biomimetics

Article

Pose Estimation of a Cobot Implemented on a Small AI-Powered Computing System and a Stereo Camera for Precision Evaluation

Marco-Antonio Cabrera-Rufino *,†, Juan-Manuel Ramos-Arreguín *,†, Marco-Antonio Aceves-Fernandez †, Efren Gorrostieta-Hurtado †, Jesus-Carlos Pedraza-Ortega † and Juvenal Rodríguez-Resendiz †

Facultad de Ingeniería, Universidad Autónoma de Querétaro, Cerro de las Campanas, Las Campanas, Queretaro 76010, Mexico; marco.aceves@uaq.mx (M.-A.A.-F.); efren.gorrostieta@uaq.mx (E.G.-H.); carlos.pedraza@uaq.mx (J.-C.P.-O.); juvenal@uaq.edu.mx (J.R.-R.)

* Correspondence: mcabrera23@alumnos.uaq.mx (M.-A.C.-R.); juan.ramos@uaq.mx (J.-M.R.-A.); Tel.: +52-558-057-8823 (M.-A.C.-R.); +52-442-250-0031 (J.-M.R.-A.)

† These authors contributed equally to this work.

Abstract: The precision of robotic manipulators in the industrial or medical field is very important, especially when it comes to repetitive or exhaustive tasks. Geometric deformations are the most common in this field. For this reason, new robotic vision techniques have been proposed, including 3D methods that made it possible to determine the geometric distances between the parts of a robotic manipulator. The aim of this work is to measure the angular position of a robotic arm with six degrees of freedom. For this purpose, a stereo camera and a convolutional neural network algorithm are used to reduce the degradation of precision caused by geometric errors. This method is not intended to replace encoders, but to enhance accuracy by compensating for degradation through an intelligent visual measurement system. The camera is tested and the accuracy is about one millimeter. The implementation of this method leads to better results than traditional and simple neural network methods.

Keywords: precision degradation; robot arm; robotic vision; stereo camera; convolutional neural network; 3D vision

check for updates

Citation: Cabrera-Rufino, M.-A.; Ramos-Arreguín, J.-M.; Aceves-Fernández, M.-A.; Gorrostieta-Hurtado, E.; Pedraza-Ortega, J.-C.; Rodríguez-Resendiz, J. Pose Estimation of a Cobot Implemented on a Small AI-Powered Computing System and a Stereo Camera for Precision Evaluation. *Biomimetics* **2024**, *9*, 610. https://doi.org/10.3390/biomimetics9100610

Academic Editor: Andrew Adamatzky

Received: 26 August 2024
Revised: 20 September 2024
Accepted: 26 September 2024
Published: 9 October 2024

Copyright: © 2024 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

Robotic vision and intelligent algorithms are of great importance for robotic manipulators, as they enable more accurate and precise task execution. As a result, the use of these methods has begun to find applications in metrology. Robotic manipulators are indispensable tools for the automotive or aerospace industries. Among other things, they perform repetitive, complex, or strenuous tasks, so robots become increasingly inaccurate and imprecise over time. Normally, the measurement of a degree-of-freedom (DoF), i.e., the number of independent joints or axes of movement that give it the ability to move, is achieved in angular units (degrees or radians). Sensors such as encoders, tachometers, gyroscopes, and others are used for this purpose.

A robotic manipulator is a mechanically articulated arm consisting of links connected by joints that allow relative motion between two consecutive links. The number of independent or minimum coordinates required to describe the robot position defines the number of DoF. Robotic arms move and manipulate tools called end effectors. These can be grippers, welding wires, drills, suction cups, cameras, or others. Figure 1 shows a prototype robot arm with 3 DoF, where (x, y, z) is the position of the end effector and θ_1, θ_2 and θ_3 are the angular position of each DoF. The robot Tool Center Point (TCP) is the point used for positioning the robot in any robot program where targets are defined in Cartesian space. TCP is defined as a transformation from the robot mounting flange [1].

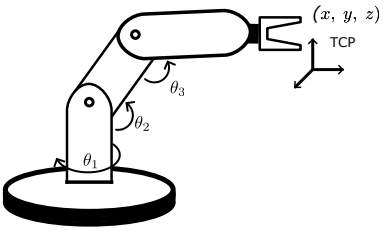


Figure 1. A 3-DoF robot architecture and its principal components, this model is simplified to understand the main concepts [1].

In the field of robotics, the precision of the robot is to reach a certain point in the working area with a minimum position error. Precision helps in evaluating the performance of a robot, and its absence is called precision degradation. It is divided into three groups, which are caused by structural deformations of the joints or geometric errors. Then there are the non-geometric errors such as friction and gear backlash. And, finally, the errors caused by computational processing. Typical geometric errors include tolerances, elasticities, eccentricities, and wear [2–5].

On the path to precision degradation analysis, computer vision devices and artificial intelligent techniques are methodologies implemented in robotic applications [6], such as control [7–11], inverse kinematics [12,13], and object segmentation [14–16]. Stereo cameras have become more precise in the last ten years [14,17–20], and actual microcomputers like Raspberry Pi model 4 or Jetson Nano 280 are powerful enough to perform real-time measuring tasks. However, there is little background on the use of these methods as an alternative measurement system [21], and even less on the specific use of stereo cameras and convolutional neural networks (CNNs), which are a type of deep learning (DL) model particularly effective for image analysis [22].

Table 1 contains state-of-the-art for this work. It has the reference of the author, the methodology, and improvements.

Table 1. Summary of state-of-the-art. The methodology implemented in every work was checked with a ✓.

Reference	Conventional Techniques	AI Techniques	Robotic Vision	ME before [mm]	ME after [mm]
[23]	✓				
[24]	✓	✓		2.613	0.31
[25]	✓	✓		0.7411	0.1007
[26]		✓	✓		27.4
[27]		✓	✓		
[28]	✓			1.2248	0.2678
[29]	✓	✓			0.1
[30]	✓	✓			
[31]	✓	✓			

We consider that, given the current state-of-the-art and its results (see Table 1), a better system is needed to compensate for the precision. The industry spends a lot of money on the maintenance and purchase of new robotic systems. Some authors have looked at the measurement of precision degradation. It is important to say that despite the advances in robotic vision, there remains a need for more accurate and industry-applicable methods of measuring angular positions in robotic arms. The gap between the existing need and the specific methods that have not yet been implemented in depth are being adsressed.

The methods mentioned in Table 1 are not small enough to meet industry requirements (less than 0.1 millimiters (mm)), or they are so complex that they cannot be applied to other robot morphologies. In this paper, we present a robust angle measurement method that uses one stereo camera, Intel RealSense model D435, and a CNN algorithm to capture the color and depth information of the robot. The red, green, and blue (RGB), i.e., a color space commonly used to color a pixel in images and videos, is useful for segmentation and is mapped to the depth channel with an accuracy of 0.001 mm. The algorithm splits the 3D robot morphology into separate 2D spaces, with one 2D space representing each DoF. The combination of these methods improves the position of the robot with considerable precision. Section 2 describes the materials and methods used in this work. Section 3 shows all the results of this work, Section 4 is the discussion of the findings. Finally, Section 5 refers to the conclusions of this work.

2. Materials and Methods

These are the materials and methods used in this work, including the camera and the collaborative robot, also known as a cobot, is an industrial robot that can work safely alongside humans in a shared workspace. This is in addition to the necessary theoretical concepts such as convolutional neural networks, stereoscopic configuration, and others.

2.1. Stereo Vision

Stereo vision is a technique for estimating three-dimensional structures of the physical world using two images taken from different angles [6]. Figure 2 shows how the stereo view is composed, which consists of two cameras with the same parameters. At a reference point $P(X_p, Y_p, Z_p)$, the projection onto the right image plane is $p_1(x_1, y_1)$ and onto the left image plane is $p_2(x_2, y_2)$, where f is the focal length and b is the distance between the optical centers of the two cameras. Referring to the Figure 3 and the principle of similar triangles, we define the disparity d in the Equation (1), then we can obtain the depth Z of the point P from the Equations (1) and (2) [16]:

$$d = x_1 - x_2, \tag{1}$$

$$Z = b \times f/d, \tag{2}$$

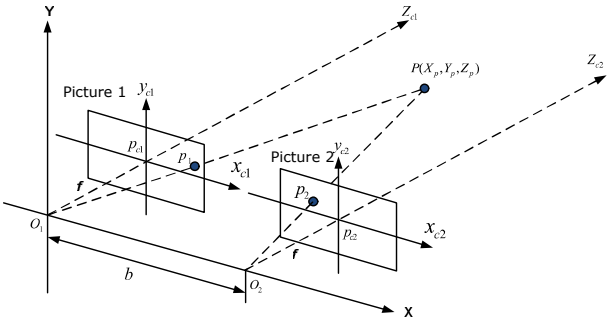


Figure 2. Schematic of stereo vision configuration in three dimensions [16,32].

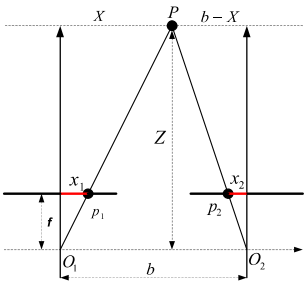


Figure 3. Triangulation scheme analysis for stereo vision a two dimensional perspective [16].

To estimate the disparity in the digital image pair, the neighborhood of a pixel p in the left image is compared to the neighborhood of the pixel image q in the right image, where q is shifted by a possible disparity δ_p compared to p. In Figure 4, for each pixel p, N possible disparities $\left(\delta_p^1, \delta_p^2, \ldots, \delta_p^N\right)$ are tested ($N = 256$ and 65,536 for 8 and 16 bits of depth mapping) and the resulting possible disparity with the lowest fitting cost is assigned to the pixel p [33].

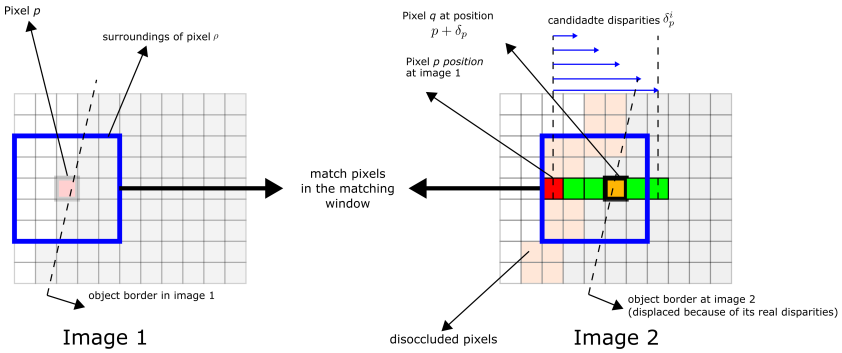


Figure 4. Matching windows in stereo vision. This represents the digital functionality of a commercial stereo camera [33].

Accuracy of the D435 Stereo Camera

Using [16,33], the accuracy of the stereo camera was calculated, some images were taken with a ruler on the background to measure 1 centimeter (cm), the distance between two pixels was drawn manually, it was not necessary to obtain 1 cm accuracy. Then, color was added to the two pixels with the information of coordinates $(x,\ y,\ z)$. The distance between the pixels was calculated with eq. Then, the same task was performed with the next pixels and the distance between two consecutive pixels was determined, which represents the accuracy of the stereo camera.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}, \tag{3}$$

Figure 5 shows the distance between 2 pixels. First, pixels 1 and 2 are measured, then pixels 1 and 3, and so on up to pixel 5. Then, the smallest distance, the distance from pixel 3 minus 2, is calculated, and so on up to pixel 5. The results are shown in Tables 2 and 3.

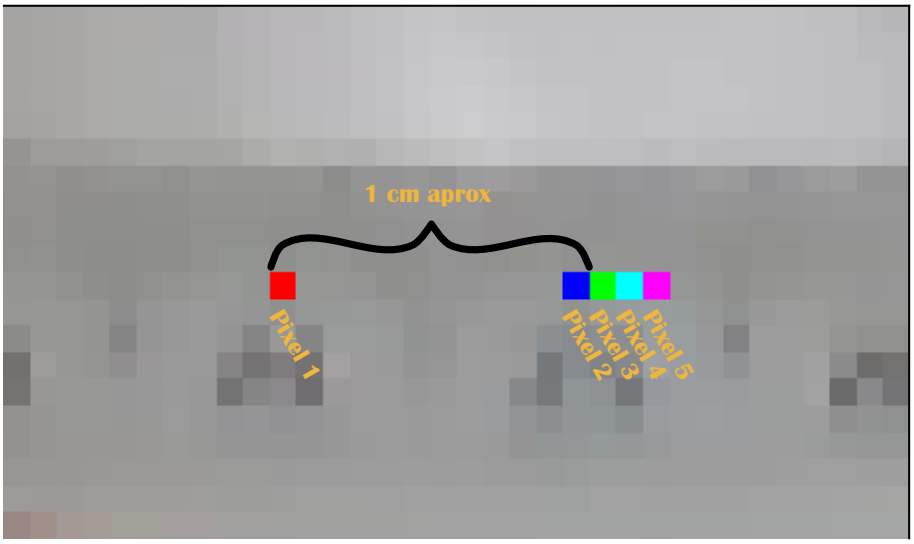


Figure 5. Distance of 1 cm in pixels, and the subsequent pixel to measure the D435 camera.

Table 2. Distance between pixels for measuring the accuracy of the camera according to Figure 5. The distance between pixel 1 and the following consecutive pixels.

Pixels	Distance [mm]
1 to 2	13.9
1 to 3	15.0
1 to 4	16.3
1 to 5	17.7

Table 3. The distance between two consecutives pixels is the resolution for distance measurement.

Pixels	Distance between Two Consecutive Pixels [mm]
3–2	1.1
4–3	1.3
5–4	1.4

As it can be observed, the robot has an accuracy of about 1 mm, which varies a little depending on the distance and the angle of measurement.

2.2. Materials

The current work implemented the ISO-9283 [34] for prototype testing. The list of materials and its characteristics used for this work are listed below.

- 1 MyCobot 280 nano, its main features are:
  - Workspace 280 mm.
  - Joint range $\pm 170^{\circ}$.
  - Repetitively ± 5 mm.
  - A micro-computer Jetson nano 280, with 2 GB of RAM and NVIDIA processor
  - Servomotors of high precision.
  - 6 DoF.
- 1 depth camera model Intel RealSense D435. Its main features are:
  - Focal length of 1.88 mm.
  - Depth resolution 1280×720 pixels, ± 0.006 mm
  - Workspace 0.2–10 m.
  - 6 DoF para calibración.
- A generic monitor.
- Mouse and keyboard.

The system is set up on a 150×56 cm table, as shown in Figure 6. The Cobot is placed at a distance of 63.5 cm from the top and 32 cm from the left side of the table. The dimensions of the robot are 15×11 cm and the camera is 2.5×9 cm. The colored circle represents the robot's working area, while the triangular is the camera's viewing area. One camera is used at the lower end of the table length and another at the upper end.

As shown in Figure 7, each DoF is measured sequentially from q_1 to q_6. Each DoF is enumerated with colored labels to make it easier for the camera to segment them. Also, the dimensions of every link of the Cobot are shown.

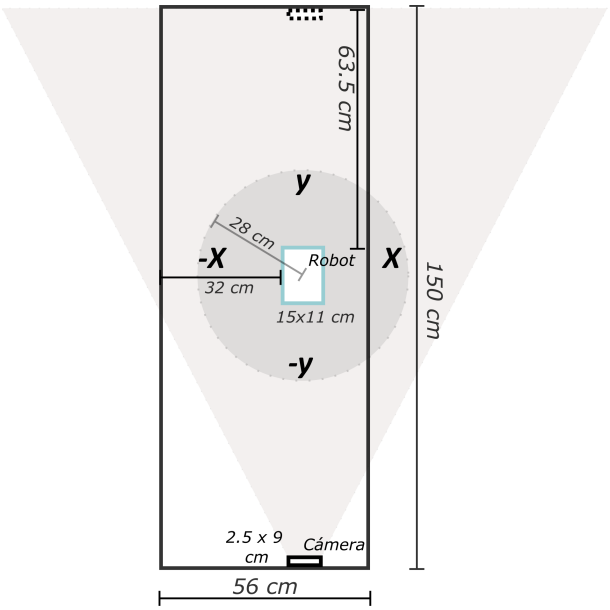


Figure 6. Configuration of the materials, the rectangle represents a table of 56 × 150 cm, the triangle is the view of the stereo camera, the circle is the workspace of the robot.

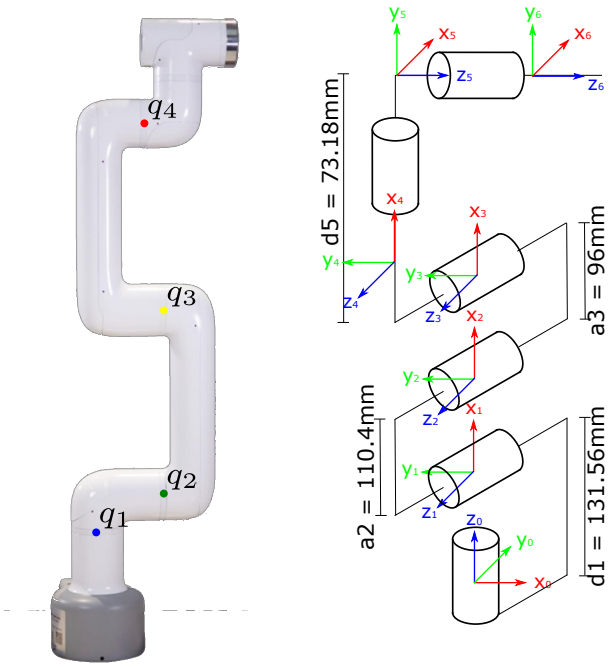


Figure 7. 6 DoF Cobot morphology. Only 4 DoF are labeled with a different color.

2.3. Graphical User Interface (GUI)

To enhance the precision when managing cobot manipulation, a Graphical User Interface (GUI), used to create an interaction between the user and the computer, is developed to exhibit the fundamental motions associated with each DoF (θ_1, θ_2, θ_3, θ_4, θ_5, θ_6) and the position values of the end effector (x, y, z, r_x, r_y, r_z). The interface also checks whether the robot is connected and has a button to bring the robot to the home position (all angles to zero) and a button to switch off the servomotors as a precaution. This interface is constantly being improved, so more test modes for positioning will be implemented in the future. Figure 8 shows the GUI.

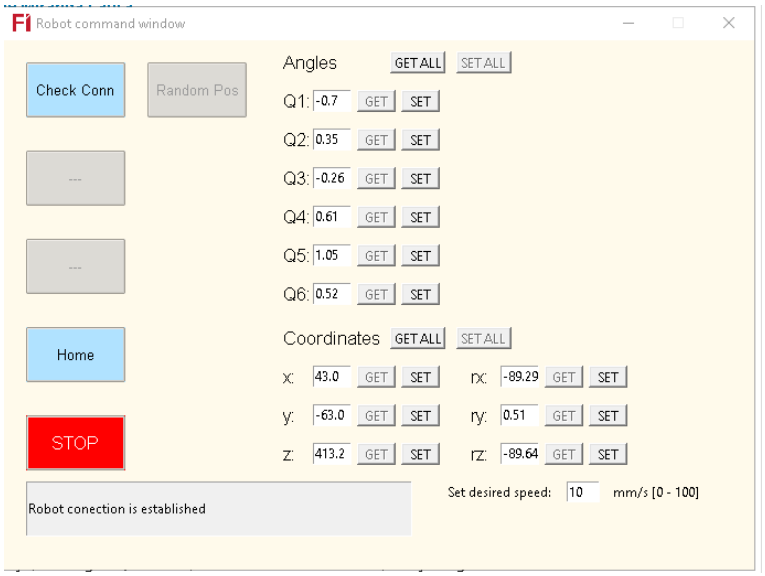


Figure 8. GUI used for better manipulation of the robot arm, for communication; RPC (Remote Procedure Call) protocol is implemented, RPC communication is based on TCP (Transfer Control Protocol) architecture.

2.4. Measurement of Position Accuracy

Position accuracy is the deviation between the commanded position of the TCP point (N) and the mean (barycenter, G) calculated from the cluster of repeatedly reached TCP positions. Figure 9 shows the position accuracy of 1 DoF, where G is the barycentre $[\bar{x}, \bar{y}, \bar{z}]$, and N is the commended position $[x, y, z]$. The position accuracy of 1 DoF a_p is given by Equation (4).

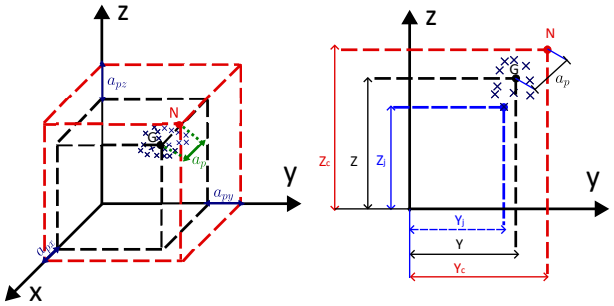


Figure 9. Graphical displaying of 1 DoF position accuracy according to ISO 9283-1998 [34].

$$a_p = \sqrt{a_{px}^2 + a_{py}^2 + a_{pz}^2}, \tag{4}$$

where $a_{px} = \bar{x} - x_{cmd}$, $a_{py} = \bar{y} - y_{cmd}$, and $a_{pz} = \bar{z} - z_{cmd}$. The position accuracy a_p of a robot manipulator is defined as the distance between the commanded pose x_{cmd} and the mean $\bar{\mathbf{x}}$ of N reached positions, as described in Equation (5)

$$x_{cmd} = \begin{bmatrix} x_{cmd} \\ y_{cmd} \\ z_{cmd} \end{bmatrix}, \ \bar{x} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix}, \tag{5}$$

The mean coordinates are given in Equation (6).

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \ \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \ \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i, \tag{6}$$

x_i, y_i, and z_i are the position coordinates of the i-th measured pose. The positioning accuracy a_p is obtained using the Euclidean norm (7).

$$a_p = \|(\bar{\mathbf{x}} - \bar{x}_{cmd})\| = \sqrt{(\bar{x} - x_{cmd})^2 + (\bar{y} - y_{cmd})^2 + (\bar{z} - z_{cmd})^2}, \quad (7)$$

A measurement plan is required to control the robot movements. The controlled robot movements for the evaluation of accuracy degradation cannot be arbitrary. The measurement plan requires that the robot TCP moves over the entire working area and is evenly distributed in both articulated and Cartesian space. The even distribution of the samples prevents the analysis algorithm from overlooking or overweighting errors, which would falsify the results. Covering the entire joint space and Cartesian space means that the measurement plan trains the robot over a subset of joints or work zones. This allows the performance of the joints to be recorded over the entire motor and encoder range. Covering the entire workspace helps to evaluate different stiffness conditions. When creating the measurement plan, a collision check is performed to minimize possible interruptions during movement and measurement. A visibility check is also performed to ensure that the planned positions are not obscured by the measuring instrument [2].

A Cobot 280 Nano from Elephant Robotics was used for this work. The fixed-loop motion generated for the cobot is shown in Figure 10. As the cobot was mounted on a table, only poses above the table are valid for this use case. The robot movement assumes an unobstructed working volume above the table.

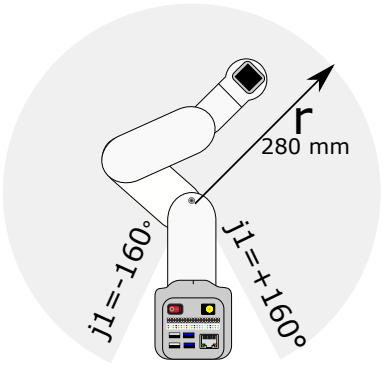


Figure 10. Workspace of the Cobot 280 Nano.

Based on [2], a grid moment of the Cobot 280 Nano is planned. The tool center positions (TCP) (x, y, z) of the robot are measured. The differences between the nominal positions and the measured positions are compared. Control-level measurement data are also collected from each joint to understand the influences of temperature, payload, and velocity on position changes. Control-level information provides clues about the causes of robot performance degradation by providing information on the actual and joint positions, velocities, currents, accelerations, torques, and temperatures.

2.5. Convolutional Neural Networks

Also known as CNNs, these are the standard neural network architecture used for predictions when the input are images, which is the case in a variety of neural network applications, such as this work.

The implemented architectures for this projects are the followed:

LeNet-5 is one of the earliest pre-trained models proposed by Yann LeCun and others in 1998 in the research paper Gradient-Based Learning Applied to Document Recognition. They used this architecture to recognize handwritten and machine-printed characters. The main reason for the popularity of this model was its simple architecture. It is a multilayer convolutional neural network for image classification, Figure 11 shows the description of its architecture [35].

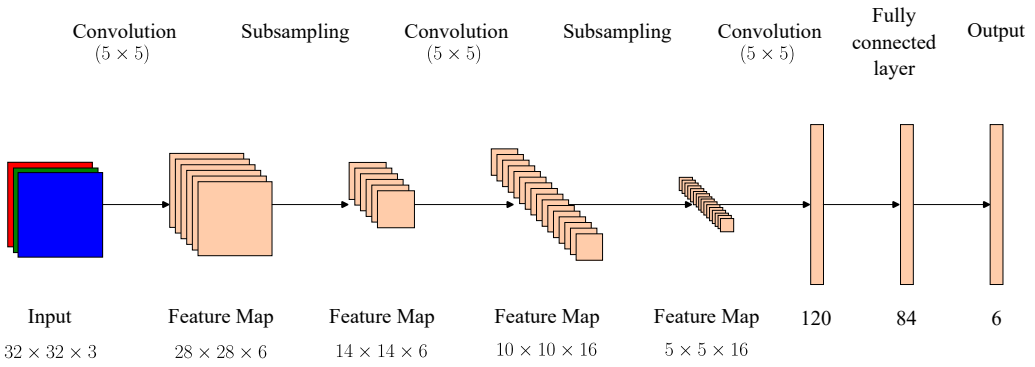


Figure 11. LENET-5 architecture [35].

AlexNet is a convolutional neural network model proposed by K. Simonyan and A. Zisserman. The model achieves a top-5 test accuracy of 92.7% on ImageNet, a dataset with over 14 million images belonging to 1000 classes, Figure 12 displays the architecture of this network [36].

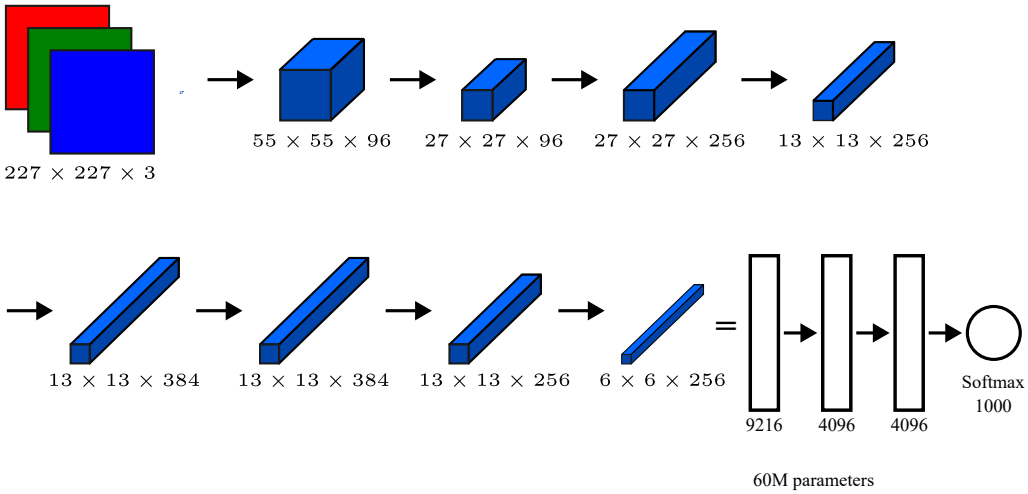


Figure 12. AlexNet architecture [36].

YOLO−v1 "You Only Look Once" is a system for recognizing objects on the Pascal VOC 2012 dataset. It can recognize the 20 Pascal object classes (most of them are animals). All previous recognition systems use classifiers or localizers to perform recognition. They apply the model to an image at multiple locations and scales. Regions with high scores are considered recognitions. The YOLO designers use a completely different approach. We apply a single neural network to the entire image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted according to the predicted probabilities [37].

Using all the methods and materials presented in this section, the results achieved are presented below.

3. Results

First, inverse kinematics was obtained using CNN, as a first test of the Deep Learning techniques, the angle of each DoF $(\theta_1, \theta_2, \cdots, \theta_6)$ and the geometric position of the TCP (x, y, z) were recorded and stored to run the algorithm. The robot moved into its workspace with 16 different positions or setpoints; finally, a matrix of 3070 × 9 data was created. Figure 13 shows the position of every axis (x, y, z), the setpoint and actual value of every one, as it can be observed precision is observable in the graphs the fourth graph is all the points in a three-dimensional plot.

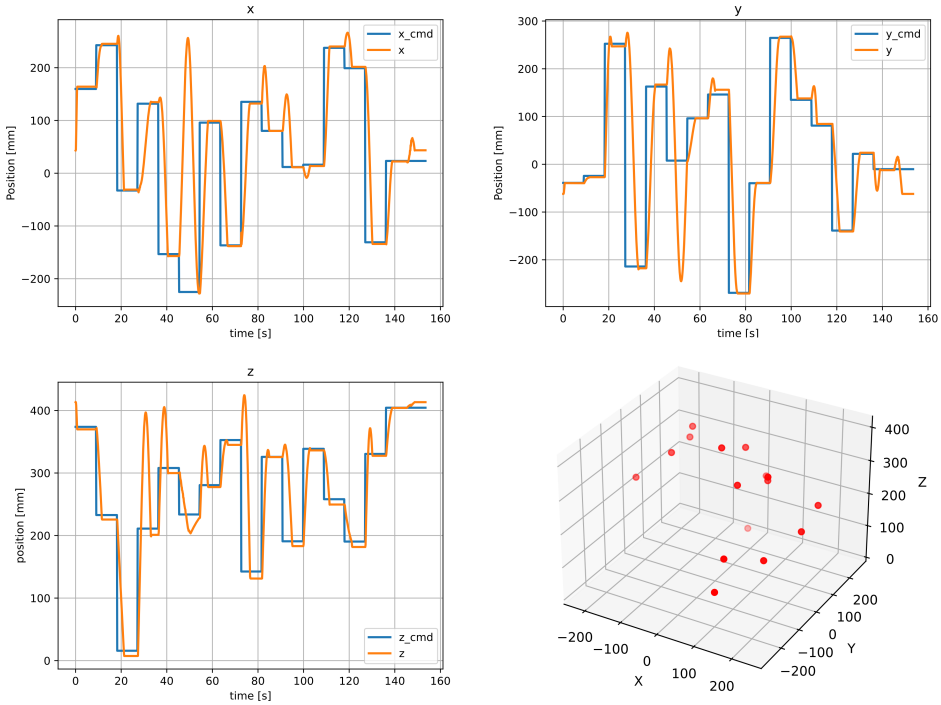


Figure 13. Every three-dimensional position of TCP position (x, y, z), the fourth graph is the points that the Cobot follows.

The algorithm for predicting the inverse kinematic solution is a CNN of 1 dimension that takes the TCP points (x, y, z) as the inputs and the angles of the DoF as the outputs $(\theta_1, \theta_2, \cdots, \theta_6)$. The results show that this is a powerful task, traditional methods are more complicated. Figure 14 shows the model of the one-dimensional CNN developed for this task, and Figure 15 shows the training results of the CNN. For this test, 3024 samples were used, with 2456 for training and 614 for validation. The Adam optimizer was used, including the mean squared error loss, and the accuracy metric was plotted.

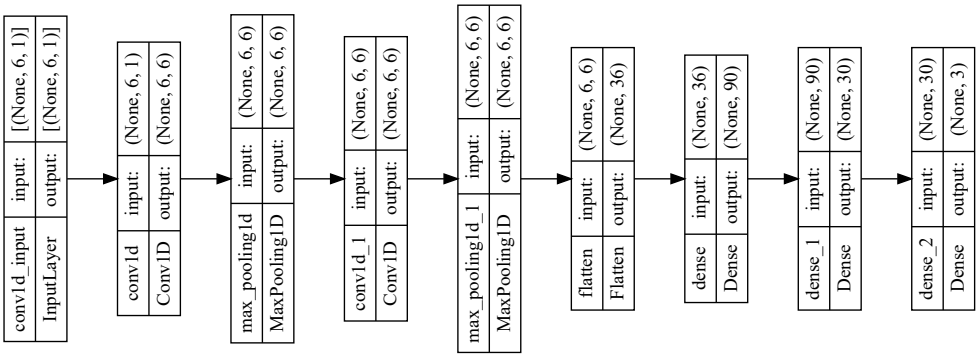


Figure 14. Design of the 1-dimensional CNN used for predicting inverse kinematics, highlighting the input layer, convolutional layers, and output.

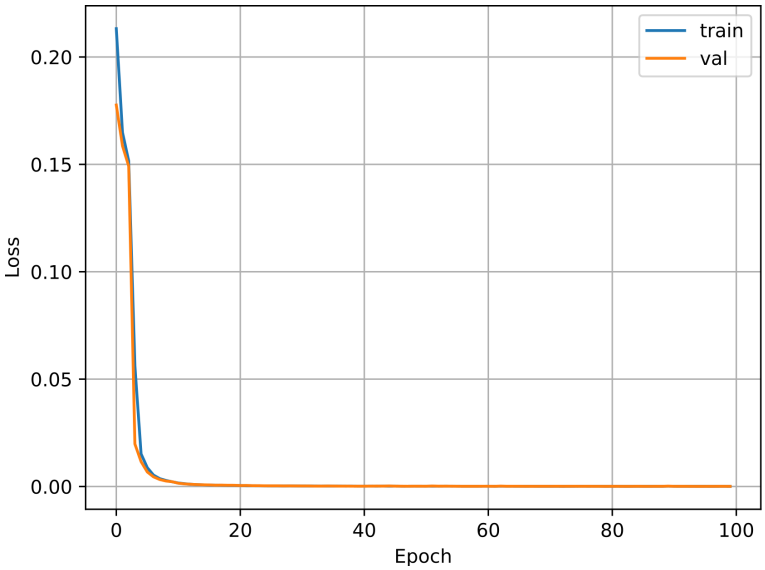


Figure 15. Training vs. validation loss function of the 1 dimensional CNN.

The next step in testing the DL techniques was to use the depth information of the cobot image to predict the results of the inverse kinematics. The D435 stereo camera provides color and depth information, 680 × 460 pixels for the RGB image, and the depth information has the same resolution. Thus, each color pixel can be represented in a depth pixel with a 16-bit value.

A sample of RGB images of the robot can be found in Figure 16. Figure 17 shows the same images but with depth information. For training this model, 120 images were used. The information of the angle of each DoF was replaced by a depth image of the robot, which makes the design of the CNN more complicated. LENET-5 and AlexNet designs were implemented.



Figure 16. Nine samples of RGB pictures of the Cobot, the pixels are 480 × 640 pixels.

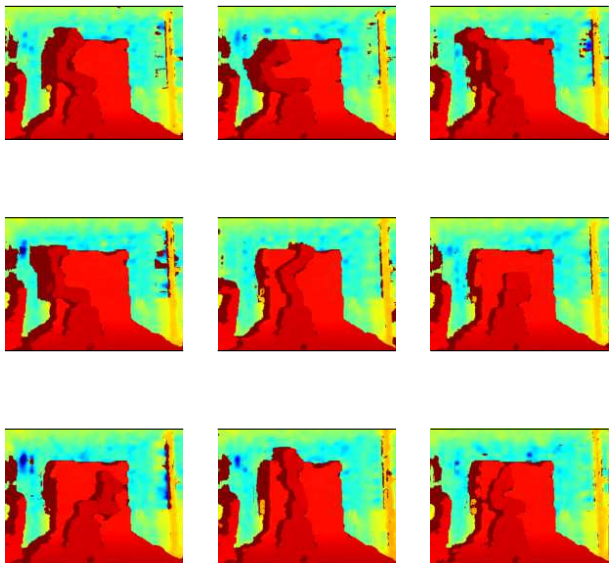


Figure 17. Nine samples of depth information for the Cobot, it has the same size of a color image, every pixel has a resolution of 16 bits.

Figure 18 shows the results of the training of LENET-5 and Figure 19 for AlexNet. For these tests 160 samples were used, 128 for training and 32 for validation. For LENET-5, the Adam optimizer was used, including the mean squared error loss and the accuracy metric. AlexNet was configured with the AdamMax optimizer.

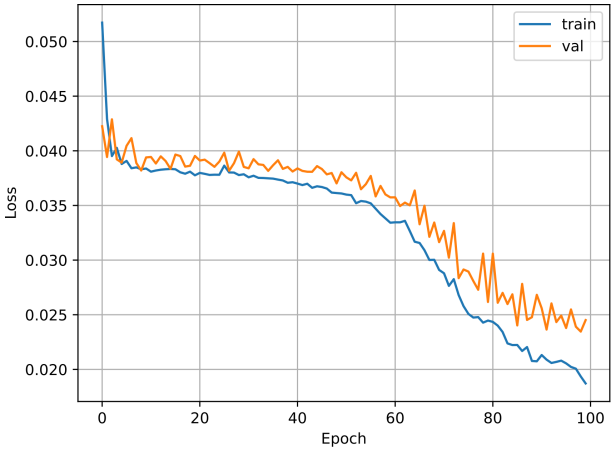


Figure 18. Training vs. validation loss function of LENET-5.

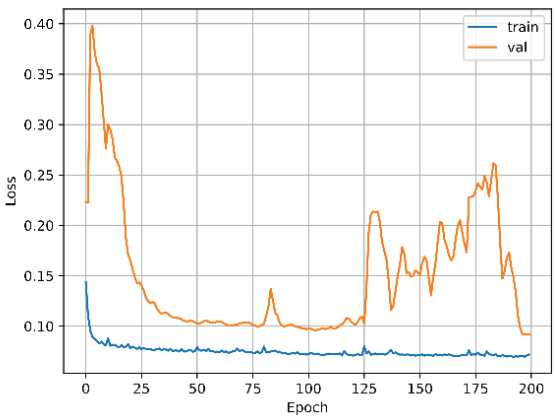


Figure 19. Training vs. validation loss function of AlexNet.

The next steps were to put the colored labels on the first four DoF of the cobot (4 out of 6 DoF were used) and train a network to recognize the labels through learning; this is shown in Figure 20. A blue, green, yellow, and red label for DoF 1 to 4, respectively, was attached to a specific point on the robot. The label was rounded and had a radius of 0.6 cm.



Figure 20. Images of the robot with colored labels, using YOLO-V1.

For label annotation, the CVAT platform was used, it allowed for converting raw images into a custom trained computer vision model. It supports object recognition and classification models.

For label recognition, YOLO-V1 was implemented, running in the Roboflow platform. The results are shown in Figure 21. For this task, 215 images of the robot in different posses were implemented. The algorithm divided the images into 151 images for training, 42 for validation and 22 for testing.

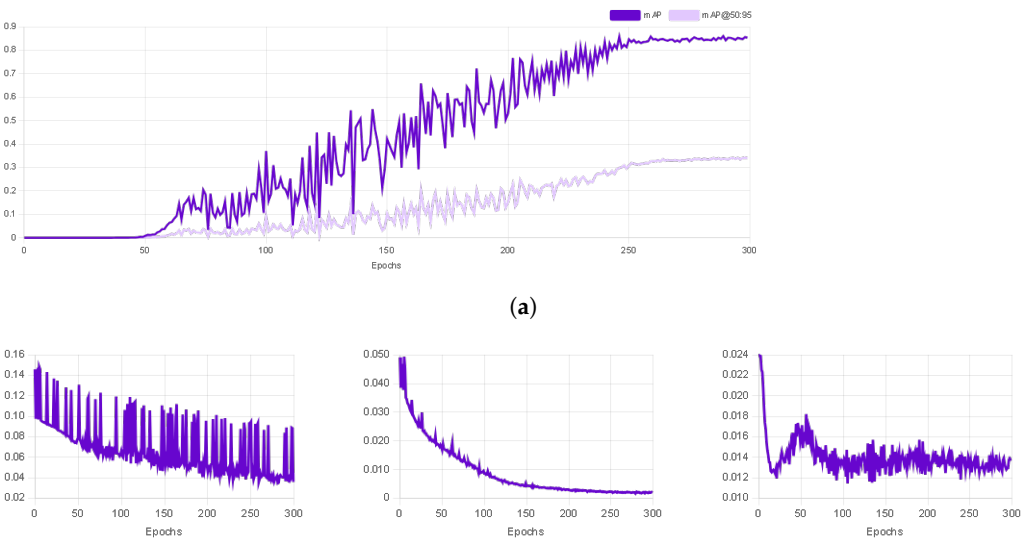


Figure 21. Yolo-V1 training results for label identification. (**a**) **Mean average precision (mAP)** is used to measure the performance of computer vision models. (**b**) **Box Loss** refers to how well the model predicts the positions and sizes of bounding boxes around objects in an image. (**c**) **Class Loss** refers to the measure of how accurately the model is predicting the correct class or label of the objects it detects. (**d**) **Object Loss** measures how well the model recognizes the presence or absence of an object in a particular region of an image.

Figure 22 shows some images of the Cobot that were classified, none of them were used for training, validation, or testing stage, the results given are satisfactory, the next step in the experimentation is to interpolate labels information with depth information of an image to locate the tridimentional position of every DoF.



Figure 22. Results with a new set of images to evaluate the Yolo-V1 algorithm. The red dots are the recognized labels.

Depth camera D435 provided the tridimensional information *(x, y, z)* of every pixel. Figure 23 shows the depth information, a similar set of data with x and y orientation was also obtained.



Figure 23. Depth information of the color images.

Yolo-V1 provided information for every label in the Cobot, information can be interpolated to the depth image so we can obtain the tridimensional information of every DoF. This can be observed in Figure 24.

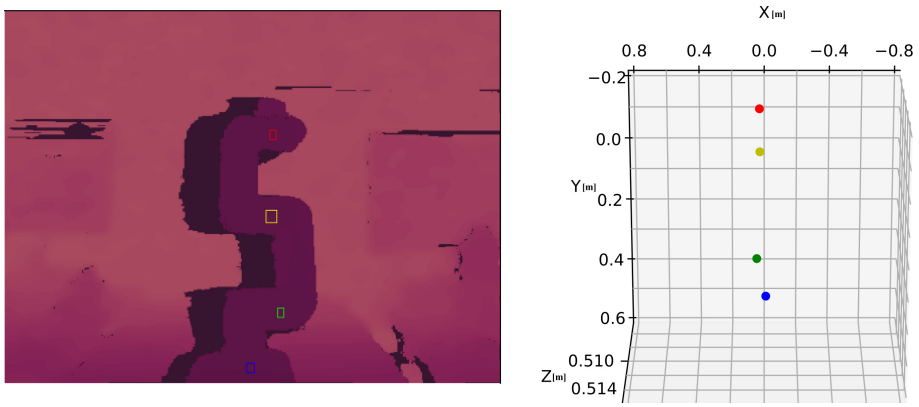


Figure 24. The image on the left is the depth information of the color images with the location of every label. The color box represents the workspace where labels are projected. The dots in the color box are the marked points detected by the YOLO-V1 algorithm that helps measure each DoF angular position, where blue dot is the first DoF label, green dot is the second DoF label, yellow dot is the third DoF label and finally, the red dot is the fourth DoF label.

This methodology outperforms all work with an average error of 1 mm. Moreover, the set of our implemented methods leads to clearer and reproducible work. It is the first time that stereo cameras and CNN are used to solve this particular problem.

4. Discussion

In this work, a measurement position with an accuracy of about 1 mm was achieved (limited by the standard camera resolution of 1 mm). Compared with previous work, conventional techniques are more accurate but more difficult to reproduce as they involve geometric analysis. Other previous works that present an intelligent algorithm and/or robot vision have an error of less than 1 mm and above. Otherwise, other works do not indicate the degree of improvement. Our work presents a methodology to measure the degradation of the positioning of the links of a Cobot, using a combination of artificial intelligence and computer vision. This combination makes it possible to obtain results that are reproducible in other architectures of a manipulator robot. It is important to mention that the reduction of the error in the measurement of the link position can be reduced by using a more precise camera and basically the same methodology presented. This opens up a line of work that can benefit the industry by extending the life of a robot while ensuring the quality of a product [2]. Based on our results, we should consider technologies with higher measurement accuracy in the future, such as a KEA camera with Time-of-Flight (ToF) technology, or replacing the depth technology with a lidar system. As mentioned earlier, there were certain limitations in terms of the resolution of the camera, which is the main drawback of this work. This system is suitable for applications where an accuracy of 1 mm is required and where the integration of non-invasive sensors such as cameras is necessary. However, it is necessary to work on the use of two chambers to avoid occlusions.

5. Conclusions

With the implemented methods, convolutional neural networks and stereo depth cameras, there is an intuitive adaptation that makes experimentation relatively comfortable. Preliminary results show that the camera tends to have a precision of 1 mm resolution, which is not sufficient for our application. However, in the future, we will work with newer stereo cameras or other depth technology. Cobot is a good tool to work with as its accuracy is in the millimeter range, which is an important requirement for our work.

LENET-5 and AlexNet as well as a proprietary CNN are used to first determine the inverse kinematics of the robot, initially with pure data from the encoders, then depth images are included with good learning results, for labeling with YOLO V1 the accuracy was quite acceptable with 91%. Finally, the methods used to analyze our positioning improvement are the most appropriate and advanced. Compared with previous methods, our results are more reproducible, with a maximum resolution of 1 mm. Depth cameras bring us closer to better results, although not the best, which is why we intend to use higher resolution stereo cameras. These deep vision methods could be the standard of the future.

Author Contributions: Conceptualization, M.-A.C.-R. and J.-M.R.-A.; Methodology, M.-A.C.-R., M.-A.A.-F. and J.R.-R.; Software, M.-A.C.-R. and J.-C.P.-O.; Validation, M.-A.A.-F., E.G.-H., J.-C.P.-O. and J.R.-R.; Formal Analysis, J.-M.R.-A., M.-A.A.-F., E.G.-H., J.-C.P.-O. and J.R.-R.; Investigation, M.-A.C.-R. and E.G.-H.; Resources, M.-A.C.-R. and J.-M.R.-A.; Data Curation, M.-A.A.-F., J.R.-R. and J.-C.P.-O.; Writing – Original Draft Preparation, M.-A.C.-R., J.-M.R.-A. and E.G.-H.; Writing – Review and Editing, J.-M.R.-A., M.-A.A.-F., E.G.-H., J.-C.P.-O. and J.R.-R.; Supervision, J.-M.R.-A., M.-A.A.-F., E.G.-H., J.-C.P.-O. and J.R.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Data Availability Statement: The original contributions presented in the study are available on request for the corresponding author.

Acknowledgments: I would like to express my sincere gratitude to the Consejo Nacional de Humandades, Ciencias y Tecnologías (CONAHCYT) for the financial support I received through the scholarship during the development of my studies Postgraduate Faculty of Engineering by the 219 Universidad Autonoma de Querétaro,under Grant CVU 957109.

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

References

1. Kelly, R.; Santibáñez, V. *Control de Movimiento de Robots Manipuladores*; Pearson Educación: London, UK , 2003.
2. Qiao, G.; Weiss, B. Industrial robot accuracy degradation monitoring and quick health assessment. *J. Manuf. Sci. Eng.* **2019**, *141*, 071006 . [CrossRef]
3. Zhang, D.; Peng, Z.; Ning, G.; Han, X. Positioning accuracy reliability of industrial robots through probability and evidence theories. *J. Mech. Des.* **2021**, *143*, 011704. [CrossRef]
4. Wang, Z.; Liu, R.; Sparks, T.; Chen, X.; Liou, F. *Industrial Robot Trajectory Accuracy Evaluation Maps for Hybrid Manufacturing Process Based on Joint Angle Error Analysis*; OMICS International: Hyderabad, India, 2018.
5. Pérez, L.; Rodrıíguez, Í; Rodríguez, N.; Usamentiaga, R.; García, D. Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors* **2016**, *16*, 335. [CrossRef]
6. Corke, P. *Robotics, Vision and Control: Fundamental Algorithms in MATLAB® Second, Completely Revised*; Springer: Berlin/Heidelberg, Germany, 2017.
7. Zeng, R.; Liu, M.; Zhang, J.; Li, X.; Zhou, Q.; Jiang, Y. Manipulator control method based on deep reinforcement learning. In Proceedings of the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020 ; pp. 415–420.
8. Sangeetha, G.; Kumar, N.; Hari, P.; Sasikumar, S. Implementation of a stereo vision based system for visual feedback control of robotic arm for space manipulations. *Procedia Comput. Sci.* **2018**, *133*, 1066–1073.
9. Liang, B.; Li, T.; Chen, Z.; Wang, Y.; Liao, Y. Robot arm dynamics control based on deep learning and physical simulation. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 2921–2925.
10. Li, C.; Chang, Y. Automated visual positioning and precision placement of a workpiece using deep learning. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 4527–4538. [CrossRef]
11. Li, Q.; Pang, Y.; Wang, Y.; Han, X.; Li, Q.; Zhao, M. CBMC: A Biomimetic Approach for Control of a 7-Degree of Freedom Robotic Arm. *Biomimetics* **2023**, *8*, 389. [CrossRef] [PubMed]
12. Athulya, P.; George, N. Others A Computer Vision Approach for the Inverse Kinematics of 2 DOF Manipulators Using Neural Network. In Proceedings of the 2020 IEEE Recent Advances In Intelligent Computational Systems (RAICS), Thiruvananthapuram, India, 3–5 December 2020; pp. 80–85.
13. Cavalcanti, S.; Santana, O. Self-learning in the inverse kinematics of robotic arm. In Proceedings of the 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), Curitiba, Brazil, 8–11 November 2017; pp. 1–5.

14. Wang, C.; Wang, C.; Li, W.; Wang, H. A brief survey on RGB-D semantic segmentation using deep learning. *Displays* **2021** *70*, 102080. [CrossRef]
15. Wu, Z.; Allibert, G.; Stolz, C.; Ma, C.; Demonceaux, C. Depth-adapted CNNs for RGB-D semantic segmentation. *arXiv* **2022**, arXiv:2206.03939.
16. Du, Y.; Muslikhin, M.; Hsieh, T.; Wang, M. Stereo vision-based object recognition and manipulation by regions with convolutional neural network. *Electronics* **2020**, *9*, 210. [CrossRef]
17. Halmetschlager-Funek, G.; Suchi, M.; Kampel, M.; Vincze, M. An empirical evaluation of ten depth cameras: Bias, precision, lateral noise, different lighting conditions and materials, and multiple sensor setups in indoor environments. *IEEE Robot. Autom. Mag.* **2018**, *26*, 67–77. [CrossRef]
18. Fischer, M.; Mylo, M.; Lorenz, L.; Böckenholt, L.; Beismann, H. Stereo Camera Setup for 360° Digital Image Correlation to Reveal Smart Structures of Hakea Fruits. *Biomimetics* **2024**, *9*, 191. [CrossRef] [PubMed]
19. Priorelli, M.; Pezzulo, G.; Stoianov, I. Active Vision in Binocular Depth Estimation: A Top-Down Perspective. *Biomimetics* **2023**, *8*, 445. [CrossRef] [PubMed]
20. Zhang, H.; Lee, S. Advancing the Robotic Vision Revolution: Development and Evaluation of a Bionic Binocular System for Enhanced Robotic Vision. *Biomimetics* **2024**, *9*, 371. [CrossRef] [PubMed]
21. Pérez, R.; Gutiérrez, S.; Zotovic, R. A study on robot arm machining: Advance and future challenges. *Ann. Daaam Proc.* **2018**, *29*, 0931.
22. Wang, T.; Tao, Y.; Liu, H. Current researches and future development trend of intelligent robot: A review. *Int. J. Autom. Comput.* **2018**, *15*, 525–546. [CrossRef]
23. Wu, J.; Zhang, D.; Liu, J.; Han, X. A moment approach to positioning accuracy reliability analysis for industrial robots. *IEEE Trans. Reliab.* **2019**, *69*, 699–714. [CrossRef]
24. Hsiao, J.; Shivam, K.; Lu, I.; Kam, T. Positioning accuracy improvement of industrial robots considering configuration and payload effects via a hybrid calibration approach. *IEEE Access* **2020**, *8*, 228992–229005. [CrossRef]
25. Jiang, Y.; Yu, L.; Jia, H.; Zhao, H.; Xia, H. Absolute positioning accuracy improvement in an industrial robot. *Sensors* **2020**,*20*, 4354. [CrossRef]
26. Lee, T.; Tremblay, J.; To, T.; Cheng, J.; Mosier, T.; Kroemer, O.; Fox, D.; Birchfield, S. Camera-to-robot pose estimation from a single image. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 9426–9432.
27. Abdelaal, M.; Farag, R.; Saad, M.; Bahgat, A.; Emara, H.; El-Dessouki, A. Uncalibrated stereo vision with deep learning for 6-DOF pose estimation for a robot arm system. *Robot. Auton. Syst.* **2021**, *145*, 103847. [CrossRef]
28. Chen, T.; Lin, J.; Wu, D.; Wu, H. Research of calibration method for industrial robot based on error model of position. *Appl. Sci.* **2021**, *11*, 1287. [CrossRef]
29. Sellami, S.; Klimchik, A. A deep learning based robot positioning error compensation. In Proceedings of the 2021 International Conference "Nonlinearity, Information and Robotics" (NIR), Innopolis, Russia, 26–29 August 2021; pp. 1–5.
30. Galan-Uribe, E.; Morales-Velazquez, L.; Osornio-Rios, R. FPGA-Based Methodology for Detecting Positional Accuracy Degradation in Industrial Robots. *Appl. Sci.* **2023**, *13*, 8493. [CrossRef]
31. Galan-Uribe, E.; Amezquita-Sanchez, J.; Morales-Velazquez, L. Supervised Machine-Learning Methodology for Industrial Robot Positional Health Using Artificial Neural Networks, Discrete Wavelet Transform, and Nonlinear Indicators. *Sensors* **2023**, *23*, 3213. [CrossRef] [PubMed]
32. Khanafer, M.; Shirmohammadi, S. Applied AI in instrumentation and measurement: The deep learning revolution. *IEEE Instrum. Meas. Mag.* **2020**, *23*, 10–17. [CrossRef]
33. Chellappa, R.; Theodoridis, S. *Academic Press Library in Signal Processing, Volume 6: Image and Video Processing and Analysis and Computer Vision*; Academic Press: Cambridge, MA, USA, 2017.
34. *ISO 9283:1998* ; International Organization for Standardization Manipulating Industrial Robots—Performance Criteria and Related Test Methods. ISO: Geneva, Switzerland, 1998.
35. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE.* **1998**, *86*, 2278-2324. [CrossRef]
36. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [CrossRef]
37. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

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.