

Universidad de Lima  
Facultad de Ingeniería  
Carrera de Ingeniería Civil



# **EVALUACIÓN DE LA EFECTIVIDAD DE LAS TÉCNICAS DE AUMENTO DE DATOS PARA LA DETECCIÓN Y CLASIFICACIÓN DE TRABAJADORES EN OBRAS DE CONSTRUCCIÓN**

Tesis para optar el Título Profesional de Ingeniero Civil

**Eimi Canahualpa Nakamatsu**

**Código 20170260**

**Nelsa Alejandra Pachas Pacheco**

**Código 20171120**

**Asesor**

Ana Felicita Luna Torres

Lima – Perú

Enero de 2024



**EVALUATION OF THE EFFECTIVENESS OF  
DATA AUGMENTATION TECHNIQUES FOR  
THE DETECTION AND CLASSIFICATION  
OF WORKERS ON CONSTRUCTION SITES**

# TABLA DE CONTENIDO

<b>RESUMEN .....</b>	<b>x</b>
<b>ABSTRACT.....</b>	<b>xi</b>
<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>CAPÍTULO I: REVISIÓN DE LITERATURA .....</b>	<b>4</b>
1.1. Construcción 4.0.....	4
1.2. Inteligencia Artificial, Machine Learning y Deep Learning .....	6
1.3. Métricas de evaluación .....	13
1.4. Métodos de aprendizaje: supervisado, no supervisado y semi-supervisado.....	14
1.5. Redes Neuronales de Convolución.....	15
1.6. Algoritmo YOLO .....	18
1.7. Computer Vision Technology .....	20
1.8. Data Augmentation.....	22
<b>CAPÍTULO II: OBJETIVOS E HIPÓTESIS.....</b>	<b>27</b>
2.1 Objetivos.....	27
2.1.1 Objetivo General .....	27
2.1.2 Objetivos Específicos.....	27
2.2 Hipótesis .....	27
<b>CAPÍTULO III: MÉTODO .....</b>	<b>28</b>
3.1. Diseño de investigación.....	28
3.2. Materiales .....	28
3.3 Procedimiento.....	29
3.3.1 Generación de imágenes .....	30
3.3.2 Preparación del conjunto de datos.....	31

3.3.3. Aumento de Datos.....	32
3.3.4 Detección de mano de obra.....	36
<b>CAPÍTULO IV: RESULTADOS Y DISCUSIÓN .....</b>	<b>39</b>
4.1 Resultados de entrenamiento.....	40
4.2 Resultados de validación de cada técnica.....	44
4.2.1 Red neuronal con conjunto de datos original.....	47
4.2.2 Red neuronal con conjunto de datos aumentado con la técnica Brillo-Contraste .....	47
4.2.3 Red neuronal con conjunto de datos aumentado con la técnica Difuminación .....	49
4.2.4 Red neuronal con conjunto de datos aumentado con la técnica Lluvia .....	50
4.2.5 Red neuronal con conjunto de datos aumentado con todas las técnicas combinadas.....	51
4.3 Resultados de validación en una obra distinta.....	53
4.3.1 Red neuronal con conjunto de datos no aumentado.....	54
4.3.2 Red neuronal con conjunto de datos aumentado con la técnica Brillo-Contraste .....	55
4.3.3 Red neuronal con conjunto de datos aumentado con la técnica Difuminación .....	56
4.3.4 Red neuronal con conjunto de datos aumentado con la técnica Lluvia .....	57
4.3.5 Red neuronal con conjunto de datos aumentado con todas las técnicas combinadas.....	58
<b>CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>62</b>
<b>REFERENCIAS.....</b>	<b>65</b>
<b>ANEXOS.....</b>	<b>71</b>

## ÍNDICE DE TABLAS

Tabla 3.1 Conjunto de datos creados .....	33
Tabla 3.2 Hiperparámetros de entrenamiento .....	38
Tabla 4.1 Resultados de entrenamientos de la red neuronal con el conjunto de datos original .....	40
Tabla 4.2 Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con la técnica Brillo-Contraste .....	41
Tabla 4.3 Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con la técnica de Difuminación .....	42
Tabla 4.4 Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con la técnica de Lluvia.....	43
Tabla 4.5 Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con todas las técnicas de Aumento de Datos.....	43
Tabla 4.6 Clasificación de cada trabajador en la obra CEBUL de la Universidad de Lima en Lima, Perú.....	46
Tabla 4.7 Niveles de confianza en cada detección de las redes neuronales durante la etapa de validación en la obra CEBUL de la Universidad de Lima en Lima, Perú .....	53
Tabla 4.8 Clasificación de cada trabajador de una edificación ubicada en el distrito de Miraflores de Lima, Perú .....	54
Tabla 4.9 Niveles de confianza en cada detección de las redes neuronales durante la etapa de validación en una obra de una edificación ubicada en el distrito de Miraflores de Lima, Perú .....	60

## ÍNDICE DE FIGURAS

Figura 1.1 Innovaciones de la Construcción 4.0 con la Inteligencia Artificial .....	6
Figura 1.2 Tendencias de publicaciones realizadas de IA en el sector AEC entre los siglos 20 y 21 .....	7
Figura 1.3 Composición de una ANN .....	10
Figura 1.4 Ecuación de IoU .....	14
Figura 1.5 Aplicaciones y algoritmos de AI, ML y DL.....	17
Figura 1.6 Arquitectura típica de una CNN .....	17
Figura 3.1 Posiciones de las cámaras en el sitio de construcción.....	29
Figura 3.2 Proceso para la detección y clasificación de trabajadores utilizando las técnicas de Aumento de Datos.....	30
Figura 3.3 Imagen etiquetada con las categorías Person y Leaning Person .....	31
Figura 3.4 Imagen original (IMG-680) sin técnica de Aumento de Datos .....	34
Figura 3.5 Imagen aumentada utilizando la técnica de Brillo – Contraste .....	35
Figura 3.6 Imagen aumentada utilizando la técnica de Difuminación .....	35
Figura 3.7 Imagen aumentada utilizando la técnica de Lluvia .....	36
Figura 4.1 Resultados de entrenamiento de las redes neuronales con y sin técnicas de Aumento de Datos .....	44
Figura 4.2 Ejemplo de Bounding Box .....	46
Figura 4.3 Nueva imagen (IMG-44) para la validación de las redes neuronales.....	46
Figura 4.4 Validación de la red neuronal entrenada con el conjunto de datos original..	47
Figura 4.5 Validación de la red neuronal entrenada con el segundo conjunto de datos .	49
Figura 4.6 Validación de la red neuronal entrenada con el tercer conjunto de datos .....	50
Figura 4.7 Validación de la red neuronal entrenada con el cuarto conjunto de datos .....	51
Figura 4.8 Validación de la red neuronal entrenada con el quinto conjunto de datos .....	52
Figura 4.9 Gráfico de máximos niveles de confianza de cada red neuronal .....	53
Figura 4.10 Imagen de una obra de construcción de una edificación ubicada en el distrito de Miraflores de Lima, Perú .....	54
Figura 4.11 Validación de la red neuronal entrenada con el conjunto de datos original en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú .....	55

Figura 4.12 Validación de la red neuronal entrenada con el segundo conjunto de datos en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú .....	56
Figura 4.13 Validación de la red neuronal entrenada con el tercer conjunto de datos en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú .....	57
Figura 4.14 Validación de la red neuronal entrenada con el cuarto conjunto de datos en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú .....	58
Figura 4.15 Resultados de validación de la red neuronal con todas las técnicas de Aumento de Datos en una obra de una edificación ubicada en el distrito de Miraflores de Lima, Perú.....	59
Figura 4.16 Gráfico de máximos niveles de confianza de cada red neuronal en una obra distinta.....	61

## ÍNDICE DE ANEXOS

Anexo 1: Gráficas de Resultados de la Tabla N° 4.1.....	72
Anexo 2: Gráficas de Resultados de la Tabla N° 4.2.....	74
Anexo 3: Gráficas de Resultados de la Tabla N° 4.3.....	76
Anexo 4: Gráficas de Resultados de la Tabla N° 4.4.....	78
Anexo 5: Gráficas de Resultados de la Tabla N° 4.5.....	80

## RESUMEN

Las aplicaciones de la Inteligencia Artificial dentro de la industria de la construcción para la detección de trabajadores, equipos y herramientas requieren de una gran cantidad de datos variados para lograr un alto nivel de precisión (*mean Average Precision*-, mAP) y nivel de confianza en las detecciones y clasificaciones, y evitar los problemas de sobreentrenamientos, los cuales podrían perjudicar a los modelos de aprendizaje profundo. Sin embargo, los conjuntos de datos publicados en línea carecen de una variedad de imágenes que capturen los diferentes escenarios en una construcción, tales como las diversas actividades en obras, los cambios de estación, la variación en la iluminación en el transcurso del día, entre otros. Además, la creación de estos suele conllevar un proceso largo y monótono.

Con el fin de abordar este problema, la presente investigación plantea el uso de 3 técnicas de Aumento de Datos aplicadas en imágenes de construcción reales a fin de generar datos variados y, seguidamente, medir la variación de mAP y nivel de confianza en la detección y clasificación de trabajadores en obras de construcción, y con ello mejorar el desempeño de los modelos de aprendizaje profundo. En ese sentido, las técnicas de Aumento de Datos utilizadas fueron Brillo-Contraste, Difuminación y Lluvia de la galería Albumentations. Para ello, se recolectaron 1212 imágenes provenientes de 4 cámaras ubicadas estratégicamente en diferentes puntos del edificio del Centro de Bienestar Universitario (CEBUL) de la Universidad de Lima en Lima, Perú.

Los resultados de la investigación mostraron que las técnicas de Aumento de Datos propuestas mejoraron el mAP de entrenamiento hasta en un 4%. Además, el uso de todas las técnicas en conjunto mejoró los niveles de confianza hasta en un 9% durante la validación de resultados en una imagen de una obra de construcción distinta. Esta investigación experimental ha demostrado la importancia de considerar las técnicas de Aumento de Datos para detectar y clasificar correctamente trabajadores en la industria de la construcción, lo cual es el primer paso para automatizar de manera certera el monitoreo de trabajadores y diversas aplicaciones futuras.

**Palabras clave:** Inteligencia Artificial, detección y clasificación de objetos, trabajadores en obras de construcción, monitoreo basado en la visión, Aumento de Datos.

## ABSTRACT

The applications of Artificial Intelligence within the construction industry for the detection of workers, equipment and tools require a large amount of varied data to achieve a high mean Average Precision (mAP) and confidence level in detections and avoid overfitting, which could harm the deep learning models. Unfortunately, publicly existing data sets lack a variety of images that capture the different scenarios in a building, such as the various activities in construction, the changes of season, the variation in lighting over the course of the day, among others. In addition, the creation of these usually entails a long and monotonous process.

Therefore, to address this problem, this investigation proposes the use of 3 Data Augmentation techniques applied on real construction images to generate varied data and, subsequently, measure the variation of mAP and confidence level in detections and classification of workers in construction sites in order to achieve the best performance from deep learning models. In that sense, the Data Augmentation techniques used were “Brightness-Contrast”, “Blur” and “Rain” from the Albumentations gallery. For this study 1212 images were collected from 4 cameras strategically located at different points in the building of the Centro de Bienestar Universitario (CEBUL) of the University of Lima in Lima, Peru.

The research results showed that the proposed Data Augmentation techniques improved training mAP by up to 4%. In addition, the use of all the techniques together obtained better levels of confidence by up to 9% during the validation of results in an image of a different construction site. This experimental research has shown the importance of considering Data Augmentation techniques to correctly detect and classify workers in the construction industry, which is the first step to accurately automate worker monitoring and various future applications.

**Keywords:** Artificial Intelligence, object detection and classification, workers in construction sites, vision-based monitoring, Data Augmentation.

# INTRODUCCIÓN

Desde hace varios años el monitoreo de los sitios de construcción es una actividad fundamental para los gerentes de proyectos para el análisis de la productividad y la gestión de la seguridad (Yang et al., 2015, Kim et al., 2016, Yang et al., 2016, Ding et al., 2018, Fang et al., 2018, Fang et al., 2018, Luo et al., 2018, Luo et al., 2018, Kim et al., 2018, Luo et al., 2019, Seo et al., 2019 Kim et al., 2020). Tradicionalmente, las formas de monitoreo se llevan a cabo visitando las construcciones en persona. Sin embargo, hoy en día estas construcciones son más complejas y dinámicas, por lo que el seguimiento en persona requiere de mucho tiempo (Bohn et al., 2010, Seo et al., 2015, Teizer et al., 2015, Yang et al., 2015, Ham et al., 2016., Zhong et al., 2019, Kim et al., 2020, Kim et al., 2020, Xiao et al., 2021). Por lo tanto, diversos estudios se han enfocado en investigar sistemas de monitoreo automatizados (Kim et al., 2020). Uno de los sistemas más utilizados es el basado en el Internet de las Cosas (IoT), el cual consiste en conectar sensores a los objetos de construcción para analizar los movimientos físicos y evaluar el desempeño operativo (Montaser et al., 2014 & Nath et al., 2017). A pesar de los resultados favorables, existen limitaciones en la aplicación de ese tipo de sistema; por ejemplo, cuando los sitios de construcción son complejos y tienen numerosos objetos de construcción lo cual hace imposible conectar sensores a todos los tipos de objetos como equipos, herramientas y personas (Park et al., 2012 y Cheng et al., 2017). En ese sentido, el monitoreo basado en la visión por computadora está tomando lugar en el campo de los sistemas de monitoreo automatizados. Debido a que las cámaras son equipos de bajo costo, estas se integran fácilmente en las obras de construcción y ofrecen un amplio rango de visión, por lo que son adecuadas para el monitoreo en dichas obras de construcción (Xiao et al., 2020 y Xiao et al., 2021).

Asimismo, con el fin de utilizar las imágenes del sitio de construcción en los sistemas de monitoreo automatizados, primero es fundamental desarrollar el reconocimiento automático y certero de los recursos operativos de construcción, tales como los trabajadores, equipos y materiales (Xiao et al., 2020 y Xiao et al., 2021); una vez realizado esto las tareas de monitoreo de la construcción se pueden automatizar. En cuanto a la detección de trabajadores de la construcción, este es un paso esencial en los sistemas de monitoreo basados en la visión por computador para controlar el avance en las obras de construcción, que por lo general requieren de una supervisión de estas in situ.

Por otro lado, el entrenamiento para la detección de trabajadores utilizando los métodos de Aprendizaje Profundo requieren grandes conjuntos de datos de imágenes de construcción. Sin embargo, no existe una gran variedad de base de datos de obras de construcción publicados en línea debido a la dificultad y alto costo que representa elaborarlos y etiquetarlos manualmente (Bang et al., 2020; Xiao et al., 2020). Además, el entrenamiento de los métodos de Aprendizaje Profundo requiere de una gran variedad de imágenes para evitar problemas de sobreentrenamiento de los conjuntos de datos.

En ese sentido, diversos estudios en diferentes áreas han demostrado que el uso de las técnicas de Aumento de Datos aplicados en la visión por computador puede resolver los problemas de sobreentrenamiento al aumentar el volumen y la diversidad de los datos, transformando las imágenes y las etiquetas correspondientes (Kim and Jeong, 2020; Prusty et al., 2021; Bang et al., 2020; Qin et al., 2021; Dung et al., 2019). Sin embargo, actualmente existen pocos estudios en la industria de la construcción que evalúen la efectividad en la adopción de estas técnicas en la detección y clasificación de trabajadores en obras de construcción. Con el fin de abordar este problema, en el presente trabajo se evaluó el desempeño en el uso de 3 técnicas de Aumento de datos aplicadas en imágenes de construcción reales con el fin de determinar los resultados de *mean Average Precision* (mAP) y niveles confianza de cada una de las técnicas y compararlos con los resultados del entrenamiento utilizando solo las imágenes originales.

La estructura del presente estudio es la siguiente: primero, en el capítulo I se presenta una revisión de la literatura sobre la inteligencia artificial y sus diversas aplicaciones dentro de la industria de la construcción. Luego, en el capítulo II se presentan los objetivos e hipótesis. En el capítulo III, se explica detalladamente el método de estudio, el cual incluye el diseño de la investigación, los materiales requeridos y el procedimiento realizado. Posteriormente, en el capítulo IV se comparan los resultados obtenidos al utilizar las imágenes originales y cada una de las técnicas de Aumento de Datos seleccionadas. Finalmente, en el capítulo V se exponen las conclusiones, recomendaciones del estudio y los trabajos futuros.

# CAPÍTULO I: REVISIÓN DE LITERATURA

## 1.1. Construcción 4.0

La cuarta revolución industrial, denominada con el término Industria 4.0, es definida como una transformación digital impulsada por las Tecnologías de la Información (TI), la cual tiene como objetivo mejorar la correlación entre la máquina-humano para lograr avances en la auto-optimización de los sistemas (Aichholzer et al., 2015). De igual manera, Cañas et al. (2021, p.11) definen a la Industria 4.0 como “sistemas de producción flexibles, los cuales con o sin interacción humana, son capaces de clonar los conocimientos y experiencia humana utilizando algoritmos, modelos, heurísticas, metaheurísticas, matemáticas e hiper heurísticas para permitirse a sí mismos ser autoconfigurables, auto-mantenibles, autoconscientes, autodiseñados, sustentables y a la vez inteligentes para tomar decisiones”.

Estos conceptos se basan en ciertos principios de esta industria identificados por distintos autores (Cañas et al., 2021; Dikhanbayeva et al., 2020; Hermann et al., 2016):

- Interconexión o Interoperabilidad, la cual se refiere a la capacidad de los dispositivos, personas, equipos y otros tipos de objetos de compartir información mediante el uso de diferentes herramientas como el Internet de las cosas (IoT), Internet de las personas (IoP) e Internet de todo (IoE), los cuales permiten una comunicación sistemática entre humanos con humanos, máquinas con humanos y máquinas con máquinas.
- Transparencia de la información, la cual establece que toda información del mundo real y virtual deberá ser accesible a todos sujetos que la requieran para cumplir con sus tareas y tomar decisiones adecuadas.
- Virtualización, qué es la capacidad de los sistemas de convertir la información del mundo real en virtual, lo cual mejoraría la percepción de los sujetos en tomar decisiones analizando distintos escenarios posibles. Esta información puede enriquecerse mediante el uso del sistema de realidad aumentada y realidad virtual.
- Decisiones descentralizadas, la cual describe la capacidad de los sistemas de realizar tareas y tomar decisiones autónomas mediante la aplicación de

tecnologías como los Sistemas ciber físicos (CPS), IoT, actuadores y sensores inteligentes, la inteligencia artificial, entre otros.

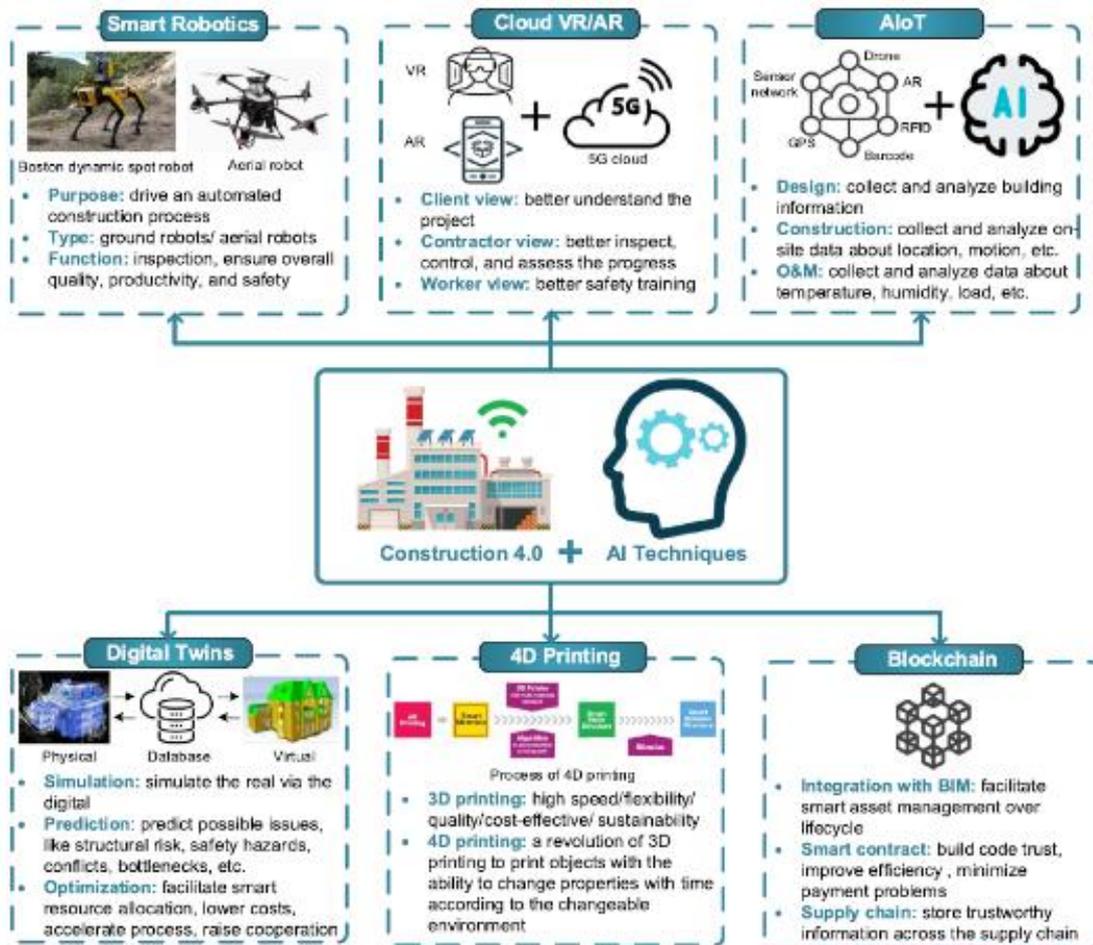
- Asistencia técnica, la cual hace referencia al objetivo de cada sistema, que es el de realizar tareas complejas, resolver problemas y tomar decisiones de manera autónoma que sirvan de soporte a las personas.
- Capacidad en tiempo real, la cual se atribuye a la capacidad de cada sistema de recolectar, transferir, analizar, monitorear y compartir información en tiempo real con el fin de acelerar los procesos de producción, identificar anomalías, identificar oportunidades de mejora y observar tendencias.

Dentro del sector de la construcción este último término es nombrado Construcción 4.0, cuya definición hace referencia a los nuevos desarrollos que tienen como fin evolucionar el sector, incluidas sus prácticas y técnicas, tales como los monitoreos manuales y las tareas de construcción reiterativas (Turner, 2020).

En los últimos años se han desarrollado diversas innovaciones por parte de empresas emergentes, capitalistas e investigadores académicos, los cuales han buscado realizar una fuerte contribución al sector de la construcción con la aplicación de una variedad de Tecnologías de la Información y Automatización (Sacos, 2020). En ese sentido, algunas innovaciones desarrolladas incluyen el uso de la robótica para automatizar procesos de construcción, el desarrollo de dispositivos con tecnología de Internet de las Cosas (IoT), tales como sensores inteligentes en herramientas, equipos de protección personal o maquinarias; la creación de vehículos autónomos que permiten que las maquinarias de construcción operen por sí mismas y con una mejor comunicación; el uso de drones con fines de monitoreo, levantamientos topográficos y modelado de información de construcción (BIM); el desarrollo de tecnologías inmersivas, representadas por la Realidad Virtual (VR) y la Realidad Aumentada (AR), los cuales simulan espacios de construcción reales; entre otros. (Okpala et al., 2020; Mohammadpour et al., 2019). En la Figura 1.1 se describe un resumen de las innovaciones de la Construcción 4.0 con el uso de la Inteligencia Artificial.

**Figura 1.1**

*Innovaciones de la Construcción 4.0 con la Inteligencia Artificial*



*Nota. De Roles of artificial intelligence in construction engineering and management: A critical review and future trends, por Y. Pan, L. Zhang, 2021, Automation in Construction, 122 (<https://doi.org/10.1016/j.autcon.2020.103517>).*

**1.2. Inteligencia Artificial, Machine Learning y Deep Learning**

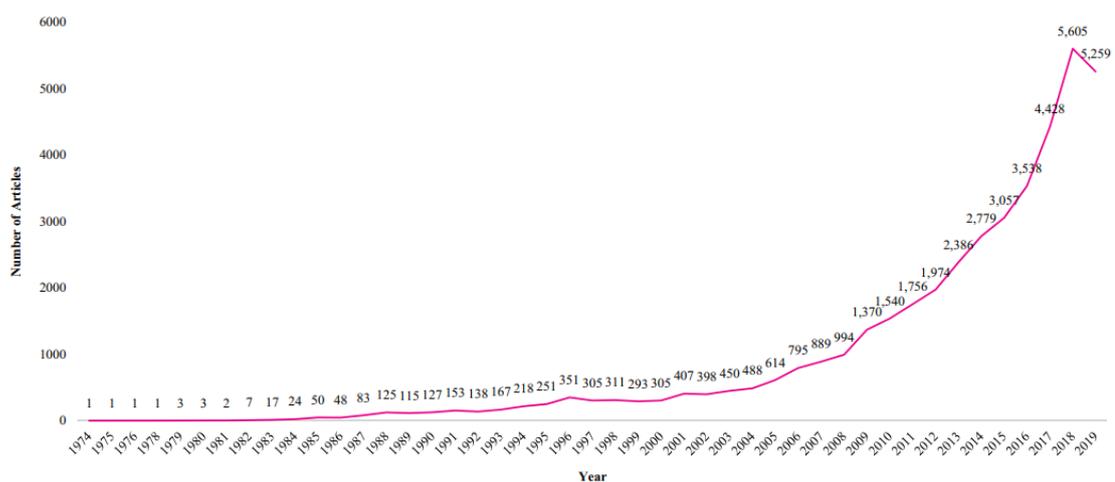
La Inteligencia Artificial (IA) hace referencia a la capacidad de cualquier dispositivo o sistema de recibir información del entorno, procesarla y transformarla internamente, y realizar acciones internas para lograr con éxito sus objetivos impuestos (Mohammadpour et al., 2019). En los últimos años, gran cantidad de las innovaciones en desarrollo se han apoyado en la aplicación de técnicas de IA para lograr su evolución, ya que permite el uso de grandes cantidades de datos y su procesamiento informático para realizar tareas que, de otra forma, exigirían una gran experiencia e inteligencia humana (Okpala et al.,

2020). Asimismo, la IA se encarga de impulsar a las computadoras a aprender como un ser humano y desarrollar habilidades como la percepción, la representación del conocimiento, el razonamiento, la resolución de problemas y la planificación para abordar situaciones complicadas de manera inteligente (Pan & Zhang, 2021).

Como mencionado, uno de los principales propósitos de la IA es optimizar procesos, decisiones, problemas y actividades; por lo tanto, esta ha sido ampliamente utilizada en diferentes industrias como la agrícola, financiera, salud y bancaria (Mohammadpour et al., 2019). Dentro del sector de la arquitectura, ingeniería y construcción (AEC), las investigaciones relacionadas a las aplicaciones de IA tuvieron sus inicios en el año 1974 y fueron incrementando gradualmente con el tiempo, como se muestra en la Figura 1.2 (Darko et al., 2020). En comparación con el siglo 20, en el presente siglo se muestra un mayor incremento en el número de investigaciones realizadas por el sector. Este incremento es explicado por Bilal et al. (2016) como la necesidad de la industria de procesar una gran cantidad de información heterogénea existente para obtener de ella nuevos conocimientos que ayuden a una mejor toma de decisiones y mejoren el estado del arte existente sobre un problema en común.

**Figura 1.2**

*Tendencias de publicaciones realizadas de IA en el sector AEC entre los siglos 20 y 21*



*Nota.* De *Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities*, por A. Darko, A. P. C. Chan, M. A. Adabre, D. J. Edwards, M. R. Hosseini, E. E. Ameyaw, 2021, *Automation in Construction*, 112 (<https://doi.org/10.1016/j.autcon.2020.103081>).

De acuerdo a las investigaciones realizadas, el uso de la IA en el sector AEC tiene el potencial de incrementar el número de actividades automatizadas e integrar diferentes fases de un proyecto durante todo su ciclo de vida, recopilando y organizando datos para un equipo de proyecto, brindando rápidamente información sobre los datos recogidos que un profesional de la construcción realizaría en mayor tiempo y procesaría con menor precisión debido a los errores humanos (Mohammadpour et al., 2019). Entre otros beneficios específicos, se incluye la supervisión automatizada de actividades, la mitigación de riesgos mediante la predicción de problemas críticos, el aumento de la eficiencia mediante la optimización de actividades repetitivas, y el aprovechamiento de imágenes tomadas en sitios de construcción reales para la producción de nueva información mediante su procesamiento utilizando la visión por computador (Pan & Zhang, 2021).

Dentro de los subcampos de la IA encontramos al Aprendizaje Automatizado, conocido como *Machine Learning* (ML), el cual se define como cualquier sistema que utiliza modelos y algoritmos matemáticos y estadísticos para lograr el aprendizaje y predicción a través de la entrada de datos o de su experiencia, al igual que un ser humano (Mohammadpour et al., 2019). Actualmente, los algoritmos de ML son aplicados en el aprendizaje de una gran cantidad de datos de distintas fuentes para el descubrimiento de conocimientos ocultos, los cuales son de gran interés en los softwares de gestión de proyectos para facilitar el análisis automático de la información y la toma de decisiones (Pan y Zhang, 2021)

En esta línea de investigación, el Aprendizaje Profundo o *Deep Learning* (DL), que es una forma especializada de aprendizaje del ML, está compuesto de estructuras conocidas como Redes Neuronales Artificiales o *Artificial Neural Networks* (ANN), las cuales permiten que el sistema aprenda por sí mismo a través de operaciones complejas, lo cual no podría ser logrado por una red neuronal tradicional del ML. Generalmente, las ANN suelen trabajar con grandes cantidades de datos como imágenes, videos y audios debido a la presencia de más de una capa oculta de neuronas artificiales en su composición, como se muestra en la Figura 1.3 (Baduge et al., 2022).

En la Figura 3.1 se observa la estructura básica de una ANN, conocida como el algoritmo *Multi-Layer Perceptrons* (MLP), la cual está compuestas por tres tipos de capas, la capa de entrada, una o más capas ocultas y la capa de salida, en donde los círculos representan las neuronas. Cada capa del MLP contiene una o más neuronas, las

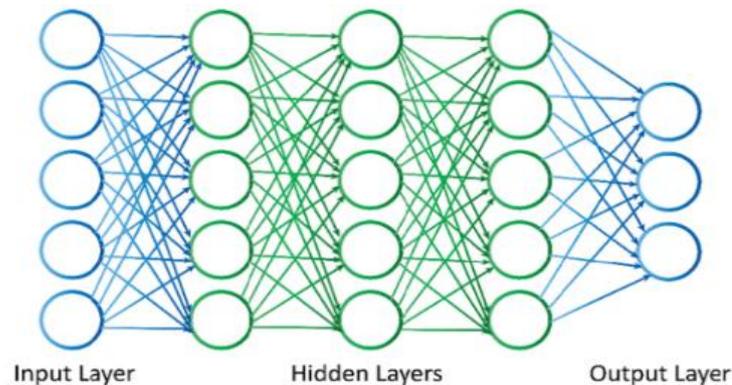
cuales están interconectadas con las neuronas de las capas previas o subsiguientes. La primera capa recibe la información de entrada o inicial para la construcción de los vectores de características de la data introducida. Luego, esta información es procesada por cada neurona de las capas ocultas, las cuales calculan los pesos de entrada de las neuronas previas. Los pesos de entrada representan el nivel de intensidad entre cada neurona interconectada. Además, cuanto mayor sea la cantidad de capas ocultas, mejores resultados de clasificación se obtendrán. Seguidamente, a través de las funciones de activación como ReLu, la función Gaussiana o Sigmoidea, las cuales se encargan de transmitir el resultado generado por la combinación lineal de los pesos y las entradas, las neuronas ocultas transfieren la información a la capa de salida, que representa el resultado visible de la red (Baduge et al., 2022).

Una vez realizado el entrenamiento, los resultados obtenidos de la red neuronal son comparados con los resultados reales esperados. Luego, el nivel de pérdida o error es calculado a través de funciones como pérdida logarítmica, error cuadrático medio, error absoluto medio, entre otros. A través de los errores calculados, se procede a obtener la función de costo, la cual deberá ser minimizada ajustando los pesos de entrada de cada neurona mediante el método backpropagation. El método backpropagation se encarga de calcular la gradiente entre los pesos y errores procesados, y utiliza algoritmos de optimización como Adadelta, Adam y Descenso de gradiente para calcular nuevos pesos que minimicen la pérdida. Finalmente, el mismo conjunto de datos es procesado varias veces ajustando en cada iteración los pesos de cada data de entrada para obtener un modelo entrenado con un error mínimo (Baduge et al., 2022).

Por lo tanto, el DL es un método efectivo de aprendizaje para la resolución de problemas complicados dentro de la industria AEC (Sircar et al., 2021). De esta forma, se ha clasificado las aplicaciones de IA, ML y DL en la industria en 7 categorías, según Kristombu et al. (2022):

### Figura 1.3

#### Composición de una ANN



Nota. De *Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications*, por S. K. Baduge, S. Thilakarathna, J. S. Perera, M. Arashpour, P. Sharafi, B. Teodosio, A. Shringi, P. Mendis, 2022, *Automation in Construction*, 141 (<https://doi.org/10.1016/j.autcon.2022.104440>).

- Diseño arquitectónico y visualización

El diseño arquitectónico dentro de la construcción es un proceso complejo que requiere de *expertise*, conocimientos teóricos previos y creatividad por parte de los arquitectos. Por lo tanto, la aplicación de técnicas de DL ha sido ampliamente estudiado para la construcción de diseños innovadores tanto en planos 2D como en 3D. Estas técnicas, conocidas como *Generative Adversarial Networks* (GAN) y *Variational autoencoders* (VAE), han revolucionado el diseño generativo automatizado de los elementos arquitectónicos, tales como la construcción de masas, la generación de planos de planta y de interiorismo desde videos, la generación de formas para la construcción urbana y diseños arquitectónicos conceptuales, el reconocimiento automático de espacios, la generación de elementos arquitectónicos decorativos en 3D, entre otros.

- Diseño y optimización de materiales

El uso de los materiales adecuados para la construcción de una estructura es un aspecto importante por su comportamiento, tiempo y costo para la construcción. En los últimos años, se han desarrollado diversas investigaciones sobre la optimización en el uso de distintos materiales y su comportamiento, entre los más investigados se encuentra el concreto, la

madera y el acero. En ese sentido, se han construido algoritmos de ML y DL que asistan en los procesos de predicción utilizando información previamente recolectada. Estas predicciones generalmente están enfocadas en el cálculo del costo del proyecto, información sobre el comportamiento de la estructura como la resistencia a la compresión y corte, y el comportamiento ambiental como la cantidad de carbono y emisiones incorporadas.

- Diseño y análisis estructural

Dentro del diseño y análisis estructural, la IA se ha usado como soporte para la resolución de incertidumbres en los análisis estadísticos y teorías probabilísticas del diseño estructural. Dentro del área sismológica, la IA ha ayudado a revelar ocurrencias sísmicas que están por debajo de los niveles de detección regulares, a través de datos extraídos de información de difícil procesamiento. Además, la IA ha servido como apoyo en el análisis de pandeo y fatiga de componentes estructurales con geometría irregular, la predicción de la capacidad de carga y nivel de daño en estructuras existentes, y el diseño y análisis estructural automatizado de componentes pretensados.

- Fabricación y automatización ex situ

Actualmente, la fabricación ex situ es ampliamente usada para la construcción de paneles o módulos prefabricados que luego son transportados y ensamblados in situ. En los últimos años, esta práctica se ha relacionado estrechamente con la automatización de los procesos de fabricación de los componentes a través del uso de diversidad innovaciones, prácticas y tecnologías como los robots inteligentes, la IA, las herramientas BIM, los principios de *Lean Construction* y los gemelos digitales.

- Monitoreo de la seguridad, el progreso y la gestión de la construcción

Gran parte de las actividades de construcción que se realizan de manera tradicional, como el monitoreo del progreso diario, el control de la seguridad y la gestión de los proyectos a través del planeamiento, la coordinación, la supervisión y el control del presupuesto y cronograma, pueden ser automatizadas mediante la aplicación de técnicas de IA para mejorar la eficiencia y precisión. Dentro de la gestión de proyectos, las técnicas de IA han ayudado en la optimización de los costos de los mismos, la predicción

del éxito de los estos basándose en distintos factores como la productividad de los trabajadores, y la reducción de los riesgos asumidos mediante el pronóstico del índice de costos y la evaluación de subcontratistas. De igual manera, se ha mejorado el monitoreo de la seguridad en obra mediante la predicción de futuros accidentes y la evaluación de las condiciones de seguridad de cada actividad en tiempo real. Por último, el monitoreo del progreso de la construcción se ha optimizado mediante diversas técnicas de IA como la predicción de la productividad de los trabajadores y máquinas para el pronóstico del progreso.

- Mantenimiento y operación de construcciones inteligentes

Se define a una construcción inteligente como un ambiente en donde se ha logrado la optimización en la construcción de la estructura, en la operación de los servicios que lo conforman y en el mantenimiento de sus sistemas. Respecto al mantenimiento y operación, dispositivos con tecnología IoT son utilizados para la resolución de problemas y toma de decisiones por parte de la IA a través de la visualización, la recolección, el almacenamiento y el análisis de los datos. Actualmente, el objetivo principal del uso de la IA y los dispositivos IoT en las construcciones inteligentes está enfocado en la optimización del uso energético y mejora de la seguridad.

- Sostenibilidad, análisis del ciclo de vida y circularidad

En una economía circular, la IA es capaz de proporcionar los criterios para la recirculación de los materiales a través del análisis de los componentes y condiciones de cada uno, y generando recomendaciones para su reutilización, reventa, reparación o reciclaje. Asimismo, permite la automatización del proceso de reciclaje de desperdicios, la cual aumenta la capacidad de recuperación de materiales a un menor costo. Para ello, se utilizan diversos sensores y cámaras que detectan los materiales y sus componentes, y entregan la información al software de control.

A pesar de todos los beneficios mencionados y estudios existentes de la IA en el sector de la construcción y el acceso a las nuevas tecnologías, la industria carece enormemente de desarrollo en esta área, donde los temas con mayor relevancia e implementación han sido el BIM, el *Cloud Computing* y la Modularización. Sin embargo,

tecnologías como la Realidad Aumentada, Virtual y Mixta, la Visión por Computador, el uso de Gemelos Digitales y la robótica presentan un desarrollo lento dentro de la industria (Schönbeck, Löfsjögård y Ansell, 2021; Salah Alaloul et al., 2020)

### 1.3. Métricas de evaluación

El *mean Average Precision* (mAP) es una métrica que evalúa el rendimiento de la detección del modelo de DL basado en parámetros como los valores obtenidos de Precisión (*Precision*) y Recuperación (*Recall*) (Lin et al., 2014; Kuznetsova et al., 2018). El valor del mAP se obtiene calculando, primeramente, ambos parámetros mencionados, como se muestra en las ecuaciones uno (1) y dos (2). Luego, la Precisión Promedio (AP) es calculada promediando la precisión en diferentes niveles de Recuperación para cada clase de objeto, como se muestra en la ecuación tres (3). Finalmente, el mAP es calculado por medio de la AP de todas las clases de objetos, como se muestra en la ecuación cuatro (4) (Xiao et al., 2021). Otra métrica utilizada es el valor-F1 (*F1-score*), el cual es representado como el promedio armónico de los valores de Precisión y Recuperación, como presentado en la ecuación (5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{AP} = \frac{1}{11} \sum_{\text{recall} \in \{0,0.1,0.2,\dots,1\}} \text{Precision@recall} \quad (3)$$

$$\text{mAP} = \frac{1}{k} \sum \text{AP} \quad (4)$$

$$\text{F1-score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

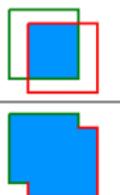
Donde, TP significa Verdadero Positivo, el cual es el número de detecciones correctas, FP significa Falso Positivo, el cual es el número objetos que han sido detectados incorrectamente como positivos y FN significa Falso Negativo, que es el número de objetos positivos que han sido identificados como negativos. Asimismo, la constante  $k$  es el número de clases de objetos predefinidos.

En ese sentido, se dice también que una detección correcta ocurre cuando se tiene un valor de *Intersection over Union* (IoU) mayor que 0.5. Dentro del campo de la detección de objetos, se define como IoU a una medida basada en la intersección de las áreas entre el *Bounding Box* detectado y el *Bounding Box* real de los objetos, dividido

entre la suma de áreas resultado de la unión de ambos *bounding box*, como se muestra en Figura 1.4 (Padilla et al., 2020).

#### Figura 1.4

*Ecuación de IoU*

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{área de superposición}}{\text{área de unión}}$$


*Nota.* De *A survey on Performance Metrics for Object – Detection Algorithms*, por R. Padilla, S. L. Netto, E. A. B. da Silva, 2020, *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* ([10.1109/IWSSIP48289.2020](https://doi.org/10.1109/IWSSIP48289.2020)).

#### 1.4. Métodos de aprendizaje: supervisado, no supervisado y semi-supervisado

El ML aborda una amplia gama de técnicas de aprendizaje utilizando tres métodos: el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje semi supervisado. Respecto al aprendizaje supervisado o aprendizaje discriminativo, este es un método en el que se brinda al sistema un conjunto de datos de entrada con la información de salida requerida con el fin de entrenar sus patrones de clasificación. Es decir, se proporciona un conjunto de datos etiquetados que están siempre disponibles de manera directa o indirecta para el sistema (Ahmad, Farman y Jan, 2019).

Por otro lado, el aprendizaje no supervisado o aprendizaje generativo, es un método cuyo criterio es crear una correlación entre el conjunto de datos de entrada proporcionados con el fin de analizar los patrones existentes, sin ninguna información de salida disponible. Por último, el aprendizaje semi supervisado o aprendizaje híbrido, es una combinación de los componentes de ambos métodos: supervisado y no supervisado, donde se logra la clasificación de la información con el apoyo de los resultados obtenidos del aprendizaje no supervisado (Ahmad, Farman y Jan, 2019).

Algunos estudios realizados en el sector de la construcción respecto al uso de los métodos de aprendizaje incluyen la destacada investigación realizada por Xiao et al. (2021), en la cual se hizo uso de la técnica de aprendizaje semi supervisado, para lo cual se desarrolló un método basado en la red neuronal *Teacher-Student* para la transferencia

de datos y el *Data Augmentation* para la generación de nuevas imágenes que incrementen el conjunto de datos original. En esta investigación se aplicó la detección a maquinarias de construcción, las cuales incluyeron grúas torre, excavadoras, camiones de residuos, niveladoras, compactadoras, mezcladoras de concreto, entre otras. Según los resultados obtenidos por los investigadores, el método desarrollado logró alcanzar un mAP de 92.7% utilizando solo el 50% de las imágenes etiquetadas y el resto sin etiqueta, a comparación del método supervisado, el cual alcanzó un mAP de 89.1% utilizando solo las imágenes etiquetadas. Este estudio nos indica que tanto los métodos de aprendizaje supervisado y semi supervisado pueden lograr resultados favorables para la detección de maquinarias utilizando un *dataset* amplio en la etapa de entrenamiento.

### **1.5. Redes Neuronales de Convolución**

Dentro de los tipos de ANN, encontramos varios tipos de arquitectura tales como el Perceptrón Multicapa (MLP), las Redes Adversas Generativas (GAN), las Redes Neuronales Recurrentes (RNN), las Redes Neuronales de Convolución (CNN), las Redes de Memoria a largo y corto plazo (LSTM) y las Redes de Función de Base Radial (RBFN). Las principales arquitecturas utilizadas pertenecientes a la IA, ML y DL se resumen en la Figura 1.5

Respecto a las CNN, estas son las más usadas dentro de los algoritmos de aprendizaje supervisado de DL, ya que son capaces de identificar automáticamente las características más relevantes de los datos de entrada por sí misma (Alzubaidi et al., 2021). Este tipo de arquitectura es principalmente utilizado para el procesamiento de imágenes y generalmente cuenta con 3 tipos de capas: la Capa de Convolución o *Convolutional Layer* (CL), la Capa de Agrupación o *Pooling Layer* (PL) y la capa totalmente conectada o *Fully Connected Layer* (FCL), las cuales se ordenan una tras otra, respectivamente, como se muestra en la Figura 1.6. Asimismo, debido a las técnicas de reducción de neuronas utilizadas, las CNN requieren de menos preprocesamiento en relación con otros algoritmos de clasificación. (Mathew, Arul y Sivakumari, 2021)

Las CNN tienen una amplia gama de aplicaciones en la clasificación de imágenes, el reconocimiento de acciones, el monitoreo de objetos, el etiquetado de diferentes escenas, el pronóstico de series de tiempo, la detección de textos y objetos, entre otras (Gu et al., 2018). Recientemente, en la industria AEC se ha desarrollado y aplicado

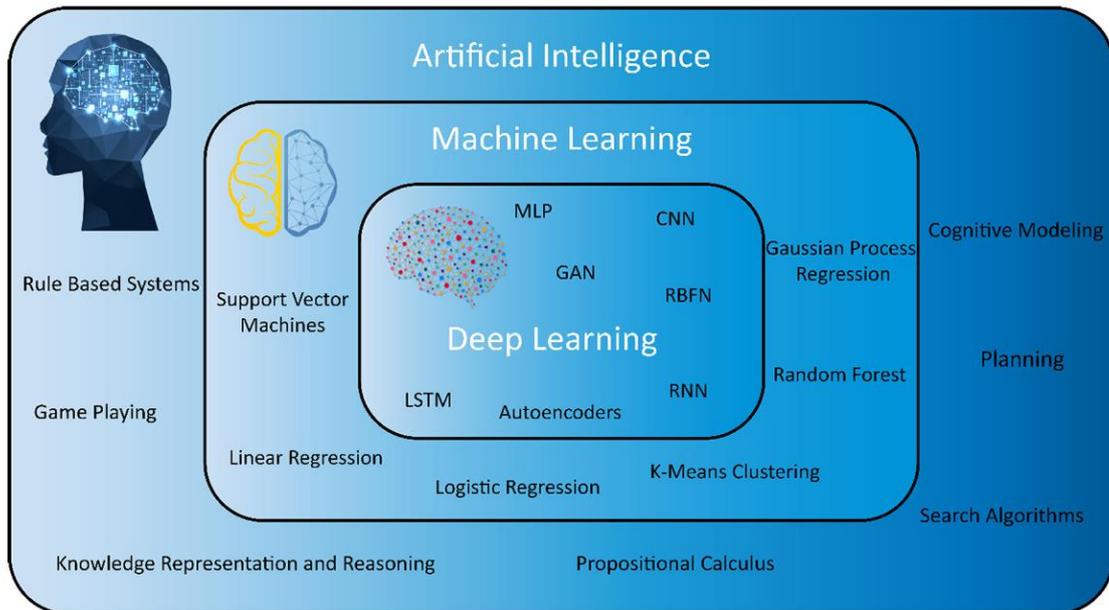
diversos algoritmos de CNN como métodos basados en la visión y aprendizaje para la resolución de problemáticas comunes como: la detección de los daños en una estructura, la gestión y operación de instalaciones, el monitoreo de la seguridad en sitios de construcción, el monitoreo de la salud estructural, estimación de la fuerza de compresión del concreto, entre otras. Entre las problemáticas nombradas, la detección de daños en estructuras ha recibido mayor atención. Asimismo, estos estudios se han centrado en los daños presentes en estructuras de concreto, particularmente en las fallas de dichas estructuras.

En ese sentido, en el estudio realizado por Cha et al. (2017), los investigadores propusieron una técnica de Visión por Computador empleando una arquitectura de CNN para la detección de fallas en estructuras de concreto de túneles. La evaluación de los resultados se realizó mediante un análisis comparativo en el que se demostró que la aplicación de la técnica propuesta tiene mejores resultados que los métodos tradicionales de detección de fallas en situaciones reales. Otro estudio realizado por Modarres et al. (2018), presentó un nuevo método basado en CNN para la detección de fallas en puentes de concreto. El resultado de mAP obtenido por el método presentado fue comparado con otros algoritmos de ML, a través de lo cual se demostró una mejor precisión en la detección de fallas con el método propuesto de CNN. Estos estudios demuestran la capacidad de las CNN de convertirse en una técnica efectiva y poderosa tanto para la detección de daños estructurales como la detección de maquinarias y mano de obra.

Por otro lado, el rendimiento de las CNN en las aplicaciones en la industria AEC depende de la cantidad de datos usados en la etapa de entrenamiento. Es decir, se requiere de una amplio y variado dataset en los entrenamientos para evitar problemas de sobreentrenamiento. Dentro de la industria, este problema puede ser abordado utilizando diversos métodos de recolección de datos y enriqueciendo las CNN con tecnología IoT, el uso de los drones, los escáneres láser, el Modelado de la Información de la Construcción o *Building Information Modeling* (BIM) y los métodos de minería de datos (Darko et al., 2020).

**Figura 1.5**

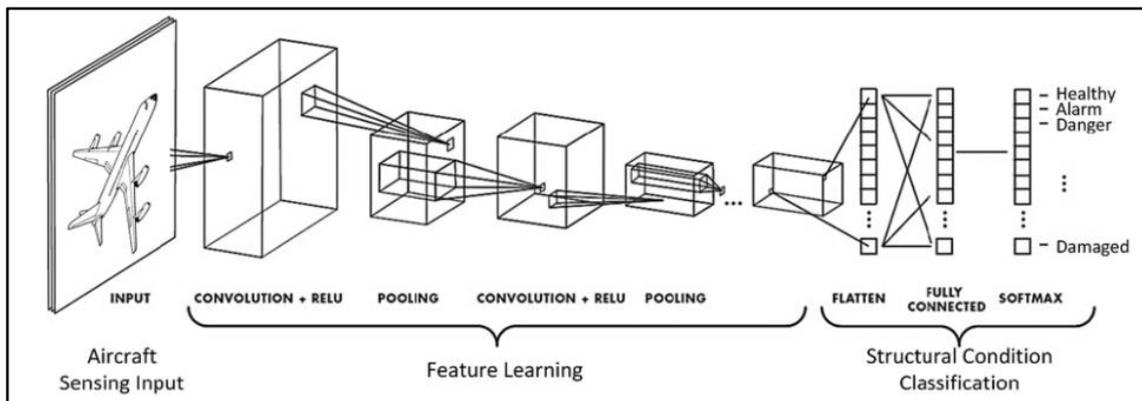
*Aplicaciones y algoritmos de AI, ML y DL*



*Nota. De Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications, por S. K. Baduge, S. Thilakarathna, J. S. Perera, M. Arashpour, P. Sharafi, B. Teodosio, A. Shringi, P. Mendis, 2022, Automation in Construction, 141 (https://doi.org/10.1016/j.autcon.2022.104440).*

**Figura 1.6**

*Arquitectura típica de una CNN*



*Nota. De Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications, por S. K. Baduge, S. Thilakarathna, J. S. Perera, M. Arashpour, P. Sharafi, B. Teodosio, A. Shringi, P. Mendis, 2022, Automation in Construction, 141 (https://doi.org/10.1016/j.autcon.2022.104440).*

## 1.6. Algoritmo YOLO

El algoritmo YOLO (siglas en inglés de *You Only Look Once*) es un tipo de CNN y está compuesto por una sola red neuronal convolucional que predice de manera simultánea múltiples Bounding Box y las probabilidades de estas últimas de pertenecer a cada clase de objeto creada. La arquitectura de YOLO está basada en la de GoogLeNet, la cual cuenta con 24 capas convolucionales, seguida por 2 capas completamente conectadas (FC). A diferencia de GoogLeNet, YOLO utiliza capas de reducción de 1x1, seguidas de capas de convolución de 3x3 (Redmon et al., 2016).

Actualmente existen diversas versiones del algoritmo *YOLO: You Only Look Once: Unified, Real-Time Object Detection* (Yolov1); *YOLO9000: Better, Faster, Stronger* (Yolov2); *YOLOv3: An Incremental Improvement* (Yolo v3); *YOLO v4: Optimal Speed and Accuracy for object detection* (Yolo v4), entre otras versiones. Asimismo, YOLO puede ser implementado en distintos entornos de trabajo como Darknet, la cual fue creada por los mismos desarrolladores del algoritmo YOLO, escrito en C con CUDA y permitiendo el cómputo en GPU. Otro entorno de trabajo utilizado es AlexeyAB/Darknet, el cual es una bifurcación del Darknet original que puede ser utilizado en Windows y Linux. Finalmente, también es utilizado el entorno Darkflow, el cual es una adaptación de Darknet a la biblioteca Tensorflow (Ameijeiras, Gonzales y Hernandez, 2020).

Respecto a las primeras versiones de YOLO (Yolov1 y Yolov2), estas presentan dificultades en la detección de objetos pequeños, es por ello que en la tercera versión (Yolov3) se agregó la detección de características multiescala, lo cual convirtió la red en un sistema más eficiente. Además, se hizo uso de 53 capas convolucionales, conocida como la arquitectura Darknet-53, lo cual mejoró el tiempo de respuesta y la precisión obtenida (Srivastava et al., 2021). Por otro lado, en la versión cuatro (Yolov4), se realizaron múltiples optimizaciones que mejoraron el AP y la cantidad de cuadros por segundo (FPS) en un 10 y 12%, respectivamente, a comparación de la versión tres (Jiang et al., 2022). Las optimizaciones realizadas involucraron al método de procesamiento de datos, la red troncal, el entrenamiento de la red, las funciones de activación y pérdida, entre otros aspectos (Xuelong et al., 2021)

En suma, se presentan algunos beneficios de YOLO según Redmon et al. (2016):

- YOLO es un algoritmo extremadamente rápido, dado que se aborda el problema de la detección desde un análisis de regresión para lo cual no necesita un código complejo. La red neuronal más simple de YOLO funciona a 45 fotogramas por segundo (fps), sin considerar un procesamiento por lotes, en una GPU Titan X y la versión más rápida de esta red funciona a 145 fps. Además, YOLO puede lograr más del doble del mAP que otros sistemas de procesamiento en tiempo real.
- En comparación con otros algoritmos de CNN, YOLO evalúa las imágenes enteras durante las etapas de entrenamiento y validación, por lo que implícitamente realiza la tarea de codificar la información contextual y la apariencia de cada clase de objeto. Otros algoritmos como el Faster R-CNN suelen confundir los objetos de fondo en una imagen, ya que no son capaces de entender el contexto general de cada imagen,
- YOLO es un algoritmo altamente generalizable, por lo que logra representaciones más certeras cuando se aplican nuevos datos de entrada.

Respecto a las investigaciones realizadas sobre la aplicación de YOLO en el sector construcción, el estudio realizado por Hyojoo y Changwan (2021) comparó cuatro arquitecturas de DL para la detección de trabajadores de construcción, entre ellas Yolov3, Yolov4, Faster R-CNN con ResNet-152 y un método propuesto el cual involucra las arquitecturas de Yolov4 y la CNN Siamese. Los resultados de este estudio mostraron que Yolov4 presenta mayor mAP, Precisión y Recuperación que su tercera versión (92.89%, 96.6% y 96.02%, respectivamente). Por otro lado, la arquitectura de Faster R-CNN con ResNet-152 tuvo mejores resultados en términos de mAP (94.30%) y Recuperación (98.13%), pero no en Precisión (96.03%). Asimismo, las arquitecturas de Yolov3 y Yolov4 mostraron mejores resultados que Faster R-CNN con ResNet-152 en términos de velocidad. Finalmente, el método propuesto por los autores logró mejorar los resultados de las tres últimas arquitecturas mencionadas en términos de mAP (96.04%), Precisión (97.50%) y Recall (98.47%), lo cual representa la viabilidad del uso de arquitecturas de CNN y sus variantes para la detección y el monitoreo continuo de trabajadores.

Otro estudio realizado por Zeng et al. (2021), desarrolló una versión mejorada de Yolov3 para la detección y localización de maquinarias y sitios en una construcción, los cuales normalmente son representados en múltiples escalas al capturar una imagen o grabar videos ínsita. Asimismo, el modelo optimizado de Yolov3 también utilizó una versión mejorada del algoritmo Grey Wolf Optimizer (GWO) para la localización de las

maquinarias. En la etapa de resultados de este estudio, se compararon las métricas de mAP y Recuperación obtenidas al entrenar el modelo original de Yolov3 y el modelo optimizado, con imágenes de objetos es escala pequeña, mediana y grande. Los valores de mAP mejoraron notablemente al utilizar el modelo optimizado (69.7%, 93.1% y 78.7%) y el modelo original (52.5%, 88.9%, 78.3%) para cada escala, respectivamente. De igual manera, los valores de Recuperación mostraron mejoría en los objetos a pequeña y mediana escala para el modelo optimizado (72% y 95.2%) y original (58.7% y 93.4%), de manera respectiva. Este estudio demuestra que tanto una versión optimizada de Yolov3 y el modelo original son capaces de detectar objetos en diferentes escalas en un sitio de construcción de manera asertiva, siendo los objetos a mediana escala, aquellos que obtienen el mayor mAP y valor de Recuperación en ambos modelos.

### **1.7. Computer Vision Technology**

Las nuevas tecnologías de la Visión por Computador o *Computer Vision* (CV), la cual es un campo dentro de la IA, se enfocan en la creación de sistemas digitales que sean capaces de procesar, analizar y dar coherencia a la información obtenida visualmente, como lo hacen las personas (Babich, 2020). De acuerdo con lo expuesto por Brian H.W. et al. (2021), el desarrollo y aplicación de CV en la industria de la construcción se clasifica en tres niveles: N1 detección, reconocimiento y monitoreo, N2 evaluación y N3 predicción. A continuación, se describe las características de cada nivel:

- N1: Detección, reconocimiento y monitoreo

En este primer nivel se busca el reconocimiento de los objetos de interés ubicados en sus posiciones naturales a través del uso de imágenes y videos. Así mismo, los métodos para la detección y reconocimiento se dividen en tres categorías: métodos basados en la geometría, métodos basados en la apariencia y métodos basados en las características de la imagen o video. Por otro lado, el monitoreo es otra de las etapas más importantes debido a su alto valor en el control de la productividad y seguridad de los trabajadores y maquinarias.

- N2: Evaluación

El segundo nivel está basado en la información recolectada en el primer nivel de detección, reconocimiento y monitoreo, y busca evaluar las acciones realizadas por el objeto de estudio o las condiciones sobre la cual se encuentra el mismo.

Los objetos de estudio pueden ser trabajadores, maquinarias, estructuras construidas, entre otros. Estas evaluaciones se realizan basadas en reglas o conocimientos que permitan al modelo comparar y clasificar la información visualizada.

- N3: Predicción

Por último, en el tercer nivel, la tarea de predicción hace referencia a la capacidad del modelo de CV de anticipar las acciones o movimientos de los objetos de estudio.

Asimismo, la visión por computador ha sido ampliamente estudiada utilizando métodos de DL. El estudio realizado por Kim et al. (2018) utilizó CNN para la detección de equipos de construcción, logrando alcanzar un 96.33% de mAP. En este estudio se utilizó un tipo de aprendizaje llamado Transfer Learning para el entrenamiento de su red neuronal, lo cual contribuyó a mejorar notablemente los resultados. Asimismo, en el estudio realizado por Xiao y Kang (2021), se creó un conjunto de datos de imágenes de maquinarias de construcción, el cual se utilizó para el entrenamiento de cuatro algoritmos de DL para la detección de objetos. Los algoritmos utilizados por los investigadores se clasifican en dos tipos: algoritmos de una etapa (YOLO-v3 e Inception-SSD) y algoritmos de dos etapas (R-FCB-ResNet101 y Faster-ResNet101). Los resultados obtenidos en la investigación mostraron que los primeros se desenvuelven mejor en términos de rapidez, mientras que los segundos en términos de precisión. De igual forma, ambos tipos de algoritmos lograron un mAP mayor al 95%, lo cual también validó la viabilidad del dataset creado.

Otro estudio similar realizado por Arabi et al. (2020) desarrolló una versión mejorada del algoritmo *Single Shot Detector* (SSD) en la arquitectura MobileNet para la detección de maquinaria de construcción utilizando sistemas embebidos o microprocesadores. En este estudio se trabajó desde la etapa de desarrollo del algoritmo hasta la implementación del mismo, tomando en cuenta la naturaleza restrictiva del hardware de los dispositivos. El resultado de esta investigación mostró un mAP de 90%.

En un siguiente estudio realizado por Yan, Zhang y Li (2020) se presentó un método de CV para la estimación de la proximidad y aglomeración en el espacio 3D entre trabajadores y maquinarias pesadas, a través de la detección de objetos en el plano 2D, con el fin de evitar la distorsión entre ambos en este último plano. La investigación obtuvo

como resultado una precisión de 100% y un valor de Recuperación de 81.6%, lo cual al mismo tiempo validó la factibilidad del método de reconstrucción de *Bounding Boxes* en 3D para la detección de maquinarias y trabajadores.

Otro de los temas con mayor relevancia entre los investigadores ha sido la detección de equipos de protección personal para trabajadores. Huang et al. (2021) desarrollaron un algoritmo optimizado de YOLO-v3 para la identificación de cascos y detectar su posición, diferenciando si estos están siendo usados o no. El algoritmo optimizado funcionó mejor que el normal, aumentando el mAP en 3.5% y la cantidad de fotogramas por segundo (FPS) en 3f/s. Del mismo modo, Shen et al (2020) implementó una novedosa metodología para la identificación del uso seguro de los cascos a través de la detección de los rostros utilizando CNN y la técnica de *Bounding Box Regression*. Con el fin de mejorar la precisión en el reconocimiento de los cascos, se utilizó un tipo de aprendizaje llamando *Deep Transfer Learning* (DTL) con la cual se entrenó la arquitectura DenseNet, logrando alcanzar un 94.47% de mAP, 96.2% de *recall* y *precision*.

Como mencionado anteriormente, los métodos de DL dentro del sector construcción han tenido avances en el campo de la CV enfocado en la seguridad, evaluación de la productividad y monitoreo del progreso de los trabajos. Asimismo, cabe mencionar que la detección de sujetos y objetos de construcción (trabajadores, maquinarias y materiales) es el primer paso para el desarrollo de un modelo de monitoreo de las tareas reiterativas (Xiao et al., 2020). En ese sentido, con la finalidad de asegurar un óptimo reconocimiento de los sujetos y objetos, se requiere un amplio conjunto de datos o *Dataset* en diferentes escenarios para evitar el sobreentrenamiento u *Overfitting* del modelo entrenado (Cogswell et al., 2015), cuyo concepto se explicará más adelante. Se le conoce como *Dataset* al conjunto de imágenes que han sido manualmente anotadas de acuerdo con una clasificación de objetos (Xiao et al., 2021). Asimismo, la anotación de imágenes es un paso esencial para la preparación del dataset, lo cual se refiere al proceso de etiquetar o anotar las regiones de interés (ROI) previamente definidas.

### **1.8. Data Augmentation**

Una de las mayores complejidades de los métodos de aprendizaje profundo, específicamente en las CNN, es que pueden surgir problemas de *Overfitting*. El

sobreentrenamiento es una complejidad que ocurre cuando una red neuronal se alinea a ciertas características de los puntos de datos de una función generando una gran varianza entre los resultados de entrenamiento y validación (Shorten y Khoshgoftaar, 2019).

Los modelos de DL tienen una alta posibilidad de sufrir problemas de sobreentrenamiento en la etapa de entrenamiento debido a la gran cantidad de parámetros involucrados que se correlacionan de manera compleja, lo cual reduce el buen desempeño que puede lograr el modelo en la etapa de validación de datos. En ese sentido, es necesario que los modelos cuenten con errores de validación y entrenamiento que decrezcan de manera paralela para la construcción de modelos sólidos (Alzubaidi et al., 2021)

El Aumento de Datos o *Data Augmentation* (DA) es uno de los métodos que puede lograr evitar el sobreentrenamiento de los modelos y mejorar el mAP. Este método crea diferentes escenarios de la data introducida que representan un dataset más complejo, disminuyendo la brecha entre los errores de validación y entrenamiento (Alzubaidi et al., 2021). La creación de los distintos escenarios es lograda mediante la transformación de las imágenes de entrenamiento y sus respectivas etiquetas (Xiao et al., 2021). Algunas de estas transformaciones incluyen el volteo horizontal y vertical de las imágenes, el recorte de ciertas partes, la rotación, la traslación en los ejes horizontal y vertical, la inserción de ruido, la transformación de los colores, entre otras (Alzubaidi et al., 2021). A continuación, se listan las características de las transformaciones más comunes de DA, explicadas por Shorten y Khoshgoftaar (2019):

- Volteado (*Flipping*)

Generalmente es más común el uso de la técnica de volteado en el eje horizontal que en el vertical. Esta técnica es una de las menos complicadas de implementar y se ha demostrado su utilidad en distintos conjuntos de datos de imágenes en línea; sin embargo, en aquellos conjuntos de datos que contienen textos, no es posible preservar las etiquetas.

- Transformación de color (*Color transformation*)

Realizar transformaciones en los espacios de canales de colores es otra de las estrategias más prácticas de implementar. Los aumentos de color simples incluyen aislar las imágenes en un solo canal de color como R, G, o B. Adicionalmente los valores RGB se pueden manipular fácilmente con simples operaciones de

matrices para incrementar o disminuir el brillo de cada imagen. Una transformación de color más avanzada proviene de la derivación de un histograma de color que describe a la imagen, cambiando los valores de intensidad de estos histogramas, lo cual genera la alteración de la iluminación.

- Recorte (*Cropping*)

Recortar las imágenes es utilizado como un procesamiento práctico de aumento de los conjuntos de datos de imágenes con mezcla de dimensiones en ancho y alto, recortando solo la parte central de cada imagen, Además, el recorte aleatorio puede ser usado para proporcionar un efecto similar a la técnica de traslación. El contraste entre estas dos técnicas es que el recorte aleatorio reducirá el tamaño de la imagen, mientras que la técnica de traslación lo preservará. En función del tamaño de la reducción elegida, esta podría afectar significativamente a las etiquetas de cada imagen.

- Rotación (*Rotation*)

La técnica de rotación se realiza girando las imágenes hacia la derecha o izquierda en un solo eje, entre  $1^\circ$  a  $359^\circ$ . Las rotaciones ligeras entre  $1^\circ$  a  $20^\circ$  y  $-1^\circ$  a  $-20^\circ$  son útiles en entrenamientos para el reconocimiento de dígitos. Sin embargo, rotaciones mayores pueden ocasionar que no se preserve la etiqueta de cada imagen.

- Traslación (*Translation*)

La traslación de las imágenes hacia la izquierda, derecha, arriba o abajo es una muy práctica de utilizar para evitar el sesgo posicional en los datos. Por ejemplo, si un dataset está compuesto de imágenes centradas, lo cual es muy común en el reconocimiento de rostros, el modelo entrenado funcionará perfectamente en la detección de solo imágenes centradas. A medida que la imagen es trasladada en una dirección, el espacio sobrante de este proceso puede ser rellenado con una constante de valor de RGB entre 0 a 255, manteniendo de esta forma la dimensión de la imagen después de la aplicación de la técnica.

- Inserción de ruido (*Noise injection*)

La inserción de ruido consiste en inyectar a una matriz valores aleatorios usualmente obtenidos de una distribución de Gaussian. Esta técnica puede ayudar a las CNN a aprender características más robustas de cada imagen.

- Kernel (*Kernel filters*)

Los filtros Kernel es una técnica de DA muy comúnmente utilizada en la transformación de imágenes para lograr un enfoque y desenfoque de estas. Esta técnica funciona aplicando una matriz de  $n \times n$  con un filtro de desenfoque de Gaussian sobre la imagen original, lo cual dará como resultado una imagen más borrosa o, por el contrario, aplicando un filtro de alto contraste resultará en una imagen más nítida. Intuitivamente, el uso de imágenes borrosas ayudará a los modelos a tener una mayor resistencia a los desenfoques ocasionados por movimientos. Asimismo, el uso de imágenes más nítidas mejorará la capacidad de los modelos de capturar más detalles en los objetos de interés.

- Borrado aleatorio (*Random Erasing*)

La oclusión es un fenómeno que ocurre cuando algunas de las partes del objeto de interés se encuentran ocultas en la imagen. La técnica del borrado aleatorio fue específicamente diseñada para combatir estos problemas de reconocimiento debido a la oclusión. El borrado aleatorio forzará al modelo a aprender características más descriptivas de una imagen, evitando el sobreentrenamiento de una determinada característica visual. Esta técnica funciona aplicando una matriz de  $n \times m$  en la imagen y rellenándola con un valor constante de RGB entre 0 a 255.

En esta línea, se han desarrollado estudios respecto al uso de técnicas de DA para mejorar la clasificación de objetos, trabajadores, maquinarias, entre otros. Kim and Jeong (2020) utilizaron la técnica de *Jittering* para mejorar la clasificación de los estados de los sistemas hidráulicos, los cuales no cuentan con suficiente data. Los resultados del estudio mostraron que el mAP aumentó en un 12% cuando la técnica de DA fue aplicada. Asimismo, otro estudio desarrolló un modelo de detección de cubrebocas en personas aplicando las técnicas de *Grayscale* y *Gaussian blurred*. Los resultados mostraron un aumento de 3% de mAP en la detección de personas individuales, grupos de personas y videos con grupos de personas (Prusty et al., 2021).

Respecto a las aplicaciones de DA en la industria de la construcción, Bang et al. (2020) utilizaron técnicas de DA para mejorar la detección de los materiales; entre las técnicas utilizadas se aplicó *Revolving-and-Painting* utilizando GAN, *Cut-and-paste*, Variación de la intensidad de iluminación, difuminado de las imágenes mediante la aplicación del filtro de Gaussian y transformación de la escala. Esta investigación tuvo como resultado una mejora de 11.5% de mAP aplicando una combinación de todas las técnicas mencionadas.

En otro estudio desarrollado por Qin et al. (2021), el cual se enfocó en la detección de vigas de acero curvadas, huecos o vacíos y el material de revestimiento inicial en la construcción de túneles, información obtenida a través de un reloj de penetración del suelo, se demostró que el mAP incrementó en 4% utilizando las técnicas de *Finite-Difference Time-Domain (FDTD)* y *Deep Convolutional Generative Adversarial Network (DCGAN)* para crear nuevas imágenes sintéticas. Finalmente, la investigación realizada por Dung et al. (2019) para la detección de grietas ocasionadas por fatigas en las uniones con placas Gusset pertenecientes a puentes de acero, mostró que el uso de las diversas técnicas de DA como la rotación, el acercamiento, el volteo horizontal y vertical, entre otras, generaron un aumento de hasta 5% en las tres técnicas de DL aplicadas.

## **CAPÍTULO II: OBJETIVOS E HIPÓTESIS**

A continuación, se presenta el objetivo general, los objetivos específicos y la hipótesis de la investigación:

### **2.1 Objetivos**

#### **2.1.1 Objetivo General**

Evaluar la efectividad del uso de técnicas de Aumento de Datos en la detección y clasificación de trabajadores en una obra de construcción.

#### **2.1.2 Objetivos Específicos**

- Comparar los resultados de mAP obtenidos al aplicar las diferentes técnicas de Aumento de Datos establecidas para la detección y clasificación de trabajadores en una obra de construcción.
- Comparar los niveles de confianza obtenidos al aplicar las diferentes técnicas de Aumento de Datos establecidas para la detección y clasificación de trabajadores en una obra de construcción.
- Determinar el modelo de aprendizaje profundo con mejor desempeño al aplicar las diferentes técnicas de Aumento de Datos para la detección y clasificación de trabajadores en una obra de construcción.

### **2.2 Hipótesis**

El uso de técnicas de Aumento de Datos presenta un mejor desempeño en términos de mAP y niveles de confianza en la detección y clasificación de trabajadores en una obra de construcción.

## CAPÍTULO III: MÉTODO

En el presente capítulo se explicará el método utilizado para la investigación. Para ello, se presenta el diseño de investigación, los materiales utilizados y el procedimiento para el procesamiento de los datos.

### 3.1. Diseño de investigación

La investigación tiene una metodología cuantitativa, ya que se desarrollaron análisis numéricos para probar la hipótesis del estudio (Hernández et al., 2014). Este análisis numérico incluyó la preparación y el procesamiento de los datos recolectados para realizar la comparación de los resultados obtenidos utilizando las diferentes técnicas de Aumento de Datos.

Adicionalmente, la investigación utilizó un diseño experimental, pues las variables independientes, las técnicas de Aumento de Datos, fueron manipuladas deliberadamente para obtener diferentes resultados sobre las variables dependiente, el mAP y los niveles de confianza. Finalmente, el estudio tiene un nivel descriptivo, a través del cual se busca analizar la influencia del uso de las técnicas de Aumento de Datos en los resultados de mAP y los niveles de confianza en la detección y clasificación de trabajadores en una obra de construcción, con el objetivo de describir la relación entre estas variables.

### 3.2. Materiales

Los materiales e imágenes utilizados en el presente estudio se obtuvieron a partir de la investigación realizada por Del Savio et al. (2022). Estos materiales incluyeron 4 cámaras posicionadas estratégicamente en diferentes zonas del edificio en construcción, de tal manera que se obtengan imágenes de diferentes ambientes con grupos de trabajadores en distintas escalas, como se muestra en la Figura 3.1. Los tipos de cámaras utilizadas en la investigación en mención fueron 3 IP tipo PTZ motorizadas (Dahua Technology, 4K 48x Starlight + IR WizMind Network PTZ Camera) y 1 tipo Bullet (Dahua Technology, 8MP Lite IR Vari-focal Bullet Network Camera), las cuales fueron posicionadas entre los 12 y 35 metros.

Asimismo, los grupos de trabajadores captados por las cámaras fueron peones, operarios, ingenieros, arquitectos, supervisores y toda aquella mano de obra presente en la obra de construcción. Por otro lado, el procesamiento de las imágenes recolectadas y los entrenamientos posteriores, se realizaron en una CPU Intel(R) Core (TM) i7-8700.

### **Figura 3.1**

*Posiciones de las cámaras en el sitio de construcción*

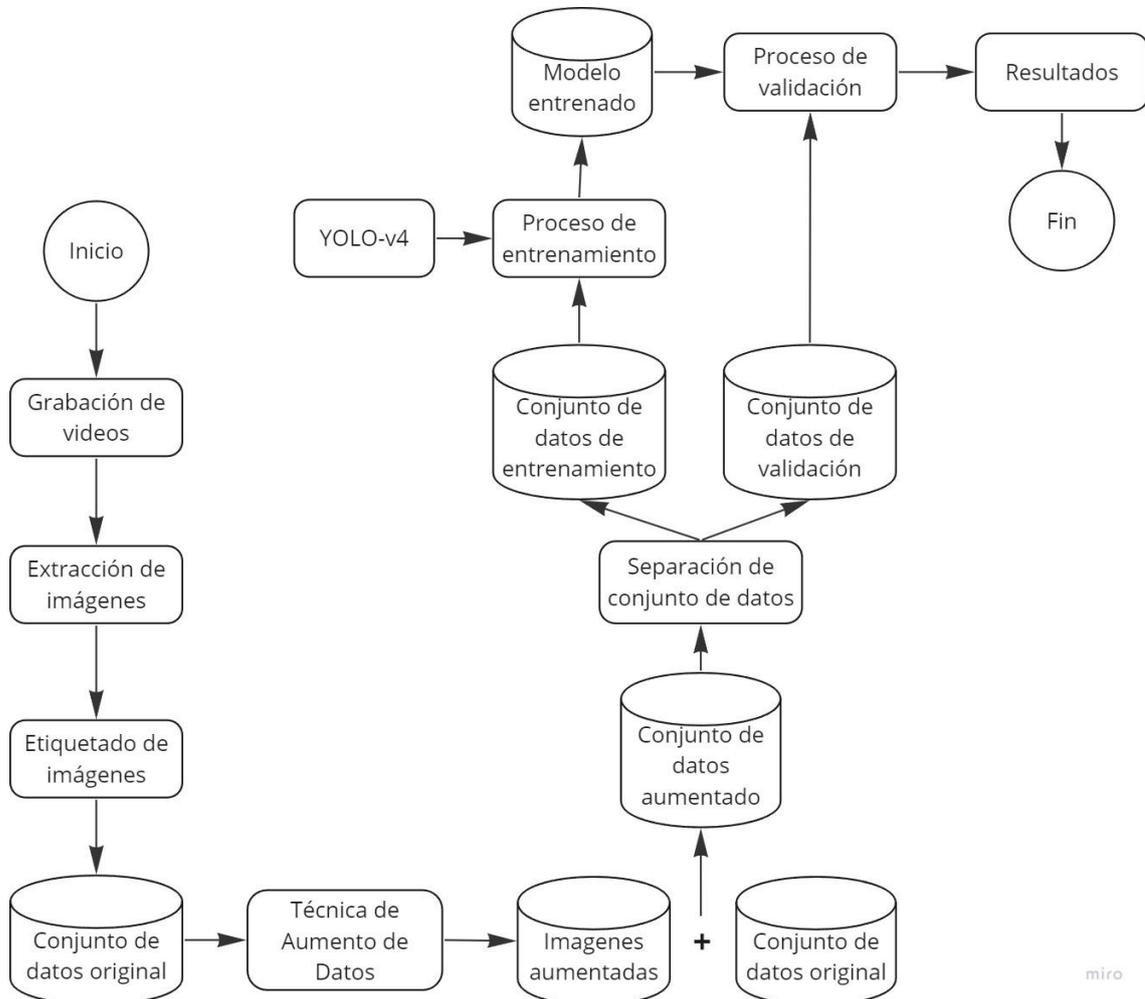


### **3.3 Procedimiento**

A continuación, se explica el procedimiento realizado para la detección de mano de obra utilizando las técnicas de Aumento de Datos propuestas, como se muestra en la Figura 3.2.

**Figura 3.2**

*Proceso para la detección y clasificación de trabajadores utilizando las técnicas de Aumento de Datos*



### 3.3.1 Generación de imágenes

El primer paso de la investigación consiste en generar las imágenes requeridas para el entrenamiento. Para ello, en la investigación realizada por Del Savio et al. (2022), durante los meses de noviembre de 2021 y abril de 2022 se grabó el proceso de construcción de la edificación, obteniendo 1212 imágenes extraídas de las grabaciones en las diferentes zonas establecidas. Las grabaciones se realizaron en este periodo de tiempo con el fin de asegurar que se cubran las distintas etapas de un proceso constructivo, como el movimiento de tierras, colocación de acero, el vertido de concreto, entre otras actividades.

### 3.3.2 Preparación del conjunto de datos

Una vez que se ha creado el primer conjunto de datos con las 1212 imágenes extraídas, el segundo paso consiste en etiquetar manualmente las imágenes utilizando la herramienta de anotación *LabelImg*. Esta herramienta te permite crear anotaciones en los formatos de PASCAL VOC XML y YOLO, las cuales son conocidas como “Bounding Box” (Tzutalin, 2015). En la presente investigación se escogieron dos categorías de anotaciones para la clasificación de la mano de obra: *Person* y *Leaning Person*. Estas categorías se seleccionaron siguiendo la investigación realizada por Del Savio et al. (2023) sobre la detección de mano de obra de construcción como parte de las medidas de seguridad luego de la pandemia de COVID-19. En ese sentido, la categoría *Person* o persona se refiere a cualquier trabajador que se encuentre parado o ligeramente inclinado, mientras que la categoría *Leaning Person* o persona inclinada se refiere a cualquier trabajador cuya posición se encuentre muy inclinada, agachada o en cuclillas (Del Savio et al. 2023). De esta manera, se forma el primer conjunto de datos para el entrenamiento de la red neuronal. El resultado del proceso de anotación se muestra en la Figura 3.3.

**Figura 3.3**

*Imagen etiquetada con las categorías Person y Leaning Person*



### 3.3.3. Aumento de Datos

En el tercer paso se procederá a crear cuatro conjuntos de datos adicionales utilizando tres técnicas de Aumento de Datos escogidas. Estas técnicas provienen de la biblioteca Albumentations, realizada por Buslaev et al. (2020), la cual contiene más de 30 tipos de transformaciones disponibles de manera abierta. Esta biblioteca te permite realizar estas transformaciones tanto en imágenes, como en máscaras de imágenes y los Bounding Box. En esta investigación se escogieron las técnicas de *Brightness - Contrast* (Brillo - Contraste), *Blur* (Difuminación) y *Rain* (Lluvia), las cuales se explican en los puntos 3.3.3.1, 3.3.3.2 y 3.3.3.3. Cabe añadir que estas transformaciones se seleccionaron con el objetivo de simular diferentes ambientes en un sitio de construcción tomando en cuenta las condiciones meteorológicas del entorno y técnicas de los dispositivos de grabación y detección.

En ese sentido, los cinco conjuntos de datos totales están formados como se muestra en la Tabla 3.1. El primero de ellos está compuesto por las 1212 imágenes originales extraídas de las grabaciones. El segundo, contiene tanto las imágenes originales, como el aumento de estas utilizando la técnica de Brillo - Contraste. El tercero y cuarto, de igual forma, están compuestas por las imágenes originales y las aumentadas con las técnicas de Difuminación y Lluvia, respectivamente. Finalmente, el quinto de ellos contiene tanto las imágenes originales, como las aumentadas con las tres técnicas mencionadas. Luego de la creación de los conjuntos de datos, estos son divididos de manera aleatoria entre las carpetas de entrenamiento (70%) y validación (30%), para ambas etapas correspondientes.

**Tabla 3.1***Conjunto de datos creados*

Conjunto de datos	Cantidad de imágenes	Imágenes Originales	Imágenes transformadas con técnicas de Aumento de Datos		
			Brillo - Contraste	Difuminación	Lluvia
1	1212	✓	×	×	×
2	2424	✓	✓	×	×
3	2424	✓	×	✓	×
4	2424	✓	×	×	✓
5	4848	✓	✓	✓	✓

A continuación, se explica el uso de las tres técnicas de Aumento de Datos escogidas para esta investigación:

### 3.3.3.1. Brillo - Contraste

La técnica de Brillo - Contraste, la cual te permite cambiar aleatoriamente ambos parámetros en una imagen, fue elegida debido a la variación de la iluminación en los ambientes de construcción, lo cual dependerá de la ubicación de la cámara y el tipo de construcción; es decir, si la cámara se encuentra ubicada en el exterior de una edificación, las imágenes o videos se capturarán con mayor brillo. Por el contrario, si la cámara está posicionada dentro de la construcción de un edificio, la luminosidad no será tan intensa. Además, esta técnica permite simular condiciones de clima reales, lo cual brinda a los conjuntos de datos una mayor similitud a la realidad permitiendo que la red neuronal se desempeñe mejor (Asif et al., 2023).

### 3.3.3.2. Difuminación

La segunda técnica utilizada fue la difuminación, la cual fue escogida debido al constante movimiento de los trabajadores dentro de los ambientes de la construcción, así como los

efectos del viento, lo cual puede influir en la captura de imágenes claras. Además, esta técnica simula escenarios en donde las cámaras pueden perder el enfoque, lo que permite que la red neuronal pueda lograr una mayor precisión en la detección al utilizar esta técnica (Asif et al., 2023).

### 3.3.3.3. Lluvia

Finalmente, debido a que los sitios de construcción se encuentran expuestos a escenarios de cambios meteorológicos, se escogió la lluvia como una de las estaciones que pueda influenciar significativamente en la calidad de una imagen y agregar ruido (Asif et al., 2023). Además, se busca que la red neuronal sea capaz de reconocer trabajadores en escenarios de oclusión, en donde cierta porción de la imagen es cegada por otros objetos, retando a la red neuronal a no confiar en características frecuentes en las imágenes.

Según lo explicado, las técnicas de Aumento de Datos mencionadas se utilizaron con la finalidad de reflejar diversos escenarios que pueden experimentar las obras de construcción. Además, estas técnicas garantizan la variedad de imágenes e incrementan el tamaño de los conjuntos de datos, evitando futuros problemas de sobreentrenamiento. A continuación, en la Figura 3.4 se muestra una de las imágenes originales extraídas y en las Figuras 3.5, 3.6, 3.7 imágenes transformadas con las técnicas de Aumento de Datos.

### Figura 3.4

*Imagen original (IMG-680) sin técnica de Aumento de Datos*



**Figura 3.5**

*Imagen aumentada utilizando la técnica de Brillo – Contraste*



**Figura 3.6**

*Imagen aumentada utilizando la técnica de Difuminación*



### Figura 3.7

Imagen aumentada utilizando la técnica de Lluvia



#### 3.3.4 Detección de mano de obra

Luego de la creación de los 5 conjuntos de datos, se procede a realizar la etapa de entrenamiento. En esta etapa se hizo uso de la red neuronal YOLO-V4 y el entorno de trabajo AlexeyAB/Darknet, utilizados en la investigación realizada por Del Savio et al. (2023), el cual es ideal para implementaciones en Windows y Linux. Además, esta red neuronal ha sido utilizado en distintas investigaciones debido a los buenos resultados obtenidos en términos de rapidez y precisión en comparación a otras redes (Bochkovskiy A, Wang C y Liao HM, 2020; Del Savio et al., 2021).

En total, se realizaron 4 entrenamientos por cada conjunto de datos, variando aleatoriamente en cada uno de ellos la distribución de las imágenes en las carpetas de entrenamiento (70%) y validación (30%); es decir, en total se realizaron 20 entrenamientos. Asimismo, los hiperparámetros mostrados en la Tabla 3.2 fueron igualmente aplicados en cada uno de ellos. Según lo explicado en la investigación de YOLO-V4 realiza por Bochkovskiy et al. (2020), *Batch size* representa el número de datos utilizados para cada iteración del entrenamiento. En otras palabras, el conjunto de datos de la carpeta de entrenamiento es dividido en varios *Batches* con el objetivo de evitar sobrecargar la memoria del CPU y mantener la rapidez del procesamiento. De igual

manera, los *Batches* se dividen en la cantidad de *Subdivisions* o *mini batches* establecida, los cuales son procesados de manera paralela por el GPU del computador.

Por otro lado, el número de *Max batches* sigue la fórmula de la cantidad de categorías de clasificación multiplicado por 2, cuyo valor no debe ser menor que la cantidad de imágenes de entrenamiento. Los valores de *Step* representan el 80% y 90% del valor de *Max batches* y los parámetros de *Saturation*, *Exposure* y *Hue* se establecieron en 1, pues estos cambian los niveles de saturación, brillo y color en las imágenes. Por último, el *Learning rate* es un hiperparámetro que representa el porcentaje de cambio de los pesos de entrada durante cada iteración para poder brindar una mejor aproximación; este valor fue establecido debido a los óptimos resultados obtenido en la investigación realizada por Quin et al. (2021).

Luego de establecer los hiperparámetros, se recuperaron los resultados de mAP de cada entrenamiento y se realizó el proceso de validación con las redes neuronales con mejores resultados para cada conjunto de datos. En este último proceso, se escogió una misma imagen de cada conjunto de datos no perteneciente al proceso de entrenamiento y, posteriormente se validó y se recolectó los resultados de los niveles de confianza sobre los 5 trabajadores de la imagen, con el fin de realizar la comparación de cada red neuronal entrenada.

**Tabla 3.2***Hiperparámetros de entrenamiento*

<b>Hiperparámetros</b>	<b>Valores</b>
Batch size	64
Subdivision	16
Max batches	4000
Steps	3200, 3600
Saturation	1
Exposure	1
Hue	1
Learning rate	0.001

## CAPÍTULO IV: RESULTADOS Y DISCUSIÓN

La presente investigación tiene como objetivo determinar la efectividad en el uso de las técnicas de Aumento de Datos para mejorar la precisión en términos de mAP y los niveles de confianza en la detección y clasificación de trabajadores de una obra construcción. En ese sentido, se elaboraron 5 conjuntos de datos los cuales contienen imágenes etiquetadas con 2 clases: *Person* y *Leaning Person*, las cuales fueron establecidas siguiendo la investigación de Del Savio et al. (2023) sobre la detección de mano de obra de construcción. Las imágenes recolectadas provienen del estudio realizado por Del Savio et al. (2022), las cuales fueron extraídas a través de grabaciones de 4 cámaras posicionadas alrededor de la edificación en construcción, cada una con una resolución de 3840 x 2160 píxeles. Asimismo, el primer conjunto de datos está compuesto por las 1212 imágenes originales obtenidas en la Sección 3.3, mientras que los restantes se crearon transformando las imágenes originales con tres técnicas de Aumento de Datos por separado: Brillo - Contraste, Difuminación y Lluvia, siendo el último de los conjuntos la suma del primer conjunto de datos más los tres aumentos anteriores.

Los resultados del método de investigación propuesto se dividen en 2 partes: resultados de entrenamiento y resultados de validación. Para ellos, las imágenes se mezclaron de manera aleatoria entre las carpetas de entrenamiento (70%) y validación (30%) para cada etapa. Para los resultados de entrenamiento, se recolectaron y compararon los valores de Precisión, Recuperación, mAP, valor-F1 e IoU obtenidos de cada entrenamiento por cada conjunto de datos. El resultado de Precisión nos indica la habilidad de la red neuronal de clasificar correctamente los casos identificados, mientras que el valor de Recuperación nos muestra la habilidad de identificar todos los casos existentes. En esta investigación, un bajo nivel de Recuperación nos indica un gran error al tener trabajadores faltantes por detectar; de igual forma, un bajo nivel de Precisión nos muestra un gran error en detectar correctamente a los trabajadores. Asimismo, el mAP es el promedio de la precisión obtenida sobre todos los tipos de clases o etiquetas de objetos (*Person* y *Leaning Person*) definidos para los conjuntos de datos. Además, el valor-F1 es un indicador que combina los valores de Precisión y Recuperación, y asume que ambos

valores son igualmente relevantes. Por otra parte, otra variable relevante para determinar si una detección es correcta es el IoU, el cual debe tener un valor mayor a 0.5.

Luego de realizar 4 entrenamientos para cada conjunto de datos, se seleccionó la red neuronal con el mejor resultado para cada uno y se procedió a la segunda etapa de validación, a través de la cual se obtuvieron los niveles de confianza de cada red neuronal. Por último, se utilizó una nueva imagen en una obra de construcción distinta con el fin de comparar estos niveles en un nuevo escenario.

#### 4.1 Resultados de entrenamiento

Como mencionado anteriormente, se dice que la detección fue un éxito cuando el indicador de IoU es mayor que 0.5 en cada entrenamiento (Xiao et al., 2021). De manera general, se puede observar que, en todos los entrenamientos realizados, este valor supera el 0.5 o 50%, por lo que cada red neuronal entrenada es igualmente apropiada para la comparación de sus resultados.

Respecto a los entrenamientos realizados con el conjunto de datos original, en donde el 70% (848) de las imágenes se utilizaron para esta etapa, los resultados se muestran en la Tabla 4.1. De los 4 entrenamientos ejecutados, se puede observar que el primero de ellos obtuvo el mejor valor de mAP, con 77.04%. Sin embargo, el entrenamiento 3 obtuvo el mejor resultado de F1-score (0.77) y un valor muy cercano en mAP, logrando un 77.02%. Dado que, en este estudio, tanto el valor de la Precisión como el de Recuperación son igual de importantes, se considera el valor-F1 como indicador para la evaluación de ambos parámetros. En ese sentido, dada la cercanía de los resultados de ambos indicadores, se ha elegido el primer entrenamiento de los 4 realizados con el conjunto de datos de imágenes originales para la comparación con los siguientes resultados utilizando las técnicas de Aumento de Datos propuestas.

**Tabla 4.1**

*Resultados de entrenamientos de la red neuronal con el conjunto de datos original*

Nº entrenamiento	mAP	Precision	Recall	F1-score	IoU
1	<b>77.04%</b>	<b>0.73</b>	<b>0.80</b>	<b>0.76</b>	<b>55.03%</b>
2	73.83%	0.71	0.78	0.74	53.39%
3	77.02%	0.73	0.81	0.77	54.90%
4	76.96%	0.72	0.81	0.76	54.77%

El segundo, tercer y cuarto conjunto de datos están compuestos por 2424 imágenes; de igual manera, el 70% (1697) de las imágenes fueron usadas para el entrenamiento. Los resultados del entrenamiento de la red neuronal con el segundo conjunto de datos, en el cual se utilizó la técnica de Brillo - Contraste, se muestran en la Tabla 4.2. Sobre los entrenamientos ejecutados, el cuarto de ellos contiene los mejores valores, los cuales son muy similares al segundo entrenamiento, por lo que ambos son igualmente válidos para la detección de trabajadores. En comparación con los resultados obtenidos en el entrenamiento con el conjunto de datos original, se muestra una mejoría de 4.39% en el valor de mAP y 3% en el valor-F1.

Asimismo, en todos los entrenamientos realizados, se observa que los 5 indicadores tuvieron un valor superior al realizado con el conjunto de datos original. En ese sentido, se afirma que la técnica empleada mejora el mAP en la detección de trabajadores en una obra de construcción. Asimismo, se espera que estas redes neuronales entrenadas con la técnica en mención sean capaces de realizar detecciones tanto en días o zonas con gran intensidad luminosa, como en lugares opacos.

**Tabla 4.2**

*Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con la técnica Brillo-Contraste*

N° entrenamiento	mAP	Precision	Recall	F1-score	IoU
1	80.19%	0.74	0.82	0.78	56.30%
2	81.42%	0.75	0.83	0.79	57.02%
3	79.89%	0.74	0.82	0.78	56.33%
<b>4</b>	<b>81.43%</b>	<b>0.75</b>	<b>0.83</b>	<b>0.79</b>	<b>57.05%</b>

Los resultados de entrenamiento con el tercer conjunto de datos se muestran en la Tabla 4.3. En este conjunto se utilizó la técnica de Difuminación, la cual está relacionada con la claridad y el enfoque en una imagen. Los mejores valores obtenidos en los indicadores de los 4 entrenamientos fueron alcanzados por el primero de ellos. En comparación con los resultados de entrenamiento con el segundo conjunto de datos, se aprecia que estos últimos fueron ligeramente superiores. Esta diferencia se puede deber a lo explicado por Na et al. (2022), en cuya investigación se menciona que altos valores de difuminación disminuyen la capacidad del modelo en detectar objetos. Por otro lado,

todos los resultados de mAP y valor-F1 continúan siendo mayores a los resultados del entrenamiento con el conjunto de datos original, siendo el mayor aumento en aproximadamente 4% en el valor de mAP y en 2% el valor-F1. Por lo tanto, ambas redes neuronales entrenadas con el segundo y tercer conjunto de datos pueden ser utilizadas para la detección de obra.

**Tabla 4.3**

*Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con la técnica de Difuminación*

<b>N° entrenamiento</b>	<b>mAP</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>IoU</b>
<b>1</b>	<b>81.21%</b>	<b>0.74</b>	<b>0.83</b>	<b>0.78</b>	<b>56.38%</b>
2	77.42%	0.73	0.80	0.76	55.02%
3	80.50%	0.74	0.83	0.78	54.90%
4	77.46%	0.74	0.81	0.77	54.77%

Respecto al entrenamiento de la red neuronal con el cuarto conjunto de datos, en el que se utilizó la técnica de Lluvia, los resultados se muestran en la Tabla 4.4. La técnica de lluvia se caracteriza por influenciar en la resolución de las imágenes y ocultar parcialmente los objetos de detección; por lo tanto, si no se realiza un entrenamiento considerando dicho fenómeno, esto podría influenciar negativamente en los resultados. El mejor de los valores resultantes de los 4 entrenamientos los obtuvo el segundo de ellos, siendo menores a los resultados de entrenamientos con el segundo y tercer conjunto de datos. En comparación con el entrenamiento con el conjunto de datos original, todos los entrenamientos realizados con la técnica empleada continúan mejorando los valores de mAP y valor-F1.

**Tabla 4.4**

*Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con la técnica de Lluvia*

<b>N° entrenamiento</b>	<b>mAP</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>IoU</b>
1	80.43%	0.74	0.83	0.79	56.99%
<b>2</b>	<b>81.03%</b>	<b>0.74</b>	<b>0.83</b>	<b>0.78</b>	<b>56.09%</b>
3	80.15%	0.74	0.83	0.78	56.01%
4	78.82%	0.74	0.81	0.77	55.75%

Por último, se realizaron los entrenamientos de la red neuronal con el conjunto de datos 5, el cual contiene las imágenes originales y los 3 aumentos de datos anteriormente ejecutados; es decir, 4848 imágenes de las cuales 3394 (70%) se separaron para esta etapa. De los resultados obtenidos, mostrados en la Tabla 4.5, se puede observar que el primero de los entrenamientos obtuvo mayor mAP y valor-F1 que los restantes, por lo que esta red neuronal es la más ideal para las detecciones. Además, se observa que estos resultados son mayores a los entrenamientos realizados con los 4 conjuntos de datos anteriores, por lo que se infiere que el incremento de la cantidad de imágenes usando todas las técnicas de Aumento de Datos en conjunto mejora el porcentaje de mAP y valor-F1 de las redes neuronales en las etapas de entrenamiento. Esto último es congruente con los resultados obtenidos por Bang et al. (2020), en cuyo estudio el uso de todas las técnicas de Aumento de Datos propuestas mejoró ambos indicadores, a pesar de las deficiencias existentes de cada técnica por separado.

**Tabla 4.5**

*Resultados de entrenamientos de la red neuronal con el conjunto de datos aumentado con todas las técnicas de Aumento de Datos*

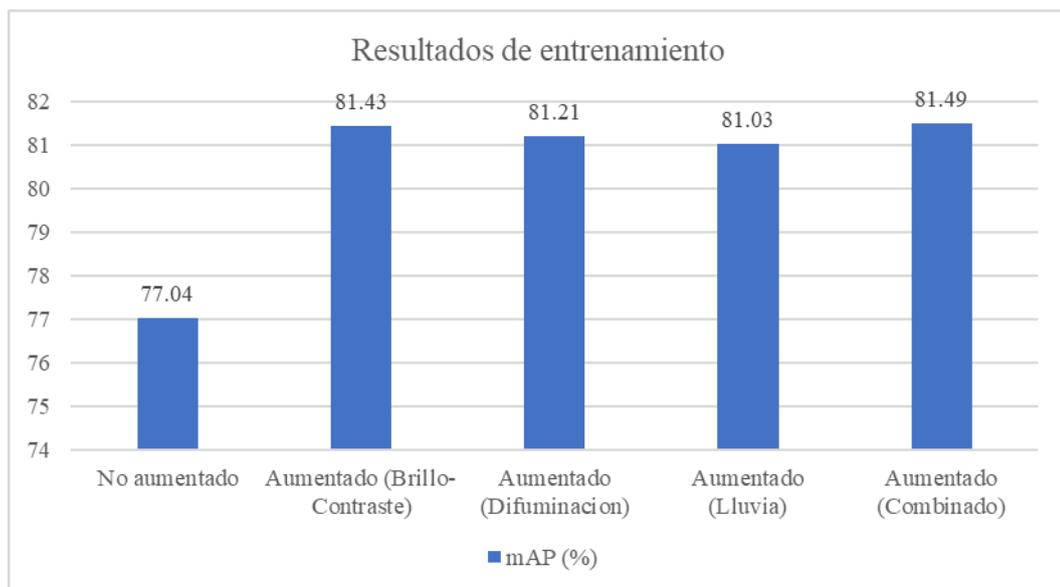
<b>N° entrenamiento</b>	<b>mAP</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>IoU</b>
<b>1</b>	<b>81.49%</b>	<b>0.75</b>	<b>0.83</b>	<b>0.79</b>	<b>57.07%</b>
2	80.08%	0.73	0.83	0.78	55.88%
3	80.93%	0.74	0.83	0.78	56.21%
4	78.64%	0.72	0.83	0.77	54.58%

En suma, como se muestra en la Figura 4.1, se puede observar que todas las técnicas de Aumento de Datos obtuvieron valores de mAP superiores al 81% en sus

mejores resultados; es decir, mejoraron la precisión media en 4% con respecto a la red neuronal entrenada sin ninguna técnica de Aumento de Datos. Además, se muestra una ligera diferencia entre los valores de mAP para cada técnica empleada, siendo la técnica de Brillo - Contraste la que generó mejores resultados que las técnicas de Difuminación y Lluvia. Esto se podría explicar debido a que un alto valor de estas últimas en las imágenes generaría una mayor distorsión y, por lo tanto, incrementan la dificultad del entrenamiento de los modelos. Finalmente, se observa que existe una tendencia en el incremento del mAP al aumentar la cantidad de imágenes de entrenamiento utilizando las técnicas de Aumento de Datos.

**Figura 4.1**

*Resultados de entrenamiento de las redes neuronales con y sin técnicas de Aumento de Datos*



#### **4.2 Resultados de validación de cada técnica**

Luego de obtener los resultados de entrenamiento y comparar los valores de cada indicador, se procedió a realizar la etapa de validación; es decir, se procesó una nueva imagen en cada red neuronal con los mejores resultados para cada conjunto de datos. En otras palabras, se realizó la detección de trabajadores en las categorías de *Person* y *Leaning Person* utilizando los entrenamientos o redes neuronales entrenadas 1, 4, 1, 2 y 1 de los conjuntos de datos 1, 2, 3, 4 y 5, respectivamente, en una imagen no antes

procesada por cada red neuronal durante la etapa de entrenamiento. Las detecciones realizadas por cada red neuronal entrenada se muestran en cada trabajador encerrado en cuadros delimitadores denominados *Bounding Boxes*, mostrado en la Figura 4.2, en donde en el lado superior izquierdo se muestra la categoría a la que pertenecen y en el lado superior derecho el nivel de confianza de pertenencia a la clase asignada. La nueva imagen evaluada corresponde a la Figura 4.3, en donde visiblemente se observa 5 trabajadores, los cuales han sido enumerados y clasificados en sus categorías correspondientes en la Tabla 4.6.

**Figura 4.2**

*Ejemplo de Bounding Box*



**Figura 4.3**

*Nueva imagen (IMG-44) para la validación de las redes neuronales*



**Tabla 4.6**

*Clasificación de cada trabajador en la obra CEBUL de la Universidad de Lima en Lima, Perú*

Nº de trabajador	Categoría
1	Leaning person
2	Leaning person
3	Person
4	Leaning person
5	Person

#### 4.2.1 Red neuronal con conjunto de datos original

Los resultados de validación de la red neuronal del conjunto de datos original se muestran en la Figura 4.4, en donde se detectaron 3 de los 5 trabajadores presentes en la imagen. En ese sentido, el trabajador N°2 fue detectado en la categoría correcta con un nivel de confianza bastante alto de 0.98. Luego, el trabajador N°4 fue detectado correctamente en la categoría *Leaning Person*, con un alto nivel de confianza de 0.88. Por último, el trabajador N°5 se detectó de igual forma correctamente en la categoría correspondiente con un nivel de confianza de 0.97. De estos resultados, se puede observar que la red neuronal entrenada con el primer conjunto de datos no es muy confiable ya que, en una primera instancia, solo se ha detectado 3 de trabajadores, por lo que el nivel de recuperación es deficiente.

**Figura 4.4**

*Validación de la red neuronal entrenada con el conjunto de datos original*



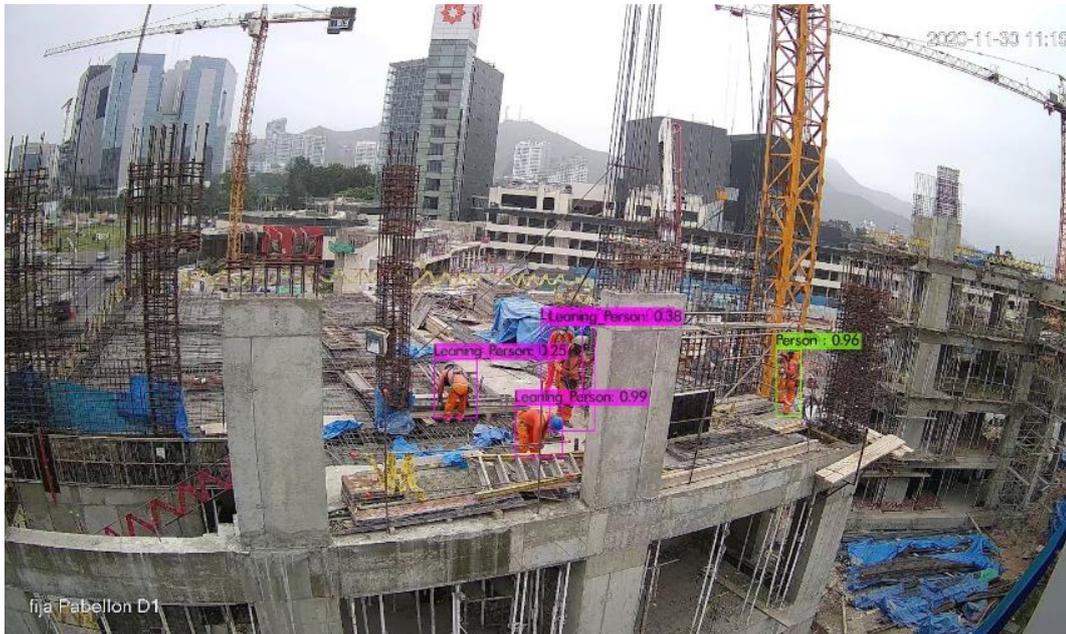
#### 4.2.2 Red neuronal con conjunto de datos aumentado con la técnica Brillo-Contraste

Respecto a los resultados de validación de la red neuronal entrenada con el segundo conjunto de datos con la técnica de *Brillo-Contraste*, estos se muestran en la Figura 4.5,

en donde se detectaron 4 de los 5 trabajadores presentes en la imagen. En ese sentido, el trabajador N°1 fue detectado en la categoría correcta; sin embargo, con un bajo nivel de confianza de 0.30. Seguidamente, el trabajador N°2 fue detectado en la categoría correcta con un nivel de confianza bastante alto de 0.99. Por otro lado, el trabajador N°3 fue detectado de manera incorrecta con un nivel de confianza bajo de 0.28. Luego, el trabajador N°4 fue detectado correctamente en la categoría correspondiente; sin embargo, con un nivel de confianza bajo de 0.38. Por último, el trabajador N°5 se detectó de igual forma en la categoría correspondiente con un nivel de confianza alto de 0.96. De estos resultados se puede observar una mejoría en el nivel de Recuperación o en la cantidad de detecciones de trabajadores y en el nivel de confianza utilizando la técnica de aumento de datos Brillo-Contraste. Sin embargo, se observa que la red neuronal presentó un error en la clasificación de un trabajador. Esto se puede deber a un error en el etiquetado manual del trabajador para la etapa de entrenamiento.

## Figura 4.5

Validación de la red neuronal entrenada con el segundo conjunto de datos



### 4.2.3 Red neuronal con conjunto de datos aumentado con la técnica Difuminación

Respecto a los resultados de validación de la red neuronal entrenada con el tercer conjunto de datos con la técnica de Difuminación, estos se muestran en la Figura 4.6, en donde se detectaron 4 de los 5 trabajadores presentes en la imagen. Se puede observar que el trabajador N°1 fue detectado en la categoría correcta y con un bajo nivel de confianza de 0.35. Igualmente, el trabajador N°2 fue detectado en la categoría correcta y con un nivel de confianza bastante alto de 0.97. Luego el trabajador N°4 fue detectado de manera correcta en la categoría *Leaning Person*, con un nivel de confianza de 0.45. Por último, el trabajador N°5 se detectó correctamente en la categoría correspondiente con un nivel de confianza alto de 0.98. A partir de los resultados mencionados, de manera similar que lo obtenido con la técnica de Brillo-Contraste, se obtuvo un mejor nivel de Recuperación con respecto a la validación de la red neuronal entrenada con el conjunto de datos original.

## Figura 4.6

Validación de la red neuronal entrenada con el tercer conjunto de datos



### 4.2.4 Red neuronal con conjunto de datos aumentado con la técnica Lluvia

Respecto a los resultados de validación de la red neuronal entrenada con el cuarto conjunto de datos con la técnica de Lluvia, estos se muestran en la Figura 4.7, en donde se detectaron 3 de los 5 trabajadores presentes en la imagen. Se puede observar que el trabajador N°2 fue detectado en la categoría correcta y con un alto nivel de confianza de 0.99. Luego, el trabajador N°4 fue detectado igualmente de manera correcta en la categoría *Leaning Person*, con un nivel de confianza de 0.52. Por último, se observa que el trabajador N°5 se detectó correctamente en la categoría correspondiente con un nivel de confianza alto de 0.97. A partir de los resultados mencionados, se infiere que la técnica de Aumento de Datos Lluvia presenta un nivel similar de recuperación y clasificación de trabajadores que la red neuronal sin técnica de aumento de datos, por lo que se puede deducir que el entrenamiento con esta técnica no genera cambios significativos en los resultados de validación de la red neuronal.

## Figura 4.7

Validación de la red neuronal entrenada con el cuarto conjunto de datos

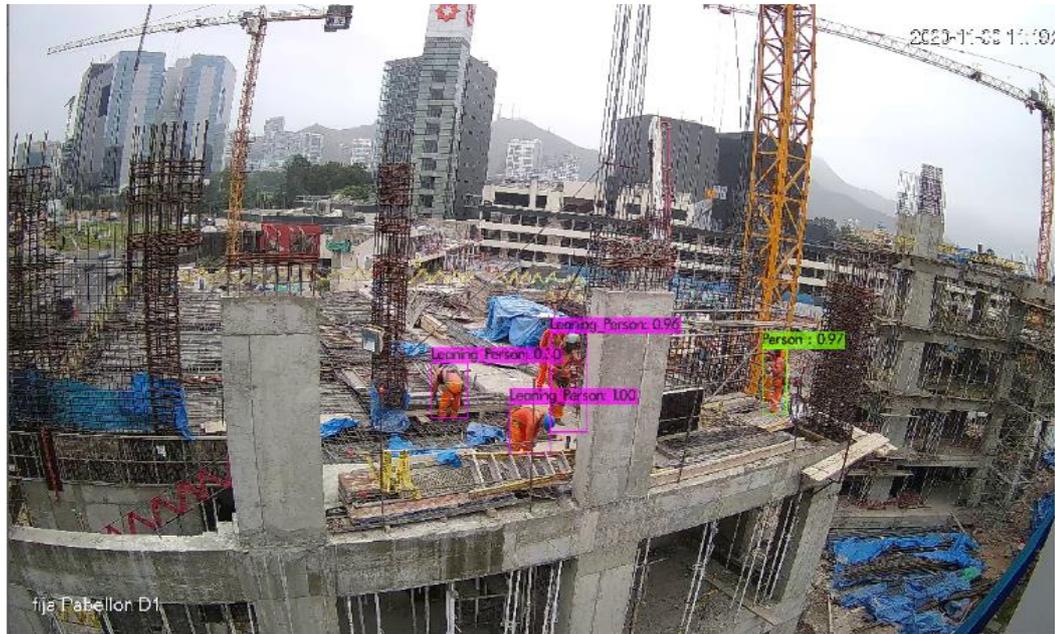


### 4.2.5 Red neuronal con conjunto de datos aumentado con todas las técnicas combinadas

Respecto a la validación con el último conjunto de datos entrenado utilizando todas las técnicas de Aumento de Datos, los resultados obtenidos se muestran en la Figura 4.8, en donde se detectaron 4 de los 5 trabajadores presentes en la imagen. Se puede observar que el trabajador N°1 fue detectado en la categoría correcta; sin embargo, con un nivel de confianza de 0.30. Luego, el trabajador N°2 se detectó correctamente en la categoría correspondiente con el mejor nivel de confianza de 1.00. De igual manera, el trabajador N°4 fue detectado de manera correcta en la categoría *Leaning Person*, con un nivel de confianza alto de 0.96. Por último, se observa que el trabajador N°5 se detectó también de manera correcta en la categoría *Person* con un nivel de confianza alto de 0.97. De los resultados obtenidos, se concluye que el uso de todas las técnicas de aumento de datos generó mejores resultados que los conjuntos de datos restantes.

## Figura 4.8

*Validación de la red neuronal entrenada con el quinto conjunto de datos*



A continuación, en la Tabla 4.7 se muestra la comparación de los niveles de confianza de cada trabajador detectado por cada red neuronal. En esta tabla se puede visualizar que el conjunto de datos aumentado con el uso de todas las técnicas obtuvo mejores niveles de confianza que los casos restantes. Por otro lado, en la Figura 4.9, según lo desarrollado en el estudio de Asif et al. (2023), se muestran los valores máximos de niveles de confianza obtenidos para cada red neuronal. De estos resultados, se observa que con el uso de técnicas de Aumento de Datos se mejoró los valores máximos de niveles de confianza detectados en cada imagen, siendo la red neuronal con todas las técnicas combinadas la que obtuvo el mayor valor de 100%, mientras que la red neuronal sin ninguna técnica de Aumento de Datos obtuvo un valor máximo de 97%.

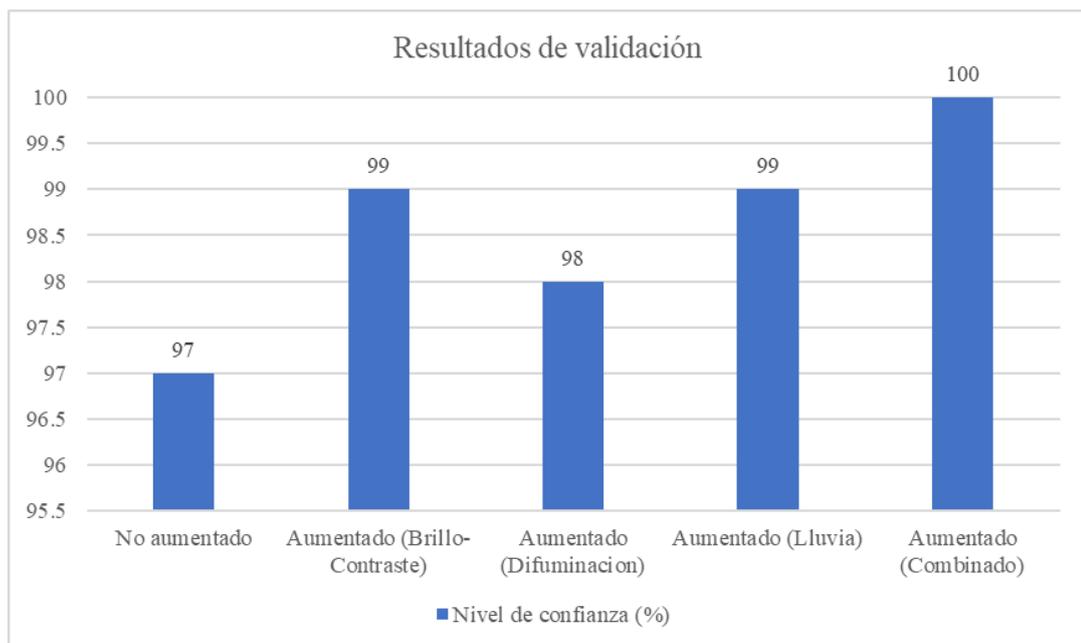
**Tabla 4.7**

*Niveles de confianza en cada detección de las redes neuronales durante la etapa de validación en la obra CEBUL de la Universidad de Lima en Lima, Perú*

Nº detección	Red neuronal con cada conjunto de datos				
	No aumentado	Brillo- Contraste	Difuminación	Lluvia	Combinado
1	No detectado	0.25	0.35	No detectado	0.3
2	0.98	0.99	0.97	0.99	1
3	0.88	0.38	0.45	0.62	0.96
4	No detectado	0.28	No detectado	No detectado	No detectado
5	0.97	0.96	0.98	0.97	0.97

**Figura 4.9**

*Gráfico de máximos niveles de confianza de cada red neuronal*



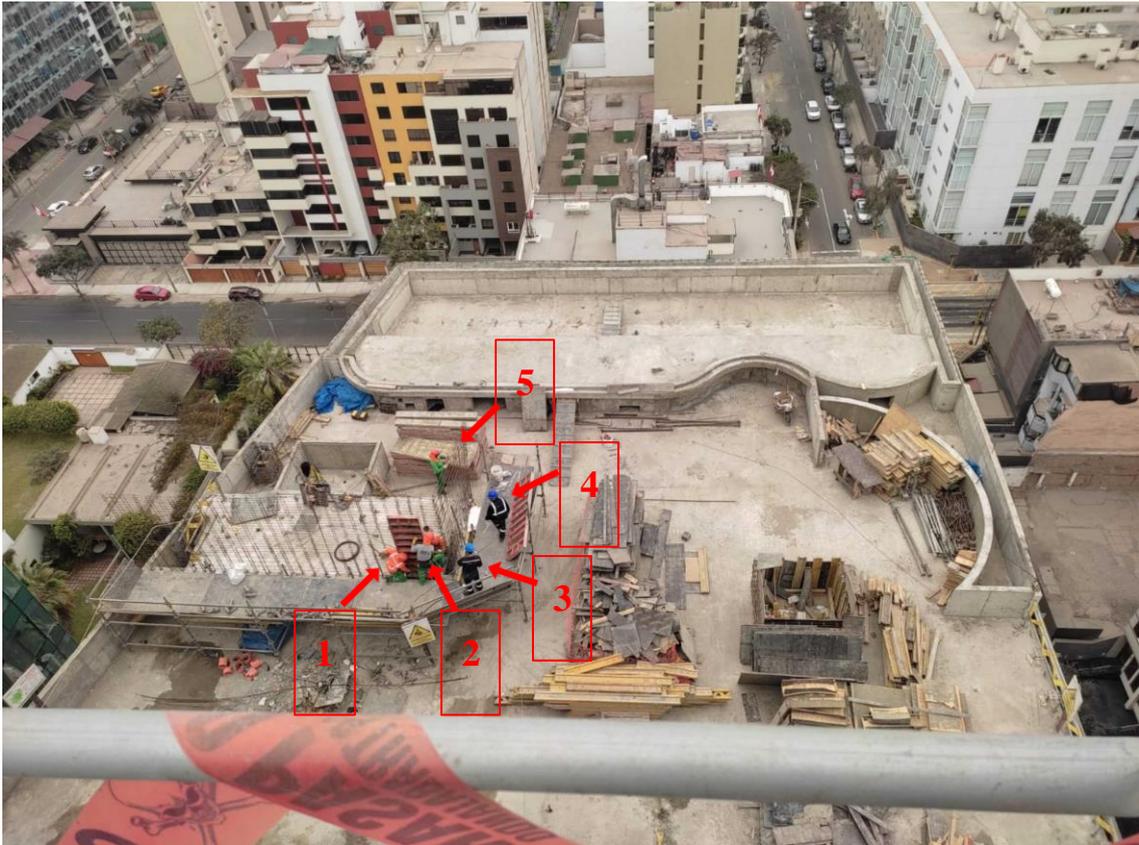
### 4.3 Resultados de validación en una obra distinta

Luego de haber realizado la validación de cada red neuronal, se procedió a evaluar los resultados de cada una en una misma imagen tomada de una obra de construcción distinta, mostrada en la Figura 4.10, en donde se ha asignado un número de detección a cada trabajador. Asimismo, la categoría correcta de clasificación de cada trabajador se muestra en la Tabla 4.8. Luego, la imagen tomada fue transformada con las cuatro técnicas de

Aumento de Datos escogidas con el fin de evaluar el desempeño de cada red neuronal sometiénolas a un mayor grado de dificultad.

**Figura 4.10**

*Imagen de una obra de construcción de una edificación ubicada en el distrito de Miraflores de Lima, Perú*



**Tabla 4.8**

*Clasificación de cada trabajador de una edificación ubicada en el distrito de Miraflores de Lima, Perú*

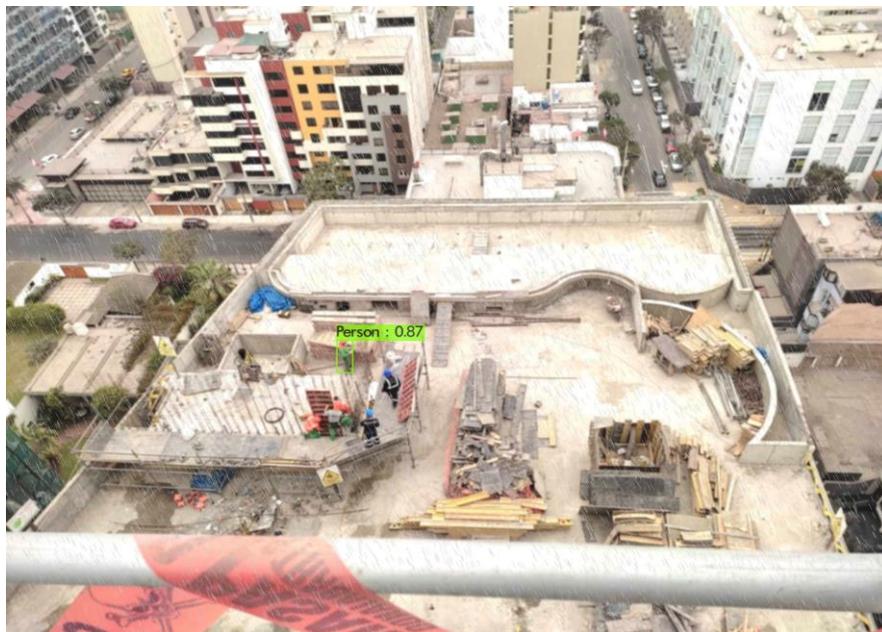
N° de trabajador	Categoría
1	Leaning person
2	Leaning person
3	Person
4	Person
5	Person

#### 4.3.1 Red neuronal con conjunto de datos no aumentado

De los resultados obtenidos, en la Figura 4.11 se puede observar que la red neuronal entrenada con el conjunto de imágenes originales solo realizó 1 detección, lo cual indicaría que esta red no es confiable para realizar detecciones en diferentes escenarios. De igual forma, este resultado es congruente con los resultados de entrenamiento obtenidos para esta red, en la que se observa un menor valor-F1 respecto a los entrenamientos con los conjuntos de datos restantes.

#### **Figura 4.11**

*Validación de la red neuronal entrenada con el conjunto de datos original en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú*



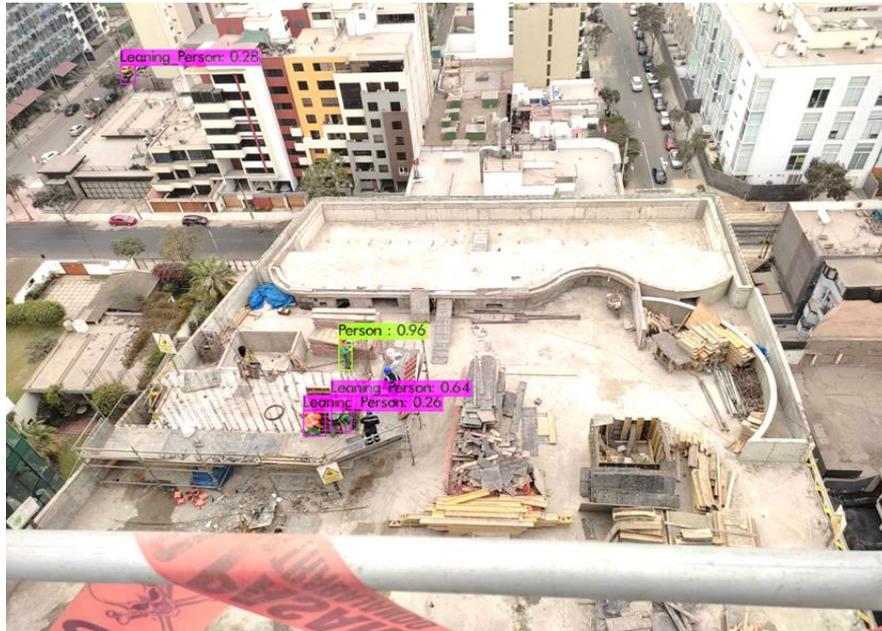
#### **4.3.2 Red neuronal con conjunto de datos aumentado con la técnica Brillo-Contraste**

Por otro lado, los resultados de la aplicación de las técnicas de Aumento de Datos muestran mejores resultados que a comparación del caso anterior. Respecto a los resultados del uso de la técnica de Brillo - Contraste, los cuales se muestran en la Figura 4.12, se detectaron 3 trabajadores clasificados correctamente; sin embargo, no se detectaron dos trabajadores pertenecientes a la categoría *Person* (3 y 4) y se puede observar que la red neuronal realizó una detección incorrecta en el extremo de la imagen. El primer trabajador se detectó con un nivel de confianza de 0.26, el segundo con 0.64 y el quinto con un nivel de confianza de 0.96. En ese sentido, se puede afirmar que esta

red neuronal es capaz de detectar ambas categorías, *Leaning Person* y *Person*; aunque aún presente dificultades de recuperación cuando se tiene imágenes nuevas.

#### Figura 4.12

Validación de la red neuronal entrenada con el segundo conjunto de datos en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú

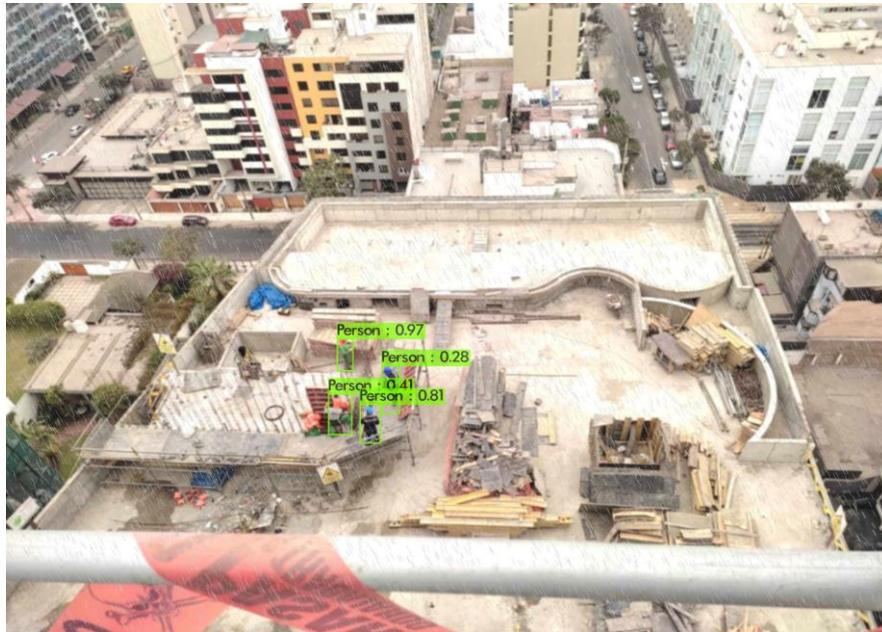


#### 4.3.3 Red neuronal con conjunto de datos aumentado con la técnica Difuminación

En el tercer caso, se presentan mejores resultados que en los casos anteriores, mostrados en la Figura 4.13. La técnica de Difuminación logró mejorar el número de detecciones realizadas; es decir, se mejoró el valor de recuperación a 4 de 5 trabajadores. Así mismo, 3 de las 4 detecciones realizadas pertenecen a la categoría correcta, siendo errónea la segunda detección, la cual pertenece a la categoría *Leaning Person*. Esta última categoría no ha sido detectada en ninguno de los trabajadores correspondientes (1 y 2). Respecto a los niveles de confianza, el segundo trabajador se detectó con un nivel de 0.41, el tercero con 0.81, el cuarto con 0.28 y el quinto con 0.97. Por lo tanto, por los resultados anteriores, se considera que la red neuronal es más confiable para detectar trabajadores en contextos distintos de una obra de construcción.

### Figura 4.13

Validación de la red neuronal entrenada con el tercer conjunto de datos en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú



#### 4.3.4 Red neuronal con conjunto de datos aumentado con la técnica Lluvia

Luego, según lo mostrado en la Figura 4.14, en el cuarto caso se obtuvieron resultados similares al segundo ya que se detectaron solo 3 trabajadores en la categoría correcta y en el extremo de la imagen se detectó dos objetos que no pertenecen a ninguna de las categorías. Respecto al nivel de confianza, la técnica de Lluvia mejoró notablemente este valor a comparación del segundo y primer caso. El tercer trabajador se detectó con un nivel de confianza de 0.68, el cuarto con 0.84 y el quinto con 0.97. Por último, de igual manera que en el caso anterior, la categoría *Leaning Person* no ha sido detectada en ninguno de los trabajadores correspondientes (1 y 2). En este sentido, esta red se considera parcialmente confiable para la detección de trabajadores en situaciones nuevas, pues se debe mejorar la red neuronal para que presente un mayor nivel de recuperación de trabajadores.

#### Figura 4.14

Validación de la red neuronal entrenada con el cuarto conjunto de datos en una obra de edificación ubicada en el distrito de Miraflores de Lima, Perú

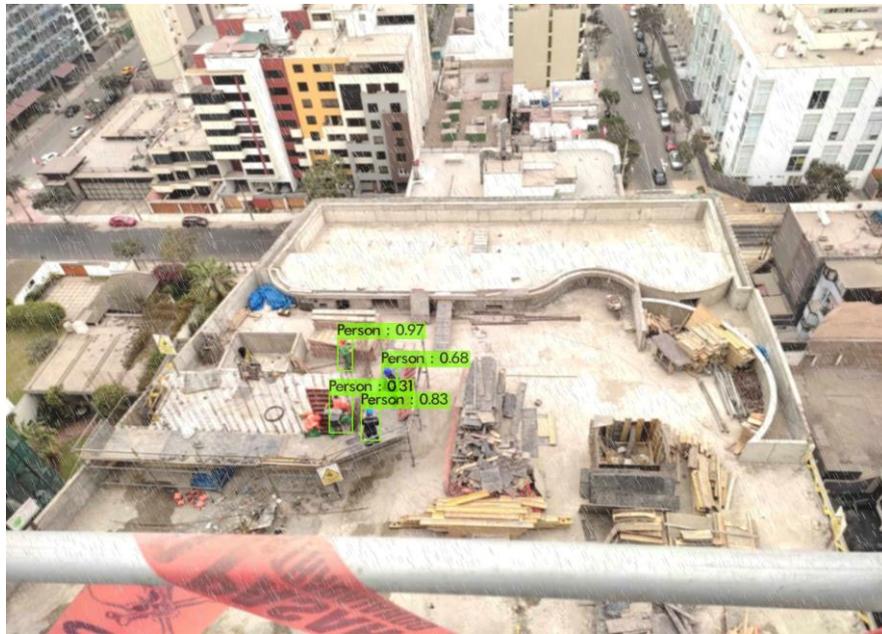


#### 4.3.5 Red neuronal con conjunto de datos aumentado con todas las técnicas combinadas

Respecto a los resultados del último caso, mostrados en la Figura 4.15, la red neuronal entrenada con todas las técnicas presenta resultados más favorables que en todos los casos anteriores. De los 5 trabajadores presentes, la red neuronal detectó 4 de ellos con niveles de confianza altos, excepto por la segunda detección la cual correspondería a la categoría *Leaning Person*. Esta última no ha sido correctamente detectada en la mayoría de los casos, por lo que se asume que esta categoría no ha estado tan presente en las imágenes durante los entrenamientos a comparación de la categoría *Person*. En ese sentido, el segundo trabajador se detectó con un nivel de confianza de 0.31, el tercero con 0.83, el cuarto con 0.68 y el quinto con 0.97. En contraste con el tercer caso la mayoría de las detecciones realizadas tienen un mejor nivel de confianza, por lo que se considera que esta red neuronal es la más precisa para detección y clasificación de trabajadores en distintas construcciones.

### Figura 4.15

*Resultados de validación de la red neuronal con todas las técnicas de Aumento de Datos en una obra de una edificación ubicada en el distrito de Miraflores de Lima, Perú*



En suma, en la Tabla 4.9 se muestra la comparación de los niveles de confianza de cada trabajador detectado por cada red neuronal. En esta tabla se puede visualizar que las 4 redes neuronales entrenadas con los conjuntos de datos aumentados obtuvieron mejores niveles de confianza que la red neuronal sin técnica de aumento de datos. Asimismo, en la Figura 4.16, según lo desarrollado en el estudio de Asif et al. (2023), se muestran los valores máximos de nivel de confianza obtenidos para cada red neuronal. De estos resultados, al igual que en los resultados de validación obtenidos en la Figura 4.9, se observa que con el uso de técnicas de Aumento de Datos se mejoró los valores máximos de niveles de confianza de cada red neuronal, siendo estos valores similares para cada red con técnica de Aumento de Datos. En ese sentido, el uso de las técnicas de Aumento de Datos mejoró en aproximadamente 9% el nivel de confianza cuando se realizó la validación en una obra de edificación distinta.

Por otro lado, se puede inferir que la segunda y cuarta red neuronal, las cuales contienen las técnicas de aumento de Brillo - Contraste y Lluvia respectivamente, presentan el problema de sobreentrenamiento. Esto se muestra en las imágenes de

validación del segundo y cuarto caso en donde no se han detectado dos trabajadores visibles, en cambio se detectaron objetos no pertenecientes a ninguna categoría. Este problema puede deberse a que las técnicas no han transformado de manera significativa la imagen, generando imágenes similares, lo cual puede acostumar al modelo entrenado a detectar solo ciertos parámetros. Por otro lado, la tercera red neuronal presenta resultados similares a la quinta, lo cual demuestra la efectividad de la aplicación de esta técnica en imágenes nuevas.

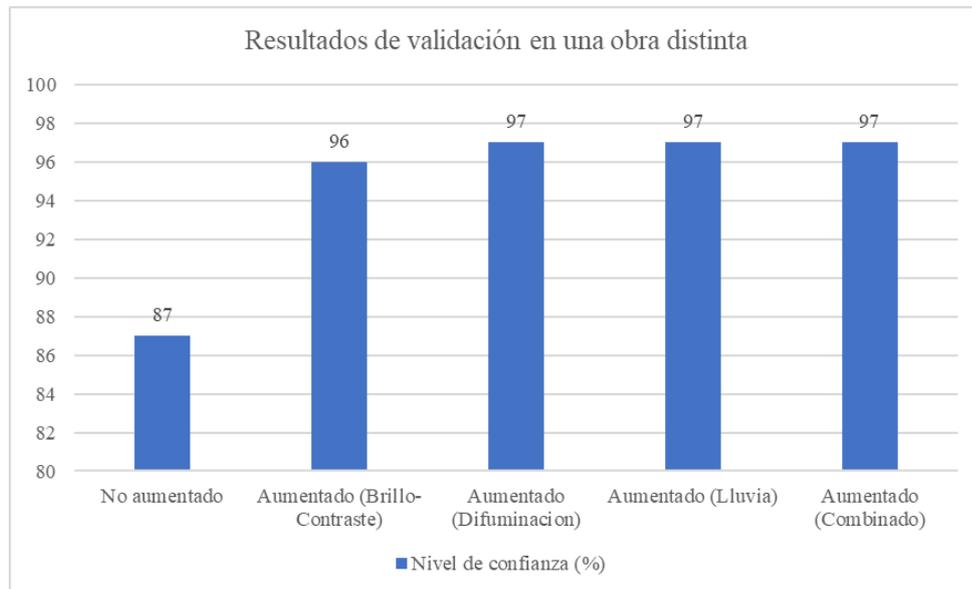
**Tabla 4.9**

*Niveles de confianza en cada detección de las redes neuronales durante la etapa de validación en una obra de una edificación ubicada en el distrito de Miraflores de Lima, Perú*

Nº detección	Red neuronal con cada conjunto de datos				
	No aumentado	Brillo- Contraste	Difuminación	Lluvia	Combinado
1	No detectado	0.26	No detectado	No detectado	No detectado
2	No detectado	0.64	0.41	No detectado	0.31
3	No detectado	No detectado	0.81	0.68	0.83
4	No detectado	No detectado	0.28	0.84	0.68
5	0.87	0.96	0.97	0.97	0.97

**Figura 4.16**

*Gráfico de máximos niveles de confianza de cada red neuronal en una obra distinta*



## CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

En la presente investigación se hizo uso de 3 técnicas de Aumento de Datos para evaluar la efectividad de estas en términos de mAP y los niveles de confianza en la detección y clasificación de trabajadores en obras de construcción, lo cual nos permite llegar al siguiente conjunto de conclusiones. Como primer punto, dado los resultados de entrenamiento obtenidos en la sección 4.1, se validó la hipótesis de que la transformación de imágenes utilizando las técnicas de Aumento de Datos mejoró el valor de mAP en 4% en la detección de trabajadores de construcción. Estos resultados son similares a los obtenidos por Asif et al. (2023), en donde se realizó un estudio similar utilizando diferentes técnicas de aumento de datos y el conjunto de datos *Alberta Construction Image Dataset* (ACID).

Como segundo punto, respecto a los resultados de validación de la sección 4.2, se observa cómo cada técnica influye de manera distinta en los niveles de confianza y capacidad de clasificación de los trabajadores. En esta sección, se puede observar que la red neuronal entrenada con el conjunto de datos aumentados utilizando todas las técnicas de aumento de datos obtuvo el mejor resultado en términos de valores de confianza en comparación a la red neuronal entrenada con el conjunto de datos sin ninguna técnica de aumento.

Como tercer punto, respecto a los resultados de validación en la imagen de una obra distinta en la sección 4.3, tanto la segunda red neuronal como la cuarta, a los cuales se les aplicó las técnicas de Brillo - Contraste y Lluvia respectivamente, mostraron aparentemente problemas de sobreentrenamiento u *Overfitting* al realizar detecciones imprecisas o no detectar trabajadores visibles en distintas posiciones. Esto demuestra que al utilizar estas técnicas se debe tomar en cuenta el grado de transformación de las imágenes para evitar que sean similares a las originales.

Asimismo, en la misma sección, la tercera red neuronal, a la cual se le aplicó la técnica de Difuminación, mostró resultados más favorables y muy similares al último de ellos. Esta técnica aplicada pudo haber influido significativamente en la transformación

de las imágenes, lo cual entrenó correctamente al modelo para nuevas detecciones en distintos ambientes. De igual manera, en una futura investigación se recomienda tomar en cuenta el grado de transformación que realiza esta técnica para estimar el rango óptimo para su aplicación. A pesar de los defectos de algunas técnicas, la aplicación en conjunto de todas estas generó mejores resultados que los casos anteriores, incrementando los niveles de confianza hasta en un 9%, los cuales pueden ser optimizados considerando el grado de transformación de cada una.

Por otro lado, la categoría de clasificación *Person* obtuvo mejores niveles de confianza que *Leaning Person* en cada resultado de validación. Esto se pudo deber a la cantidad de cada una presente en las imágenes, lo cual ocasionó un mejor entrenamiento en la primera de ellas. En futuras investigaciones, con el fin de evitar estas disparidades, se recomienda utilizar categorías de clasificación que se encuentren en proporciones similares en una imagen o uniformizar las categorías en una misma en caso cuenten con características similares. Asimismo, con respecto a la cantidad de entrenamientos, es recomendado realizar al menos 4 de ellos para cada conjunto de datos, variando aleatoriamente la distribución de las imágenes entre las carpetas de entrenamiento y validación, pues en esta investigación se demostró que se presentan resultados distintos de Precisión, Recuperación y mAP en cada variación. De igual manera, se recomienda realizar un preentrenamiento de cada red neuronal mediante técnicas como *Transfer Learning* para optimizar los resultados posteriores.

Por último, a pesar de los buenos resultados obtenidos al aplicar todas las técnicas de Aumentado de Datos en conjunto, lo cual simuló distintos escenarios de una obra de construcción, se recomienda realizar este mismo procedimiento utilizando imágenes provenientes de distintas obras para mejorar los resultados en la etapa de validación y con el fin de evitar problemas de sobreentrenamientos de las redes neuronales, pues los conjuntos de datos no cubren al 100% los diferentes escenarios de los proyectos.

El principal aporte de este estudio es el de contribuir a la poca literatura existente sobre el análisis de la efectividad del uso de técnicas de Aumento de Datos para la detección y clasificación de trabajadores de la construcción, las cuales pueden mejorar el mAP y los niveles de confianza cuando se tiene un número limitado de conjunto de datos. Además, dado que estas técnicas son muy variadas, este estudio contribuyó a analizar individualmente y en conjunto 3 de ellas, con el fin de encontrar las más adecuadas para

aplicarlas en los escenarios de construcción. Se espera que estos hallazgos sean replicados para la detección de trabajadores en futuros proyectos de construcción para los análisis que se consideren convenientes en la gestión de proyectos.

## REFERENCIAS

- Ahmad, J., Farman, H., & Jan, Z. (2019). Deep Learning Methods and Applications. *Deep Learning: Convergence to Big Data Analytics*. SpringerBriefs in Computer Science. Springer, Singapore. [https://doi.org/10.1007/978-981-13-3459-7\\_3](https://doi.org/10.1007/978-981-13-3459-7_3)
- Aichholzer, G., Rhomberg, W., Gudowsky, N., Saurwein, F., & Weber, M. (2015). *Industry 4.0. Background Paper on the pilot project 'Industry 4.0. Foresight & Technology Assessment on the social dimension of the next industrial revolution'*. Austrian Academy of Sciences.
- Alaloul, W. S., Liew, M. S., Zawawi, N. A. W. A., & Kennedy, I. B. (2020). Industrial Revolution 4.0 in the construction industry: Challenges and opportunities for stakeholders. *Ain Shams Engineering Journal*, 11 (1), 225-230. <https://doi.org/10.1016/j.asej.2019.08.010>.
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(53). <https://doi.org/10.1186/s40537-021-00444-8>
- Ameijeiras Sánchez, D., González Diez, H. R., & Hernández Heredia, Y. (01 de septiembre de 2020). Revisión de algoritmos de detección y seguimiento de objetos con redes profundas para videovigilancia inteligente. *Revista Cubana de Ciencias Informáticas*, 14(3), 165-195. [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S2227-18992020000300165](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992020000300165)
- Angah, O., & Chen, A.Y. (2020). Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy. *Automation in Construction*, 119. <https://doi.org/10.1016/j.autcon.2020.103308>.
- Arabi, S., Haghghat, A., & Sharma, A. (2020). A deep-learning-based computer vision solution for construction vehicle detection. *Computer-Aided Civil and Infrastructure Engineering*, 35(7), 753-767. <https://doi.org/10.1111/mice.12530>
- Babich, N. (2020). *What Is Computer Vision & How Does it Work? An Introduction*. <https://xd.adobe.com/ideas/principles/emerging-technology/what-is-computer-vision-how-does-it-work/>
- Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A., & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141. <https://doi.org/10.1016/j.autcon.2022.104440>

- Bang, S., Park, S., Kim, H., & Kim, H. (2019). Encoder-decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering*, 34(8), 713-727. <https://doi.org/10.1111/mice.12440>
- Bang, S., Baek, F., Park, S., Kim, W., & Kim, H. (2020). Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques. *Automation in Construction*, 115, <https://doi.org/10.1016/j.autcon.2020.103198>
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S., Akinade, O., Owolabi, H., Alaka, H., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced engineering informatics*, 30(3), 500-521. <https://doi.org/10.1016/j.aei.2016.07.001>
- Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. <https://doi.org/10.48550/arXiv.2004.10934>
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. *Information*, 11(2), 125. <https://doi.org/10.3390/info11020125>
- Caña, H., Mula, J., Díaz-Madroñero, M., & Campuzano-Bolarín, F. (2021). Implementing Industry 4.0 principles. *Computers & Industrial Engineering*, 158. <https://doi.org/10.1016/j.cie.2021.107379>
- Guo, B. H., Zou, Y., Fang, Y., Goh, Y. M., & Zou, P. X. (2021). Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety Science*, 135. <https://doi.org/10.1016/j.ssci.2020.105130>
- Cheng, C.F., Rashidi, A., Davenport, M.A., & Anderson, D.V. (2017). Activity analysis of construction equipment using audio signals and support vector machines. *Automation in Construction*, 81, 240–253. <https://doi.org/10.1016/j.autcon.2017.06.005>
- Cogswell, M., Ahmed, F., Girshick, R., Zitnick, L., & Batra D. (2015). Reducing overfitting in deep networks by decorrelating representations. *arXiv preprint arXiv:1511.06068*. <https://doi.org/10.48550/arXiv.1511.06068>
- Darko, A., Chan, A. P. C., Adabre, M. A., Edwards, D. J., Hosseini, M. R., & Ameyaw, E. E. (2020). Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, 112, 103081. <https://doi.org/10.1016/j.autcon.2020.103081>
- Del Savio, A., Luna, A., Cárdenas-Salas, D., Vergara, M., & Urday, G. (2022). Dataset of manually classified images obtained from a construction site. *Data in Brief*, 42. <https://doi.org/10.1016/j.dib.2022.108042>
- Del Savio, A., Luna Torres, A., Cárdenas-Salas, D., Vergara Olivera, M. A., & Urday Ibarra, G. T. (2023). Artificial Intelligence Applied to the Control and Monitoring of Construction Site Personnel. *Advances in Mechanics of Materials for*

*Environmental and Civil Engineering* (pp. 19-29). Revista: Springer International Publishing. [https://doi.org/10.1007/978-3-031-37101-1\\_2](https://doi.org/10.1007/978-3-031-37101-1_2)

- Dikhanbayeva, D.; Shaikholla, S.; Suleiman, Z., & Turkyilmaz, A. (2020). Assessment of Industry 4.0 Maturity Models by Design Principles. *Sustainability*, 12(23). <https://doi.org/10.3390/su12239927>
- Dung, C. V., Sekiya, H., Hirano, S., Okatani, T., & Miki, C. (2019). A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Automation in Construction*, 102, 217-229. <https://doi.org/10.1016/j.autcon.2019.02.013>
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., & Rose, T. M. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 85, 1–9. <https://doi.org/10.1016/j.autcon.2017.09.018>.
- Fang, W., Ding, L., Love, P. E. D., Luo, H., Li, H., Peña-Mora, F., Zhong, B., & Zhou, C. (2020). Computer vision applications in construction safety assurance. *Automation in Construction*, 110. <https://doi.org/10.1016/j.autcon.2019.103013>
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X, Wang, G. Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377. <https://doi.org/10.1016/j.patcog.2017.10.013>
- Hermann, M., Pentek T., & Otto, B. (2016, 10 de marzo). *Design Principles for Industrie 4.0 Scenarios* [Conferencia]. 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA. <https://doi.org/10.1109/HICSS.2016.488>
- Hernández, R., Fernández C., & Baptista, P. (2014). *Metodología de la investigación* (6.<sup>a</sup> Ed.). McGraw-Hill
- Hu, X., Liu, Y., Zhao, Z., Liu, J., Yang, X., Sun, C., Chen, S., Li, B., & Zhou, C. (2021). Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network. *Computers and Electronics in Agriculture*, 185. <https://doi.org/10.1016/j.compag.2021.106135>
- Huang, L., Fu, Q., He, M., Jiang, D., & Hao, Z. (2021). Detection algorithm of safety helmet wearing based on deep learning. *Concurrency and Computation: Practice and Experience*, 33(13). <https://doi.org/10.1002/cpe.6234>
- Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022, 9-11 de julio). *A Review of Yolo algorithm developments*. [Conferencia]. The 8th International Conference on Information Technology and Quantitative Management. (ITQM 2020 & 2021), Chengdu, China. <https://doi.org/10.1016/j.procs.2022.01.135>
- Kim, H., Kim, H., Hong, Y. W., & Byun, H. (2018). Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. *Journal of Computing in Civil Engineering*, 32(2). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000731](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731)

- Kim, J., Hwang, J., Chi, S., & Seo, J. (2020). Towards database-free vision-based monitoring on construction sites: A deep active learning approach. *Automation in Construction*, 120. <https://doi.org/10.1016/j.autcon.2020.103376>
- Kim, K., & Jeong, J. (2020, 9-12 de agosto). *Deep Learning-based Data Augmentation for Hydraulic Condition Monitoring System*. [Conferencia]. The 17th International Conference on Mobile Systems and Pervasive Computing (MobiSPC), Leuven, Belgica. <https://doi.org/10.1016/j.procs.2020.07.007>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521 (7553): 436–444. <https://doi.org/10.1038/nature14539>.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., & Berg, A. C. (2016). *SSD: Single Shot Multibox Detector* [Conferencia]. European Conference on Computer Vision. Tel Aviv, Israel. <https://doi.org/10.48550/arXiv.1512.02325>
- Lu, Y., & Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2020.105760>.
- Liu, Y., Yeoh, J., & Chua, D. (2020). Deep Learning-Based Enhancement of Motion Blurred UAV Concrete Crack Images. *Journal of Computing in Civil Engineering*, 34. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000907](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000907)
- Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep Learning Techniques: An Overview. *Advances in Intelligent Systems and Computing*. [https://doi.org/10.1007/978-981-15-3383-9\\_54](https://doi.org/10.1007/978-981-15-3383-9_54)
- Modarres, C., Astorga, N., Droguett, E. L., & Meruane, V. (2018). Convolutional neural networks for automated damage recognition and damage type identification. *Structural Control and Health Monitoring*, 25(10). <https://doi.org/10.1002/stc.2230>
- Mohammadpour, A., Karan, E., & Asadi, S. (2019). *Artificial Intelligence Techniques to Support Design and Construction* [Conferencia]. Proceedings of the International Symposium on Automation and Robotics in Construction. Banff, Canadá. <https://doi.org/10.22260/ISARC2019/0172>
- Montaser, A., & Moselhi, O. (2014). Truck + for earthmoving operations. *Journal of Information Technology in Construction*, 19, 412-433. <http://www.itcon.org/2014/25>.
- Na, S., Heo, S., Han, S., Shin, Y., & Lee, M. (2022). Development of an Artificial Intelligence Model to Recognise Construction Waste by Applying Image Data Augmentation and Transfer Learning. *Buildings*, 12 (2), 175. <https://doi.org/10.3390/buildings12020175>
- Nath, N.D., Akhavian, R., & Behzadan, A.H. (2017). Ergonomic analysis of construction worker's body postures using wearable mobile sensors. *Applied Ergonomics*. 62, 107-117. <https://doi.org/10.1016/j.apergo.2017.02.007>

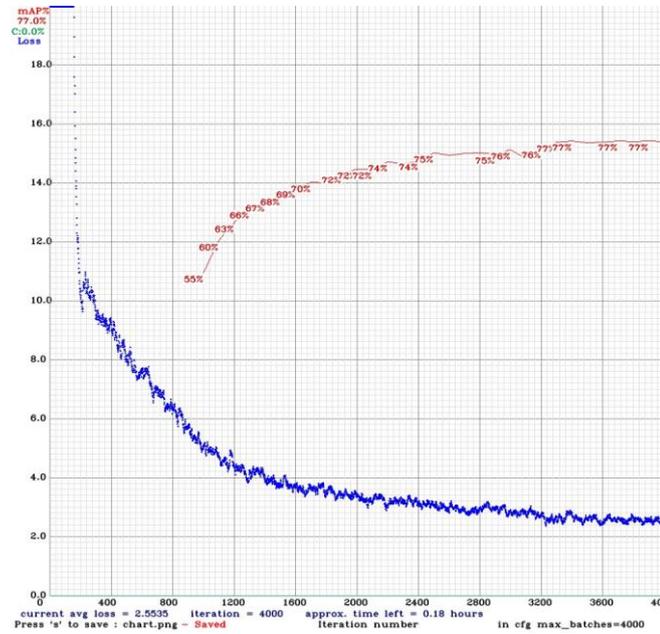
- Okpala, I., Nnaji, C., & Karakhan, A. A. (2020). Utilizing Emerging Technologies for Construction Safety Risk Mitigation. *Practice Periodical on Structural Design and Construction*, 25(2). [https://doi.org/10.1061/\(ASCE\)SC.1943-5576.0000468](https://doi.org/10.1061/(ASCE)SC.1943-5576.0000468)
- Padilla, R., Netto, S. L., & Da Silva, E. A. B. (2020). *A Survey on Performance Metrics for Object-Detection Algorithms* [Conferencia]. International Conference on Systems, Signals and Image Processing (IWSSIP). Niterói, Brazil. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>.
- Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122. <https://doi.org/10.1016/j.autcon.2020.103517>
- Park, M.W., & Brilakis, I. (2012). Construction worker detection in video frames for initializing vision trackers. *Automation in Construction*, 28, 15-25. <https://doi.org/10.1016/j.autcon.2012.06.001>.
- Prusty, M. R., Tripathi, V., & Dubey, A. (2021). A novel data augmentation approach for mask detection using deep transfer learning. *Intelligence-Based Medicine*, 5. <https://doi.org/10.1016/j.ibmed.2021.100037>
- Qin, H., Zhang, D., Tang, Y., & Wang, Y. (2021). Automatic recognition of tunnel lining elements from GPR images using deep convolutional networks with data augmentation. *Automation in Construction*, 130. <https://doi.org/10.1016/j.autcon.2021.103830>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection* [Conferencia]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Estados Unidos. <https://doi.org/10.48550/arXiv.1506.02640>
- Ren, S., He, K., R. Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Sacks, R., Girolami, M., & Brilakis, I. (2020). Building Information Modelling, Artificial Intelligence and Construction Tech. *Developments in the Built Environment*, 4. <https://doi.org/10.1016/j.dibe.2020.100011>
- Schönbeck, P., Löfsjögård, M., & Ansell, A. (2021). Collaboration and knowledge exchange possibilities between industry and construction 4.0 research. *Procedia Computer Science*, 192, 129-137. <https://doi.org/10.1016/j.procs.2021.08.014>.
- Shen, J., Xiong, X., Li, Y., He, W., Li, P., & Zheng, X. (2020). Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. *Computer-Aided Civil and Infrastructure Engineering*, 36(1), 180-196. <https://doi.org/10.1111/mice.12579>
- Shorten, C., & Khoshgoftaar, T. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*. 6. <https://doi.org/10.1186/s40537-019-0197-0>

- Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, 4, 379-391. <https://doi.org/10.1016/j.ptlrs.2021.05.009>
- Sohn, K., Berthelot, D., Li, C.-L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., & Raffel, C. (2020). FixMatch: Simplifying semi-supervised learning with consistency and confidence. *ArXiv*. <http://arxiv.org/abs/2001.07685>
- Son, H., & Kim, C. (2021). Integrated worker detection and tracking for the safe operation of construction machinery. *Automation in Construction*, 126. <https://doi.org/10.1016/j.autcon.2021.103670>
- Srivastava, S., Divekar, A.V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, 8(66). <https://doi.org/10.1186/s40537-021-00434-w>.
- Turk, Ž. (2023). Structured analysis of ICT adoption in the European construction industry. *International Journal of Construction Management*, 5, 756-762. <https://doi.org/10.1080/15623599.2021.1925396>.
- Turner, C., Oyekan, J., Stergioulas, L., & Griffin, D. (2020). Utilizing Industry 4.0 on the Construction Site: Challenges and opportunities. *IEEE Transactions on Industrial Informatics*, 17, 746-756. <https://doi.org/10.1109/TII.2020.3002197>
- Tzatalin, D. (2015). *LabelImg*. GitHub repository. <https://github.com/heartexlabs/labelImg>.
- Xiao, B., & Kang, S.-C. (2021). Development of an Image Data Set of Construction Machines for Deep Learning Object Detection. *Journal of Computing in Civil Engineering*, 35(2). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000945](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945).
- Xiao, B., Zhang, Y., Chen, Y., & Yin, X. (2021). A semi-supervised learning detection method for vision-based monitoring of construction sites by integrating teacher-student networks and data augmentation. *Advanced Engineering Informatics*, 50. <https://doi.org/10.1016/j.aei.2021.101372>.
- Yan, X., Zhang, H., & Li, H. (2020). Computer vision-based recognition of 3D relationship between construction entities for monitoring struck-by accidents. *Computer-Aided Civil and Infrastructure Engineering*, 35 (9), 1023-1038. <https://doi.org/10.1111/mice.12536>
- Yang, J., Park, M.-W., Vela, P.A., & Golparvar-Fard, M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. *Advanced Engineering Informatics*, 29, 211–224. <https://doi.org/10.1016/j.aei.2015.01.011>
- Zhao, Z.-Q., Zheng, P., S.-T. X., & X. Wu. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30 (11), 3212–3232. <https://doi.org/10.1109/TNNLS.2018.2876865>.

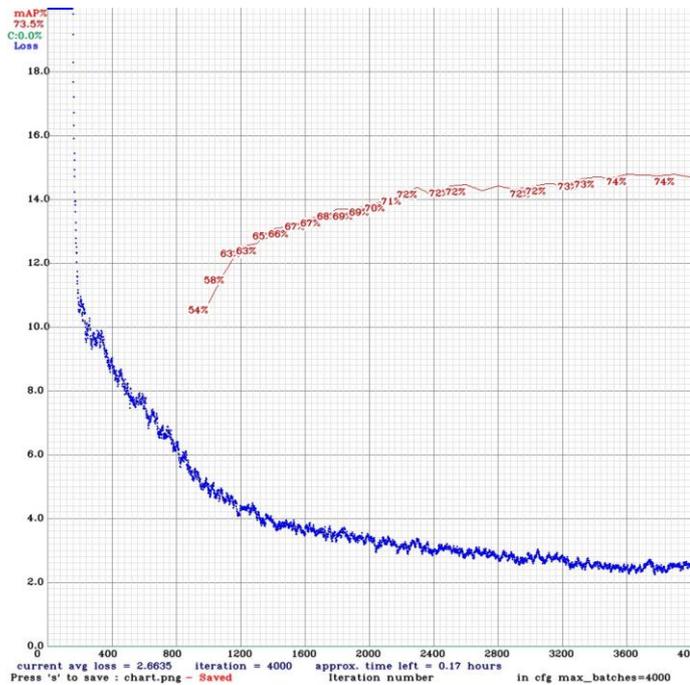
## **ANEXOS**

# Anexo 1: Gráficas de Resultados de la Tabla N° 4.1

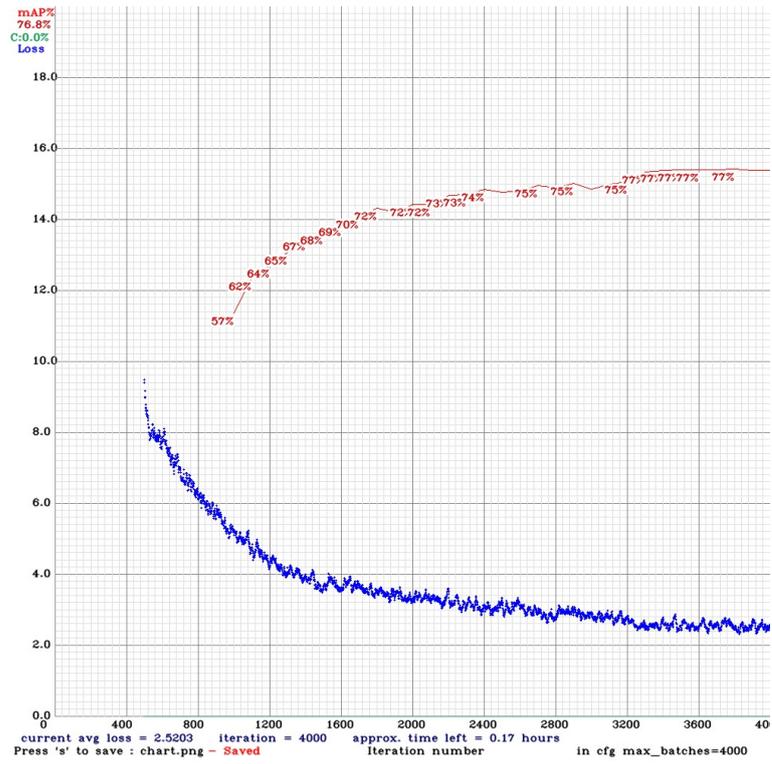
- Primer entrenamiento



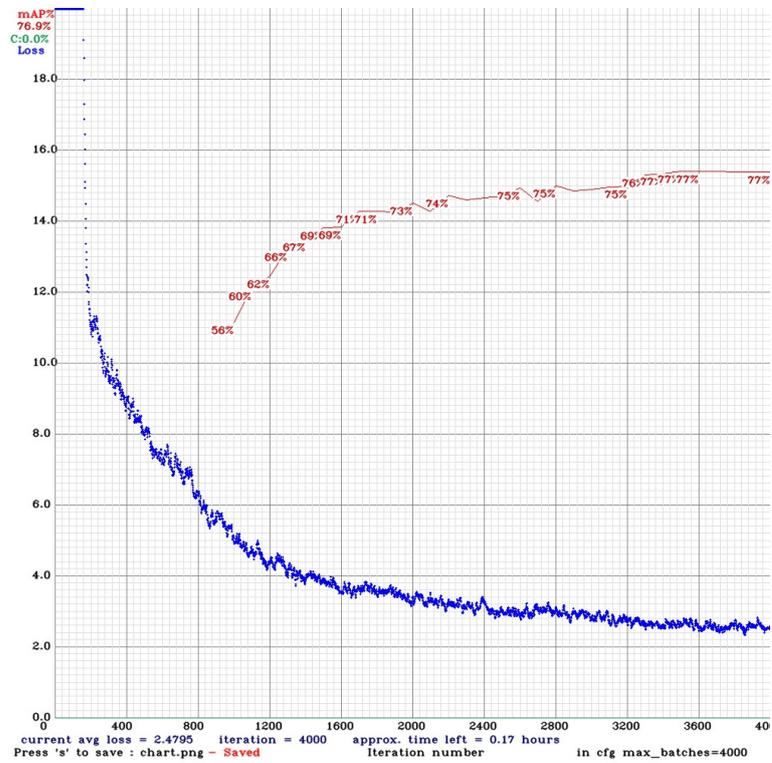
- Segundo entrenamiento



- Tercer entrenamiento

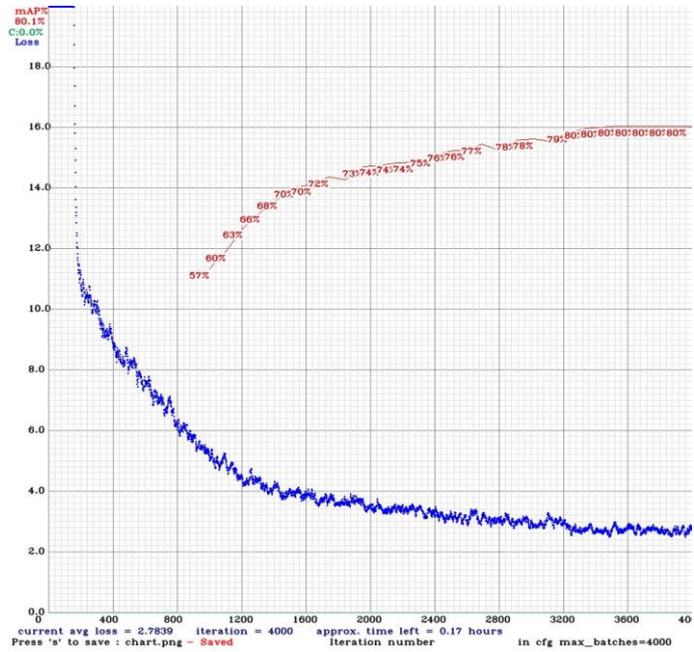


- Cuarto entrenamiento

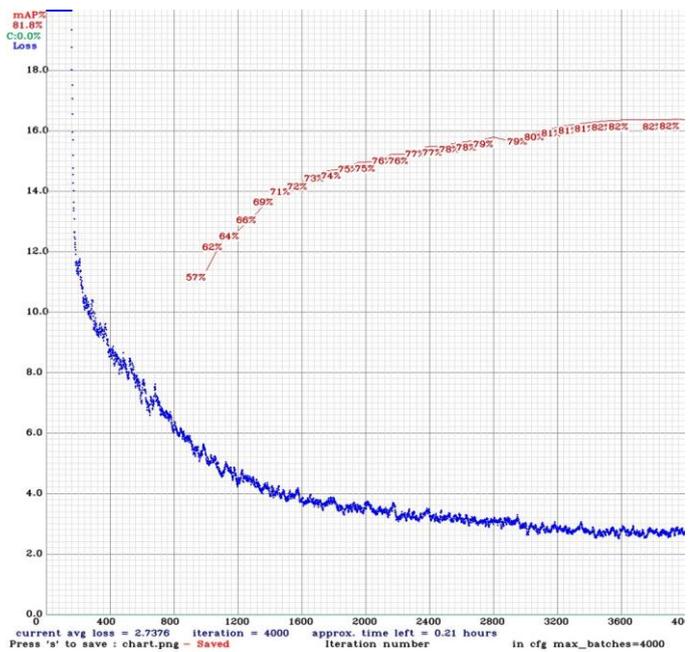


## Anexo 2: Gráficas de Resultados de la Tabla N° 4.2

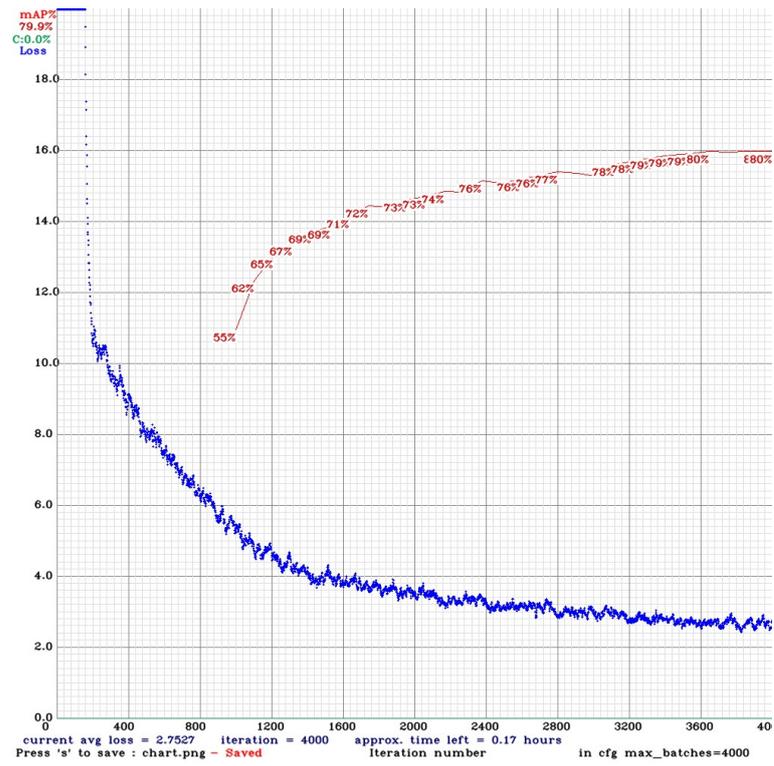
- Primer entrenamiento



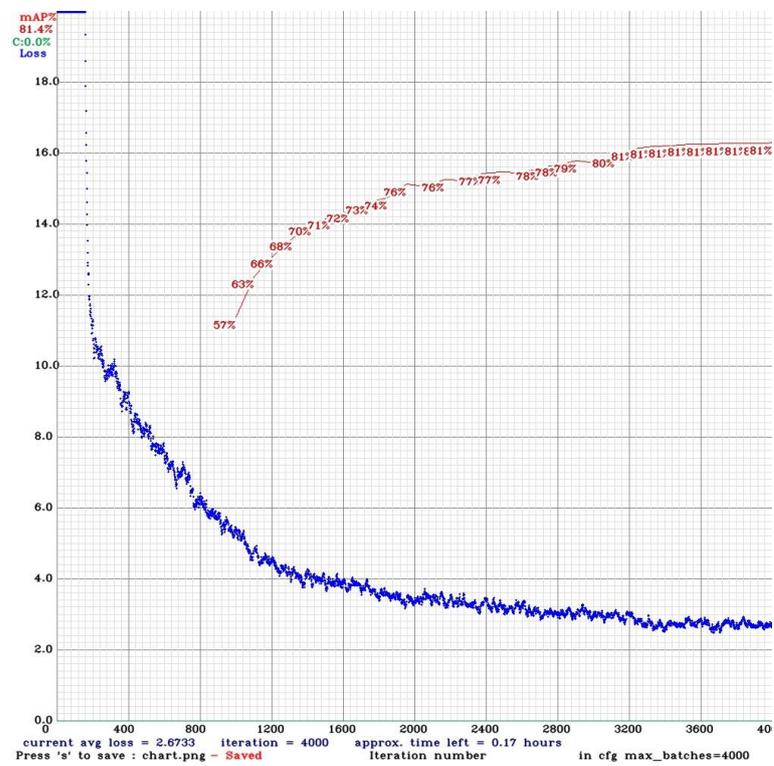
- Segundo entrenamiento



- Tercer entrenamiento

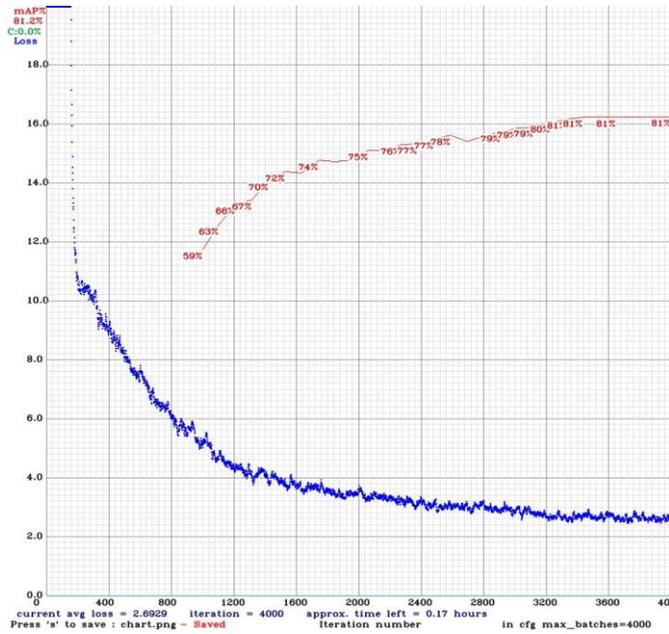


- Cuarto entrenamiento

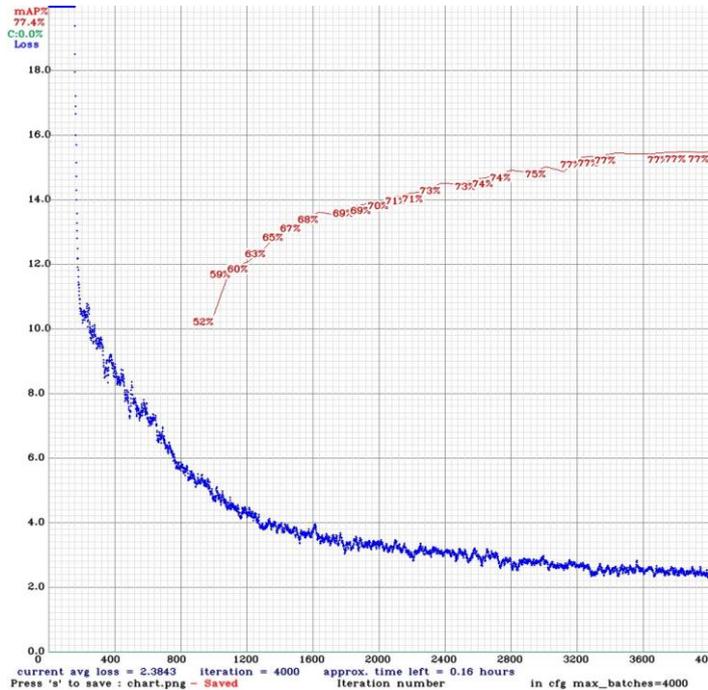


## Anexo 3: Gráficas de Resultados de la Tabla N° 4.3

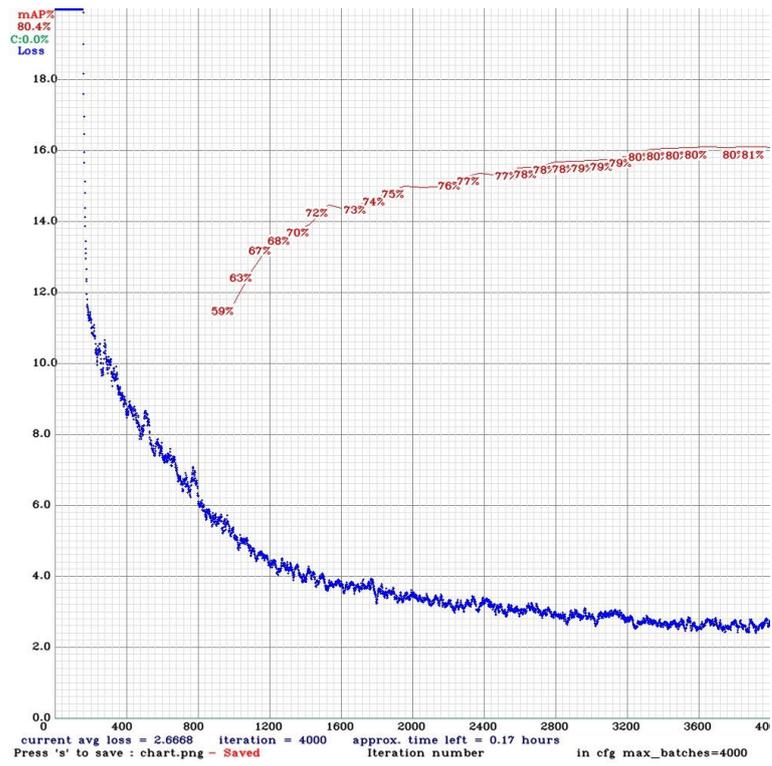
- Primer entrenamiento



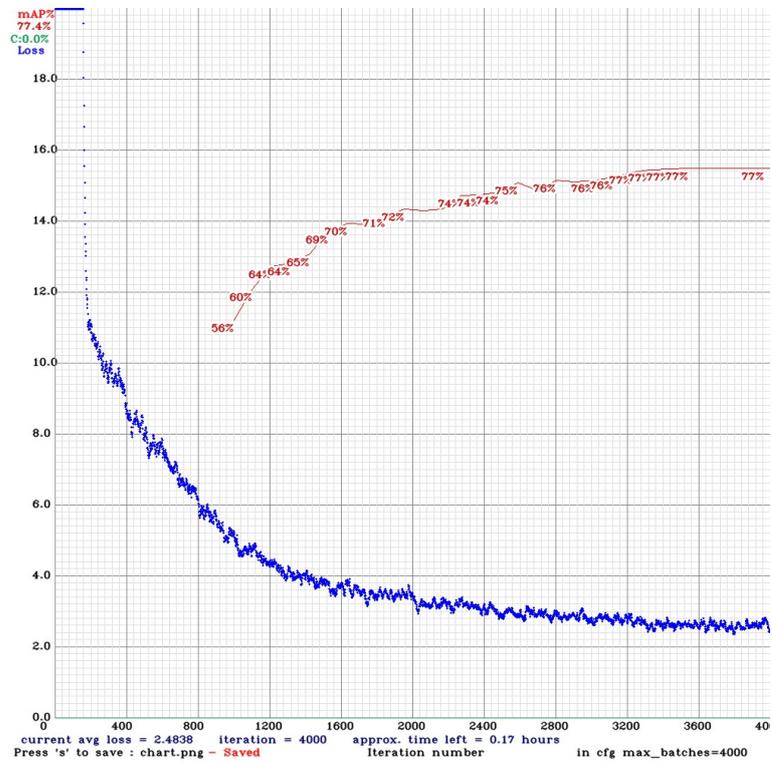
- Segundo entrenamiento



- Tercer entrenamiento

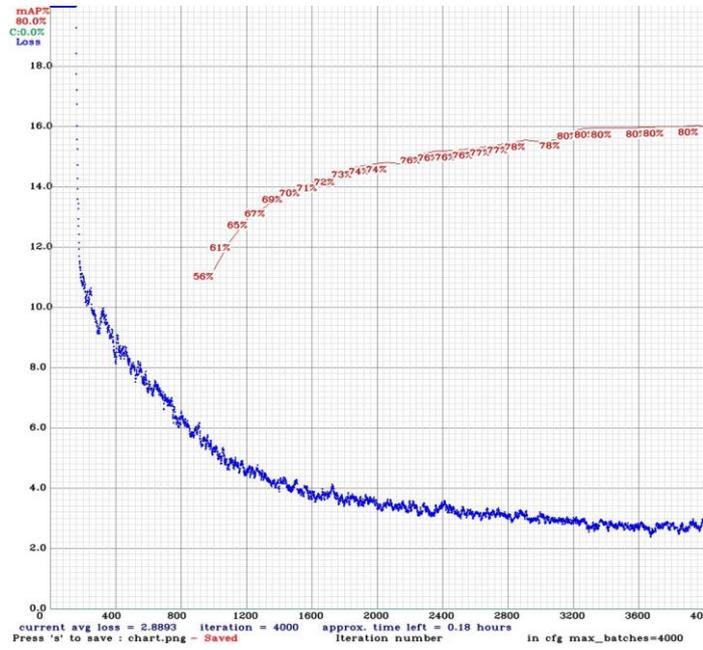


- Cuarto entrenamiento

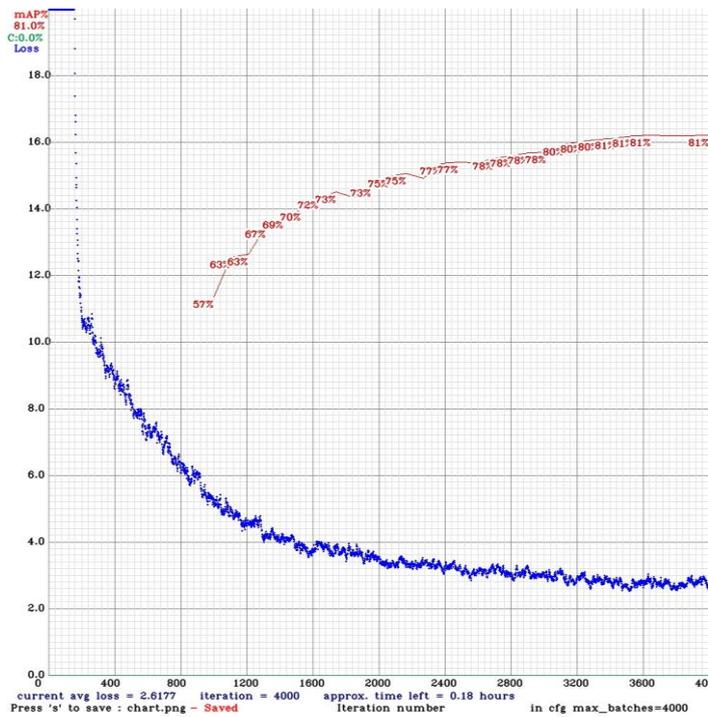


## Anexo 4: Gráficas de Resultados de la Tabla N° 4.4

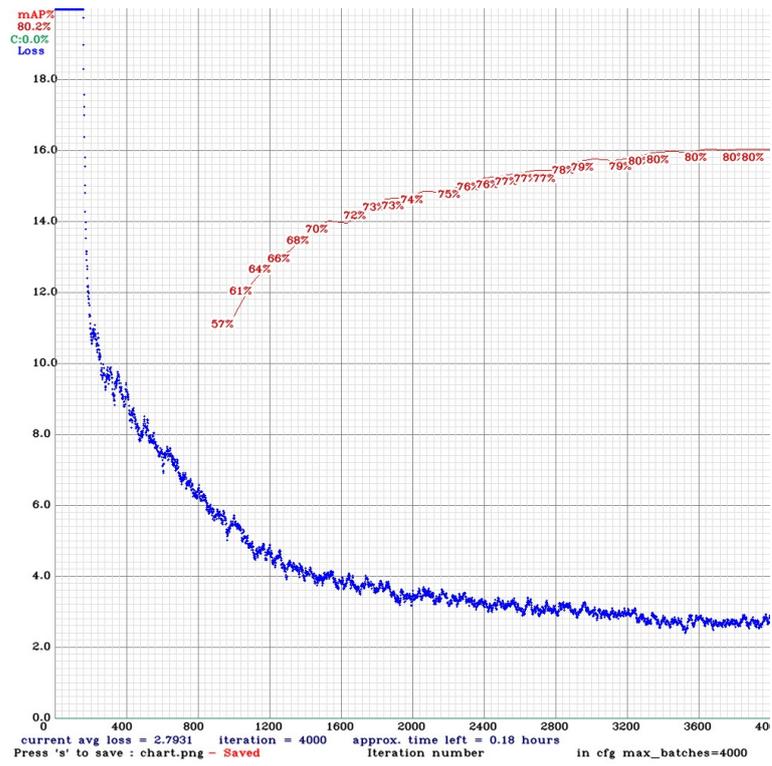
- Primer entrenamiento



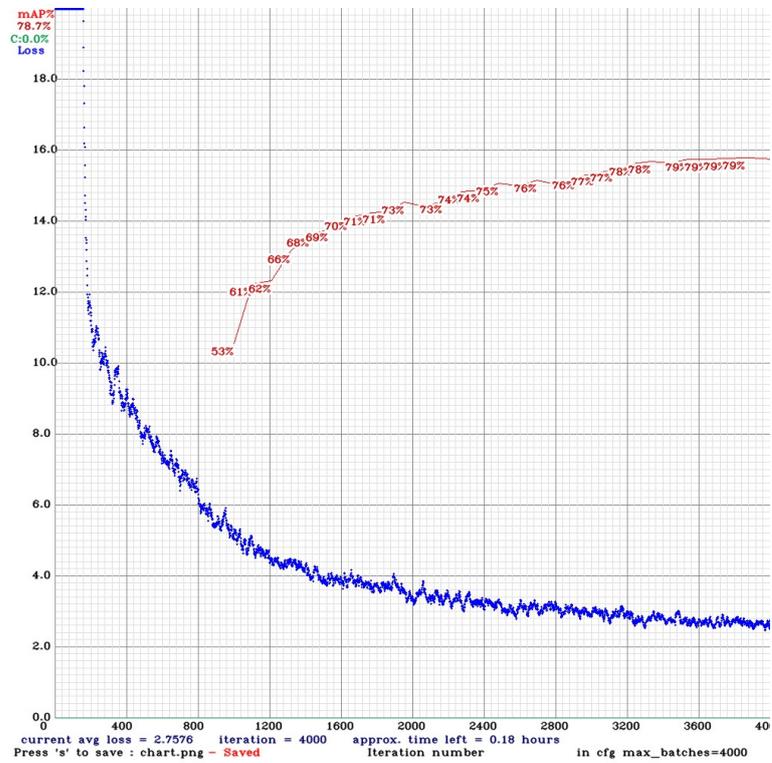
- Segundo entrenamiento



- Tercer entrenamiento

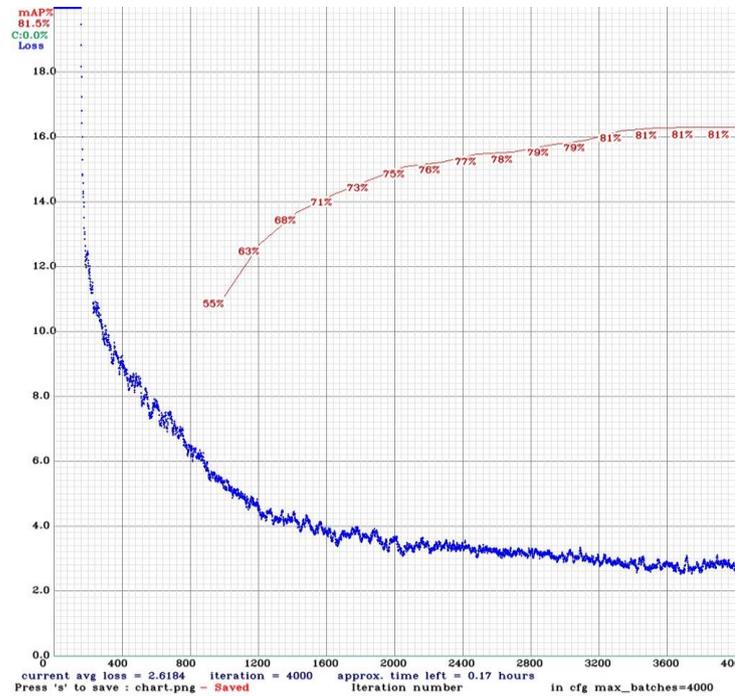


- Cuarto entrenamiento

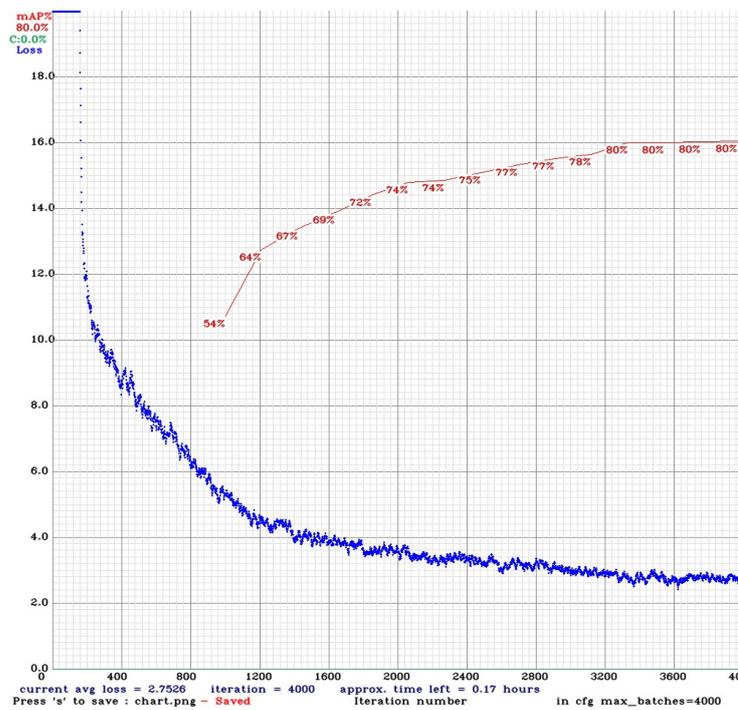


## Anexo 5: Gráficas de Resultados de la Tabla N° 4.5

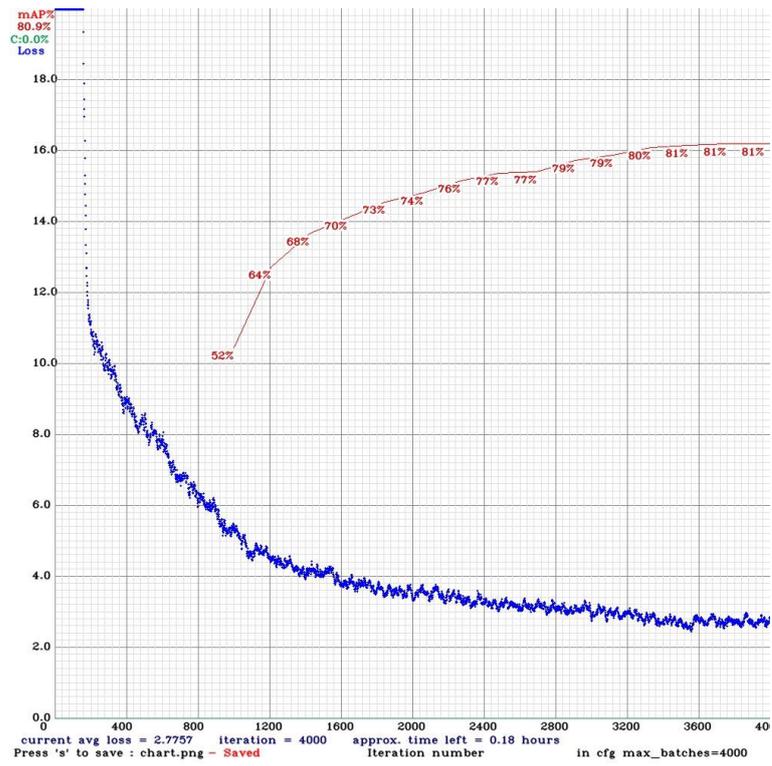
- **Primer entrenamiento**



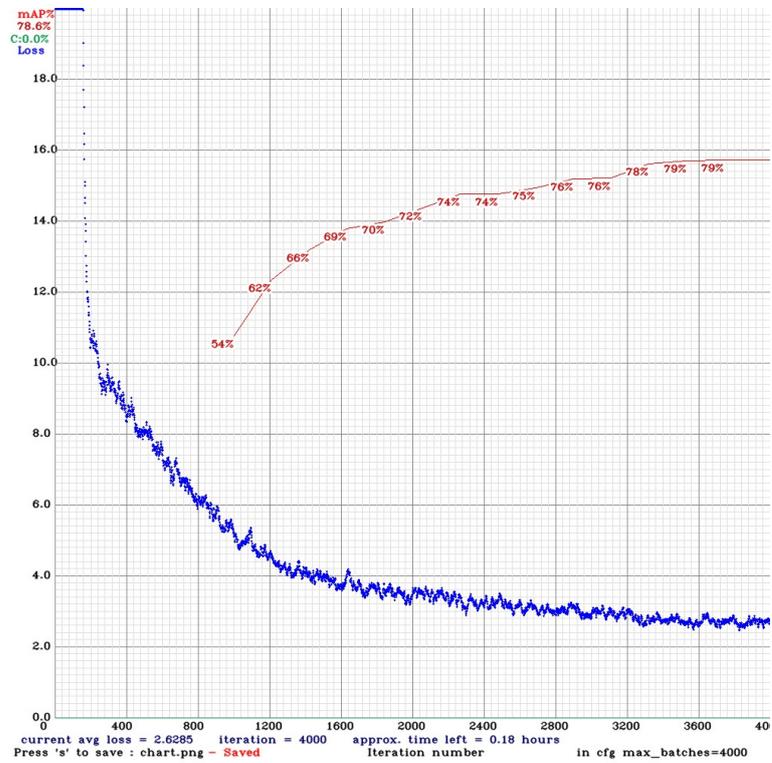
- **Segundo entrenamiento**



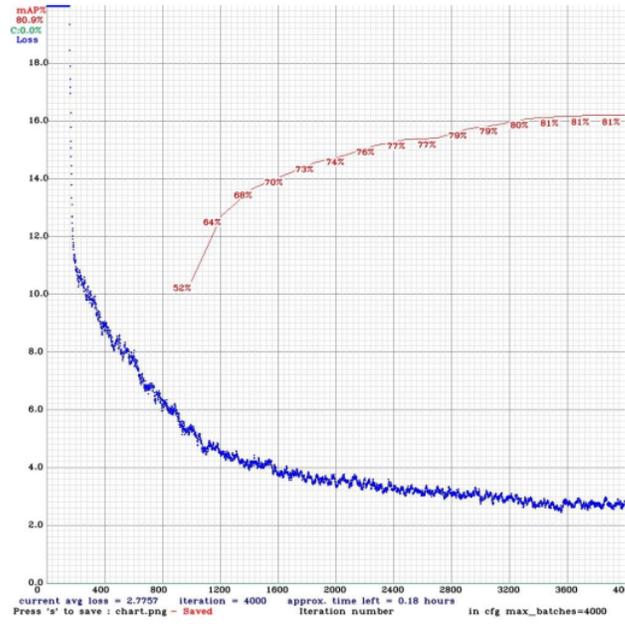
- Tercer entrenamiento



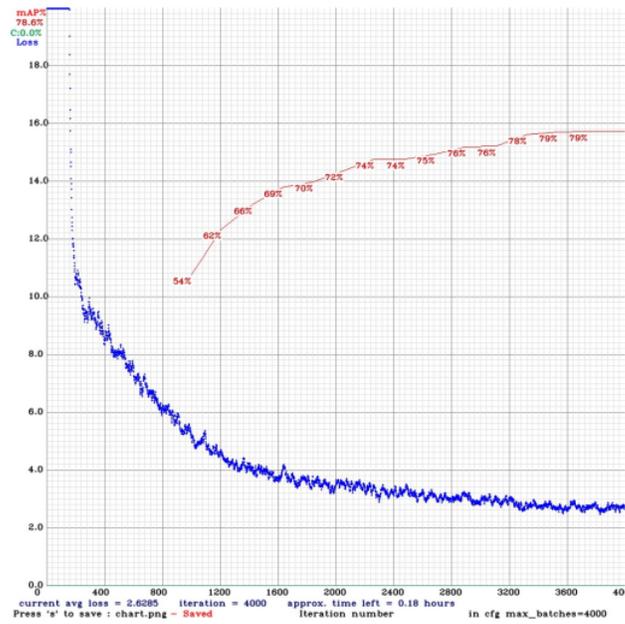
- Cuarto entrenamiento



- Tercer entrenamiento



- Cuarto entrenamiento



# Revisión de originalidad

## INFORME DE ORIGINALIDAD

10%

INDICE DE SIMILITUD

10%

FUENTES DE INTERNET

5%

PUBLICACIONES

%

TRABAJOS DEL ESTUDIANTE

## FUENTES PRIMARIAS

1	<a href="https://repositorio.uniandes.edu.co">repositorio.uniandes.edu.co</a> Fuente de Internet	1%
2	<a href="http://www.mrp.org.mx">www.mrp.org.mx</a> Fuente de Internet	1%
3	<a href="http://findanexpert.unimelb.edu.au">findanexpert.unimelb.edu.au</a> Fuente de Internet	1%
4	P. García-Aunon, J.J. Roldán, J. De León, J. Del Cerro, A. Barrientos. "Aplicaciones practicas de los sistemas multi-UAV y enjambres aéreos", Revista Iberoamericana de Automática e Informática industrial, 2021 Publicación	<1%
5	<a href="http://bdigital.uncu.edu.ar">bdigital.uncu.edu.ar</a> Fuente de Internet	<1%
6	Pablo Martínez Fernández. "Estudio y optimización del consumo energético del ferrocarril mediante redes neuronales y algoritmos heurísticos.", Universitat Politecnica de Valencia, 2019 Publicación	<1%

7	<a href="http://eprints.ucm.es">eprints.ucm.es</a> Fuente de Internet	<1 %
8	<a href="http://riunet.upv.es">riunet.upv.es</a> Fuente de Internet	<1 %
9	<a href="http://scholarworks.uark.edu">scholarworks.uark.edu</a> Fuente de Internet	<1 %
10	<a href="http://scielo.sld.cu">scielo.sld.cu</a> Fuente de Internet	<1 %
11	<a href="http://repositorio.ucsg.edu.ec">repositorio.ucsg.edu.ec</a> Fuente de Internet	<1 %
12	<a href="http://semh.net">semh.net</a> Fuente de Internet	<1 %
13	<a href="http://repository.javeriana.edu.co">repository.javeriana.edu.co</a> Fuente de Internet	<1 %
14	<a href="http://www.researchgate.net">www.researchgate.net</a> Fuente de Internet	<1 %
15	<a href="http://uvadoc.uva.es">uvadoc.uva.es</a> Fuente de Internet	<1 %
16	<a href="http://repositorio.udec.cl">repositorio.udec.cl</a> Fuente de Internet	<1 %
17	<a href="http://alfapublicaciones.com">alfapublicaciones.com</a> Fuente de Internet	<1 %
18	<a href="http://biblio3.url.edu.gt">biblio3.url.edu.gt</a> Fuente de Internet	<1 %

19	<a href="http://ar5iv.labs.arxiv.org">ar5iv.labs.arxiv.org</a> Fuente de Internet	<1 %
20	<a href="http://dehesa.unex.es">dehesa.unex.es</a> Fuente de Internet	<1 %
21	<a href="http://riuma.uma.es">riuma.uma.es</a> Fuente de Internet	<1 %
22	<a href="http://www.dropbox.com">www.dropbox.com</a> Fuente de Internet	<1 %
23	<a href="http://www.cacic2016.unsl.edu.ar">www.cacic2016.unsl.edu.ar</a> Fuente de Internet	<1 %
24	<a href="http://www.scielo.cl">www.scielo.cl</a> Fuente de Internet	<1 %
25	<a href="http://www.mdpi.com">www.mdpi.com</a> Fuente de Internet	<1 %
26	<a href="http://gredos.usal.es">gredos.usal.es</a> Fuente de Internet	<1 %
27	<a href="http://dialnet.unirioja.es">dialnet.unirioja.es</a> Fuente de Internet	<1 %
28	<a href="http://repositorio.unal.edu.co">repositorio.unal.edu.co</a> Fuente de Internet	<1 %
29	<a href="http://dspace.ort.edu.uy">dspace.ort.edu.uy</a> Fuente de Internet	<1 %
30	<a href="http://happylibnet.com">happylibnet.com</a> Fuente de Internet	<1 %

31	<a href="http://www.nature.com">www.nature.com</a> Fuente de Internet	<1 %
32	Alexandre Del Savio, Ana Luna, Daniel Cárdenas-Salas, Mónica Vergara, Gianella Urday. "Dataset of manually classified images obtained from a construction site", Data in Brief, 2022 Publicación	<1 %
33	<a href="http://bibdigital.epn.edu.ec">bibdigital.epn.edu.ec</a> Fuente de Internet	<1 %
34	<a href="http://issuu.com">issuu.com</a> Fuente de Internet	<1 %
35	<a href="http://link.springer.com">link.springer.com</a> Fuente de Internet	<1 %
36	<a href="http://ouci.dntb.gov.ua">ouci.dntb.gov.ua</a> Fuente de Internet	<1 %
37	<a href="http://repositorio.ucv.edu.pe">repositorio.ucv.edu.pe</a> Fuente de Internet	<1 %
38	<a href="http://ris.utwente.nl">ris.utwente.nl</a> Fuente de Internet	<1 %
39	<a href="http://www.coursehero.com">www.coursehero.com</a> Fuente de Internet	<1 %
40	<a href="http://www.irsitio.com">www.irsitio.com</a> Fuente de Internet	<1 %

[www.theibfr.com](http://www.theibfr.com)

41

Fuente de Internet

&lt;1 %

42

[lanbide.net](http://lanbide.net)

Fuente de Internet

&lt;1 %

43

[pt.scribd.com](http://pt.scribd.com)

Fuente de Internet

&lt;1 %

44

[repositorioacademico.upc.edu.pe](http://repositorioacademico.upc.edu.pe)

Fuente de Internet

&lt;1 %

45

[bonnus.ulb.uni-bonn.de](http://bonnus.ulb.uni-bonn.de)

Fuente de Internet

&lt;1 %

46

[repositorio.uchile.cl](http://repositorio.uchile.cl)

Fuente de Internet

&lt;1 %

47

[repositorio.unsch.edu.pe](http://repositorio.unsch.edu.pe)

Fuente de Internet

&lt;1 %

48

[www.repositoriodigital.ipn.mx](http://www.repositoriodigital.ipn.mx)

Fuente de Internet

&lt;1 %

49

Sergio Morell Monzó. "Desarrollo de procedimientos para la detección del abandono de cultivos de cítricos utilizando técnicas de teledetección", Universitat Politecnica de Valencia, 2023

Publicación

&lt;1 %

50

Tamara Rocío Ruiz Calleja. "Funcionalización de textiles mediante la aplicación de grafeno", Universitat Politecnica de Valencia, 2021

Publicación

&lt;1 %

---

51

rus.ucf.edu.cu

Fuente de Internet

<1 %

---

52

www.ridaa.unicen.edu.ar

Fuente de Internet

<1 %

---

53

fdocuments.mx

Fuente de Internet

<1 %

---

---

Excluir citas

Activo

Excluir coincidencias < 15 words

Excluir bibliografía

Activo