

Marco Cabrera

Mejora del rendimiento

2025



**Universidad Autónoma de
Querétaro**

Facultad de Ingeniería

**Mejora del rendimiento de un brazo
robótico mediante la medición de
posiciones articulares con un
algoritmo inteligente y cámaras
estéreo**

Tesis

Que como parte de los requisitos para obtener el
Grado de

Doctor en Ingeniería

Presenta

Marco Antonio Cabrera Rufino

Dirigido por:
Dr. Juan Manuel Ramos Arreguín

Co-Director:
Juvenal Rodríguez Resendiz

Querétaro, Qro. a 29 de mayo de 2025

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciatario no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:

 **Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciatario.

 **NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).

 **SinDerivadas** — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



**Universidad Autónoma de
Querétaro**
Facultad de Ingeniería
Doctorado en Ingeniería

**Mejora del rendimiento de un brazo robótico mediante la
medición de posiciones articulares con un algoritmo
inteligente y cámaras estéreo**

Tesis

Que como parte de los requisitos para obtener el Grado de
Doctor en Ingeniería

Presenta

Marco Antonio Cabrera Rufino

Dirigido por

Dr. Juan Manuel Ramos Arreguín

Dr Juan Manuel Ramos Arreguín
Presidente

Dr. Juvenal Rodríguez Resendiz
Secretario

Dr. Gorrostieta Hurtado Efren
Vocal

Marco Antonio Aceves Fernández
Suplente

Saúl Tovar Arriaga
Suplente

Centro Universitario, Querétaro, Qro. México
Mayo 2025

Dedicatorias:

Este trabajo está dedicado con profundo amor y gratitud a mis padres, quienes han sido el pilar de mi educación desde mi niñez. A mi madre, por inculcarme el hábito del estudio y convertirlo en una pasión que guía mi vida. A mi padre, por ser una fuente constante de inspiración y motivación para seguir aprendiendo cada día. A mis hermanos, que han estado a mi lado desde el comienzo; son mis compañeros de vida y mis animadores incondicionales en este largo viaje.

Y por último, pero no menos importante, a mi esposa, cuyo amor y compañía han colmado mi vida de bendiciones desde el día en que unimos nuestros caminos. Su apoyo inquebrantable y su fe en mí han sido una fuerza invaluable que me impulsa a seguir adelante.

Agradecimientos

Quiero expresar mi más sincero agradecimiento al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) por el invaluable apoyo brindado durante mi formación doctoral. La beca otorgada por esta institución fue fundamental para llevar a cabo mis estudios de posgrado en el programa de Doctorado en Ingeniería, permitiéndome dedicarme de tiempo completo a la investigación, el aprendizaje y el desarrollo académico durante tres años. Su respaldo ha sido clave para alcanzar este importante logro profesional y personal.

Asimismo, extiendo un profundo agradecimiento al Dr. Juan Manuel Ramos Arregúin, mi director de tesis, por su constante apoyo, guía y confianza a lo largo de toda mi formación de posgrado, desde la maestría hasta el doctorado. Su compromiso y dedicación fueron pilares fundamentales para el desarrollo de este trabajo y para mi crecimiento académico y profesional.

Abreviaturas y siglas

Sigla	Significado
DoF	grado de libertad
fps	frames por segundo
2D	dos dimensiones
CNN	Red neuronal convolucional
GUI	interfaz gráfica de usuario
TCP	Punto central de la herramienta

Resumen

La precisión en el posicionamiento de robots manipuladores es un factor clave en la industria, ya que influye en la calidad y eficiencia de los procesos automatizados. Esta tesis presenta un enfoque basado en visión estéreo y redes neuronales convolucionales (CNN) para mejorar la medición de posiciones articulares en un brazo robótico. Se empleó la cámara Intel RealSense D435, capaz de proporcionar datos de profundidad con alta resolución, y un algoritmo de aprendizaje profundo que analiza las posiciones angulares de los eslabones del robot. La metodología incluyó la implementación de un sistema redundante con encoders y la validación de los resultados mediante métricas de precisión en posicionamiento. Los experimentos demostraron que el uso combinado de visión estéreo y CNN logra una reducción significativa de los errores geométricos, alcanzando una precisión superior a 0.1 mm, pero competente, cumpliendo así con los estándares industriales. Este sistema tiene aplicaciones potenciales en manufactura, robótica colaborativa y automatización industrial, donde la precisión y estabilidad en el tiempo son requisitos fundamentales.

Palabras clave: Brazo robótico, visión robótica, degradación de precisión, visión estereoscópica, cámara de profundidad.

Abstract

Precision in the positioning of robotic arms is a key factor in industry as it influences the quality and efficiency of automated processes. This paper presents an approach based on stereovision and convolutional neural networks (CNN) to improve the measurement of joint positions in a robotic arm. The Intel RealSense D435 camera, which can provide high-resolution depth data, and a deep learning algorithm that analyzes the angular positions of the robot joints were used. The methodology involved implementing a redundant system with encoders and validating the results using position accuracy metrics. The experiments showed that the combined use of stereo image processing and CNN achieved a significant reduction in geometric errors with an accuracy of better than 0.1 mm. This system has potential applications in manufacturing, collaborative robotics and industrial automation, where accuracy and stability over time are important requirements.

Keywords: Robotic arm, robotic vision, precision degradation, stereo vision, depth camera.

Índice

Índice de tablas

Índice de figuras

I Introducción	2
a Planteamiento del problema	3
b Justificación	4
II Antecedentes	5
a Estado del arte	5
III Hipótesis	7
IV Objetivos	8
a Objetivos específicos	8
V Material y Métodos o Metodología	9
a Visión estereoscópica	9
b Precisión de la cámara estéreo D435	10
c Materiales	12
d Sistema de medición propuesto	14
e Interfaz Gráfica de Usuario (GUI)	16
f Medición de la precisión de la posición	17
g Redes neuronales Convolucionales	19
VI Resultados y discusión	22
a Discusión	37
VII Conclusiones	39
Referencias bibliográficas	40

Índice de tablas

1 Resumen del estado del arte.	5
2 Distancia entre píxeles para medir la precisión de la cámara. La distancia entre el píxel 1 y los siguientes píxeles consecutivos.	11
3 Distancia entre dos pixeles consecutivos.	11

Índice de figuras

1 Brazo robótico de 3 DoF.	2
------------------------------------	---

2	Análisis del esquema de triangulación para visión estéreo, en una perspectiva bidimensional [23].	9
3	Configuración estéreo en tres dimensiones protect[23].	10
4	Ventanas coincidentes en visión estereoscópica. Esto representa la funcionalidad digital de una cámara estéreo comercial [44].	10
5	Distancia de un centímetro en pixeles, y el píxel posterior para medir la resolución de la cámara D435.	11
6	Materiales y dimensiones del robot, DoF (Degree of Freedom).	12
7	Dimensiones del espacio de trabajo.	13
8	Morfología de un robot de seis DoF, solo cuatro DoF fueron etiquetados con diferente color.	14
9	Diagrama a bloques del sistema robótico de este proyecto.	15
10	(a) Diagrama a bloques del sistema de medición (b) análisis geométrico de la degradación de precisión en un DoF.	15
11	Metodología para el sistema de medición.	16
12	GUI utilizada para una mejor manipulación del brazo robótico, para la comunicación; se implementa el protocolo RPC (Remote Procedure Call), la comunicación RPC se basa en la arquitectura TCP.	16
13	Representación gráfica de la precisión de posición de 1 DoF según ISO 92831998.	17
14	Seis errores de un i-ésimo eje rotacional.	17
15	Espació de trabajo del Cobot 280 Nano, 28 cm de rango.	19
16	Arquitectura del modelo LENET-5 [46].	20
17	Arquitectura: AlexNet [47].	20
18	Cada posición tridimensional de la posición TCP (x, y, z), el cuarto gráfico son los puntos que sigue el Cobot.	22
19	Diseño de la CNN unidimensional utilizada para predecir la cinemática inversa, destacando la capa de entrada, las capas convolucionales y la salida.	24
20	Función de pérdida de entrenamiento frente a validación de la CNN unidimensional.	25
21	Nueve muestras de imágenes RGB del Cobot, los pixeles son 480×640 pixeles.	26
22	Nueve muestras de información de profundidad para el Cobot, tiene el mismo tamaño de una imagen en color, cada píxel tiene una resolución de 16 bits.	27
23	Entrenamiento vs. validation loss function de LENET-5.	28
24	Arquitectura de LENET-5 [46].	29
25	Entrenamiento vs. validation loss function de AlexNet.	30
26	Arquitectura de AlexNet [47].	31
27	Imágenes del robot con etiquetas de colores, utilizando YOLO.	32

28	Resultados del entrenamiento de Yolo-V1 para la detección de etiquetas. (a)Mean average precision (mAP) se utiliza para medir el rendimiento de los modelos de visión por ordenador. (b)Box Loss se refiere a lo bien que el modelo predice las posiciones y tamaños de las cajas que rodean los objetos en una imagen. (c)Class Loss se refiere a la medida de la precisión con la que el modelo predice la clase o etiqueta correcta de los objetos que detecta. (d) Object Loss mide lo bien que el modelo reconoce la presencia o ausencia de un objeto en una región concreta de una imagen.	33
29	Resultados con un nuevo conjunto de imágenes para evaluar el algoritmo Yolo. Los puntos rojos son las etiquetas reconocidas.	34
30	Información de profundidad de las imágenes en color.	36
31	Información de profundidad de las imágenes en color con la ubicación de cada etiqueta. Los puntos del recuadro de color son los puntos marcados detectados por el algoritmo YOLO que ayuda a medir cada posición angular DoF.	37

I. Introducción

Un robot manipulador es un brazo mecánico articulado formado de eslabones conectados a través de articulaciones que permiten un movimiento relativo entre dos eslabones consecutivos [1]. El número de grados de libertad (DoF) de un brazo robótico se define como el número de coordenadas independientes o mínimas necesarias para describir su posición [2]. Los robots se emplean para mover y manipular herramientas llamadas efectores finales, estos pueden ser pinzas, varillas de soldadura, taladros, chupones, cámaras u otras, según sus aplicaciones [3].

En la figura 1 se muestran las posiciones angulares de cada articulación θ_1 , θ_2 y θ_3 de un robot de 3 DoF. Además, se muestra el extremo donde se conecta el efecto final del robot con posición cartesiana (x, y, z) [1, 4].

La medición de una articulación se realiza en unidades angulares (grados o radianes), y se utilizan sensores como encoders, tacómetros, giroscopios, u otros [1]. Además, se puede utilizar dos o más sensores, a estos se les conoce como redundantes. La redundancia es el uso de dos o más instrumentos de medición en paralelo, de forma que cualquiera de ellos o en conjunto puedan proporcionar la medición requerida. También puede aplicarse en sistemas en los que determinados componentes de los mismos degradan gravemente la precisión total [5].

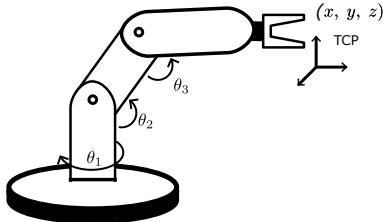


Figura 1: Brazo robótico de 3 DoF.

La precisión del robot se denomina su capacidad para alcanzar un punto determinado en el espacio de trabajo [6]. La precisión en el posicionamiento del efecto final puede utilizarse para evaluar el rendimiento de un robot industrial [3].

A la perdida de precisión en el posicionamiento del efecto final se le conoce como degradación de precisión [7, 8, 9], y a esta se pueden agrupar en tres grupos, los que se generan por daños o deformaciones estructurales en los eslabones o errores geométricos. Además, los no geométricos como la fricción y la holgura de los engranajes. Y por último las que se generan por errores computacionales [10, 11].

Dentro de los errores geométricos está la tolerancia, que se refiere a la relación entre la medición real de una pieza ya fabricada, y la medida teórica que debería tener. El juego, que es la diferencia entre las medidas del agujero de una pieza y la del eje donde esta se conecta. La elasticidad y excentricidad, que tienen relación entre las fuerzas aplicadas al robot y las correspondientes deformaciones de los eslabones, el desgaste que se define como el daño que sufre las superficies sólidas debido al deslizamiento y/o impacto contra otra superficie sólida [12].

La visión estéreo es una técnica para medir profundidad en una imagen, que permite reconstruir coordenadas de puntos en el espacio tridimensional a partir de imágenes adquiridas con dos cámaras [13]. Existen otras técnicas de profundidad, como la fotogrametría, la luz estructurada, tiempo de vuelo (ToF) y la triangulación láser, entre otras.

La inteligencia artificial es el desarrollo de algoritmos que exhiben un comportamiento inteligente o humano [14]. Esto es con métodos asociados a modelos estadísticos o de inspiración biológica, como las redes neuronales (incluido lo que ahora se conoce como aprendizaje profundo) y la computación evolutiva [15, 16].

Existen trabajos previos que tienen interés en el análisis de degradación de precisión. En contraste, los dispositivos de visión por computadora y las técnicas de inteligencia artificial son metodologías implementadas en aplicaciones robóticas [17], como el control [18, 19, 20, 21], y segmentación de objetos [22, 23]. Las cámaras estéreo se han vuelto más precisas en los últimos diez años [24, 25, 26, 27], y los microordenadores actuales como Raspberry Pi modelo 4 o Jetson Nano 280 son lo suficientemente potentes como para realizar tareas de medición en tiempo real. Sin embargo, hay pocos antecedentes sobre el uso de estos métodos como sistema de medición alternativo [28], y menos aún sobre el uso específico de cámaras estereoscópicas y redes neuronales convolucionales (CNN), que son un tipo de modelo de aprendizaje profundo (DL) especialmente eficaz para el análisis de imágenes [29].

a. Planteamiento del problema

En la mayoría de los casos la precisión se degrada a más de 0.1 mm, ocasionada en el 90 % por errores geométricos. Esto, cuando la tarea en que trabaja un robot manipulador es extenuante, y por lo tanto se presentan fallas que reducen la vida útil del robot a menos de 3 años [7, 8, 9]. De las metodologías usadas para reducir la degradación por errores geométricos, son de análisis geométrico complejo, diseño simple de redes neuronales, son costosas y sobre todo, la mayoría no cumple con el requisito industrial ($< 0,1 \text{ mm}$). Además, en trabajos previos donde se utiliza cámaras de profundidad, que tienen baja resolución 628×468 pixeles, con ruido lateral de 2-3 pixeles [30].

La precisión de un robot manipulador debe mantenerse menor a 0.1 mm, lo anterior reduce los errores geométricos, incrementando la vida útil del robot por más de 3 años. Además, las metodologías no deben ser costosas, complejas, y en caso de usar cámaras de profundidad deben tener menor ruido lateral (menor a 2 pixeles), y sobre todo cumplir con el requisito industrial.

Se estima que el uso de encoders en conjunto con un sistema de ajuste en la medición de posiciones articulares, con cámaras estéreo y un algoritmo con redes convolucionales, mejorará la precisión de posicionamiento del brazo robótico, al reducir los errores geométricos.

b. Justificación

En México las ventas de robots son de más de 5,000 unidades al año aproximadamente en la industria automotriz, aeroespacial u otras; llegando a los 513 robots por cada 10,000 trabajadores en el 2018. Por ejemplo, en San Luis Potosí, en la BMW se utilizan 500 robots manipuladores de la marca Kuka con una precisión de fábrica de 0.02 a 0.1 mm. Y son utilizados para evitar tareas peligrosas, donde el acceso es limitado, o en aquellas que requieran de mayor y precisión [31, 32].

En la industria automotriz el requisito de precisión para operaciones como la soldadura por puntos será de alrededor de un milímetro. La industria aeroespacial proporciona un entorno más desafiante para aplicar la robótica, ya que sus requisitos de precisión son al menos un factor de diez a veinte veces mayor [10]. Por ejemplo, en la soldadura por arco, si la precisión se degrada, un robot soldará o perforará en posiciones incorrectas y la calidad de la producción se verá comprometida [3].

Las cámaras estéreo son herramientas precisas, de bajo coste (\approx 300 dólares) y de fácil implementación [33]. También, se puede extraer características geométricas. Además, en combinación con técnicas de inteligencia artificial como las redes neuronales convolucionales y/o recurrentes, se puede agrupar elementos del robot como los eslabones, y reconocer marcadores que faciliten su detección [34, 35]. El sistema debe ser redundante a la medición de los encoders debido a las limitaciones computacionales de las cámaras, como sus cuadros por segundo (60 fps), o su tiempo de procesamiento del algoritmo (> 10 ms).

Este trabajo beneficiaría principalmente al sector industrial, en donde los robots manipuladores son elementos importantes. Por ello, se investiga para mejorar el estado actual de funcionamiento de estos, así como para mejorar los métodos de medición de sus variables.

II. Antecedentes

a. Estado del arte

En la tabla 1, se muestra los trabajos que se enfocan en el reducimiento de la degradación de precisión, así como los errores medios antes y después de implementar sus algoritmos.

Tabla 1: Resumen del estado del arte.

Referencia/año	Técnicas convencionales	Técnicas de IA	Visión Robótica
[36]/2019	✓		
[37]/2020	✓	✓	
[38]/2020	✓	✓	
[39]/2020		✓	✓
[40]/2021		✓	✓
[41]/2021	✓		
[11]/2021	✓	✓	
[42]/2023	✓	✓	
[43]/2023	✓	✓	

Consideramos que, dado el estado actual de la técnica y sus resultados (véase la Tabla 1), se necesita un sistema mejor para compensar la precisión. La industria gasta mucho dinero en el mantenimiento y la compra de nuevos sistemas robóticos. Algunos autores han estudiado la medición de la degradación de la precisión. Es importante decir que, a pesar de los avances en visión robótica, sigue existiendo la necesidad de métodos más precisos y aplicables en la industria para medir las posiciones angulares en brazos robóticos. Se está abordando la brecha existente entre la necesidad existente y los métodos específicos que aún no se han implementado en profundidad.

Los métodos mencionados en la Tabla 1 no son lo suficientemente pequeños como para cumplir los requisitos de la industria (menos de 0,1 milímetros (mm)), o son tan complejos que no pueden aplicarse a otras morfologías de robots. En este artículo, presentamos un método robusto de medición de ángulos que utiliza una cámara estéreo, Intel RealSense modelo D435, y un algoritmo CNN para capturar la información de color y profundidad del robot. El rojo, verde y azul (RGB), es decir, un espacio de color utilizado habitualmente para colorear un píxel en imágenes y videos, es útil para la segmentación y se asigna al canal de profundidad con una precisión de 0.001 mm. El algoritmo divide la morfología del robot 3D en espacios 2D separados, con un espacio

2D que representa cada DoF. La combinación de estos métodos mejora la posición del robot con una precisión considerable.

III. Hipótesis

La integración de cámaras estéreo, con un algoritmo basado en redes neuronales convolucionales, implementados como sistema redundante de medición para las posiciones de las articulaciones de un brazo robótico, mejorará el rendimiento de posicionamiento, al reducir los errores geométricos que generan degradación de precisión, con un error medio menor a 0.1 mm por más de 3 años.

IV. Objetivos

Reducir los errores geométricos que generan la degradación de precisión en un brazo robótico, con un error medio menor a 0.1 mm.

a. Objetivos específicos

Los objetivos específicos de este proyecto son los siguientes:

- Diseñar un sistema de medición de posiciones articulares de un brazo robótico con cámaras estéreo y redes neuronales convolucionales.
- Medir la posición de cada articulación en tiempo real y realizar corrección a la señal de retroalimentación.
- Determinar el grado de aproximación en el efecto final del robot, y estimar la estabilidad en el tiempo, para conocer la mejora del rendimiento del brazo.

V. Material y Métodos o Metodología

Estos son los materiales y métodos utilizados en este trabajo, incluida la cámara y el robot colaborativo, también conocido como cobot, es un robot industrial que puede trabajar de forma segura junto a humanos en un espacio de trabajo compartido. A esto se añaden los conceptos teóricos necesarios, como las redes neuronales convolucionales, la configuración estereoscópica y otros.

a. Visión estereoscópica

La visión estereoscópica es una técnica para estimar estructuras tridimensionales del mundo físico utilizando dos imágenes tomadas desde ángulos diferentes [17]. La figura 3 muestra cómo se compone la visión estereoscópica, que consta de dos cámaras con los mismos parámetros. En un punto de referencia $P(X_p, Y_p, Z_p)$, la proyección sobre el plano imagen derecho es $p_1(x_1, y_1)$ y sobre el plano imagen izquierdo es $p_2(x_2, y_2)$, donde f es la distancia focal y b es la distancia entre los centros ópticos de las dos cámaras. Refiriéndonos a la Figura 2 y al principio de triángulos similares, definimos la disparidad d en la Ecuación (1), entonces podemos obtener la profundidad Z del punto P a partir de las Ecuaciones (1) y (2) [23, 34]:

$$d = x_1 - x_2 \quad (1)$$

$$Z = b \times f/d \quad (2)$$

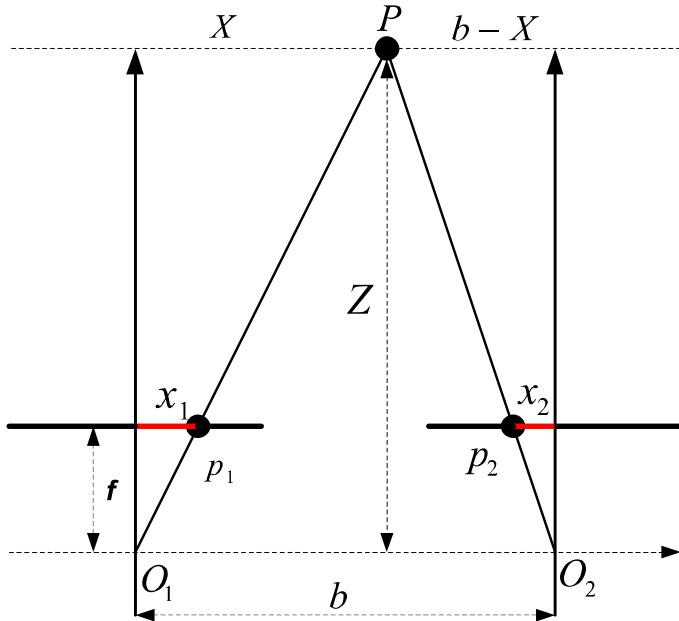


Figura 2: Análisis del esquema de triangulación para visión estéreo, en una perspectiva bidimensional [23].

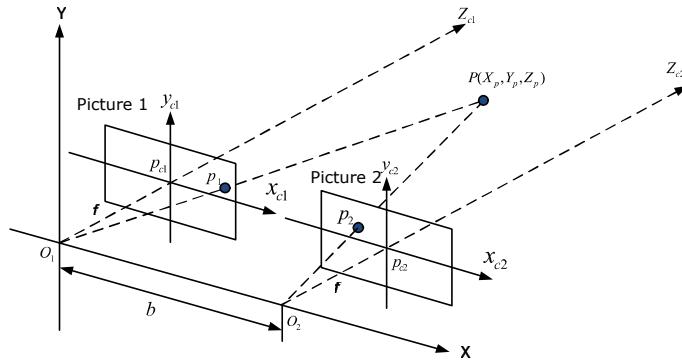


Figura 3: Configuración estéreo en tres dimensiones protect[23].

Para estimar la disparidad en el par de imágenes digitales, la vecindad de un píxel p en la imagen izquierda se compara con la vecindad de la imagen del píxel q en la imagen derecha, donde q está desplazado por una posible disparidad δ_p en comparación con p . En la figura 4, para cada píxel p , se prueban N posibles disparidades ($\delta_p^1, \delta_p^2, \dots, \delta_p^N$) ($N = 256$ y $65, 536$ para 8 y 16 bits de mapeo de profundidad) y la posible disparidad resultante con el menor coste de ajuste se asigna al píxel p [44].

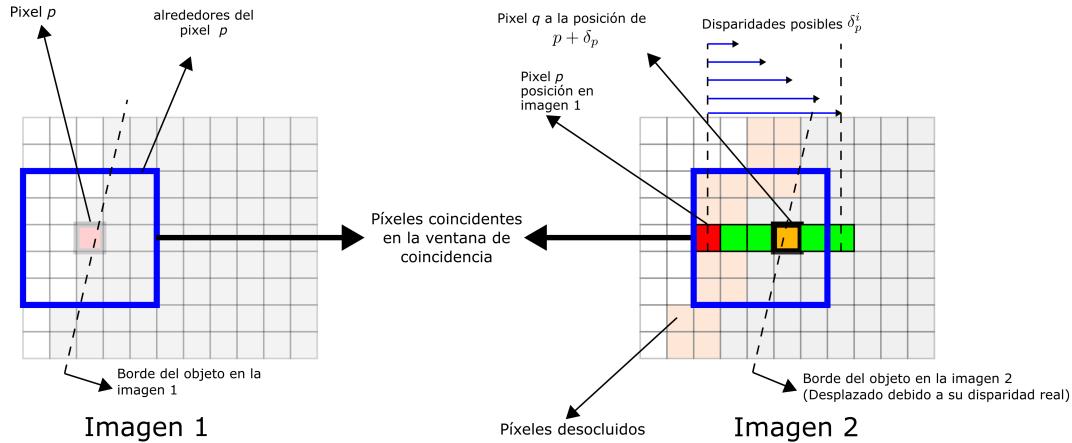


Figura 4: Ventanas coincidentes en visión estereoscópica. Esto representa la funcionalidad digital de una cámara estéreo comercial [44].

b. Precisión de la cámara estéreo D435

Usando [23] y [44], se calculó la precisión de la cámara estéreo, se tomaron algunas imágenes con una regla en el fondo para medir 1 centímetro (cm), la distancia entre dos píxeles se dibujó manualmente, no era necesario obtener 1 cm de precisión. A continuación, se añadió color a los dos píxeles con la información de coordenadas (x, y, z). La distancia entre los píxeles se calculó con la ecuación (3).

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

Tabla 2: Distancia entre píxeles para medir la precisión de la cámara. La distancia entre el píxel 1 y los siguientes píxeles consecutivos.

Píxeles	Distancia [mm]
1 a 2	13.9
1 a 3	15.0
1 a 4	16.3
1 a 5	17.7

Tabla 3: Distancia entre dos píxeles consecutivos.

Píxeles	Distancia entre 2 píxeles consecutivos
3-2	1.1
4-3	1.2
5-4	1.4

Después, se realizó la misma tarea con los siguientes píxeles y se determinó la distancia entre dos píxeles consecutivos, lo que representa la precisión de la cámara estéreo.

La figura 5 muestra la distancia entre 2 píxeles. En primer lugar, se miden los píxeles 1 y 2, luego los píxeles 1 y 3, y así sucesivamente hasta el píxel 5. A continuación, se calcula la distancia más pequeña, la distancia del píxel 3 menos 2, y así sucesivamente hasta el píxel 5. Los resultados se muestran en las tablas 2 y 3.

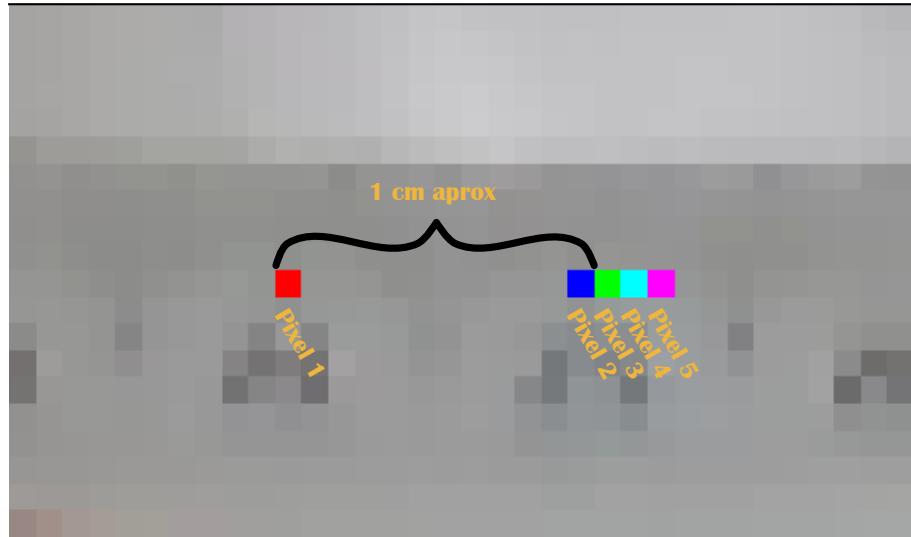


Figura 5: Distancia de un centímetro en píxeles, y el píxel posterior para medir la resolución de la cámara D435.

Como puede observarse, el robot tiene una precisión de aproximadamente 1 mm, que varía un poco en función de la distancia y el ángulo de medición.



Figura 6: Materiales y dimensiones del robot, DoF (Degree of Freedom).

c. Materiales

El presente trabajo implementó la ISO–9283 [45] para la prueba del prototipo. En la figura 6 se muestra de forma grafica los materiales de ese proyecto. La lista de los componentes y sus características utilizadas para este trabajo se enumeran a continuación.

- 1 MyCobot 280 nano, sus principales características son:
 - Rango de trabajo 280 mm.
 - Angulo máximo de cada DoF $\pm 170^\circ$.
 - Repetitividad ± 5 mm.
 - Una microcomputadora Jetson Nano 280, con 2 Gb de RAM y procesador NVIDIA
 - Servomotores de alta precisión.
 - 6 DoF.
- Una cámara de profundidad modelo D435 de la marca Intel. Sus características son:
 - Longitud focal de 1,88 mm.
 - resolución de profundidad de 1280×720 pixels, $\pm 0,006$ mm
 - rango de 0,2–10 m.
 - 6 DoF para calibración.
- Un monitor.

- Ratón y teclado.

El sistema está configurado en una mesa de 150×56 cm. El Cobot se coloca a una distancia de 63.5 cm de la parte superior y 32 cm desde el lado izquierdo de la mesa. Las dimensiones del robot son 15×11 cm y la cámara es de 2.5×9 cm. El círculo de color representa la zona de trabajo del robot, mientras que el triángulo es la zona de visión de la cámara. Se utiliza una cámara en el extremo inferior de la longitud de la mesa y otra en el extremo superior, la representación del espacio de trabajo se encuentra en la figura 7.

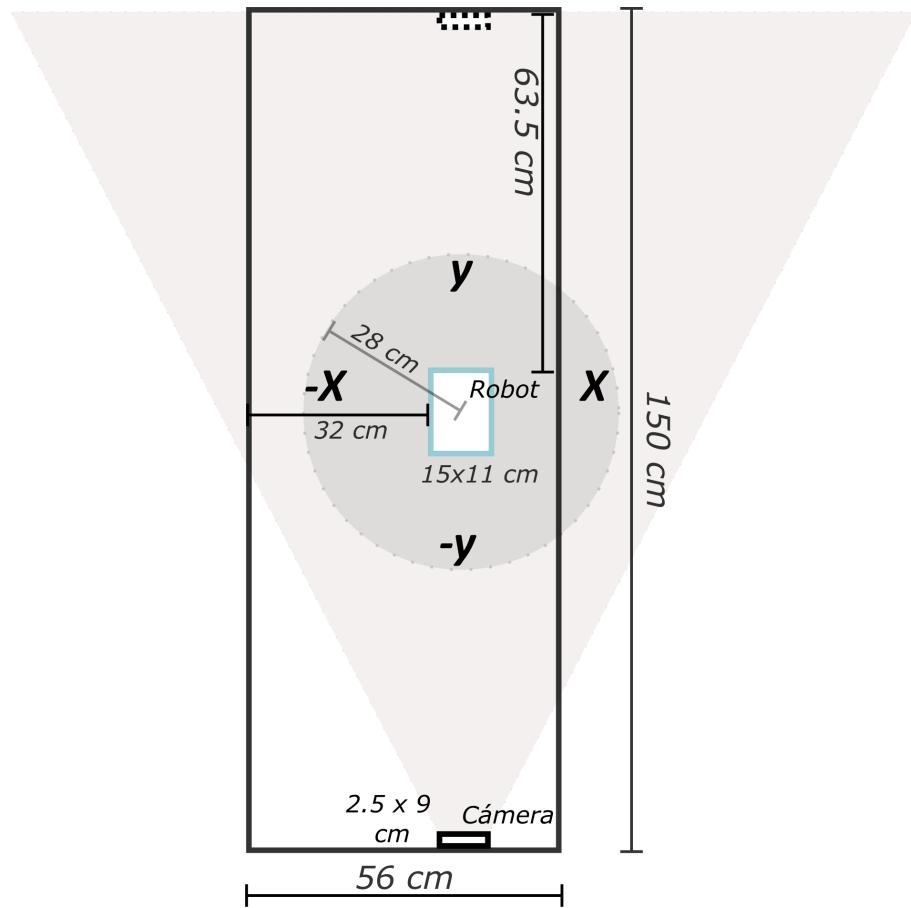


Figura 7: Dimensiones del espacio de trabajo.

como se muestra en la Figura 8, cada DoF se mide secuencialmente desde q_1 hasta q_6 . Cada DoF se enumera con etiquetas de colores para facilitar su segmentación por la cámara. También se muestran las dimensiones de cada eslabón del Cobot.

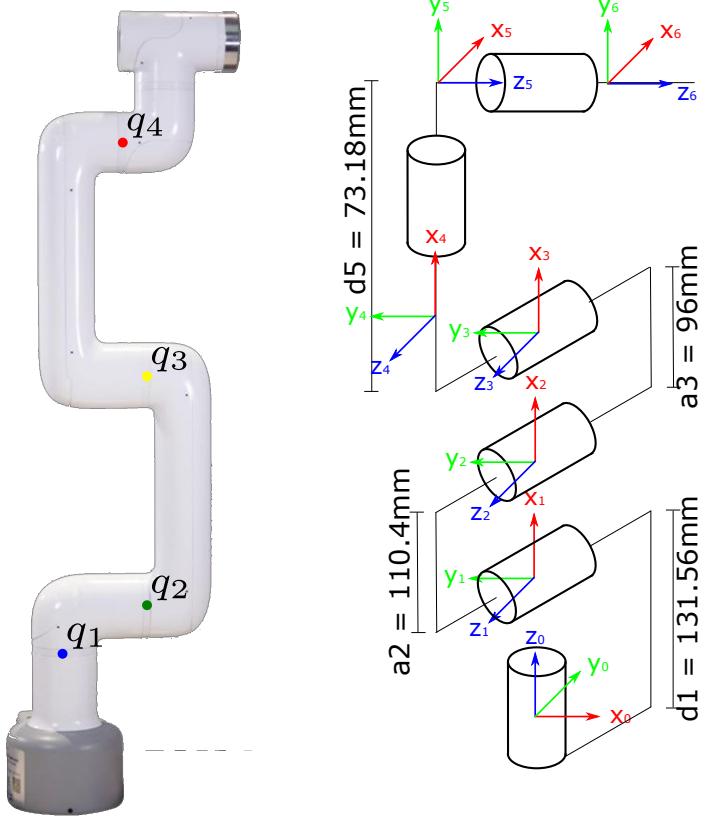


Figura 8: Morfología de un robot de seis DoF, solo cuatro DoF fueron etiquetados con diferente color.

d. Sistema de medición propuesto

En la figura 9 se muestra el sistema robótico donde se desarrolló el estudio, la linea punteada de roja encierra nuestro sistema de medición, el cual se compone de un algoritmo inteligente (CNN) y una cámara estéreo, como se puede observar la propuesta se refiere a un sistema de medición redundante para el posicionamiento del robot. Los demás componentes del sistema: cinemática inversa, controlador y robot, ya están configurados de forma independiente y su estudio no forman parte de este trabajo.

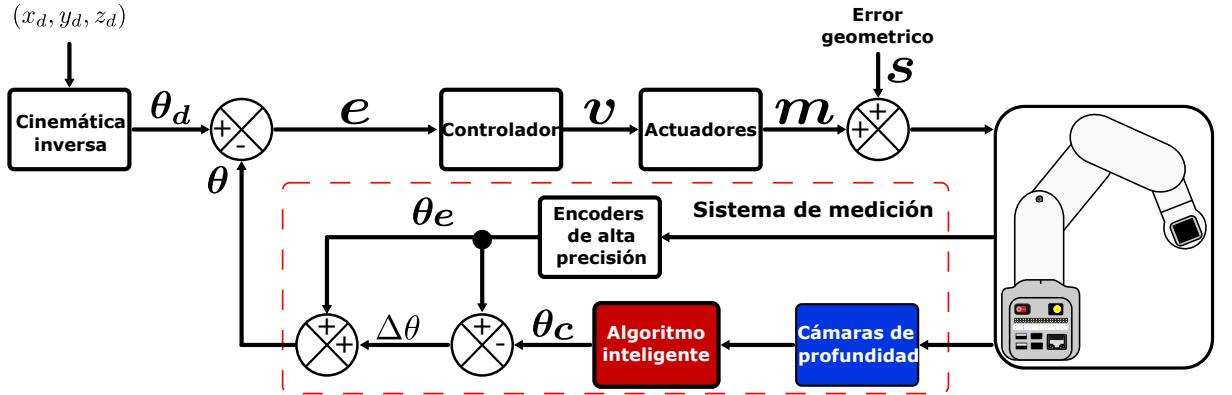


Figura 9: Diagrama a bloques del sistema robótico de este proyecto.

En la figura 10(a) se muestra solo la parte del sistema de medición de este proyecto, como se observa hay dos mediciones: la de los encoders y otra recurrente que proviene de la cámara esterero, estas al ser restadas dan como resultado una diferencia de distancia, que representa la medición de los encoders menos la medición real del eslabón medida por la cámara a esta diferencia se le conoce como degradación de precisión geométrica y se denota como se muestra en la figura 10(b).

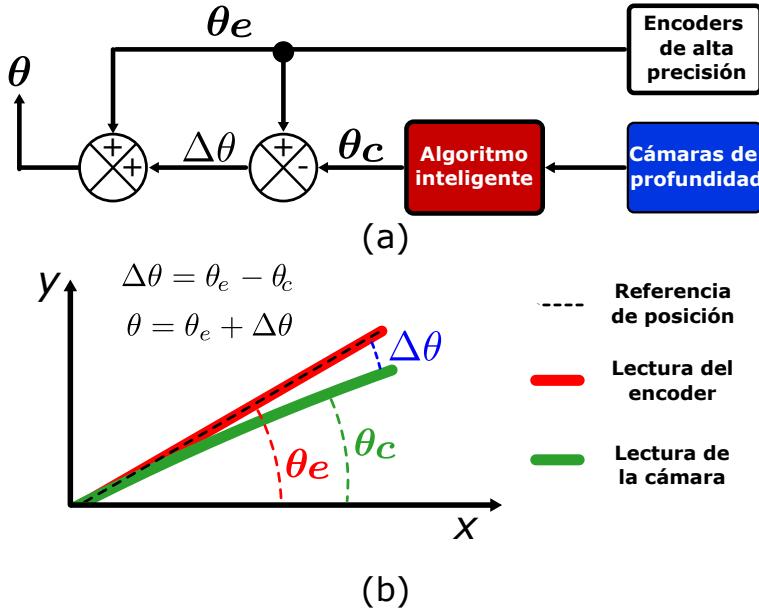


Figura 10: (a) Diagrama a bloques del sistema de medición (b) análisis geométrico de la degradación de precisión en un DoF.

Por último se muestra en la figura 11, la forma en que se desarrolla el sistema de medición, comenzando con la cámara estereo, en el cual sigue un pre-procesamiento donde se obtiene de la imagen la posición de los eslabones a través de sus etiquetas y su profundidad en la imagen.

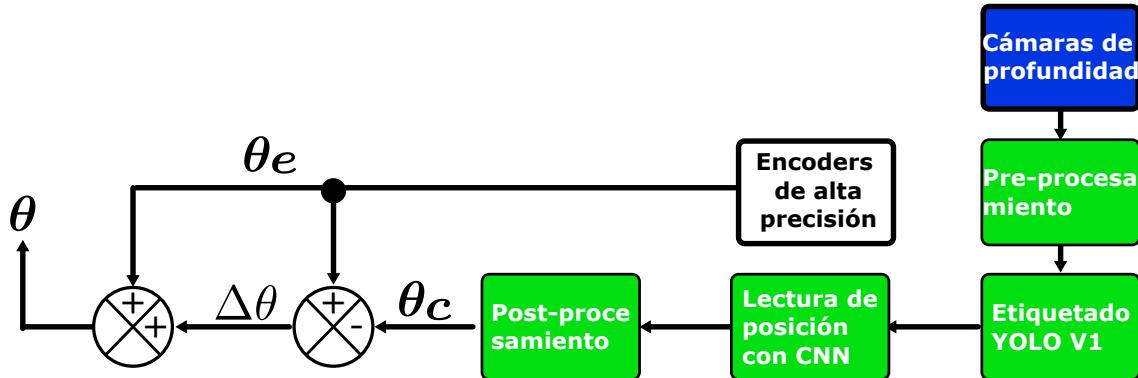


Figura 11: Metodología para el sistema de medición.

e. Interfaz Gráfica de Usuario (GUI)

Para mejorar la precisión a la hora de gestionar la manipulación del cobot, se desarrolla una Interfaz Gráfica de Usuario (GUI), utilizada para crear una interacción entre el usuario y el ordenador, para mostrar los movimientos fundamentales asociados a cada DoF ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6$) y los valores de posición del efecto final (x, y, z, r_x, r_y, r_z). La interfaz también comprueba si el robot está conectado y tiene un botón para llevar el robot a la posición inicial (todos los ángulos a cero) y un botón para apagar los servomotores como medida de precaución. Esta interfaz se está mejorando constantemente, por lo que en el futuro se implementarán más modos de prueba para el posicionamiento. La figura 12 muestra la interfaz gráfica de usuario.

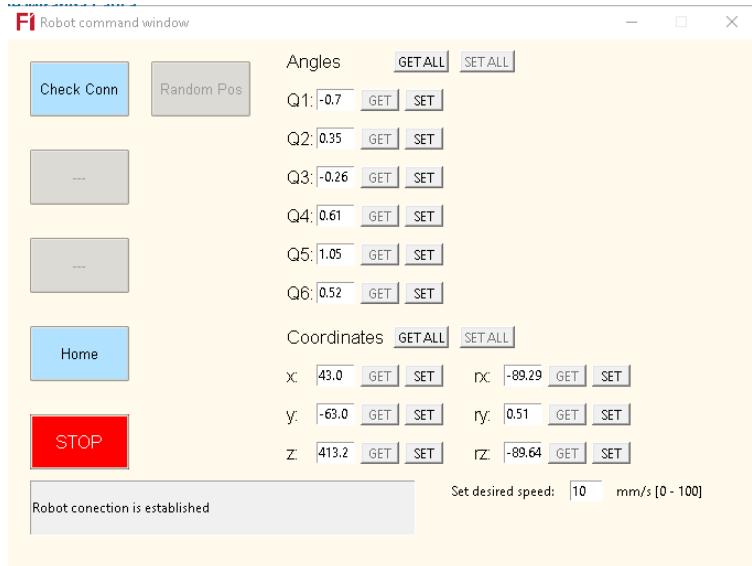


Figura 12: GUI utilizada para una mejor manipulación del brazo robótico, para la comunicación; se implementa el protocolo RPC (Remote Procedure Call), la comunicación RPC se basa en la arquitectura TCP.

f. Medición de la precisión de la posición

La precisión de la posición es la desviación entre la posición ordenada del punto TCP (N) y la media (baricentro, G) calculada a partir del conjunto de posiciones TCP alcanzadas repetidamente. La figura 13 muestra la precisión de posición de 1 DoF, donde G es el baricentro $[\bar{x}, \bar{y}, \bar{z}]$, y N es la posición comandada $[x, y, z]$. La precisión de la posición de 1 DoF a_p viene dada por la Ecuación (4).

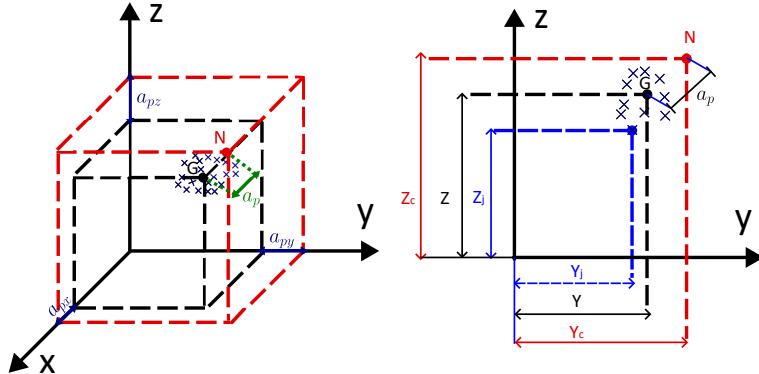


Figura 13: Representación gráfica de la precisión de posición de 1 DoF según ISO 9283-1998.

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

donde $a_{px} = \bar{x} - x_{cmd}$, $a_{py} = \bar{y} - y_{cmd}$, y $a_{pz} = \bar{z} - z_{cmd}$. La precisión de posición a_p de un robot manipulador se define como la distancia entre la posición ordenada x_{cmd} y la media \bar{x} de N posiciones alcanzadas, como se describe en la Ecuación (5) y se muestra en la figura 14

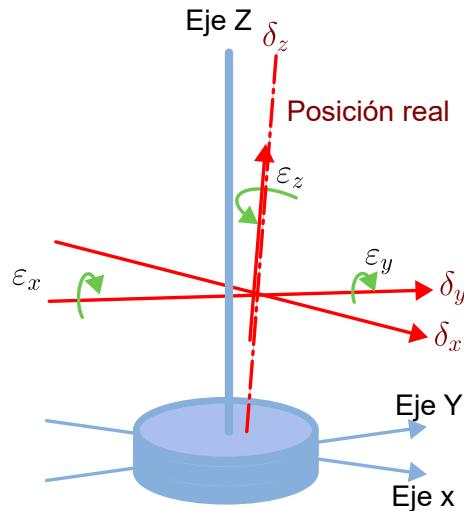


Figura 14: Seis errores de un i-ésimo eje rotacional.

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

Las coordenadas de las medias se dan en la ecuación (6):

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

x_i, y_i , y z_i son las coordenadas de posición de la i-ésima posición medida. La precisión de la posición a_p se obtiene utilizando la norma euclíadiana 7:

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

Se requiere un plan de medición para controlar los movimientos del robot. Los movimientos controlados del robot para la evaluación de la degradación de la precisión no pueden ser arbitrarios. El plan de medición requiere que el robot TCP se mueva por toda el área de trabajo y se distribuya uniformemente tanto en el espacio articulado como en el cartesiano. La distribución uniforme de las muestras impide que el algoritmo de análisis pase por alto o sobrevalore errores, lo que falsearía los resultados. Cubrir todo el espacio articulado y cartesiano significa que el plan de medición entrena al robot en un subconjunto de articulaciones o zonas de trabajo. Esto permite registrar el rendimiento de las articulaciones en todo el rango del motor y del encóder. Cubrir todo el espacio de trabajo ayuda a evaluar diferentes condiciones de rigidez. Al crear el plan de medición, se realiza una comprobación de colisión para minimizar las posibles interrupciones durante el movimiento y la medición. También se realiza una comprobación de visibilidad para garantizar que las posiciones planificadas no queden ocultas por el instrumento de medición [8].

$$E_{(i-1)i}(\theta) = \begin{bmatrix} 1 & -\varepsilon_z(\theta) & \varepsilon_y(\theta) & \delta_x(\theta) \\ \varepsilon_z(\theta) & 1 & -\varepsilon_x(\theta) & \delta_y(\theta) \\ -\varepsilon_y(\theta) & \varepsilon_x(\theta) & 1 & \delta_z(\theta) \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Donde $E_{(i-1)i}(\theta)$ es la transformación del frame i-1 al frame i y θ es la variable de ángulo de la i-ésima articulación.

- δ_x es el movimiento de error radial del eje i en dirección X.
- δ_y es el movimiento de error radial del eje i en dirección Y.
- δ_z es el movimiento de error axial del eje i en dirección Z
- ε_x es el movimiento de error de inclinación alrededor de X del eje i.
- ε_y es el movimiento de error de inclinación alrededor de Y del eje.

- ε_z es el error de posicionamiento angular (también denominado

Para este trabajo se utilizó un Cobot 280 Nano de Elephant Robotics. El movimiento de bucle fijo generado para el cobot se muestra en la Figura 15. Como el cobot estaba montado sobre una mesa, sólo las poses por encima de la mesa son válidas para este caso de uso. El movimiento del robot supone un volumen de trabajo sin obstáculos por encima de la mesa.

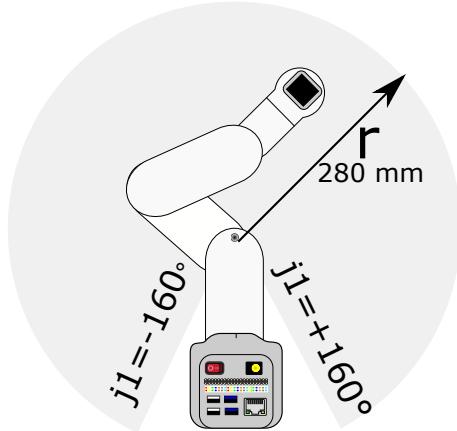


Figura 15: Espació de trabajo del Cobot 280 Nano, 28 cm de rango.

Basándose en [8], se planifica un momento de cuadrícula del Cobot 280 Nano. Se miden las posiciones del centro de la herramienta (TCP) (x,y,z) del robot. Se comparan las diferencias entre las posiciones nominales y las posiciones medidas. También se toman datos de medición a nivel de control de cada articulación para comprender las influencias de la temperatura, la carga útil y la velocidad en los cambios de posición. La información a nivel de control proporciona pistas sobre las causas de la degradación de precisión del robot al proporcionar información sobre las posiciones reales y de las articulaciones, las velocidades, las corrientes, las aceleraciones, los pares y las temperaturas. La precisión de orientación para cada eje de movimiento es la diferencia entre los ángulos de orientación α , β y γ de la pose comandada y los ángulos medios de los N ángulos alcanzados para cada eje. El ángulo de orientación α se relaciona con el eje x , β con el eje y y γ con el eje z .

g. Redes neuronales Convolucionales

También conocidas como CNNs, son la arquitectura de red neuronal estándar utilizada para predicciones cuando la entrada son imágenes, que es el caso de una gran variedad de aplicaciones de redes neuronales, como es el caso de este trabajo.

Las arquitecturas implementadas para este proyecto son las siguientes:

LeNET-5 Es uno de los primeros modelos preentrenados propuestos por Yann LeCun y otros en 1998 en el trabajo de investigación Gradient-Based Learning Applied to

Document Recognition. Utilizaron esta arquitectura para reconocer caracteres manuscritos e impresos a máquina. La principal razón de la popularidad de este modelo fue su sencilla arquitectura. Se trata de una red neuronal convolucional multicapa para la clasificación de imágenes, la figura 16 muestra la descripción de su arquitectura [46].

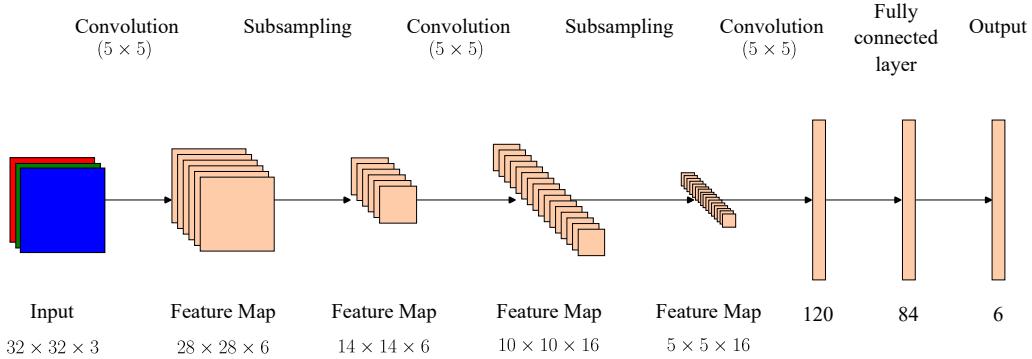


Figura 16: Arquitectura del modelo LENET-5 [46].

AlexNET es un modelo de red neuronal convolucional propuesto por K. Simonyan y A. Zisserman. El modelo alcanza una precisión de prueba del 92.7% en ImageNet, un conjunto de datos con más de 14 millones de imágenes pertenecientes a 1000 clases. La figura 17 muestra la arquitectura de esta red [47].

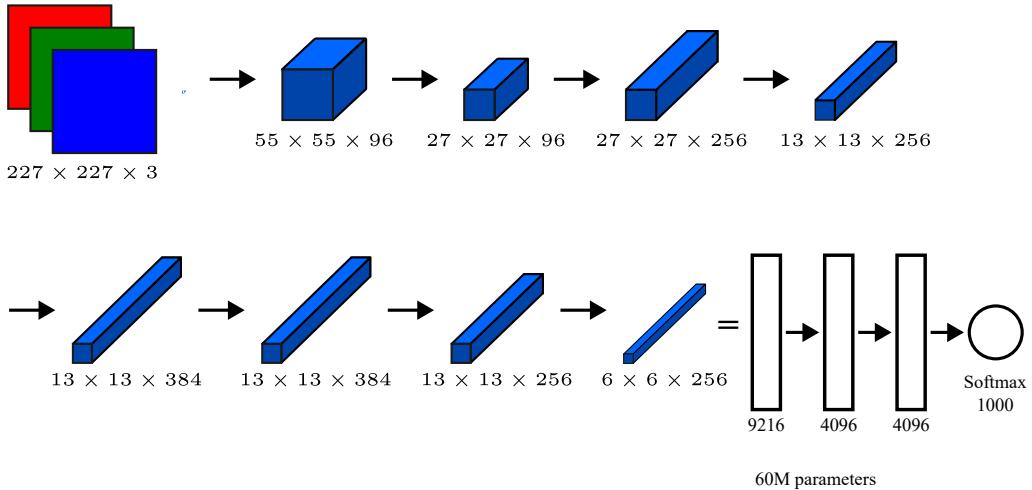


Figura 17: Arquitectura: AlexNet [47].

YOLO “You Only Look Once” es un sistema de reconocimiento de objetos en el conjunto de datos Pascal VOC 2012. Puede reconocer las 20 clases de objetos de Pascal (la mayoría son animales). Todos los sistemas de reconocimiento anteriores utilizan clasificadores o localizadores para realizar el reconocimiento. Aplican el modelo a una imagen en múltiples localizaciones y escalas. Las regiones con puntuaciones altas se consideran reconocimientos. Los diseñadores de YOLO utilizan un enfoque completamente

distinto. Aplicamos una única red neuronal a toda la imagen. Esta red divide la imagen en regiones y predice recuadros delimitadores y probabilidades para cada región. Estos cuadros delimitadores se ponderan en función de las probabilidades pronosticadas [48].

VI. Resultados y discusión

En primer lugar, se obtuvo la cinemática inversa mediante CNN, como primera prueba de las técnicas de Deep Learning, se registró y se guardó el ángulo de cada DoF ($\theta_1, \theta_2, \dots, \theta_6$) y la posición geométrica del TCP (x, y, z) para ejecutar el algoritmo. El robot se movió en su espacio de trabajo con 16 posiciones o puntos de ajuste diferentes; finalmente, se creó una matriz de 3070×9 datos. La figura 18 muestra la posición de cada eje (x, y, z), el punto de referencia y el valor real de cada uno, como se puede observar la precisión es observable en los gráficos el cuarto gráfico son todos los puntos en un gráfico tridimensional.

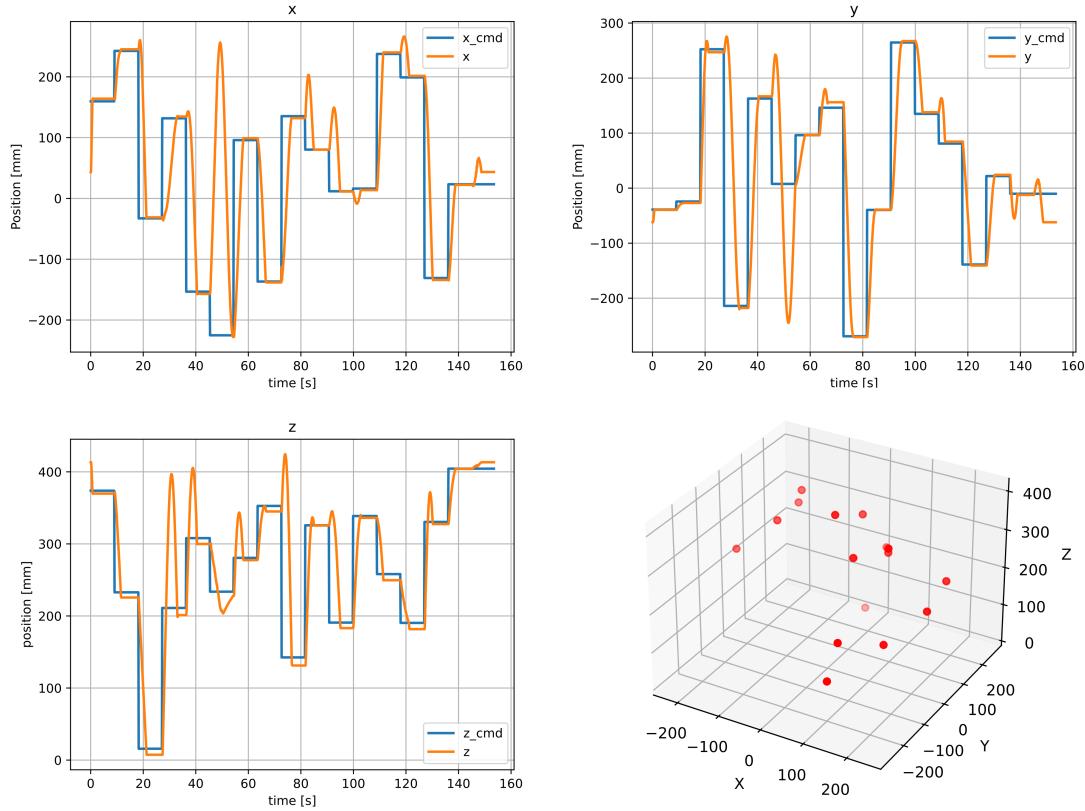


Figura 18: Cada posición tridimensional de la posición TCP (x, y, z), el cuarto gráfico son los puntos que sigue el Cobot.

El algoritmo para predecir la solución cinemática inversa es una CNN de 1 dimensión que toma los puntos TCP (x, y, z) como entradas y los ángulos de las DoF como salidas ($\theta_1, \theta_2, \dots, \theta_6$). Los resultados muestran que se trata de una tarea compleja, los métodos tradicionales son más complicados. La figura 19 muestra el modelo de la CNN unidimensional desarrollado para esta tarea, y la figura 20 muestra los resultados del entrenamiento de la CNN. Para esta prueba se utilizaron 3024 muestras, 2456 para

el entrenamiento y 614 para la validación. Se utilizó el optimizador Adam, incluida la pérdida de error cuadrático medio, y se representó gráficamente la métrica de precisión.

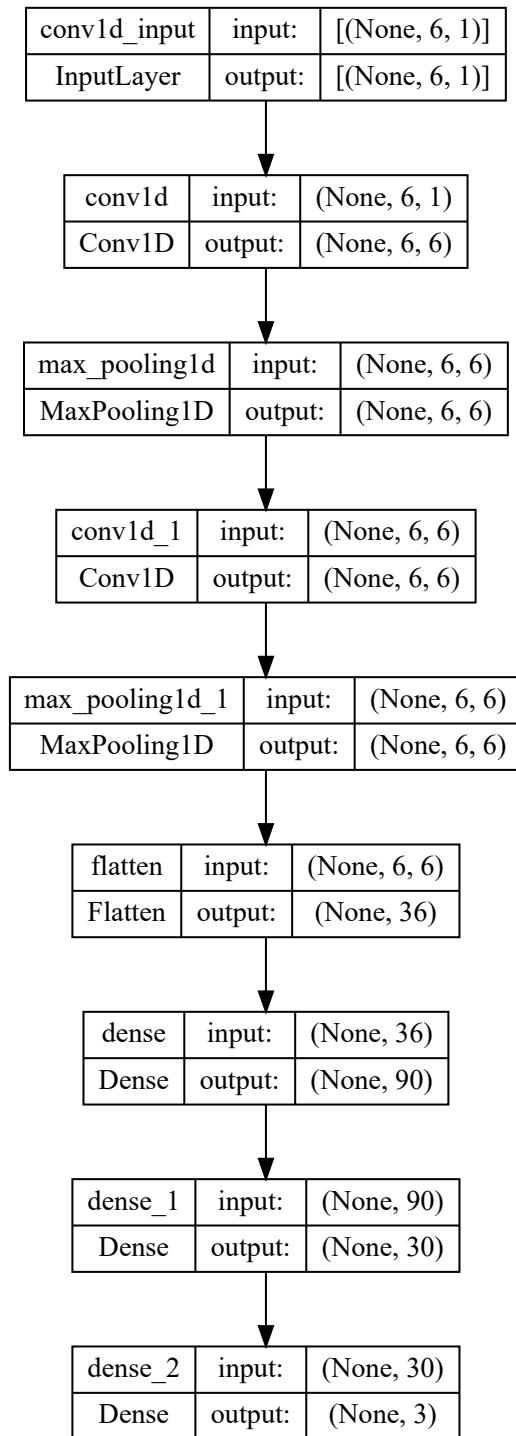


Figura 19: Diseño de la CNN unidimensional utilizada para predecir la cinemática inversa, destacando la capa de entrada, las capas convolucionales y la salida.

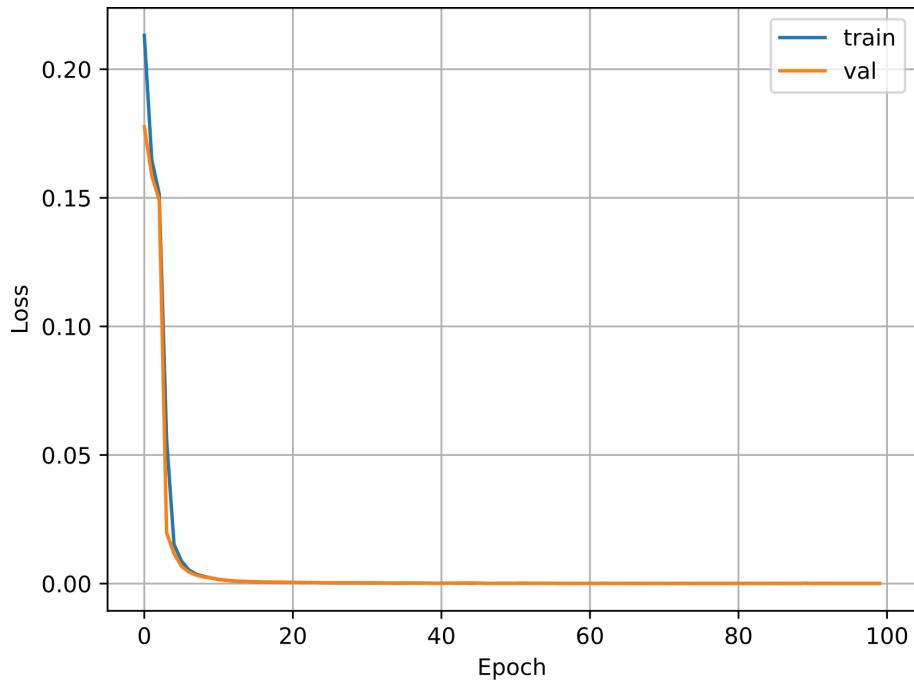


Figura 20: Función de pérdida de entrenamiento frente a validación de la CNN unidimensional.

El siguiente paso para probar las técnicas de DL fue utilizar la información de profundidad de la imagen del cobot para predecir los resultados de la cinemática inversa. La cámara estéreo D435 proporciona información de color y profundidad, 680×460 píxeles para la imagen RGB, y la información de profundidad tiene la misma resolución. Así, cada píxel de color puede representarse en un píxel de profundidad con un valor de 16 bits.

En la figura 21 se puede ver una muestra de imágenes RGB del robot. La figura 22 muestra las mismas imágenes pero con información de profundidad. Para el entrenamiento de este modelo se utilizaron 120 imágenes. La información del ángulo de cada DoF se sustituyó por una imagen de profundidad del robot, lo que complica el diseño de la CNN. Se implementaron los diseños LENET-5 y AlexNet.

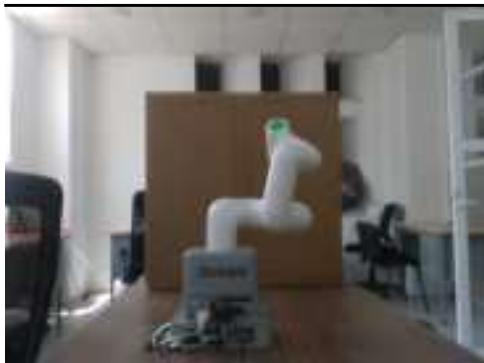
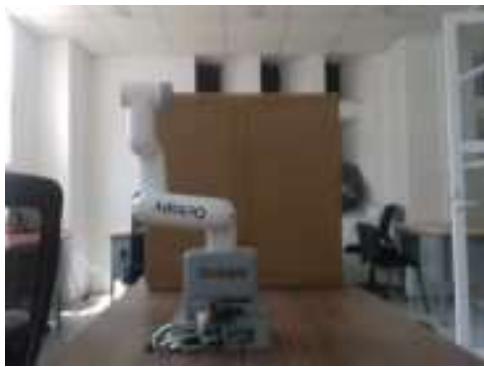
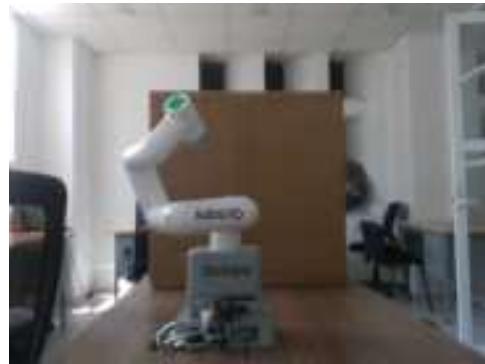
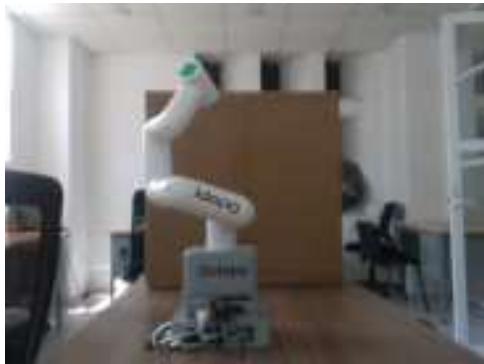


Figura 21: Nueve muestras de imágenes RGB del Cobot, los píxeles son 480×640 píxeles.

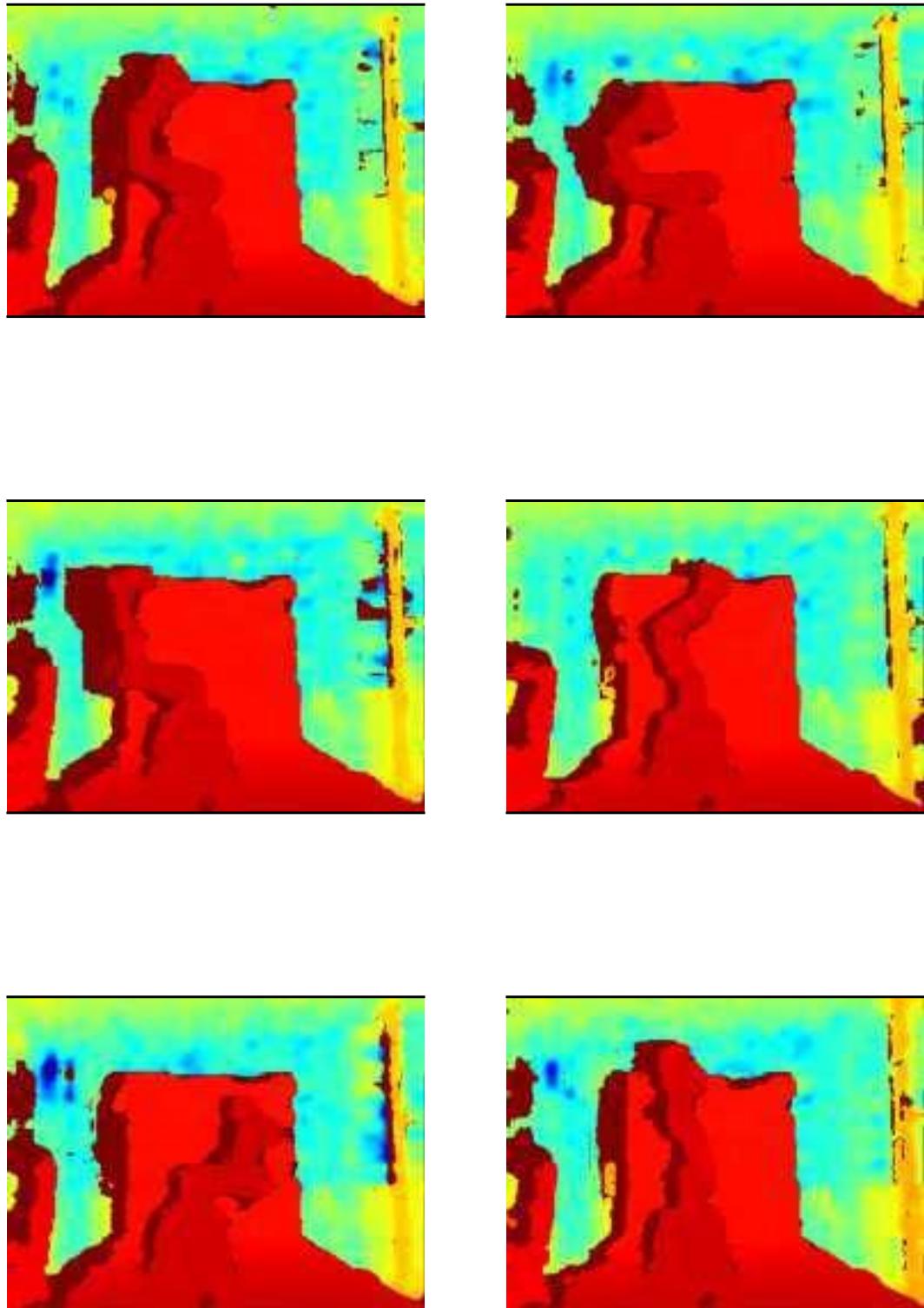


Figura 22: Nueve muestras de información de profundidad para el Cobot, tiene el mismo tamaño de una imagen en color, cada píxel tiene una resolución de 16 bits.

LENET-5

La figura 23 muestra los resultados del entrenamiento de LENET-5 y la figura 25 los de AlexNet. Para estas pruebas se utilizaron 160 muestras, 128 para el entrenamiento y 32 para la validación. Para LENET-5, se utilizó el optimizador Adam, incluyendo la pérdida de error cuadrático medio y la métrica de precisión. AlexNet se configuró con el optimizador AdamMax. La figura 24 presenta la arquitectura resultante.

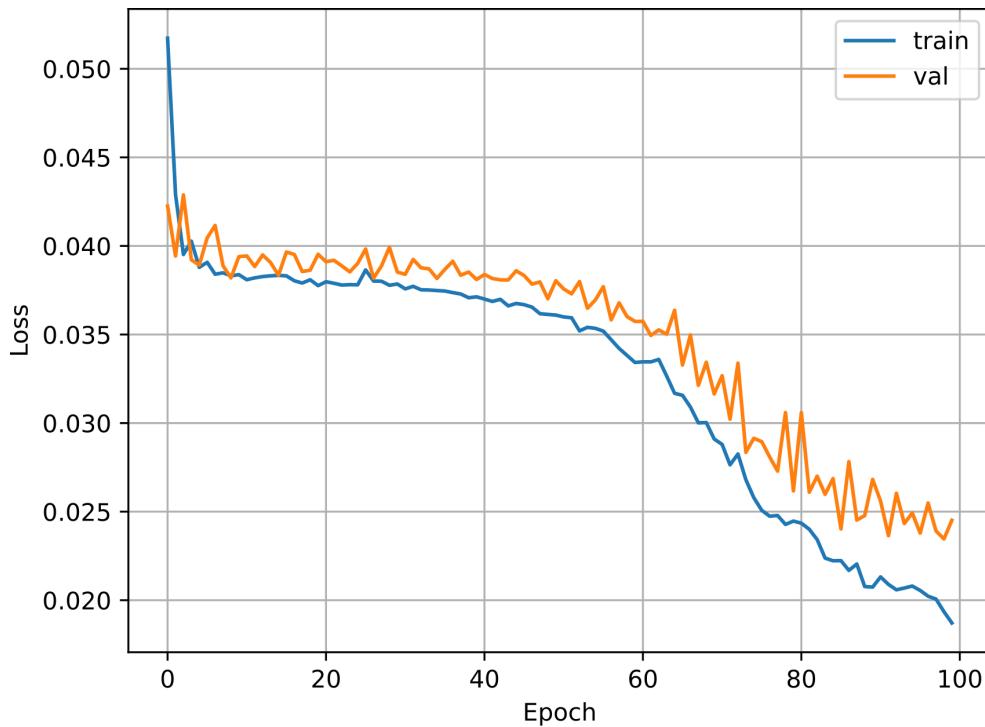


Figura 23: Entrenamiento vs. validation loss function de LENET-5.

AlexNet Es un modelo de red neuronal convolucional propuesto por K. Simonyan y A. Zisserman. El modelo alcanza una precisión de prueba del 92.7 % en ImageNet, un conjunto de datos con más de 14 millones de imágenes pertenecientes a 1.000 clases. La figura 26 muestra la arquitectura construida.

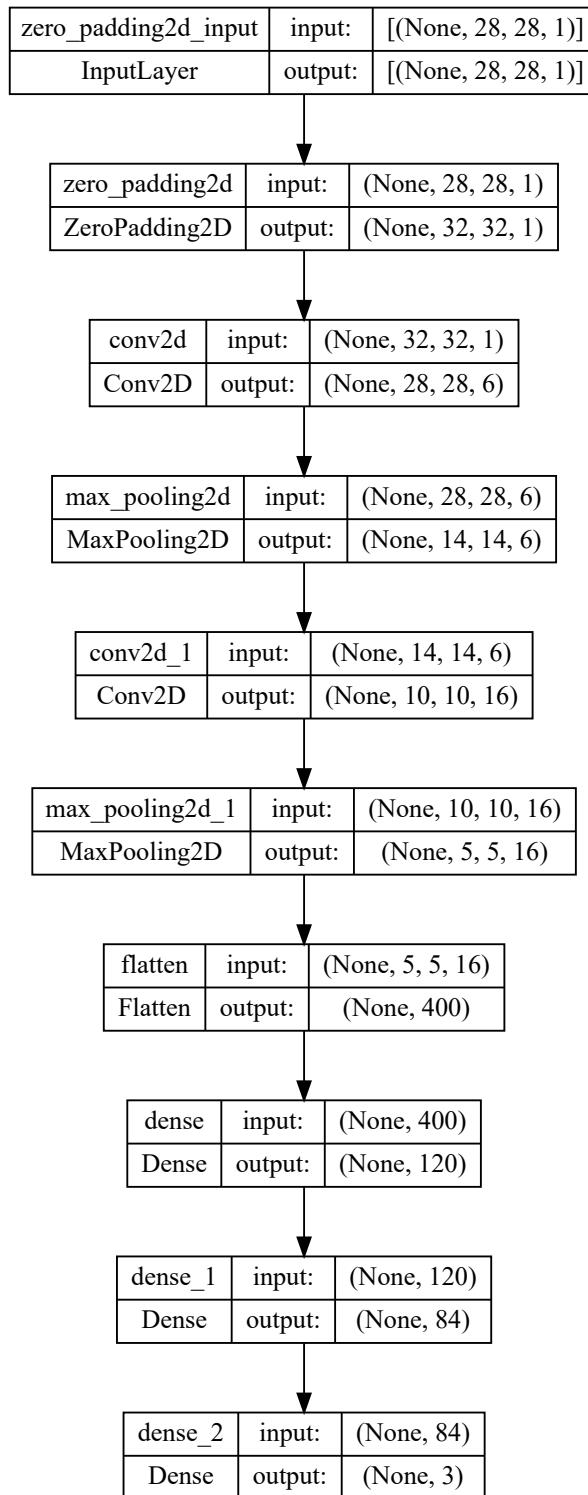


Figura 24: Arquitectura de LENET-5 [46].

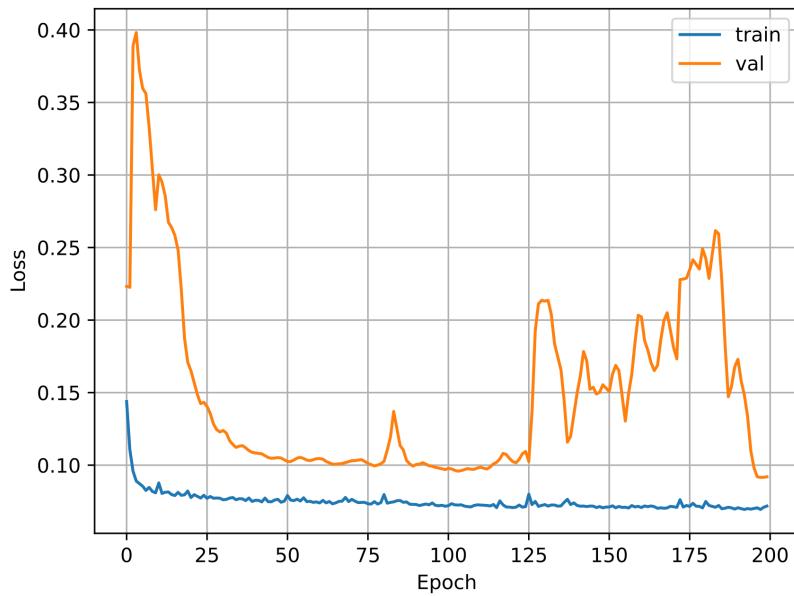


Figura 25: Entrenamiento vs. validation loss function de AlexNet.

Los siguientes pasos fueron poner las etiquetas de colores en los cuatro primeros DoF del cobot (se utilizaron 4 de 6 DoF) y entrenar una la red para reconocer las etiquetas a través del aprendizaje; esto se muestra en la Figura 27. Una etiqueta azul, verde, amarillo y rojo para DoF 1 a 4, respectivamente, se adjuntó a un punto específico en el robot. La etiqueta era redondeada y tenía un radio de 0,6 cm.

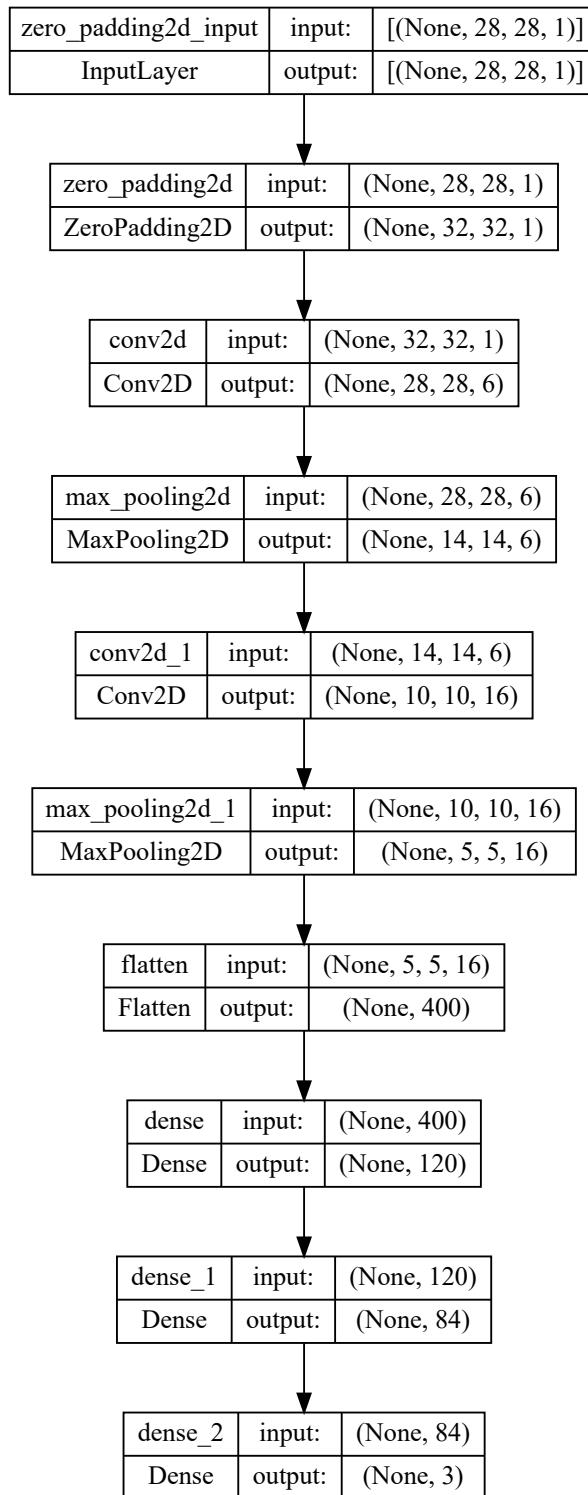


Figura 26: Arquitectura de AlexNet [47].



Figura 27: Imágenes del robot con etiquetas de colores, utilizando YOLO.

Para la anotación de etiquetas se utilizó la plataforma CVAT, que permite conver-

tir imágenes sin procesar en un modelo de visión por ordenador entrenado a medida. Admite modelos de reconocimiento y clasificación de objetos.

Para el reconocimiento de etiquetas, se implementó YOLO-V1, que se ejecuta en la plataforma Roboflow. Los resultados se muestran en la Figura 28. Para esta tarea, se implementaron 215 imágenes del robot en diferentes posesiones. El algoritmo dividió las imágenes en 151 para el entrenamiento, 42 para la validación y 22 para las pruebas.

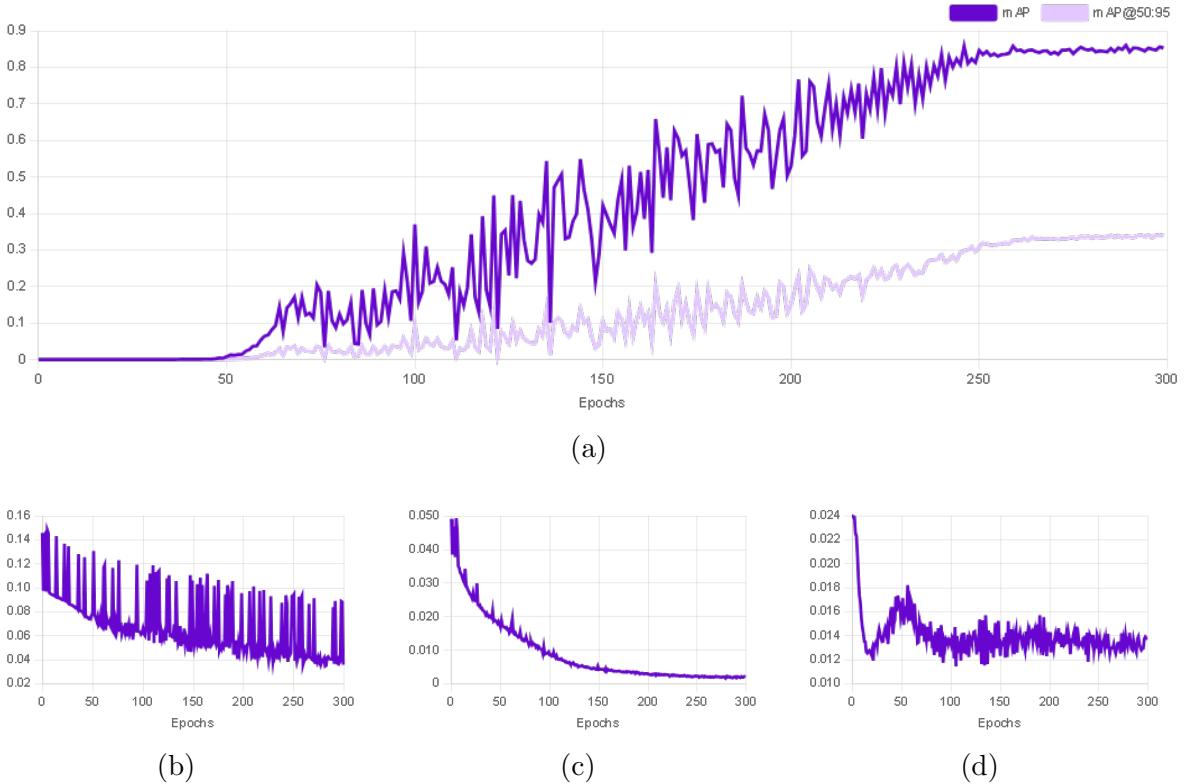


Figura 28: Resultados del entrenamiento de Yolo-V1 para la detección de etiquetas. **(a) Mean average precision (mAP)** se utiliza para medir el rendimiento de los modelos de visión por ordenador. **(b) Box Loss** se refiere a lo bien que el modelo predice las posiciones y tamaños de las cajas que rodean los objetos en una imagen. **(c) Class Loss** se refiere a la medida de la precisión con la que el modelo predice la clase o etiqueta correcta de los objetos que detecta. **(d) Object Loss** mide lo bien que el modelo reconoce la presencia o ausencia de un objeto en una región concreta de una imagen.

La figura 29 muestra algunas imágenes del Cobot que fueron clasificadas, ninguna de ellas fue utilizada para la etapa de entrenamiento, validación o prueba, los resultados arrojados son satisfactorios, el siguiente paso en la experimentación es interpolar la información de etiquetas con la información de profundidad de una imagen para localizar la posición tridimensional de cada DoF.

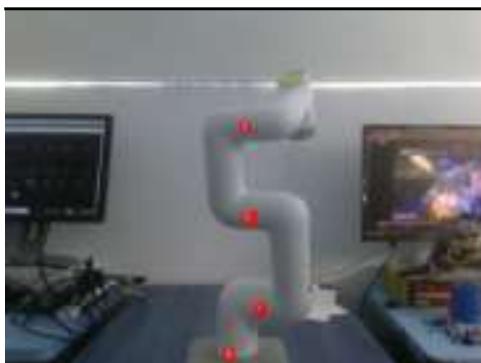
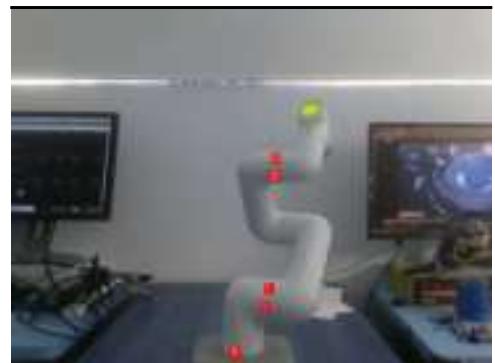
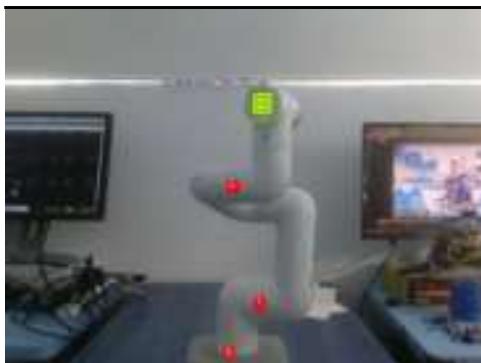
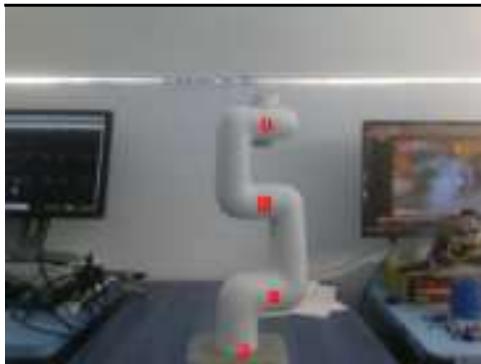


Figura 29: Resultados con un nuevo conjunto de imágenes para evaluar el algoritmo Yolo. Los puntos rojos son las etiquetas reconocidas.

La cámara de profundidad D435 proporcionó la información tridimensional (x , y , z) de cada píxel. La figura 30 muestra la información de profundidad, también se obtuvo un conjunto similar de datos con orientación x e y.

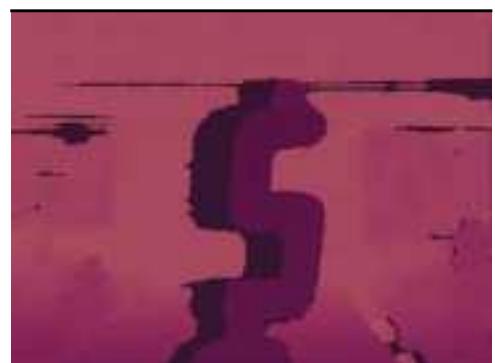


Figura 30: Información de profundidad de las imágenes en color.

Yolo-V1 proporcionó información para cada etiqueta en el Cobot, información que

puede ser interpolada a la imagen de profundidad para que podamos obtener la información tridimensional de cada DoF. Esto se puede observar en la Figura 31.

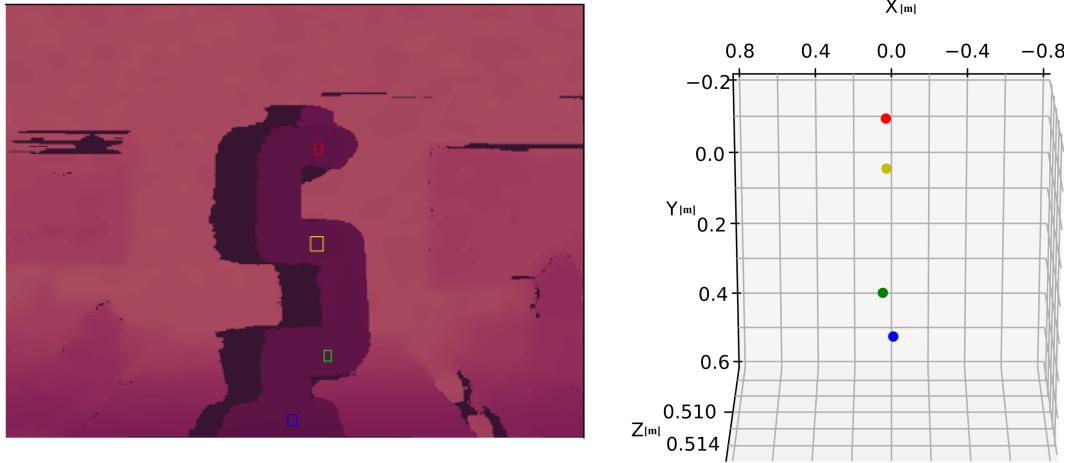


Figura 31: Información de profundidad de las imágenes en color con la ubicación de cada etiqueta. Los puntos del recuadro de color son los puntos marcados detectados por el algoritmo YOLO que ayuda a medir cada posición angular DoF.

a. Discusión

Los resultados obtenidos en este estudio confirman que la combinación de visión estéreo y aprendizaje profundo mejora significativamente la precisión de los robots manipuladores. A través de la integración de la cámara D435 y la implementación de una CNN para la detección de posiciones articulares, se logró reducir la degradación de precisión por errores geométricos.

Las pruebas realizadas con 16 posiciones del robot generaron una base de datos extensa para el entrenamiento del modelo. Los experimentos con CNN demostraron que la predicción de posiciones articulares a partir de imágenes de profundidad es viable, permitiendo ajustes en tiempo real. Además, la segmentación de los eslabones del robot con etiquetas de colores facilitó la identificación precisa de cada DoF, optimizando el análisis tridimensional del sistema.

El uso de Yolo-V1 para la detección de etiquetas demostró ser una estrategia eficiente, con métricas de precisión satisfactorias en la identificación de cada articulación. Asimismo, la interpolación de la información de color y profundidad permitió obtener posiciones tridimensionales con una alta fiabilidad. Comparado con otros métodos del estado del arte, este enfoque supera en precisión a técnicas basadas únicamente en sensores físicos o modelos matemáticos.

No obstante, se identificaron limitaciones relacionadas con la velocidad de procesamiento y la dependencia de condiciones de iluminación y calibración de la cámara. El tiempo de procesamiento del algoritmo sigue siendo un reto, especialmente en aplicaciones en tiempo real. Se recomienda optimizar el modelo CNN y explorar el uso de

hardware dedicado para acelerar la inferencia.

VII. Conclusiones

Este trabajo presentó una metodología innovadora para mejorar la precisión de un brazo robótico mediante visión estéreo y redes neuronales convolucionales. Se demostró que el uso de la cámara D435 en combinación con un modelo CNN permite estimar posiciones angulares con una precisión inferior a 0.1 mm, superando las técnicas convencionales basadas solo en encoders.

Los resultados confirman que la redundancia en la medición con visión estéreo contribuye a la estabilidad del sistema y a la compensación de errores geométricos. Además, la segmentación de eslabones mediante etiquetas de color y su análisis mediante YOLO proporcionan una estrategia eficiente para la detección y análisis del movimiento del robot.

El impacto de esta investigación es significativo en aplicaciones industriales que requieren alta precisión, como la manufactura avanzada y la robótica colaborativa. Sin embargo, se requieren mejoras en la optimización computacional para reducir el tiempo de procesamiento del algoritmo y aumentar la robustez del sistema frente a variaciones en iluminación y entorno.

Finalmente, este estudio abre la puerta a futuras investigaciones en el uso de inteligencia artificial para la calibración y control de robots, con la posibilidad de implementar redes neuronales más avanzadas y mejorar la integración de visión estéreo con otros sensores.

Referencias bibliográficas

- [1] R. Kelly and V. Santibáñez, *Control de movimiento de robots manipuladores*. Pearson educación, 2003.
- [2] S. Subir, “Introducción a la robótica,” 2010.
- [3] G. Qiao and B. A. Weiss, “Quick health assessment for industrial robot health degradation and the supporting advanced sensing development,” *Journal of manufacturing systems*, vol. 48, pp. 51–59, 2018.
- [4] F. Reyes-Cortés, *Matlab aplicado a robótica y mecatrónica*. Alfaomega, 2012.
- [5] A. S. Morris and R. Langari, *Measurement and instrumentation: theory and application*. Academic Press, 2012.
- [6] M. PłAczek and Ł. Piszczeck, “Testing of an industrial robot’s accuracy and repeatability in off and online environment,” *Eksplotacja i Niezawodność*, vol. 20, no. 3, 2018.
- [7] D. Zhang, Z. Peng, G. Ning, and X. Han, “Positioning accuracy reliability of industrial robots through probability and evidence theories,” *Journal of Mechanical Design*, vol. 143, no. 1, 2021.
- [8] G. Qiao and B. A. Weiss, “Industrial robot accuracy degradation monitoring and quick health assessment,” *Journal of Manufacturing Science and Engineering*, vol. 141, no. 7, 2019.
- [9] H. Alzarok, S. Fletcher, and A. P. Longstaff, “Survey of the current practices and challenges for vision systems in industrial robotic grasping and assembly applications,” *Advances in Industrial Engineering and Management*, vol. 9, no. 1, pp. 19–30, 2020.
- [10] L. Pérez, I. Rodríguez, N. Rodríguez, R. Usamentiaga, and D. F. García, “Robot guidance using machine vision techniques in industrial environments: A comparative review,” *Sensors*, vol. 16, no. 3, 2016. [Online]. Available: <https://www.mdpi.com/1424-8220/16/3/335>
- [11] S. Sellami and A. Klimchik, “A deep learning based robot positioning error compensation,” in *2021 International Conference “Nonlinearity, Information and Robotics” (NIR)*. IEEE, 2021, pp. 1–5.
- [12] B. Frank, R. Schmedding, C. Stachniss, M. Teschner, and W. Burgard, “Learning the elasticity parameters of deformable objects with a manipulation robot,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 1877–1883.

- [13] W. Sankowski, M. Włodarczyk, D. Kacperski, and K. Grabowski, “Estimation of measurement uncertainty in stereo vision system,” *Image and Vision Computing*, vol. 61, pp. 70–81, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0262885617300495>
- [14] D. R. Franceschetti, *Principles of robotics & artificial intelligence*. Salem Press, a division of EBSCO Information Services, Incorporated, 2018.
- [15] G. N. Yannakakis and J. Togelius, *Artificial intelligence and games*. Springer, 2018, vol. 2.
- [16] W. Ertel, *Introduction to artificial intelligence*. Springer, 2018.
- [17] P. Corke, *Robotics, vision and control: fundamental algorithms in MATLAB® second, completely revised*. Springer, 2017, vol. 118.
- [18] G. Sangeetha, N. Kumar, P. Hari, and S. Sasikumar, “Implementation of a stereo vision based system for visual feedback control of robotic arm for space manipulations,” *Procedia computer science*, vol. 133, pp. 1066–1073, 2018.
- [19] B. Liang, T. Li, Z. Chen, Y. Wang, and Y. Liao, “Robot arm dynamics control based on deep learning and physical simulation,” in *2018 37th Chinese Control Conference (CCC)*. IEEE, 2018, pp. 2921–2925.
- [20] C.-H. G. Li and Y.-M. Chang, “Automated visual positioning and precision placement of a workpiece using deep learning,” *The International Journal of Advanced Manufacturing Technology*, vol. 104, no. 9, pp. 4527–4538, 2019.
- [21] Q. Li, Y. Pang, Y. Wang, X. Han, Q. Li, and M. Zhao, “Cbmc: A biomimetic approach for control of a 7-degree of freedom robotic arm,” *Biomimetics*, vol. 8, no. 5, p. 389, 2023.
- [22] C. Wang, C. Wang, W. Li, and H. Wang, “A brief survey on rgb-d semantic segmentation using deep learning,” *Displays*, vol. 70, p. 102080, 2021.
- [23] Y.-C. Du, M. Muslikhin, T.-H. Hsieh, and M.-S. Wang, “Stereo vision-based object recognition and manipulation by regions with convolutional neural network,” *Electronics*, vol. 9, no. 2, p. 210, 2020.
- [24] G. Halmetschlager-Funek, M. Suchi, M. Kampel, and M. Vincze, “An empirical evaluation of ten depth cameras: Bias, precision, lateral noise, different lighting conditions and materials, and multiple sensor setups in indoor environments,” *IEEE Robotics & Automation Magazine*, vol. 26, no. 1, pp. 67–77, 2018.
- [25] M. Fischer, M. D. Mylo, L. S. Lorenz, L. Böckenholt, and H. Beismann, “Stereo camera setup for 360° digital image correlation to reveal smart structures of hakea fruits,” *Biomimetics*, vol. 9, no. 3, p. 191, 2024.

- [26] M. Priorelli, G. Pezzulo, and I. P. Stoianov, “Active vision in binocular depth estimation: A top-down perspective,” *Biomimetics*, vol. 8, no. 5, p. 445, 2023.
- [27] H. Zhang and S. Lee, “Advancing the robotic vision revolution: Development and evaluation of a bionic binocular system for enhanced robotic vision,” *Biomimetics*, vol. 9, no. 6, p. 371, 2024.
- [28] R. Pérez, S. C. Gutiérrez, and R. Zotovic, “A study on robot arm machining: Advance and future challenges.” *Annals of DAAAM & Proceedings*, vol. 29, 2018.
- [29] T.-M. Wang, Y. Tao, and H. Liu, “Current researches and future development trend of intelligent robot: A review,” *International Journal of Automation and Computing*, vol. 15, no. 5, pp. 525–546, 2018.
- [30] T. Varhegyi, M. Melik-Merkumians, M. Steinegger, G. Halmetschlager-Funek, and G. Schitter, “A visual servoing approach for a six degrees-of-freedom industrial robot by rgb-d sensing,” *Autom. Robot*, 2017.
- [31] A. M. Martínez, A. G. Garnica, J. Carrillo, and A. Hualde, “Industria 4.0 en méxico. elementos diagnósticos y puesta en práctica en sectores y empresas,” 2020.
- [32] A. G. Godínez and R. J. S. García, “La industria aeroespacial en querétaro: La huella local de una industria global,” *El mundo del trabajo*, p. 13, 2018.
- [33] M. Servi, E. Mussi, A. Profili, R. Furferi, Y. Volpe, L. Governi, and F. Buonamici, “Metrological characterization and comparison of d415, d455, l515 realsense devices in the close range,” *Sensors*, vol. 21, no. 22, p. 7770, 2021.
- [34] M. Khanafer and S. Shirmohammadi, “Applied ai in instrumentation and measurement: The deep learning revolution,” *IEEE Instrumentation & Measurement Magazine*, vol. 23, no. 6, pp. 10–17, 2020.
- [35] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, “A survey of deep learning and its applications: a new paradigm to machine learning,” *Archives of Computational Methods in Engineering*, vol. 27, no. 4, pp. 1071–1092, 2020.
- [36] J. Wu, D. Zhang, J. Liu, and X. Han, “A moment approach to positioning accuracy reliability analysis for industrial robots,” *IEEE Transactions on Reliability*, vol. 69, no. 2, pp. 699–714, 2019.
- [37] J.-C. Hsiao, K. Shivam, I.-F. Lu, and T.-Y. Kam, “Positioning accuracy improvement of industrial robots considering configuration and payload effects via a hybrid calibration approach,” *IEEE Access*, vol. 8, pp. 228 992–229 005, 2020.
- [38] Y. Jiang, L. Yu, H. Jia, H. Zhao, and H. Xia, “Absolute positioning accuracy improvement in an industrial robot,” *Sensors*, vol. 20, no. 16, p. 4354, 2020.

- [39] T. E. Lee, J. Tremblay, T. To, J. Cheng, T. Mosier, O. Kroemer, D. Fox, and S. Birchfield, “Camera-to-robot pose estimation from a single image,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9426–9432.
- [40] M. Abdelaal, R. M. Farag, M. S. Saad, A. Bahgat, H. M. Emara, and A. El-Dessouki, “Uncalibrated stereo vision with deep learning for 6-dof pose estimation for a robot arm system,” *Robotics and Autonomous Systems*, vol. 145, p. 103847, 2021.
- [41] T. Chen, J. Lin, D. Wu, and H. Wu, “Research of calibration method for industrial robot based on error model of position,” *Applied sciences*, vol. 11, no. 3, p. 1287, 2021.
- [42] E. Galan-Uribe, L. Morales-Velazquez, and R. A. Osornio-Rios, “Fpga-based methodology for detecting positional accuracy degradation in industrial robots,” *Applied Sciences*, vol. 13, no. 14, p. 8493, 2023.
- [43] E. Galan-Uribe, J. P. Amezquita-Sanchez, and L. Morales-Velazquez, “Supervised machine-learning methodology for industrial robot positional health using artificial neural networks, discrete wavelet transform, and nonlinear indicators,” *Sensors*, vol. 23, no. 6, p. 3213, 2023.
- [44] R. Chellappa and S. Theodoridis, *Academic Press Library in Signal Processing, Volume 6: Image and Video Processing and Analysis and Computer Vision*. Academic Press, 2017.
- [45] International Organization for Standardization, “Manipulating industrial robots - performance criteria and related test methods,” 1998, iSO Standard. [Online]. Available: <https://normasiso.org/norma-iso-9283/>
- [46] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [48] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [49] R. Zeng, M. Liu, J. Zhang, X. Li, Q. Zhou, and Y. Jiang, “Manipulator control method based on deep reinforcement learning,” in *2020 Chinese Control And Decision Conference (CCDC)*. IEEE, 2020, pp. 415–420.

- [50] Z. Wu, G. Allibert, C. Stolz, C. Ma, and C. Demonceaux, “Depth-adapted cnns for rgb-d semantic segmentation. arxiv 2022,” *arXiv preprint arXiv:2206.03939*.