AJBAR Vol 1(5), 2022: 1-11, ISSN: 2811-2881

Arid Zone Journal of Basic and Applied Research

Faculty of Science, Borno State University
Maiduguri, Nigeria

Journal homepage: https://www.azjournalbar.com

Research Article

Convolution Neural Network Based Human and Robot Collaboration for 3D Scene Perception

Abdulkadir Hamidu Alkali[1], Muhammed Zaharadeen Ahmed[1], Yusuf Ayuba[1] and Adati Elkanah Chahari[2]

[1]Department of Computer Engineering, University of Maiduguri, Borno State, Nigeria
[2]Federal Collage of Education, Yola, Adamawa State, Nigeria.
*Corresponding author: zaharadeenccna@gmail.com, doi.org/10.55639/607ztnhb

ARTICLE INFO:

Keyword:
Mobility Matrix,
Torque,
Axial force,
Effector

ABSTRACT

The approach of algorithmic operation of the Convolutional Neural Network (ConvNet/CNN) is like that of Deep Learning. The algorithm receives input instruction as an image by assigning characteristics of learnable weights and preferences on all inputs (aspects/objects) to the image. These features enable scalable input processing and effective differentiation and assignment of input and output images. For pre-processing requirements, ConvNet uses significantly low attributes as compared to other algorithms. However, in a primitive approach, filters are hand-engineered having sufficient algorithmic training. The ConvNets possess other abilities to learn image filtering and its characteristics. In this paper, an analysis of robot visual reasoning (for pick and place) is conducted. This refers to reasoning the latent meaning of visual signals or indications for future robot actions from visual observations of an HRC scene. In this paper, projection matrix estimation is computed using 2D points to represent the figure plane and 3D points to represent scene detection for initiating pick and place operation. During simulation, equations are represented using a matrix format based on figure coordinates. The design also implements automatic calibration and accuracy for the robot workspace. In addition, vision-based management for the robot end effector is also conducted based on horizontal targets, and upright targets. However, the pick and the place is articulated in the experiment without visual control. In our results, analysis of joint tangential forces is computed for the three joints, Alink1, Blink2, and Clink3. The outcome of axial force variation of the three joints is conducted in their state of mobility. Finally, the result of torque acting on the three joints increases when the simulation time is increased to achieve optimum performance.

Corresponding author: Muhammed Zaharadeen Ahmed, Email: zaharadeenccna@gmail.com
Department of Computer Engineering, University of Maiduguri, Borno State, Nigeria

INTRODUCTION

Convolution Neural Network (CNN) is a multi-layer perceptron (i.e., a neural network unit or an artificial neuron). CNN has a weight-sharing network structure that is related to the biological neural network Ansari et al (2022). CNN framework can lessen the complexity of the network model and the number of weights. CNN is applicable in several image classifications. During the image recognition process, only one object is recognized from the centre of the general image. This means the process identifies what the image is. During computer image processing, CNN imports a picture (image) using a box to precisely identify where the main object is in the image. When the main object is found in the box, the border from the image is then narrowed. The image import enables objects with similar characteristics to be processed using an adopted dimension. An image box is divided into small boxes as a trainer for the three different models Cheng et al. (2020). These are:

i. A CNN that produces image characteristics
ii. A classifier to compute image classes
iii. A regression model to tighten borders for precision.

A related application of CNN is the Human-robot collaboration (HRC). The HRC has gained attention in several areas of computer engineering, manufacturing, structures, medicine and others. This is due to the rapid and continuous increase in electronic devices, production efficiency and mass customization in industries, homes, schools, and hospitals. HRC team combines the strength and accuracy of robots with the flexibility and inventiveness of people, allowing human operators and robots to work together in a shared workspace on shared tasks Shao et al. (2022); Huang & Mok (2018).

Related Work

Convolution Layers

A convolutional layer is a set of parallel characteristics that are collected by sliding different convolution kernels. During each sliding position, a product operation results in a product and summation conducted between a convolution kernel and an input image. This process is carried out to project the information in the receptive field onto an element in the feature map. The convolution kernel's size is much smaller than the input image. Convolution layers consist of several characteristics. The first characteristic is local perception. The aim of using local perception is that researchers consider parts in the picture that are near to each other for relevancy Chen *et al*. (2022).

Robot Control Strategies

A major concern during the implementation of industrial or academic HRC is its classification of physical contact. Research by Mey et al. (2022) was conducted and concluded that the control approach cannot compensate for a poor mechanical design on its own. However, it remains a crucial characteristic in terms of performance, lessening sensitivity to indecision, and improvement of precision. The strategy of robot Control is classified into two Nagata et al. (2020) and Smys & Ranganathan (2019). These are:

1. Pre-collision: The Pre-collision control strategy is centred on preventing damaging contact between the robot and its environment. This process involved the use of collision avoidance. This uses sensory input such as a camera, laser sensors etc to adjust the velocity of the body based on the distance to independent quantity in the surroundings Ansari et al (2022).
2. Post-collision: The Post-collision strategy does not avoid possible contact between humans and robots.

Instead, it lessens the contact force and the energy exchange between the two entities. This strategy is known as ‘‘interaction control strategies’’, where the two famous sub-categories are termed direct and indirect control strategies. The direct control strategy is known as the hybrid control strategy, whereas the indirect includes Admittance and Impedance control schemes. Direct-force control approaches control the manipulator’s force along the constraint as well as the motion along with the directions of the unconstrained path by measuring the force Smys & Ranganathan (2019).

METHODOLOGY

The robot has a functional camera device to visualize at 360 degrees. Calibration involves putting together both 3D and 2D points whereas calibrated Objects are mostly planes or cubes in two or three orthogonal forms. Chess sheet configurations are mostly included to develop the surfaces in an equal distance and all through distributed precisely on the object exterior. The edges explain the well-known points within the robot coordinate system. An edge detection algorithm is used to identify the captured images. This ensures all the corners and edges are used in the process of projection matrix estimation. An alternative way to ensure this process is the self-calibration process. This approach computes the similarities from the camera adjustment when stationary.

Projection Matrix Estimation

The image plane surfaces represent the 2D points while the scene of the situation of detection, pick and place represents the 3D points which are ascribed and can be deduced by the projection matrix, P as used in benchmark [1]:

$$\begin{pmatrix} xi \\ yi \\ 1 \end{pmatrix} = P \begin{pmatrix} xw \\ yw \\ zw \\ 1 \end{pmatrix} \quad (1)$$

The Elements in matrix P are set to be computed in terms of figure coordinates xi and yi. The associated 3D domain coordinates Xw, Yw, and Zw as follows: A vector p = (p11, … p33) Tis defined, and equation (1) can be altered as Ap = 0. equation Ap=0, is defined by Matrix P having rank 11, requires at least N = 6 correspondences. 60x12 of matrix A is generated when N = 30 computed equalities from the calibration site are used. Computation errors can easily be lessened using more than the required minimum of 6 calibration points. separately from the trivial situation x = 0, no distinct solution of Ax = 0 with the inclusion of signal noise. However, it is best to search for an exact answer, but one should lessen the computation cost function, such as the norm ||Ax||. The norm ||UDVT x|| is limited using an outstanding value decomposition, which factors matrix A into A = UDVT. The Single Value Decomposition (SVD) algorithm used in MATLAB simulation is used to invert Matrix A.

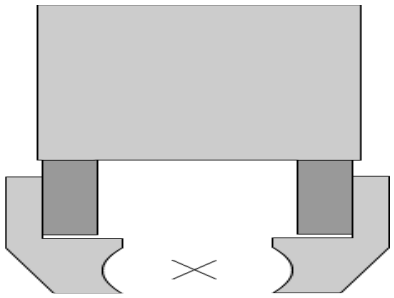

Figure 1: Front View of Gripper - Cross Hairs Cheng *et al.* (2020)

Using the proposed approach, the camera adjustment and calibration can be structured based on the following:

- Matrix approximation from 2D point intersection
- Matrix Interpretation of the camera from the initial approximated matrix
- Re-establishment of the Euclidean space using the identified control points from camera calibration (minimum of four non- coplanar control points).

Similarly, loss function is included the proposed work to compute the degree of variation between the foreseen value and a real value. The value is a non-negative real value function. It is expressed as $L(Y, f(x)$. Therefore, the standard form of the loss function can be expressed (as in equation 2) below.

$$L(Y, f(x) = |\, Y - f(x)| \qquad (2)$$

The loss function is crucial in terms of model optimization and enhancing precision. The proposed robot system is now more robust when the loss function is lesser. Also, the loss function remains a major part of empirical and structural risk function.

Automatic Calibration and Accuracy for Robot Workspace

As used in the benchmark by Cheng et al. (2020), an experimental scenario in this study includes a Fanuc M-16iB industrial robot, and two pairs of coloured Charged Coupled Device sensors as cameras are used between two ends of the workspace. Different capture of the robot used in the benchmark Cheng et al. (2020) is presented (as in figure 1) holding a calibration slant at its geometrical point (crosshairs). The precision-made calibration tip fitted on the robot end effector can also be identified in the same figure. The conical pointer of the calibration tip is painted red to make the colour unique to the calibration tip and to make it easier to identify the tip in all calibration photographs. When coupled to the robot gripper, the tip was precision built to create an exact and repeatable reference. Through the known Denavit-Hardenberg geometry of the robot arm, the spatial tip coordinates can be traced back to the fixed robot base for any robot position. The Euclidean distance between a known 3D point P0 and its triangulated spatial representation Pt derived from the calibrated scene can be stated as follows:

$$|P^{(0+i)}| = \left(\sqrt{(P^{(0+i)})x\text{^}2 + (P^{(0+i)})y\text{^}2 + (P^{(0+i)})z\text{^}2}\right) \qquad (3)$$

The precision is less towards centre of the scene of the two cameras. But the target is about equidistant from both. The precision of one observation increases as the calibration pointer moves away from center. Whereas the precision of the other decreases resulting to less precision.

Vision Based management for Robot End Effector

The process of vision control for pick operation consists of both horizontal and upright targets.

i. *Horizontal Targets:* The orientation of the object for the Pick operation must be established to grasp horizontal cylindrical targets. The angle between the two extracted surface points and the world coordinate x-axis may be readily calculated to determine the orientation.

ii. *Upright Targets:* During the target extraction phase of the robot, there is no need to distinguish between upright and horizontal items for the pick and place operation. The orientations of the objects have little influence on how the cylinder appears in the image.

Pick and Place Without visual control

A real pick-and-place experiment has been conducted using an existing testbed. The setup has been adopted from Shao et al. (2022) (as in figure 1) below.

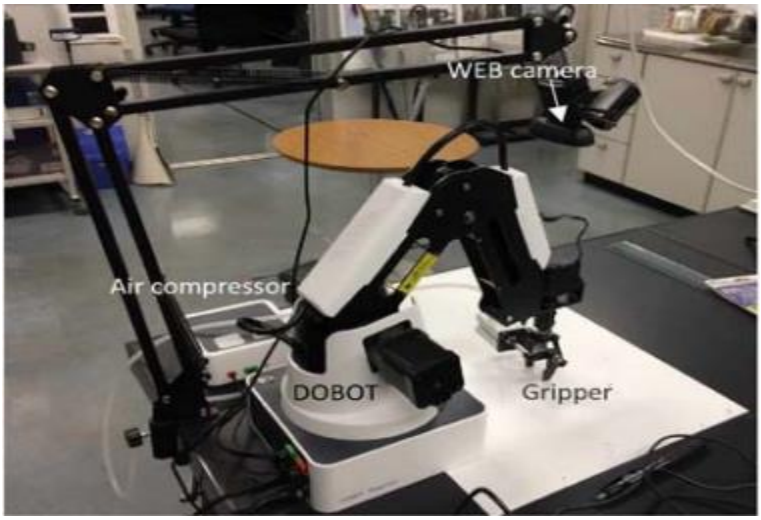

Figure 2: An Articulated Experimental Robot

Where *j* is the robot position [x y and z]^T and the angle 0 of the gripper for the robot coordinate system is managed by a function as (x_a, y_a, $z_{a=,}$ $0_{a)}$. For the robot timer interrupt, a resolution of 1600x1200 is captured. This is converted to binary in form of black and white. A connected device that has the highest area is identified as the target object. The positions are represented as [I_x $I_{y+}]^T$ for (1 less than or equal to I_{x+} -less than or equal to 1600. Therefore, the image coordinate system is then subtracted using the following equation below as follows.

$$I(x) = [\sum_{x=1}^{1600} xf(x,y)\ (\sum_{y=1}^{1200})]/O \quad (4)$$

$$I(y) = [\sum_{x=1}^{1600} xf(x,y)\ (\sum_{y=1}^{1200})]/O \quad (5)$$

Whereas [X_1 $Y_1]^T$ and [X_2 $Y_2]^T$ of the robot coordinate system represent the different positions of left top and right down of the capture.

Pick and Place with Visual Control

Using this approach, camera configuration for visual management is not needed. This is not the case without visual control. Based on Nagata et al. (2020), a lightweight endoscope camera has been placed near the gripper. Variable used during computation are varied as v(k) = [v_x (k) v_y $(k)]_T$. this is assisted for visualization of the robot

and is generated using the following formula below.

$$v(k) = \sum_{n=1}^{k} e(n)(k\, e\ (k) + ki) \qquad (6)$$

Where k represents the discrete time. K_p and K_i represent the gains of the proportional and integral operation. Finally, $e(k) = [e_x\ (k)\ e_y\ (k)]^T$ represent error section of the image coordinate system. This is then computed by the formula below as.

$$E\big(k) = Xd - I(k)\big) \qquad (7)$$

Whereas $X_d = [X_r/2\ Y_{r+/2]}^T$ and $I(k) = [I_x\ (k)\ I_y\ (k)]^T$ represent the anticipated position and the computed object position from the image coordinate system. For the resolution of the system, [X_r Y_y] represent the captured image.

Designing the End Effector

The end effector is made up of three rotating components. Nearby linkages' axes are usually perpendicular to one another. These are referred to as ALink 1, BLink 2, and CLink 3, based on their arrangement and working mechanism. The rotational direction of the axis is generally considered off from the end effector since Alink 1, Blink 2, and Clink 3 are rotated 180°, 90°, and 90°, respectively during the mobile situation. In the simulation, the payload is a cubic shape with a side length of 250 mm and a mass of 100 g. All three links may spin 360 degrees. However, the ranges of the linkages are limited, ALink 1 ranges between -1800C to 1800C, ALink 2 ranges between -900C to 900C ALink 3 ranges between -900C to 900C.

The workspace of the end effector might have two options depending on the initial location of Alink 3. In one situation, the workspaces of endpoints 1 and 2 are the same. These data points have two separate workspaces in the second example. To have the two options specified, the line connecting two endpoints must be perpendicular and parallel to the link's axis, respectively. The second circumstance is considered during the system simulation. When the workspace of the system is analysed, the end effector's axis of link 2 is parallel to the line connecting the two endpoints. For better mobility, the three links can spin one after the other or at the same time. To avoid unwanted vibrations. ALink 1, Alink 2, and Alink 3 spin one after the other to avoid affecting precision and smooth motion. This represents the main function of this end-effector.

RESULTS

This section presents the result of the dynamic analysis of the proposed research. The [X1 Y1]T and [X2 Y2]T in the robot device represent the location points of the left upper and right bottom as presented in the figure below. The distance between the two points represents the pixels in form of (0,0) and (1500 1100). Some sections of the connected areas of the robot joints are then cropped where the COG can be reflected.

After image cropping, the size of the image is changed to 237x237 thereby generating an input layer for the CNN transfer learning. This makes the image orientation to be computed using CNN and exploited for a suitable yaw angle. The robot device can precisely grasp the object or an image using a long-axis shape. In addition, the technique of a captured image extraction using our proposed convolution neural network (CNN) algorithm to achieve better orientation of a workpiece is presented (as in the figure 3) below.

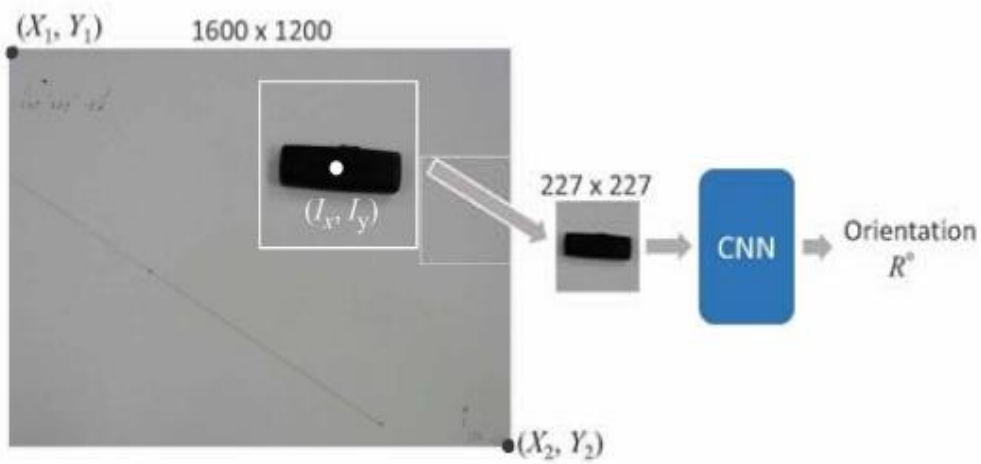

Figure 3: Extracting a Captured Image Using CNN Algorithm

The magnitudes and directions of the forces and torques imparted to each component are determined using formulas and re-presentation using graphs. In the proposed research, the robot system is initiated based on predetermined inputs analysed using formulas in equations (1-7). The forces and torques are being operated with time precision, especially during a pick-and-place operation. All form of movement scenarios of the robot is instantiated at the initial design stage with precise positions of the end effector. Similarly, the axial, Tangential forces and torques of the joints are represented as ALink 1, ALink 2, and ALink 3. These joints are enabled to rotate at angles 180°, 90°, and 45° respectively and the results can be depicted (as in figure 5) below.

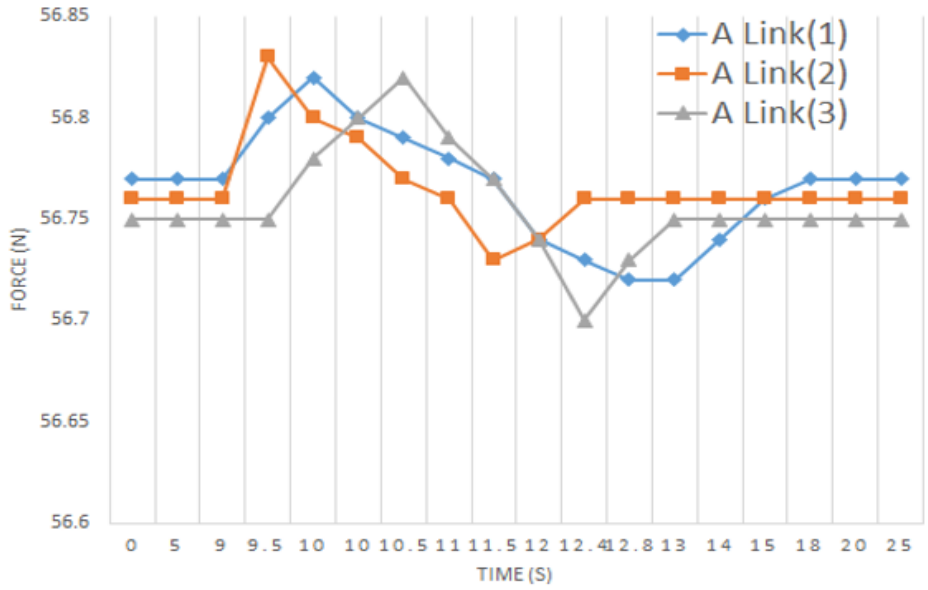

Figure 4: Joint Tangential forces

The simulation's payload is a cubic shape with a side length of 250mm and a mass of 100g. Angular acceleration and deceleration of the joints are maintained constant during the simulation. After instantiation, different scenarios of robot movement (joints) present optimum angular velocity between the three links. This means the robot joints begins movement based on its implemented algorithms and at a constant angular velocity.

The input angular displacements of the joints determine the shape of the delivered motion to the joints. The following are the inputs given to the system during the simulation. Also, the component can execute an axial force - Figure 5, tangential force - Figure 4, and torque - Figure 6 on each joint. The magnitudes and orientations of the forces and torques based on the joins with respect to time can be represented in the graph (as in figure 5) below.

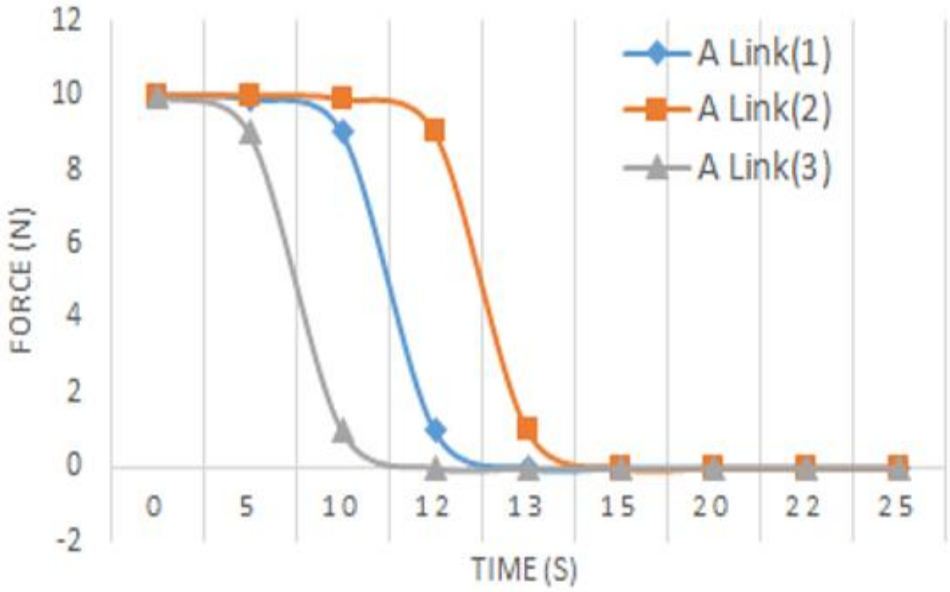

Figure 5: Axial force variation of three joints during mobility

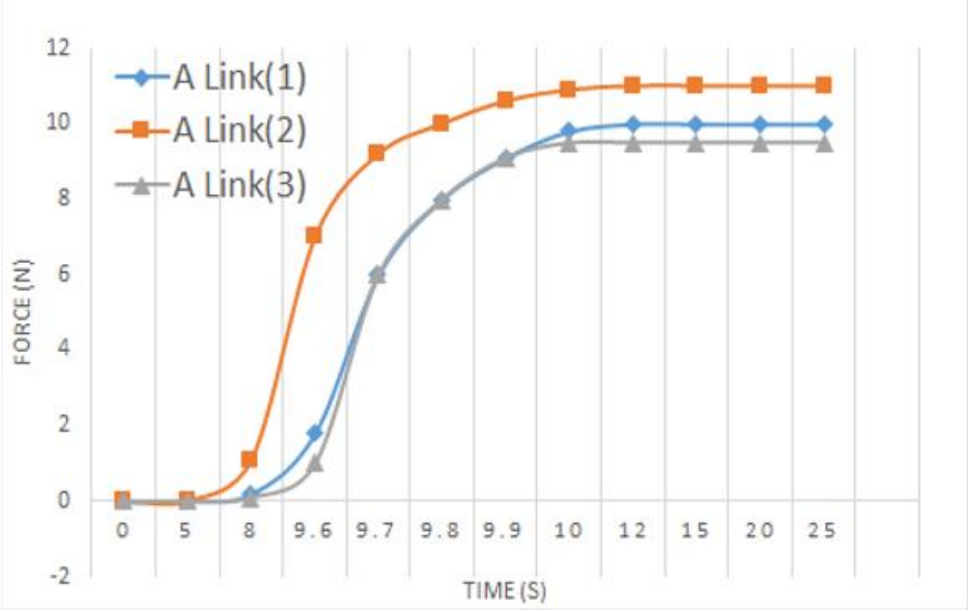

Figure 6: Joint Tangential forces

Based on mathematical computation, the position of the topmost i.e ALink3 is computed for an angular displacement for the three axes. By considering the coordinate on the plane [x_a, y_a, z_a] which is the base of the effector, and the robot is anchored and is achieved using isolation in the equation. It is important to note that all the three links are rotational. The computation is cantered based on the strategy as follows.

For ALink 1 0 less than equal to θ_1 less than equal to 2 π, where θ_1 represent the rotational joint and as rotation angle. For ALink 2, - $\pi/2$ less than equal to θ_2 less than equal to $\pi/2$; where θ_2 represent the rotational joint and as rotation angle. For ALink 3, - $\pi/2$ less than equal to θ_3 less than equal to $\pi/2$; where θ_3 represent the rotational joint and as rotation angle (when $\theta_2 = 90^0$ or 270^0. Figure 7 below analyses the effector graphically further precisely for the three axes, where ALink 1 cannot rotate towards angle 90^0 or 270^0.

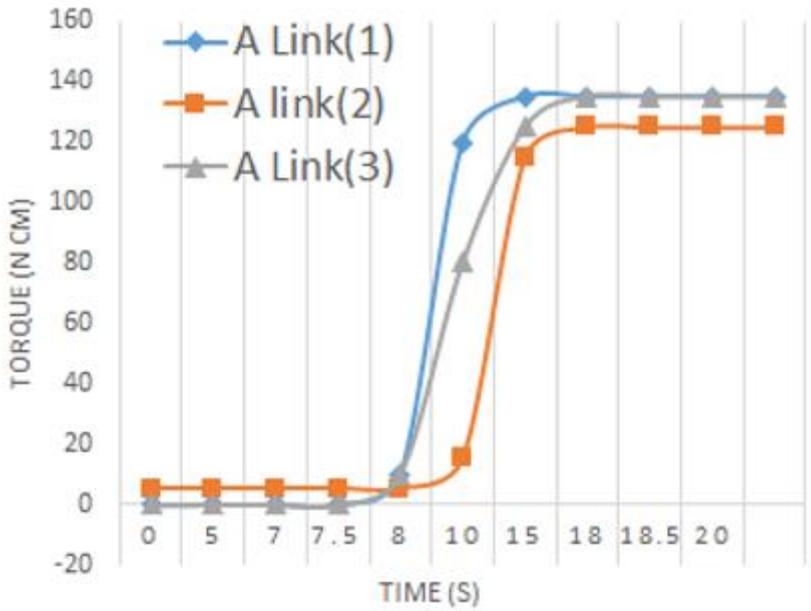

Figure 7: Torques acting on the joints during operation

CONCLUSION

After exploiting the CNN algorithm using spectrograms as input, the proposed framework has been efficient to achieve effective and precise robot operation guidance using orientation data as an input interface. Classifier Precision is generated using muscle fatigue on a long-term basis. While for the long term, a 98% precision accuracy is recorded. A Typical robotic machine for the pick and place process must comprise sensors for touch and vision. This must be attached to the robot at the appropriate location for precise function. The paper conducted an experimental analysis that ensures functional operation and efficient strategies for reliable and flexible management of robots using automated stereo vision and touch systems. Projection Matrix Estimation and Automatic Calibration and Accuracy for Robot Workspace were conducted using equations for precision. This led to the creation of Vision-Based Control of The Robot End Effector, Pick and Place Operations for both horizontal and upright targets.

In the Future, the research focus is to centre on Convolution Neural Network (CNN) Parameter optimization and enlarging the CNN architecture. Based on existing algorithms the process of fast training and orientation presents a few challenges that require immense improvement for CNN and robot workpieces. This requires the addition of appropriate devices and power ember hardware. The upper and lower limb of the robot device also requires proper examination for efficiency.

REFERENCES

Ansari, A. A., Akbar, S., & Sankaranarayanan, S. (2022). Precise Pick & Place System of 7-DoF KUKA Robot, Remote Control, Plotting & Computation of Trajectories, Cartesian Positions and Jacobian Matrices.

Cheng, C. Y., Renn, J. C., Chang, S. J., & Saputra, I. (2020) Development of a Simple Servo-Pneumatic 3-DOF Pick-and-Place Manipulator.

Shao, L., Nagata, F., Habib, M. K., & Watanabe, K. (2022). Visual Feedback Control Through Real-Time Movie Frames for Quadcopter with Object Count Function and Pick-and-Place Robot With Orientation Estimator. In *Handbook of Research on New Investigations in Artificial Life, AI, and Machine Learning* (pp. 99-116). IGI Global.

Chen, H., Kiyokawa, T., Wan, W., & Harada, K. (2022). Category-Association Based Similarity Matching for Novel Object Pick-and-Place Task. *IEEE Robotics and Automation Letters*.

Mey, J., Ebert, S., Lin, T., Nguyen, G. T., Gumhold, S., & Aßmann, U. (2022, January). Teaching Distributed and Heterogeneous Robotic Cells. In *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)* (pp. 1-2). IEEE.

Nagata, F., Miki, K., Otsuka, A., Yoshida, K., Watanabe, K., & Habib, M. K. (2020, October). Pick and place robot using visual feedback control and transfer learning-based CNN. In *2020 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 850-855). IEEE.

Ansari, A. A., Akbar, S., & Sankaranarayanan, S. (2022). Precise Pick & Place System of 7-DoF KUKA Robot, Remote Control, Plotting & Computation of Trajectories, Cartesian Positions and Jacobian Matrices.

Smys, S., & Ranganathan, G. (2019). Robot assisted sensing control and manufacture in automobile industry. *Journal of ISMAC*, *1*(03), 180-187.

Kazemi, S., & Kharrati, H. (2017). Visual processing and classification of items on moving conveyor with pick and place robot using PLC. *Intelligent industrial systems*, *3*(1), 15-21.

Huang, P. C., & Mok, A. K. (2018, August). A case study of cyber-physical system design: Autonomous pick-and-place robot. In *2018 ieee 24th international conference on embedded and real-time computing systems and applications (rtcsa)* (pp. 22-31). IEEE.

Sumit Saha (2018, December). A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53