_X J(\theta, X, y)) \quad (3)$$

where X_{adv} is the adversarial image, X is the original input image, ϵ is the perturbation magnitude (noise intensity), and $\nabla_X J(\theta, X, y)$ is the gradient of the loss function J for the input image X , θ represents the parameters of the detection model, and y is the true label of the object. We generate adversarial examples with different noise intensities by varying ϵ , testing the system's tolerance and ability to maintain accuracy. To examine how noise impacts the segmentation and localization of objects, especially during critical phases of object manipulation, these adversarial examples are fed into the Mask R-CNN model.

By experimenting with different noise intensity levels, the system can determine when it starts misidentifying or inaccurately manipulating objects. Thus, we identify possible vulnerabilities arising from adversarial attacks in real-life scenarios by assessing the system's reliability.

4. Evaluation

In this section, we summarize our experimental results to demonstrate the effectiveness of the proposed system. We evaluate the system's manipulation capabilities, including detection and handling, as well as its resistance to adversarial attacks.

All experiments were conducted in a controlled tabletop workspace. Two object types were used: a 500 mL cylindrical water bottle (21 cm height, 6.5 cm diameter, semi-transparent PET) and a standard-size sports ball (diameter 22 cm, chosen for its textured, fuzzy surface within the typical range for sports balls). The camera-to-object distance was varied between 30 cm and 60 cm. For each configuration, 10 manipulation trials were performed, starting from the same robot home position and placing the object at a fixed target location. Success criteria were (1) detection accuracy: the object was correctly segmented and localized; (2) gripping accuracy: the object was securely held without slip or drop; and (3) placement accuracy: the object was placed within 2 cm of target coordinates.

4.1. Manipulation Evaluation Through Mask R-CNN

Our evaluation focused on three critical aspects of the system: the detection capabilities of the Mask R-CNN algorithm, the performance of our gripping technique, and the overall efficiency from object detection to correct placement. Extensive testing revealed that, under optimal conditions, the system achieves a 98% detection accuracy, showcasing its robustness in controlled environments.

In our investigation into the object detection system's performance, we conducted targeted experiments to understand how detection accuracy varied with the distance between the camera-equipped robotic arm and a standard water bottle. We selected a standard water bottle as the primary test object due to its cylindrical shape, smooth surface, and semi-transparent plastic material, which challenges both RGB-based detection and depth estimation. The bottle is also easy for the gripper to grasp, resilient to drops, and can be squeezed without damage, making it ideal for repeatable experiments in realistic industrial scenarios. The object was placed at distances ranging from 30 cm to 60 cm from the camera, and at each distance, the system underwent four trials to ensure data reliability. Our findings indicated that the system's accuracy improved from 89% at 30 cm to an optimal 98% at 50 cm and 60 cm, indicating the camera's optimal operational range. These results, summarized in Figure 6, underscore the critical impact of spatial distance on detection efficiency.

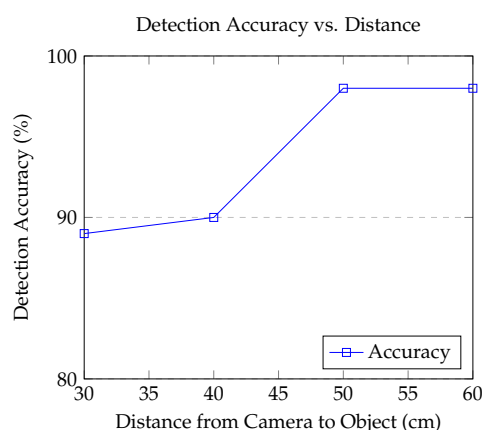


Figure 6. Variation in detection accuracy with respect to the distance of the sports ball from the camera.

With the optimal 50 cm distance incorporated, Table 3 captures the overall accuracy performance across ten trials, focusing on its detection, gripping, and transport capabilities. With detection accuracy uniformly high at 98%, the arm demonstrates reliable object recog-

nition. Gripping accuracy, however, varies between 92% and 98%, reflecting the challenges posed by the adaptive gripping strategy employed by the xArm. This variability is partly due to the arm's approach of adjusting the gripper size to closely match the detected object size, which is not always aligned precisely with the object's actual dimensions. By extending its gripper to the size of the object plus or minus 5 mm, the arm aims to secure a firm grip, accounting for potential size differences and minimizing the risk of dropping the object during transport. Placement accuracy to a defined location is consistently 100%, showcasing the arm's precision in positioning objects. The overall accuracy, calculated from detection and gripping performances, ranges from 96% to 99%, which indicates that this system is effective despite some variability in gripping efficiency.

Table 3. Overall Manipulation Accuracy (at optimal 50 cm distance).

Trial	Detection Acc. (%)	Gripping Acc. (%)	Transport Acc. (%)	Overall Acc. (%)
1	98	92	100	96
2	98	93	100	97
3	98	96	100	98
4	98	93	100	97
5	98	95	100	98
6	98	95	100	98
7	98	98	100	99
8	98	95	100	98
9	98	98	100	99
10	98	95	100	98

To evaluate the robotic arm's efficiency in manipulating objects, we standardized the testing environment. For each trial, the objects, either a bottle or a sports ball, were consistently placed at a predetermined location. Prior to beginning each task, the robotic arm was reset to its initial position, and the target destination for object placement remained consistent throughout the testing process. These measures ensured a uniform testing setup for all trials, as depicted in Figure 7.

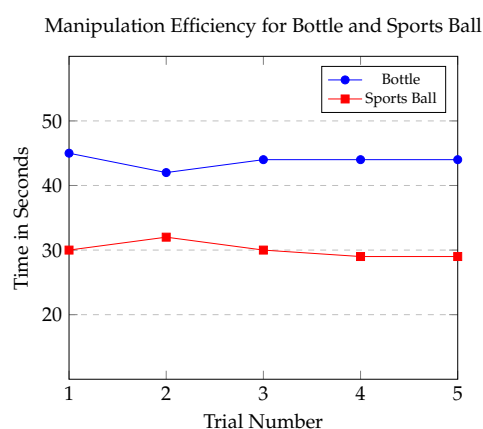


Figure 7. Comparison of manipulation efficiency over trials for bottle and sports ball.

In the experiment, the robotic arm maintained a steady manipulation time for bottles, with an average duration of 44 s across trials. This suggests consistent handling of objects with complex shapes. Conversely, the manipulation times for sports balls decreased from 30 s in the initial trial to 29 s in subsequent trials. The arm demonstrated a quicker optimization of its handling technique for spherical objects.

Notably, any trial resulting in unintended object displacement, such as a sports ball rolling away due to arm contact, was excluded from our results. This approach ensured

that only successful manipulation sequences were considered, providing a clear view of the arm's effective operational capabilities.

With the addition of depth information from the D435 camera, our object identification and manipulation system has become significantly more accurate. Table 4 illustrates how this depth data improves three-dimensional point cloud generation and enables a more thorough examination of objects in the system's working environment. Consequently, our detection algorithms have become more precise, and the robotic arm's manipulation procedure has become more reliable.

Table 4. Performance Improvement with Depth Data Integration.

Metric	Accuracy	
	Without Depth	With Depth
Detection Accuracy (%)	88	98
Manipulation Success Rate (%)	80	100

As part of our experiment, we conducted eight trials to assess the performance of the UFactory xArm 5 robotic arm. These trials were equally divided into two sets: the first set of four trials was performed without incorporating depth data, serving as our control group; the second set of four trials utilized the depth information provided by the D435 camera, which allowed us to evaluate the direct impact of depth information on the system's performance.

The accuracy of detection was measured by the robotic arm's ability to correctly identify and locate objects within its operational field in each trial. To calculate the manipulation success rate, we extended our measurement criteria beyond mere detection. We evaluated whether the system could grip objects without causing displacement or dropping and subsequently could place them precisely at a specified location. The system's ability to complete this sequence without error served as a measure of its manipulation success. The results, as shown in Table 4, show a significant improvement when depth data were used. Detection accuracy saw a 10% increase from 88% to 98%, and the manipulation success rate rose by 20% from 80% to 100% with the integration of depth data.

In addition, we computed confidence intervals that were supported by our data. For the depth ablation (Table 4), treating each trial as a Bernoulli event ($n = 4$ per condition), the Wilson 95% CI is [0.30, 0.95] without depth (3/4 successes) and [0.51, 1.00] with depth (4/4 successes). This reinforces the observed 20% improvement despite the small sample size. Finally, while our evaluation demonstrates clear gains with depth integration, the small number of trials limits formal statistical power. We therefore regard the present evaluation as a proof-of-concept, with larger-N studies planned for future work to strengthen statistical generalizability.

4.2. Security Evaluation Through Adversarial Learning

During the object detection phase, we introduced structured variations in the form of adversarial noise to test the stability of the system under adversarial conditions, resulting in insights into the system's resilience. To evaluate the impact of noise on the detection, gripping, and overall precision of manipulation, three discrete noise intensity levels (50, 100, and 150) were applied. As a form of attack simulation, the adversarial learning approach utilizes noise to assess the system's vulnerability to deliberate attempts to degrade detection and manipulation.

We fed adversarial noise directly into the RGB and depth input streams of the Mask R-CNN model. Detection accuracy was measured, representing the model's ability to correctly

identify and localize objects in the presence of perturbations. Furthermore, gripping accuracy was assessed based on the robotic arm's ability to secure objects under adverse conditions. Overall manipulation accuracy, a composite metric, measured the detection, gripping, and successful manipulation of objects under noise-augmented scenarios.

Table 5 demonstrates the effects of varying intensities of adversarial noise on system performance metrics such as detection, gripping, and transportation accuracy. Based on the noise levels tested (50, 100, and 150), overall accuracy decreases with increasing noise intensity. The table provides insight into the system's vulnerability and resistance against external manipulations intended to disrupt its operational accuracy.

Table 5. System Performance under Different Noise Levels.

Noise Level	Detection Acc. (%)	Gripping Acc. (%)	Transport Acc. (%)	Overall Acc. (%)
50	97	95	100	97
100	94	95	100	96
150	92.5	90	100	94

Figure 8 compares the robotic arm's manipulation efficiency across trials under different noise levels. The blue line, taken from Figure 7, represents the baseline performance without noise, specifically for bottle manipulation. The other lines (purple, brown, and green) correspond to noise levels of 50, 100, and 150, respectively, showing how system performance changes as noise intensity increases. As the noise level rises, manipulation time increases slightly, with a minimum increase of around 2.2% at noise level 50 and a maximum of 6.7% at noise level 150. While the system remains functional, the addition of noise causes a noticeable, though slow, decrease in performance.

Manipulation Efficiency for Bottle at Different Noise Levels

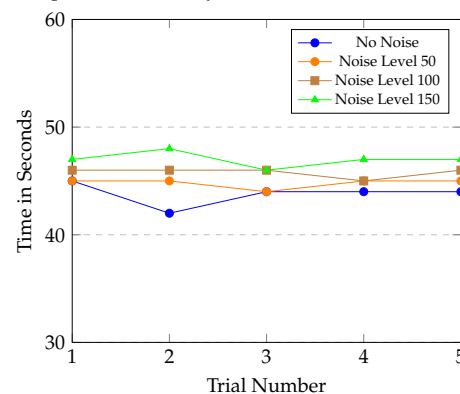


Figure 8. Comparison of manipulation efficiency over trials for different noise levels.

Figures 9 and 10 illustrate the impact of adversarial noise on object detection. Figure 9 illustrates how visual clarity and detectability are affected by noise levels decreasing from left to right (150, 100, 50, and 0). Figure 10 compares bottle images with varying noise levels arranged side-by-side: the top row represents noise levels of 150 and 100, while the bottom row represents noise levels of 50 and 0. Both figures illustrate the system's response to varying noise conditions, serving as a visual reference for the performance data. To quantify variability, we also report 95% confidence intervals (t-distribution, $df = 4$) for manipulation times. For the bottle, the mean was 43.8 s [42.4, 45.2]; for the sports ball, 30.0 s [28.5, 31.5]. Under adversarial noise, times were 44.8 s [44.2, 45.4] at level 50, 45.8 s [45.2, 46.4] at level 100, and 47.0 s [46.1, 47.9] at level 150.

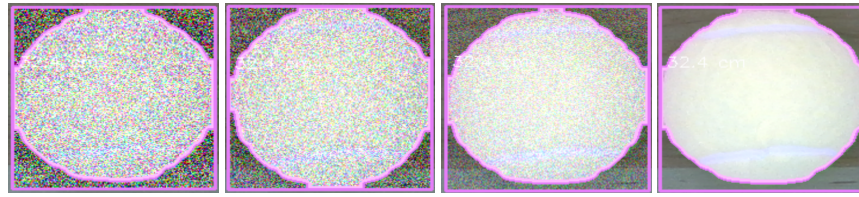


Figure 9. Comparison of ball with different noise measurements (Decreasing from left to right: 150, 100, 50, and 0).

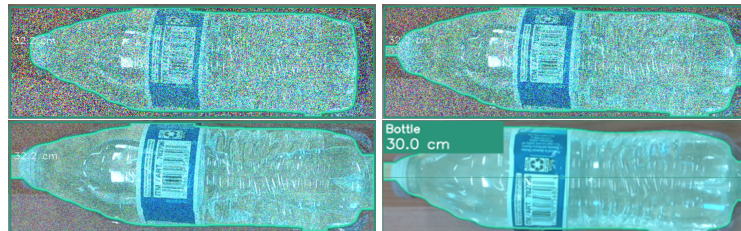


Figure 10. Comparison of bottle with different noise measurements (Top row: 150 and 100; Bottom row: 50 and 0).

5. Discussion: Cybersecurity Issues

As robotic systems integrate with cutting-edge technologies such as artificial intelligence and the Internet of Things, they also present significant cybersecurity threats. Different types of cyber threats target these platforms in different ways, and they can be categorized into physical attacks, network attacks, and operating system-based attacks [22]. Robot hardware, such as microcontrollers, can be damaged or destroyed as a result of physical attacks, resulting in reduced performance. Malicious actions can be carried out remotely through networking attacks by manipulating data transmission or corrupting sensor inputs without direct physical access to the system. An attack against an operating system, such as a Robot Operating System (ROS), exploits software vulnerabilities to manipulate or even take over robotic functions, causing serious security risks [23].

To ensure the safety of various categories of robotics, including industrial, medical, drones, humanoids, and autonomous vehicles, a variety of security measures must be implemented based on each category's specific needs [23]. Industrial robots are targeted by cyberattacks that aim to disrupt production or destroy equipment, thereby impacting the manufacturing process. For medical robots, privacy and accuracy are paramount. Patient care could be compromised by such a breach. Drones are susceptible to jamming and spoofing, which could lead to misdirected operations or data interception. Humanoid robots, which interact closely with humans and handle personal information, are subject to threats that could lead to unauthorized data access and physical harm to the user. Consequently, autonomous vehicles, being highly dependent on accurate sensor information and connectivity for safe operation, are vulnerable to attacks that may mislead navigational decisions or manipulate vehicle functions.

In our system, the most vulnerable attack surface is the RGB-D video feed from the D435 to the Mask R-CNN processing unit. Adversarial perturbations injected into the RGB channel can cause missed detections or mislocalized grasp points, leading to failed manipulations. Figure 8 illustrates how FGSM noise at different magnitudes degrades performance. Based on these findings, a potential mitigation strategy could involve an adversarial defense module that performs input frame sanitization using a secondary binary classifier trained to distinguish clean from perturbed inputs. In principle, such a module could halt grasp execution when suspicious inputs are detected and request a new frame, reducing the likelihood of error.

5.1. Specific Threats to Robotic Systems

Despite the numerous benefits that robotic systems bring across a broad range of industries, these systems are also subject to specific, serious threats that can have adverse effects on their functionality and safety. The purpose of this section is to address hijacking, sabotage (which compromises robotic operation and control), and physical harm (which directly affects their physical integrity). Protecting robotic platforms from severe and potentially catastrophic consequences presents unique challenges for all three categories. Through real-world incidents, we explore these threats as well as their broader implications.

5.1.1. Hijacking and Sabotage

A factory's robotic arms can be hijacked and sabotaged. Malicious actors can take control of these robots [22] to cause accidents, damage products, or stop production lines, resulting in significant economic losses. According to the FBI's 2023 Internet Crime Report [24], ransomware attacks on robotic systems are on the rise across a variety of sectors, including critical infrastructure, which is often integrated with robotics. Across multiple industries, ransomware incidents have caused substantial financial losses totaling over \$59 million, affecting not only financial stability but also operational continuity.

Researchers at IOActive have carried out a ransomware attack on SoftBank Robotics' NAO humanoid robot as part of a cyber intelligence operation [25]. In the IOActive experiment, robots were exposed to cyberattacks that could force them to display inappropriate content or issue threats. As a result of the researchers' ransomware installation, the robot's functions were manipulated and Bitcoin was demanded, demonstrating how ransomware can pose financial and operational risks to businesses.

In the study [26], the researchers demonstrated that they could intercept and shut down a man-in-the-middle cyberattack on an unmanned military robot, illustrating the potential for malicious actors to hijack and sabotage robotic operations. The existence of these threats necessitates the development of robust access controls, encryption, and intrusion detection systems to prevent unauthorized access and manipulation of robotic systems.

5.1.2. Physical Harm

It is extremely dangerous to manipulate robots to cause physical harm to workers or the environment. A series of serious incidents illustrates the dangers associated with the use of robotics in the workplace. A man was killed by a robot in South Korea that mistook him for cargo [27], illustrating the critical need for improved sensory technologies in robots to avoid such tragedies. A fatal accident occurred in a Thailand [28] factory when a robotic arm, incorrectly positioned during operation, crushed a worker, pointing to a failure in safety protocols and oversight as the cause of the accident. Another serious incident took place at Tesla's Texas headquarters in 2021 [29] when an engineer was allegedly pinned against a wall by a malfunctioning KUKA robot arm. A robot intended to cut aluminum inflicted severe injuries on an engineer, illustrating the severe potential of robotic malfunctions in high-risk situations.

These examples demonstrate how crucial it is for robotic systems to implement reliable cybersecurity measures. It is widely recognized that the risk of physical harm, operational disruption, and economic loss is substantial. To ensure the safety of human lives as well as the efficiency of industrial operations, these systems must be secured. Enhanced safety measures are urgently needed in robotic applications, particularly in industrial settings where humans and machines interact frequently. To prevent tragic accidents, high-tech sensory and recognition technologies must be integrated into the workplace [30]. Employees working alongside robots must also receive comprehensive safety training to ensure they are aware of potential risks and how to deal with them in an emergency [31].

Further, strict safety protocols must be established and enforced, with regular audits and updates to accommodate new technological advancements and operational insights.

5.2. Recommendations—Techniques and Strategies to Protect Robots

Protecting robotic systems from cyber threats requires the implementation of reliable security measures across multiple domains. These measures, such as securing communication channels, enhancing authentication protocols, and implementing intrusion monitoring, are fundamental building blocks for robust cybersecurity. These align directly with the requirements for secure system design and continuous monitoring emphasized by the globally recognized ISA/IEC 62443 series for Industrial Automation and Control Systems (IACS) [19]. Similarly, the integration of the D435 camera for enhanced object detection and manipulation accuracy, as demonstrated in our system, contributes to crucial aspects of physical safety, which is mandated by ISO 10218-2:2025 [32]. This standard provides detailed guidance for the design, integration, and safe operation of industrial robots, aiming to minimize hazards and protect human operators. By continuously updating software, conducting regular vulnerability assessments, and adopting secure coding practices, organizations can build resilience against threats like hijacking and sabotage, which are directly addressed by ISA/IEC 62443's focus on operational continuity and risk management.

In real-time, a machine learning-powered anomaly detection system can detect and respond to unusual threats and activities [33]. Robotic operations should be guided by a comprehensive cybersecurity policy, from development to deployment and maintenance. To foster a security-aware culture, this policy must include secure coding practices, routine security assessments, and comprehensive training for all users [34].

The integration of the UFactory xArm 5 robotic arm with the D435 camera, as discussed in the Proposed System section, highlights the necessity of robust cybersecurity measures. The robotic arm's detection, grasping, manipulation, and placement capabilities improve operational efficiency, but they also introduce significant security challenges. For instance, hijacking and sabotage pose serious threats; malicious actors could take control of the robotic arm to cause accidents, damage products, or disrupt production lines, resulting in substantial economic losses. Beyond these threats, physical harm also poses a risk if cybersecurity is not properly addressed. A robot whose control system has been compromised could cause injury to a person. To prevent such risks, strong encryption protocols, secure coding practices, and regular security assessments must be implemented. Robotic systems can become more secure and reliable if these measures are adopted to prevent unauthorized access.

The threat posed by adversarial attacks, which can reduce system performance without immediately causing catastrophic failure, now warrants consideration. We observed a reduction in detection accuracy when adversarial noise was introduced to our experiments using the Fast Gradient Sign Method (FGSM). Studies have demonstrated how small adversarial perturbations can mislead detection systems, potentially leading to dangerous physical actions [18,35]. The gripping and manipulation accuracy of the system was also affected, though less significantly than detection accuracy. However, the detection delays caused by adversarial noise could create inefficiencies, leading to delays in task completion or, in more severe scenarios, disruption of operations. Thus, robotic systems should incorporate adversarial training to build resilience against such attacks. Through training with adversarial examples, robots can learn to recognize and mitigate the effects of malicious inputs, thereby maintaining performance in compromised environments. The threat posed by adversarial attacks, as evidenced by the reduction in detection accuracy when FGSM noise was introduced, emphasizes the need for adversarial training to build resilience. This directly informs the 'Risk Management' component of ISA/IEC 62443, which

requires identifying and mitigating specific vulnerabilities to reduce risks to tolerable levels, considering not only economic impact but also health and safety factors. Future robotic systems, especially those in safety-critical industrial environments, should incorporate such resilience measures as part of their comprehensive security program, aligning with the principles of these international standards.

Organizations must prioritize cybersecurity, especially those in critical infrastructure and industrial settings. This includes the use of intrusion detection systems, regular vulnerability assessments, and constant updates to antivirus and anti-malware software. Maintaining system integrity requires updating all software related to robotic operations and patching vulnerabilities as soon as possible.

Beyond technical measures, employee training is crucial for raising awareness of cybersecurity threats. During training, employees should be taught to recognize phishing attempts and to secure operational endpoints. Encrypting data transmissions helps prevent unauthorized interception and manipulation by strengthening access controls through multi-factor authentication and enforcing effective policies.

Accessing advanced tools and insights into emerging threats requires collaboration with cybersecurity experts and firms. Vulnerabilities can be identified and addressed through regular security audits and penetration tests conducted by external experts. To improve the overall security posture of robotic systems, international cybersecurity standards and frameworks must be adopted [34]. Specifically, the ISA/IEC 62443 series is a widely used global consensus-based standard for Industrial Automation and Control Systems (IACS) cybersecurity, which directly applies to robotic systems in smart manufacturing. This comprehensive framework addresses security on an organization-wide basis, providing requirements for establishing a security program to reduce risks to tolerable levels, considering not only economic impact but also health, safety, and environmental factors [19,20]. To complement these cybersecurity frameworks and ensure comprehensive protection, strict adherence to physical safety standards is also essential. The ISO 10218-2:2025 standard is crucial for specifying requirements for the safety of industrial robot applications and robot cells. This standard provides detailed guidance for the integration, commissioning, operation, maintenance, and decommissioning of robots within industrial settings. It emphasizes the design and integration of machines and components to enhance safety and is vital for minimizing hazards and protecting human operators and other stakeholders from the unique and varied risks associated with industrial robotic systems. By implementing the guidelines outlined in ISO 10218-2:2025, organizations can effectively reduce the risk of physical harm, complementing the resilience built through cybersecurity measures [32].

5.3. Future Directions for Robotic Systems

Future research will focus on integrating Large Language Models (LLMs) into robotic arms for natural language control, which presents several challenges. Despite LLMs' ability to comprehend natural language, they struggle to maintain context over long pieces of text, potentially leading to incorrect or irrelevant responses. In addition, LLMs often lack creativity and rely heavily on patterns from their training data. Furthermore, their ability to apply logical reasoning may be limited, resulting in inaccurate or nonsensical results. Despite these limitations, LLMs are still capable of performing complex tasks such as analyzing voice commands, making logical inferences from object recognition results, and determining grasping sequences and methods.

To address these challenges, as shown in Figure 11, our future work involves leveraging the Tree of Thoughts prompt to enhance the system's logical thinking process. Tree of Thoughts (ToT) is a novel approach to prompt engineering designed to improve logical problem-solving, reasoning, and strategic thinking in LLMs like GPT-4. To enhance the

LLM's logical analysis capabilities, we will organize prompts using the Tree of Thoughts. In addition, the ToT framework will allow backtracking using depth-first and breadth-first search algorithms, enabling the LLM to quickly revert to previous states and adjust grasping strategies if necessary.

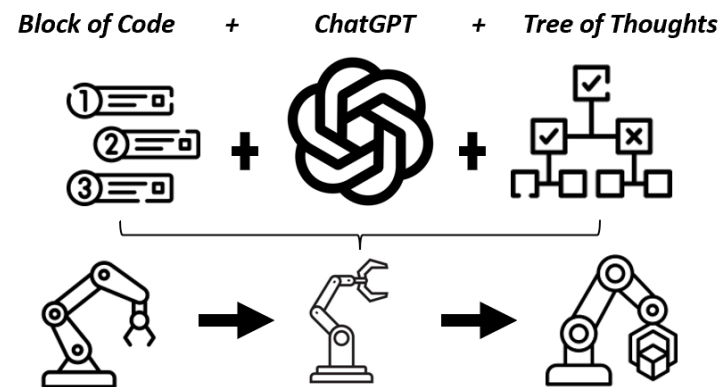


Figure 11. Future Directions for Robotic Systems: Leveraging Large Language Models with Tree of Thoughts Prompt Engineering.

6. Conclusions

In conclusion, this study has demonstrated the integration of the UFactory xArm 5 robotic arm with the D435 camera and the Mask R-CNN algorithm to develop a sophisticated object detection and manipulation system tailored to smart manufacturing applications. Our results indicate that the system is capable of identifying and handling objects with a high degree of accuracy, achieving a manipulation accuracy rate of 99%. We observed a decrease in detection performance at closer ranges, which provides further insight into the system's operational range and the importance of distance calibration in ensuring consistent results. Moreover, the system demonstrated resilience under adversarial noise conditions; its performance was only slightly affected as noise levels increased. In addition, the incorporation of cybersecurity measures ensures the system's resilience against potential cyberattacks, thereby maintaining its functionality within sensitive industrial environments. There are many practical implications of this work. By improving the flexibility and adaptability of robotic operations, it extends the frontiers of industrial automation. A system such as this illustrates the potential of integrating advanced vision systems and artificial intelligence into manufacturing to enable smarter, more responsive, and more efficient production processes. Furthermore, the results presented, particularly in precise object handling and stability against adversarial inputs, lay a crucial foundation for developing robotic systems that can adhere more comprehensively to critical international standards such as the ISA/IEC 62443 series for cybersecurity and ISO 10218-2:2025 for physical safety, thereby fostering smarter, more responsive, and more efficient production processes that prioritize both performance and security.

Author Contributions: Conceptualization, K.K., N.A.K. and F.L.; methodology, K.K., N.A.K. and F.L.; software, N.A.K. and K.K.; validation, K.K. and N.A.K.; formal analysis, K.K.; investigation, K.K.; resources, F.L.; data curation, N.A.K.; writing—original draft preparation, K.K. and N.A.K.; writing—review and editing, K.K. and F.L.; visualization, K.K.; supervision, F.L.; project administration, F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Huang, Z.; Shen, Y.; Li, J.; Fey, M.; Brecher, C. A Survey on AI-Driven Digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics. *Sensors* **2021**, *21*, 6340. [\[CrossRef\]](#) [\[PubMed\]](#)
- Malik, A.; Rajaguru, P.; Azzawi, R. Smart Manufacturing with Artificial Intelligence and Digital Twin: A Brief Review. In Proceedings of the 2022 8th International Conference on Information Technology Trends (ITT), Dubai, United Arab Emirates, 25–26 May 2022; pp. 177–182. [\[CrossRef\]](#)
- Islam, I.; Shahria, M.T.; Sunny, M.S.; Khan, M.M.R.; Ahamed, S.; Wang, I.; Rahman, M. A Vision-based Object Detection and Localization System in 3D Environment for Assistive Robots' Manipulation. In Proceedings of the 9th International Conference of Control Systems, and Robotics (CDSR'22), Niagara Falls, ON, Canada, 2–4 June 2022. [\[CrossRef\]](#)
- Shahria, M.; Arvind, A.; Iqbal, I.; Saad, M.; Ghommam, J.; Rahman, M. Vision-Based Localization and Tracking of Objects Through Robotic Manipulation. In Proceedings of the 2023 IEEE 14th International Conference on Power Electronics and Drive Systems (PEDS), Montreal, QC, Canada, 7–10 August 2023. [\[CrossRef\]](#)
- Farag, M.; Ghafar, A.N.A.; ALSIBAI, M.H. Real-Time Robotic Grasping and Localization Using Deep Learning-Based Object Detection Technique. In Proceedings of the 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 29 June 2019; pp. 139–144. [\[CrossRef\]](#)
- Çapkan, Y.; Fidan, C.B.; Altun, H. Robotic Arm Guided by Deep Neural Networks and New Knowledge-Based Edge Detector for Pick and Place Applications. In Proceedings of the 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Kocaeli, Turkey, 25–27 August 2021; pp. 1–4. [\[CrossRef\]](#)
- Wali-ur Rahman, M.; Ahmed, S.I.; Ibne Hossain, R.; Ahmed, T.; Uddin, J. Robotic Arm with Proximity and Color Detection. In Proceedings of the 2018 IEEE 7th International Conference on Power and Energy (PECon), Kuala Lumpur, Malaysia, 3–4 December 2018; pp. 322–326. [\[CrossRef\]](#)
- Srinivasamurthy, C.; SivaVenkatesh, R.; Gunasundari, R. Six-Axis Robotic Arm Integration with Computer Vision for Autonomous Object Detection using TensorFlow. In Proceedings of the 2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC), Puducherry, India, 7–8 December 2023; pp. 1–6. [\[CrossRef\]](#)
- Fina, L.; Mascarenhas, T.; Smith, C.; Sevil, H.E. Object Detection Accuracy Enhancement in Color based Dynamic Sorting using Robotic Arm. In Proceedings of the 36th Florida Conference on Recent Advances in Robotics, Virtual, 13–14 May 2021. [\[CrossRef\]](#)
- Li, H.; Dong, C.; Jia, X.; Xiang, S.; Hu, Z. Design of Automatic Sorting and Transportation System for Fruits and Vegetables Based on Dobot Magician Robotic Arm. In Proceedings of the 2023 2nd International Symposium on Control Engineering and Robotics (ISCER), Hangzhou, China, 17–19 February 2023; pp. 229–234. [\[CrossRef\]](#)
- Giang, T.T.H.; Ryoo, Y.J. Autonomous Robotic System to Prune Sweet Pepper Leaves Using Semantic Segmentation with Deep Learning and Articulated Manipulator. *Biomimetics* **2024**, *9*, 161. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ruppert, N.; George, K. Robotic Arm with Obstacle Detection Designed for Assistive Applications. In Proceedings of the 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 17–19 June 2022; pp. 437–443. [\[CrossRef\]](#)
- Hong, S.; Wei, Y.; Ma, M.; Xie, J.; Lu, Z.; Zheng, X.; Zhang, Q. Research of robotic arm control system based on deep learning and 3D point cloud target detection algorithm. In Proceedings of the 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEET), Shenyang, China, 18–20 November 2022; pp. 217–221. [\[CrossRef\]](#)
- Chen, C.W.; Tsai, A.C.; Zhang, Y.H.; Wang, J.F. 3D object detection combined with inverse kinematics to achieve robotic arm grasping. In Proceedings of the 2022 10th International Conference on Orange Technology (ICOT), Shanghai, China, 10–11 November 2022; pp. 1–4. [\[CrossRef\]](#)
- Noh, S.; Kim, J.; Nam, D.; Back, S.; Kang, R.; Lee, K. GraspSAM: When Segment Anything Model Meets Grasp Detection. *arXiv* **2024**, arXiv:2409.12521.
- A. P, J.; Anurag, A.; Shankar, A.; Narayan, A.; T R, M. Robotic and Cyber-Attack Classification Using Artificial Intelligence and Machine Learning Techniques. In Proceedings of the 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 11–12 January 2024; pp. 1–6. [\[CrossRef\]](#)
- Pu, H.; He, L.; Cheng, P.; Chen, J.; Sun, Y. CORMAND2: A Deception Attack Against Industrial Robots. *Engineering* **2024**, *32*, 186–201. [\[CrossRef\]](#)
- Jia, Y.; Poskitt, C.M.; Sun, J.; Chattopadhyay, S. Physical Adversarial Attack on a Robotic Arm. *IEEE Robot. Autom. Lett.* **2022**, *7*, 9334–9341. [\[CrossRef\]](#)
- ANSI/ISA-62443-2-1-2024; Security for Industrial Automation and Control Systems. International Society of Automation (ISA): Durham, NC, USA, 2024.

20. International Society of Automation (ISA). Update to ISA/IEC 62443 Standards Addresses Organization-Wide Cybersecurity in Industrial and Critical Infrastructure Operations. 2025. Available online: <https://www.isa.org/news-press-releases/2025/january/update-to-isa-iec-62443-standards-addresses-organi> (accessed on 1 August 2025).
21. Wang, D. Kinematic and Dynamic Parameters of UFACTORY xArm Series. 2020. Available online: <https://help.ufactory.cc/en/articles/4330809-kinematic-and-dynamic-parameters-of-ufactory-xarm-series> (accessed on 30 August 2025).
22. Yaacoub, J.P.A.; Noura, H.N.; Salman, O.; Chehab, A. Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. *Int. J. Inf. Secur.* **2022**, *21*, 115–158. [CrossRef] [PubMed]
23. Botta, A.; Rotbei, S.; Zinno, S.; Ventre, G. Cyber security of robots: A comprehensive survey. *Intell. Syst. Appl.* **2023**, *18*, 200237. [CrossRef]
24. Federal Bureau of Investigation. 2023 Internet Crime Report. Internet Crime Complaint Center. 2023. Available online: <https://www.ic3.gov> (accessed on 1 May 2023).
25. IOActive. IOActive Conducts First-Ever Ransomware Attack on Robots at Kaspersky Security Analyst Summit 2018. 2018. Available online: <https://www.ioactive.com/article/ioactive-conducts-first-ever-ransomware-attack-on-robots-at-kaspersky-security-analyst-summit-2018/> (accessed on 1 May 2023).
26. Santoso, F.; Finn, A. Trusted Operations of a Military Ground Robot in the Face of Man-in-the-Middle Cyberattacks Using Deep Learning Convolutional Neural Networks: Real-Time Experimental Outcomes. *IEEE Trans. Dependable Secur. Comput.* **2024**, *21*, 2273–2284. [CrossRef]
27. Atkinson, E. Man Crushed to Death by Robot in South Korea, 2023. Available online: <https://www.independent.co.uk/asia/east-asia/south-korea-robot-kills-man-b2444245.html> (accessed on 1 August 2025).
28. Jolly, B. Alarming Moment Robot Arm Crushes Worker to Death in Factory. 2024. Available online: <https://www.mirror.co.uk/news/world-news/shocking-moment-robot-arm-crushes-32459232> (accessed on 1 May 2024).
29. Independent Digital News and Media. Tesla Engineer Attacked by Robot at Company’s Giga Texas Factory. 2023. Available online: <https://www.independent.co.uk/news/world/americas/tesla-robot-attacks-engineer-texas-b2470252.html> (accessed on 15 November 2023).
30. Baek, W.J.; Kröger, T. Safety Evaluation of Robot Systems via Uncertainty Quantification. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 10532–10538. [CrossRef]
31. San Martin, A.; Kildal, J.; Lazkano, E. Taking Charge of One’s Own Safety While Collaborating with Robots: Enhancing Situational Awareness for a Safe Environment. *Sustainability* **2024**, *16*, 4024. [CrossRef]
32. ISO 10218:2025; Robots and Robotic Devices—Safety Requirements for Industrial Robots—Parts 1 and 2. International Organization for Standardization: Geneva, Switzerland, 2025. Available online: <https://www.iso.org/standard/73933.html> (accessed on 30 August 2025).
33. Ahsan, M.; Nygard, K.E.; Gomes, R.; Chowdhury, M.M.; Rifat, N.; Connolly, J.F. Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review. *J. Cybersec. Priv.* **2022**, *2*, 527–555. [CrossRef]
34. Fosch-Villaronga, E.; Mahler, T. Cybersecurity, safety and robots: Strengthening the link between cybersecurity and safety in the context of care robots. *Comput. Law Secur. Rev.* **2021**, *41*, 105528. [CrossRef]
35. Rodríguez-Rodríguez, J.A.; López-Rubio, E.; Ángel Ruiz, J.A.; Molina-Cabello, M.A. The Impact of Noise and Brightness on Object Detection Methods. *Sensors* **2024**, *24*, 821. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.