




Sorting algorithm for product classification using deep learning

Robinson Jiménez Moreno, Doctor en ingeniería¹ , Anny Astrid Espitia Cubillos, Magíster en ingeniería² , y Esperanza Rodríguez Carmona, Magíster en Educación³ 

^{1,2,3}Universidad Militar Nueva Granada, Colombia, robinson.jimenez@unimilitar.edu.co, anny.espitia@unimilitar.edu.co, esperanza.rodiguez@unimilitar.edu.co

Abstract - This article presents the development of a simulated product sorting environment through identification and localization using regional convolutional neural networks. This type of deep learning network allows us to identify three types of products and their location in the scene, which results in using a sorting algorithm by product type and allows us to determine the inventory level of each of them. The network presents 100% identification within the three trained classes and a robotic arm is used in a simulated environment to manipulate each of the products according to their coordinates, which facilitates the preparation of orders. The relocation of personnel dedicated to these tasks is proposed. reducing negative effects on their physical health, speeding up the preparation of orders, which allows a better and more timely response to the customer.

Keywords- Faster RCNN, robotic arm, product ordering, order preparation.

Algoritmo de ordenamiento para clasificación de producto mediante aprendizaje profundo

Sorting algorithm for product classification using deep learning

Robinson Jiménez Moreno, Doctor en ingeniería¹, Anny Astrid Espitia Cubillos, Magíster en ingeniería², y Esperanza Rodríguez Carmona, Magíster en Educación³
^{1,2,3}Universidad Militar Nueva Granada, Colombia, robinson.jimenez@unimilitar.edu.co, anny.espitia@unimilitar.edu.co, esperanza.rodiguez@unimilitar.edu.co

Resumen– El presente artículo expone el desarrollo de un ambiente simulado de ordenamiento de producto mediante la identificación y localización usando redes neuronales convolucionales por regiones. Este tipo de red de aprendizaje profundo permite identificar tres tipos de productos y su localización en la escena, lo cual da lugar a utilizar un algoritmo de ordenamiento por tipo de producto y permite determinar el nivel de inventario de cada uno de ellos. La red presenta una identificación del 100% dentro de las tres clases entrenadas y se emplea un brazo robótico en ambiente simulado para realizar la manipulación de cada uno de los productos según sus coordenadas lo que facilita el alistamiento de pedidos, se propone la reubicación del personal dedicado a dichas tareas reduciendo afectaciones negativas a su salud física agilizando el alistamiento de pedidos lo que permite una mejor y más oportuna respuesta al cliente.

Keywords– Faster RCNN, brazo robótico, ordenamiento de producto, alistamiento de pedidos.

Abstract - This article presents the development of a simulated product sorting environment through identification and localization using regional convolutional neural networks. This type of deep learning network allows us to identify three types of products and their location in the scene, which results in using a sorting algorithm by product type and allows us to determine the inventory level of each of them. The network presents 100% identification within the three trained classes and a robotic arm is used in a simulated environment to manipulate each of the products according to their coordinates, which facilitates the preparation of orders. The relocation of personnel dedicated to these tasks is proposed. Reducing negative effects on their physical health, speeding up the preparation of orders, which allows a better and more timely response to the customer.

Keywords– Faster RCNN, robotic arm, product ordering, order preparation.

I. INTRODUCCIÓN

La industria 5.0 busca la sostenibilidad industrial y el bienestar de los trabajadores, se presenta como una evolución de la industria 4.0 que es considerada una revolución industrial impulsada por la tecnología que promueve la eficiencia y la

calidad de la producción integrando análisis de big data, inteligencia artificial y gemelo digital [1]

El gobierno japonés en 2016 propuso el concepto de Sociedad 5.0. para equilibrar el avance económico con la resolución de los problemas sociales a través del alto grado de fusión entre el ciberespacio y el espacio físico [2].

Estos dos conceptos se centran en el ser humano, se ven beneficiados por las tecnologías emergentes y promueven su avance [1].

Durante años, los brazos robóticos se han venido utilizado en la industria para automatizar tareas repetitivas, extenuantes y complejas donde se requiere tanto velocidad como precisión [3].

Actualmente se están resolviendo un gran número de problemas de recogida de contenedores mediante redes de aprendizaje profundo combinadas con distintos procedimientos [4].

Gracias al desarrollo del aprendizaje profundo sumado a la industria 5.0, se prevé que los robots realicen tareas que un humano puede realizar, por ello el algoritmo propuesto aporta a estos conceptos, al pensar en la sustitución de personas en las labores de alistamiento de pedidos para ubicarlas en áreas que requieran el desarrollo de habilidades como la creatividad y no la ejecución de tareas repetitivas y mecánicas, con sus consecuentes afectaciones a nivel musculoesquelético y relacionadas, logrando una mayor productividad al reducir el tiempo de preparación de pedidos y lograr mayor exactitud al eliminar el error humano usando conceptos de inteligencia artificial y protegiendo la salud y el bienestar de los colaboradores. Adicionalmente,

Ya se ha documentado esta problemática de aumento de los trastornos musculoesqueléticos derivados del movimiento manual [5], también se puede considerar que el costo que implican las incapacidades (temporales y permanentes) asociadas a la manipulación manual de materiales puede compensar la inversión en la tecnología requerida.

Varios autores han trabajado en identificación y recolección de objetos usando aprendizaje profundo en diversos entornos, en [6] integran un sistema de reconocimiento de objetos

utilizando un enfoque de aprendizaje profundo con un sistema de agarre que tiene un manipulador en serie y una pinza para recoger objetos y trasladarlos.

De otro lado, la detección de objetos mediante el aprendizaje automático se está convirtiendo en un componente importante de la operación de tomar y ubicar, ya que brinda una alta posibilidad de detectar objetos sin recalibrar cámaras [7].

En la cadena de suministro y logística [8] usan un modelo de reconocimiento de imágenes de objetos mediante el uso de una red neuronal convolucional (CNN) sobre la base de la simulación de una mano humana que agarra objetos para almacenamiento y clasificación. Otra propuesta [9] usa un enfoque híbrido de aprendizaje profundo e ingeniería para la selección y colocación en tiempo real de múltiples partes 3D de geometrías arbitrarias, en primer lugar, utiliza detectores de objetos basados en aprendizaje profundo configurados en forma de cascada tanto para detectar piezas en un contenedor como para extraer características asociadas con las piezas individuales detectadas, éstas se introducen directamente en la estimación de la pose 3D de la pieza correspondiente en un contenedor así como su grado de oclusión.

En [10] proponen el alistamiento de objetos planos sin textura usando un sistema robótico con 7 grados de libertad de selección guiado por visión basado en imágenes de profundidad que usa un modelo de red neuronal convolucional profunda que además predice patrones de agarre óptimos para una mano con dos ventosas (copa izquierda, copa derecha o ambas copas). También [11] trabaja con objetos planos iniciando con aprendizaje profundo con imágenes 2D para clasificar y localizar el objeto.

Para la detección y clasificación de objetos [12] usan una combinación de procesamiento de nubes de puntos y técnicas de aprendizaje profundo, su modelo utiliza dos flujos que reconocen objetos en RGB y datos de profundidad por separado y combina los dos para clasificar objetos.

En [13] se presenta una célula robótica para ensamblaje que implica recolección combina un sistema avanzado de visión RGB-D y políticas de aprendizaje profundo con un robot colaborativo para el montaje de un teléfono móvil.

A nivel agrícola, [14] usa dos brazos robóticos cooperativos para recoger manzanas en un huerto no estructurado, el primero se utiliza para detectar manzanas e inspeccionar el espacio en busca de caminos despejados, el segundo cosecha la fruta, usa un algoritmo de detección de objetos basado en aprendizaje profundo personalizado, la ubicación y la ruta se codifican en un gráfico y se utilizan para la planificación del movimiento. Por su parte, [15] desarrollan un robot para recolectar automáticamente huevos del piso que consta de un detector de huevos basado en aprendizaje profundo, un brazo robótico, una pinza de dos dedos y una cámara en la mano; el modelo de aprendizaje profundo, se integró en el sistema de visión para detectar y localizar huevos. También [16] para la detección y clasificación del té propone un enfoque de detección y clasificación de una red de fusión de

dos niveles con un universo variable, este enfoque combina la capacidad de detección rápida de YOLOv3 y la de clasificación de alta precisión de DenseNet201, además, comparan la influencia del ángulo de disparo de la cámara en la detección.

Para el tratamiento de residuos de construcción y demolición (RCD) [17] propone un robot para clasificar usando un método de aprendizaje profundo para la detección que genera algunos rectángulos de captura a través de un algoritmo de búsqueda e ingresa los rectángulos en la red neuronal, luego, la red genera la postura de agarre óptima.

Redes de aprendizaje profundo como las Faster RCNN [18] son empleadas con agentes robóticos para la interacción humano maquina en tareas colaborativas así como en tareas de clasificación autónomas [19].

El presente artículo se estructura en la sección 1 con la introducción y estado del arte, la sección 2 con la metodología, la sección tres expone la base de datos empleada, el entrenamiento de la red y los resultados alcanzados, y finalmente, la sección 4 presenta las conclusiones del trabajo.

II. METODOLOGÍA

El primer paso para la validación de redes convolucionales en un ambiente de producción simulado es establecer una base de datos para la clasificación de algún tipo de producto. Dado que se pretende entrenar una red convolucional basada en regiones de tipo Faster RCNN, es necesario hacer ajustes a la base de datos de entrenamiento, que dará lugar a la detección y localización de cada objeto, variando la cantidad y ubicación de cada objeto y agrupándolos en fila y columna para la respectiva correspondencia en un estante de almacenamiento.

Para el entrenamiento de la Faster RCNN se emplea transferencia de aprendizaje [20] [21], la cual implica una arquitectura predefinida, para el caso se emplea la arquitectura ResNet-18 [22] [23], esta arquitectura requiere un volumen de entrada de 224*224 píxeles a color, es decir para los 3 canales R, G y B de la imagen.

Para el entrenamiento se compararon los resultados usando el optimizador SGDM y el Adam, se manejaron otros parámetros como 200 épocas de entrenamiento, una razón de aprendizaje de 1x10⁻³, en un equipo de cómputo con procesador Intel core I7, de 32 GB de memoria RAM y una GPU RTX3070 de 8 GB de memoria.

Tras el entrenamiento, se valida que la red identifique adecuadamente la posición de cada objeto.

Finalmente, se adiciona el algoritmo de ordenamiento el cual se basa en el reposicionamiento de los objetos mediante sus coordenadas. La ResNet18 entrega tanto la clase como la ubicación del recuadro donde se encuentra el objeto, como punto inicial en (x,y) y el ancho y el largo del recuadro, a este se le saca el punto central mediante (1) y se implementa en un escenario virtual usando MATLAB para validar la identificación de los productos, su extracción por agarre del efector del brazo y reubicación mediante la cinemática del brazo robótico.

$$Centro_{x,y} = (x + Ancho/2)_x, (y + Alto/2)_y \quad (1)$$

III. RESULTADOS

Para el establecimiento de la base de datos para la clasificación, a fin de no sesgar el ejercicio, se toman como ejemplo tres tipos diferentes de objetos, que permiten ampliar el espectro de uso del algoritmo implementado.

Para el caso a fin de representar algún tipo de decoración se emplean flores, para simbolizar objetos de entretenimiento se emplean libros y finalmente para simbolizar juguetes se emplea un objeto tipo robot, como se ilustra en la Fig. 1.

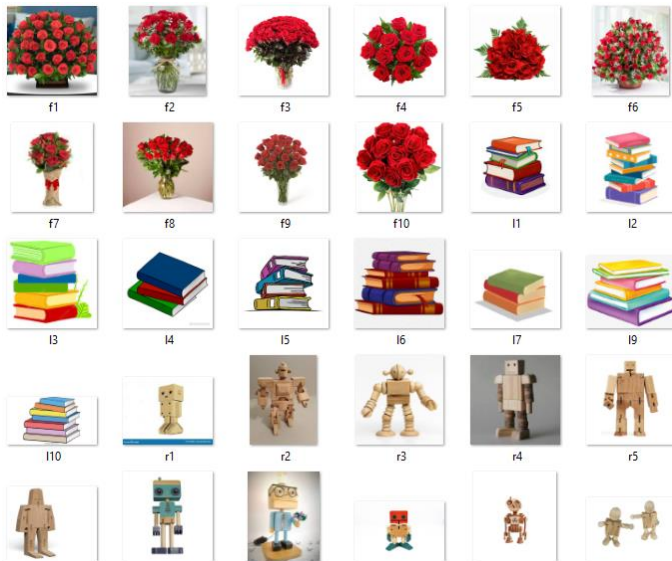


Fig. 1. Base de datos de objetos a ordenar

Se requiere un preprocesamiento para redimensionar las imágenes de la base de datos y conservar la uniformidad de la imagen de origen que para el caso son de 660 x 600 píxeles, dado que los objetos base de la Fig. 1 tenían diferente tamaño se redimensionaron a 224 *224 píxeles.

Para el entrenamiento de la red convolucional basada en regiones de tipo Faster RCNN, se crean imágenes en agrupaciones de 9 elementos con los objetos de interés, como se observa en la Fig. 2. Sobre esta base de datos se etiquetan cada uno de los objetos como robot, florero y libro.

En la Fig. 2 se presentan tres muestras de la base de datos donde se modificó tanto la cantidad como la ubicación de cada elemento. Como se señala en la metodología, se trabajan agrupaciones en fila y columna para su equivalencia en un estante de almacenamiento.

La Fig. 2a presenta los objetos ordenados en filas e igual cantidad, la Fig. 2b presenta el mismo ordenamiento, pero intercambiando fila superior e inferior y la Fig. 2c presenta todos los tipos de objetos con mayoría tipo robot. Esto ejemplifica la variedad en las distintas posiciones y cantidades que se pueden dar en un estante de almacenamiento.



Fig. 2. Base de datos de entrenamiento

La red toma la base de datos para aprender cada objeto en diferentes ubicaciones, entregando así el nivel de precisión con que reconoce un objeto y la información del recuadro dentro de

la imagen donde se localiza como un punto inicial x,y, un ancho y un largo.

Para el entrenamiento se comparó el rendimiento de optimizador SGDM y el Adam, los resultados se presentan en la Tabla 1.

TABLA 1.

MÉTRICAS DE RENDIMIENTO OPTIMIZADORES PROBADOS

Optimizador	Porcentaje de precisión	Tiempo de entrenamiento
SGDM	95.1%	34 minutos
Adam	93.83%	39 minutos

Con los demás parámetros usados, con el optimizador SGDM, el entrenamiento de la red tardó 34 minutos, presenta un RMSE (raíz de error cuadrado medio) de 0.10967 como se aprecia en la Fig. 3.

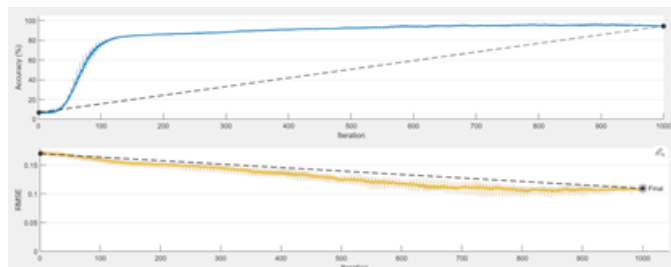


Fig. 3. Proceso de entrenamiento

Finalmente, la red genera adecuadamente la caja que demarca la posición de cada objeto, como se aprecia en la figura 4. Se puede observar que el recuadro para las dos primeras detecciones de florero no queda centrado, lo cual es propio de la detección de la red y los recuadros que emplea el proceso de detección.

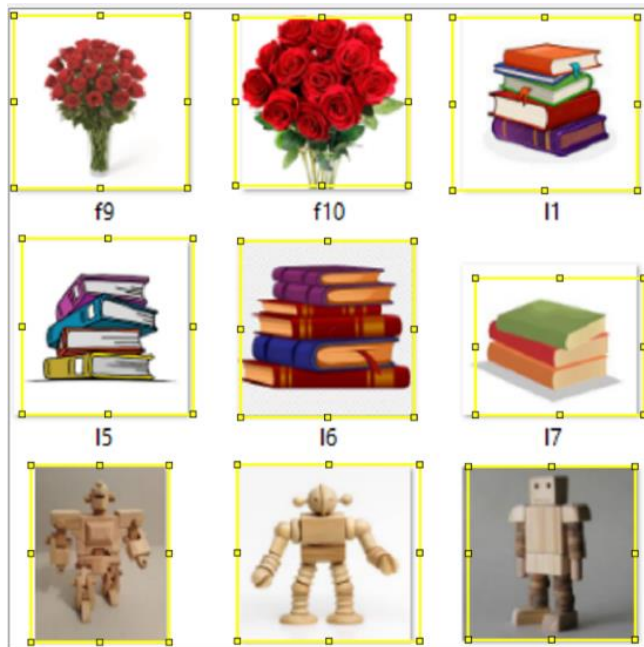


Fig. 4. RoI de objetos identificados

De otro lado, el algoritmo de ordenamiento se da en relación a la numeración de las etiquetas, para lo cual se implementa en MATLAB un escenario virtual de ordenamiento robótico con tres colores, cada uno representando un objeto, a partir de ello se determinó la identificación de los productos, que posteriormente permite su extracción y reubicación por el brazo robótico, como se explicó previamente en la metodología.

La Fig. 5 muestra el etiquetado usado para los tres objetos con los que se trabajó en MATLAB.

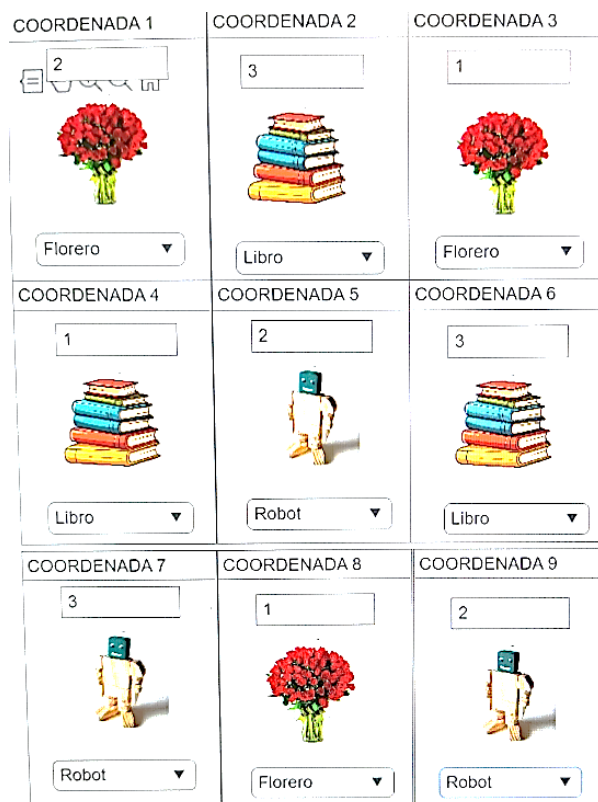


Fig. 5. Etiquetado de objetos identificados

Para el caso expuesto en la Fig. 5 se obtienen los resultados mostrados en la Tabla 2, donde se evidencia un alto reconocimiento de cada objeto.

TABLA 2. PREDICCIÓN DE LA RED

Etiqueta: florero Score: 0.9939	Etiqueta: libro Score: 0.9932	Etiqueta: florero Score: 0.9952
Etiqueta: libro Score: 0.9979	Etiqueta: robot Score: 0.9970	Etiqueta: libro Score: 0.9974
Etiqueta: robot Score: 0.9719	Etiqueta: florero Score: 0.9528	Etiqueta: robot Score: 0.9782

Para el proceso de ordenamiento se emplean colores en el escenario virtual para identificar cada objeto, para el caso se emplea rojo para el florero, verde para los libros y azul para el robot. La Fig. 6. presenta el primer escenario virtual de ordenamiento robótico con la ubicación de los objetos por color.

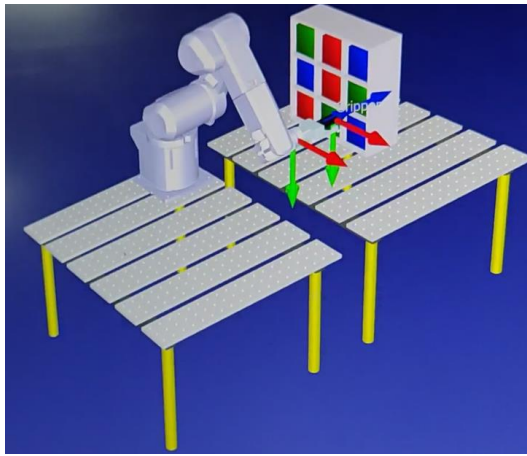


Fig. 6. Escenario virtual de ordenamiento robótico

La Fig. 7. muestra los productos extraídos de la estantería por el brazo robótico.

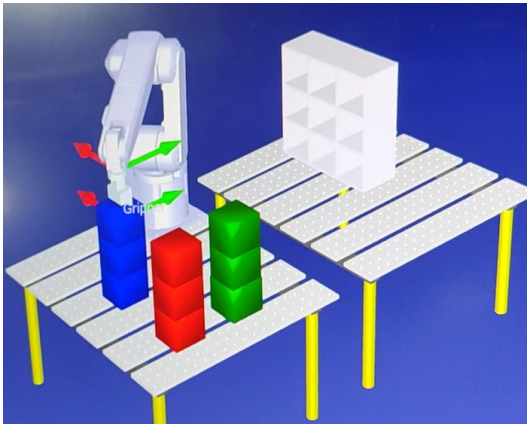


Fig. 7. Vaciado de estantería

Finalmente, la Fig. 8. permite ver los productos reubicados en la estantería tras ser organizados por el brazo robótico según su categoría.

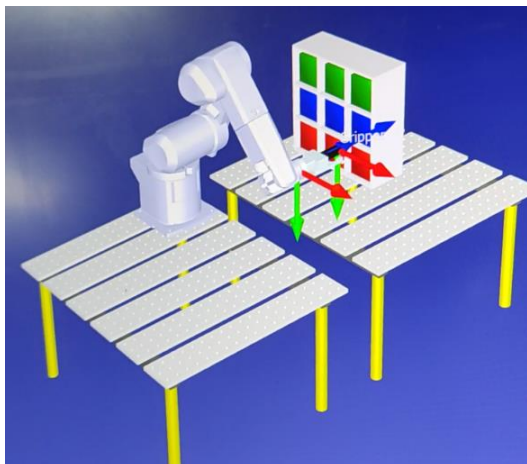


Fig. 8. Reordenamiento por categoría

IV. CONCLUSIONES

El algoritmo de ordenamiento para clasificación de producto mediante aprendizaje profundo desarrollado logra identificar los productos, lo que permite al brazo robótico extraerlos y reubicarlos usando redes neuronales convolucionales por regiones, lo que ayuda a conocer el nivel de inventario de cada producto y a agilizar el alistamiento de productos lo que se traduce en un mejor nivel de servicio al cliente y además permite proteger a los colaboradores al reubicarlos en actividades que no impliquen fatiga física y posturas con un impacto negativo en su salud física.

La propuesta permite identificar los tres productos con una precisión del 95.1%.

Finalmente, como investigación futura se propone estudiar aplicaciones reales de la metodología desarrollada, incluyendo mayor número de categorías de objetos en escenarios con más variables.

AGRADECIMIENTOS

Producto derivado del proyecto de investigación titulado “Mejoramiento de procesos de control de inventario y alistamiento de pedidos en el marco de la industria 4.0”, código INV-ING-3970 financiado por la vicerrectoría de investigaciones de la Universidad Militar Nueva Granada, vigencia 2024.

REFERENCIAS

- [1] S. Huang, B. Wang, X. Li, P. Zheng, D. Mourtzis y L. Wang, «Industry 5.0 and Society 5.0—Comparison, complementation and co-evolution,» *Journal of Manufacturing Systems*, vol. 64, pp. 424-428, 2022.
- [2] Hitachi-UTokyo Laboratory (H-UTokyo Lab.), Society 5.0 A people-centric super-smart society, Bunkyo-ku, Tokyo.; Springer Open, 2020. <https://doi.org/10.1007/978-981-15-2989-4>, p. 18.
- [3] A. Lobbezoo, . Y. Qian y H.-J. Kwo, «Reinforcement Learning for Pick and Place Operations in Robotics: A Survey,» *Robotics*, vol. 10, n° 3, p. 105, 2021. DOI: 10.3390/robotics10030105.
- [4] A. Cordeiro, L. F. Rocha, C. Costa, P. Costa y M. F. Silva, «Bin Picking Approaches Based on Deep Learning Techniques: A State-of-the-Art Survey,» de *2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Santa Maria da Feira, Portugal, 2022. DOI: 10.1109/ICARSC55462.2022.9784795.
- [5] B. Gajšek, . S. Šinko , T. Kramberger, M. Butlewski, E. Özceylan y G. Đukić, «Towards Productive and Ergonomic Order Picking: Multi-Objective Modeling Approach,» *MDPI - Publisher of Open Access Journals*, vol. 11, n° 9, p. 4179, 2021. <https://doi.org/10.3390/app11094179>.
- [6] H. Shin, H. Hwang, H. Yoon y S. Lee, «Integration of deep learning-based object recognition and robot manipulator for grasping objects,» de *16th International Conference on Ubiquitous Robots (UR)*, Korea, 2019. DOI:10.1109/URAI.2019.8768650..
- [7] E. Kee, . J. J. Chong, . Z. J. Choong y M. Lau, «A Comparative Analysis of Cross-Validation Techniques for a Smart and Lean Pick-and-Place Solution with Deep Learning,» *Electronics*, vol. 12, n° 11, p. 2371, 2023. DOI: 10.3390/electronics12112371.

- [8] H. Liu, L. Zhou, J. Zhao, F. Wang, J. Yang, K. Liang y Z. Li, «Deep-Learning-Based accurate identification of warehouse goods for robot picking operations,» *Sustainability*, vol. 14, n° 3, p. 7781, 2022. DOI: <https://doi.org/10.3390/su14137781>.
- [9] S. Lee y Y. Lee, «Real-time industrial bin-picking with a hybrid deep learning-engineering approach,» de *020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Busan, Korea (South), 2020. DOI: 10.1109/BigComp48618.2020.00015..
- [10] P. Jiang, Y. Ishihara, N. Sugiyama, J. Oaki, S. Tokura, A. Sugahara y k. Ogawa, «Depth Image-Based Deep Learning of grasp planning for textureless planar-faced objects in vision-guided robotic bin-picking,» *Sensors*, vol. 20, n° 3, p. 706, 2020. DOI: <https://doi.org/10.3390/s20030706>.
- [11] T.-T. Le y C.-Y. Lin, «Bin-Picking for planar objects based on a deep learning network: a case study of USB packs,» *Sensors*, vol. 19, n° 16, p. 3602, 2019. DOI: <https://doi.org/10.3390/s19163602>.
- [12] S. Soltan, A. Oleinikov, M. Fatih Demirci y A. Shintemirov, «Deep learning-based object classification and position estimation pipeline for potential use in robotized pick-and-place operations,» *Robotics*, vol. 9, n° 3, p. 63, 2020. DOI: <https://doi.org/10.3390/robotics9030063>.
- [13] A. S. Olesen, B. B. Gergaly, E. A. Ryberg, M. R. Thomsen y D. Chrysostomou, «A collaborative robot cell for random bin-picking based on deep Learning policies and a multi-gripper switching strategy,» *Procedia Manufacturing*, vol. 51, pp. 3-10, 2020. DOI: <https://doi.org/10.1016/j.promfg.2020.10.002>.
- [14] H. Sarabu, K. Ahlin y . A.-P. Hu, «Leveraging deep learning and RGB-D cameras for cooperative apple-picking robot arms,» de *ASABE Annual International Meeting*, 2019. DOI:10.13031/aim.201901125.
- [15] G. Li, G. Chesser, Jr., . Y. Huang, Y. Zhao y J. L. Purswell, «Development and optimization of a deep-learning-based egg-collecting robot,» *Transactions of the ASABE*, vol. 64, n° 5, pp. 1659-1669, 2021. DOI: 10.13031/trans.14642.
- [16] W. Xu, L. Zhao, J. Li, S. Shang, X. Ding y T. Wang, «Detection and classification of tea buds based on deep learning,» *Computers and Electronics in Agriculture*, vol. 192, p. 106547, 2022. <https://doi.org/10.1016/j.compag.2021.106547>.
- [17] Y. Ku, . J. Yang, . H. Fang, W. Xiao y J. Zhuang , «Deep learning of grasping detection for a robot used in sorting construction and demolition waste,» *Journal of Material Cycles and Waste Management*, vol. 23, pp. 84-95, 2021. DOI: <https://doi.org/10.1007/s10163-020-01098-z>.
- [18] X. Li, J. Li, C. Cai, B. Zhang y H. Zhou, «Location and Collaboration Robot Arm System via Multi-view,» de *2023 IEEE 9th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, Kuala Lumpur, Malaysia, 2023. DOI: 10.1109/ICSIMA59853.2023.10373487.
- [19] N. S. Patil y M. Jaybhaye, «Autonomous Sorting with 6 DOF Robotic Arm using Machine Vision,» de *International Conference on Network, Multimedia and Information Technology (NMITCON)*, Bengaluru, India, 2023. DOI: 10.1109/NMITCON58196.2023.10276195.
- [20] A. G. Eker, M. K. Pehlivanoglu, İ. İnce y N. Duru, «Deep learning and transfer learning based brain tumor segmentation,» de *2023 8th International Conference on Computer Science and Engineering (UBMK)*, Burdur, Turkiye, 2023. DOI: 10.1109/UBMK59864.2023.10286591..
- [21] B. Kumar, A. K. Singh y P. Banerjee, «A deep learning approach for product recommendation using resNet-50 CNN model,» de *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Coimbatore, India, 2023. DOI: 10.1109/ICSCSS57650.2023.10169441..
- [22] S. Jiang, C. Hua y M. Yuan, «Image classification method of bearing fault based on BOA optimization ResNet-18,» de *2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*, Jinzhou, China, 2023. DOI: 10.1109/ICSECE58870.2023.10263534..
- [23] G. K. Pandey y S. Srivastava, «ResNet-18 comparative analysis of various activation functions for image classification,» de *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023. DOI: 10.1109/ICICT57646.2023.10134464..