

The background is a dark blue gradient. In the top-left and bottom-right corners, there are clusters of 3D cubes in various shades of purple and blue, some with a metallic texture. The main title is centered in the middle of the slide.

INTRODUCCION AL PROCESAMIENTO DE IMAGENES

Facultad de Ingeniería - UNJu
2024



PROFESORA:
DRA. G. NOELIA REVOLLO SARMIENTO

INTEGRANTES:
CALA MARTIN ALEJANDRO
PELOC SERGIO ANTONINO

Facultad de Ingeniería - UNJu
2024





DEEP LEARNING APLICADO AL PROCESAMIENTO DE IMÁGENES PARA LA DETECCIONES DE OBJETOS

2022

Practica Profesional Supervisada
Barreto, Agustina Ayelén
Universidad Nacional
Arturo Jauretche



Resumen

El proyecto se realizó con el propósito de ingresar al mundo de la inteligencia artificial, porque es un tema muy interesante. Además, la idea de detectar objetos defectuosos y no defectuosos surge de haber estudiado un tema donde se estudia la calidad.

durante los procesos de producción. El objetivo del proyecto fue la adquisición de la Conocimientos necesarios para comprender la creación y uso de redes neuronales para futuros proyectos personales o laborales. Durante el proyecto se adoptó un enfoque cuantitativo, siguiendo los métodos y técnicas adquiridos como analista programador,

sumando los necesarios a través de la investigación para concluir el trabajo de manera óptima. El

Los resultados obtenidos fueron buenos porque, en los casos de detecciones correctas, el éxito los porcentajes rondaban en general el 80%. Además, al realizar evaluaciones de desempeño

De los modelos creados, una precisión promedio o mAP solía ser superior al 65%.

Se lograron los objetivos inicialmente planteados y se utilizaron diferentes técnicas para el objeto.

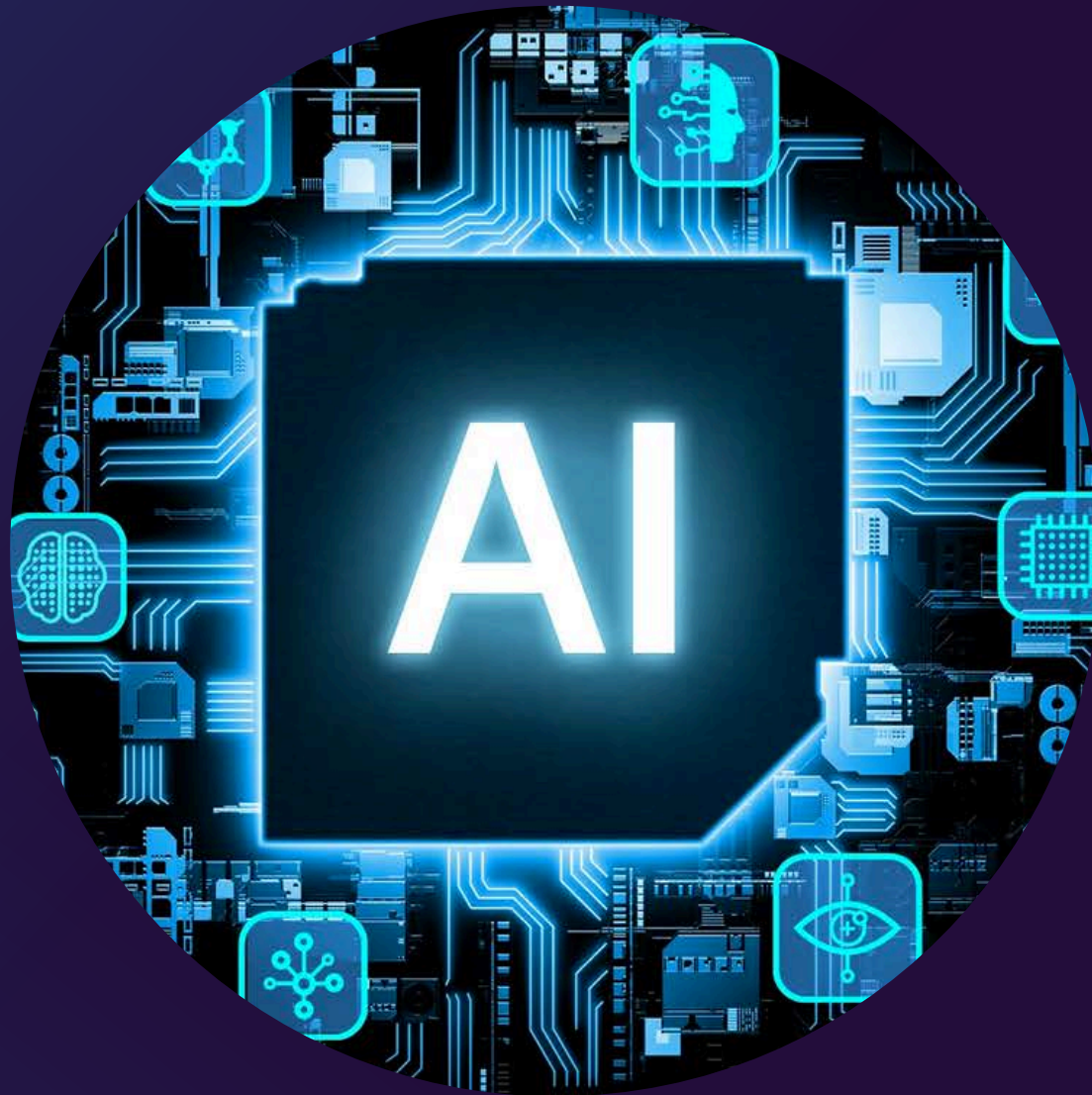
Se investigó la detección, obteniendo buenos rendimientos y realizando pruebas reales.

en un sistema integrado.



Introduccion

El presente trabajo forma parte de la Práctica Profesional Supervisada (PPS) dentro de la carrera de Ingeniería en Informática en la Universidad Nacional Arturo Jauretche (UNAJ). El mismo tiene por objetivo general enfocarse en el desarrollo de una aplicación de software encargada de realizar una búsqueda de objetos en imágenes a partir de una clasificación de las mismas en tiempo real, mediante la cual se buscará detectar la presencia de determinados objetos en una imagen. El desarrollo se enfocará específicamente en el uso de redes neuronales convolucionales, las cuales han demostrado ser las más eficientes en el área del procesamiento de imágenes. Específicamente, se espera que el software permita realizar la tarea de detección de objetos dañados y en buen estado en imágenes o video.

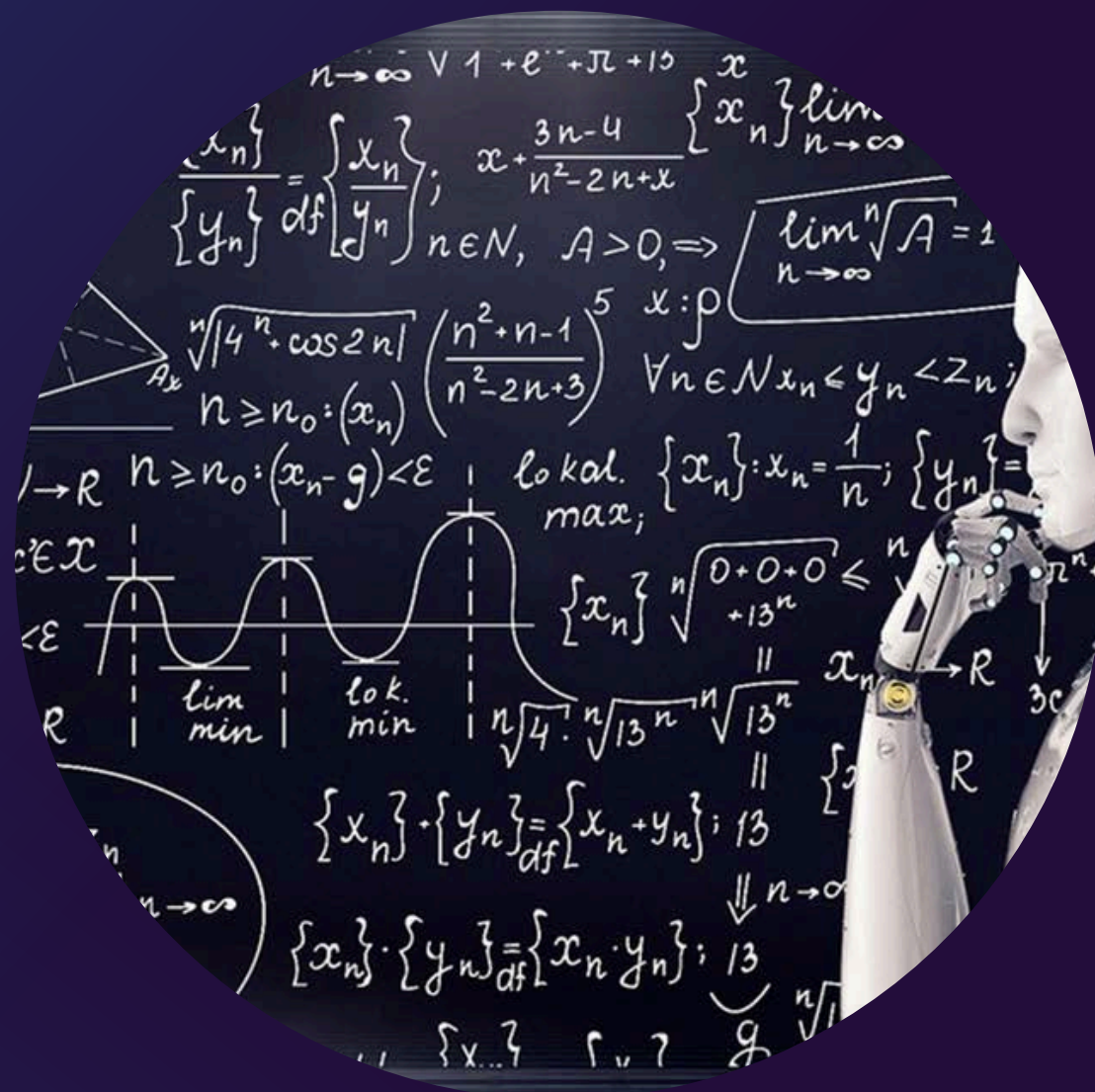


Marco teorico



Inteligencia Artificial

La Inteligencia Artificial (IA) es la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano.



Marco teorico

Machine Learning

El Aprendizaje Automático o Machine Learning (ML) es el subconjunto de inteligencia artificial (IA) que se centra en desarrollar sistemas que aprenden, o mejoran el rendimiento, en función de los datos que consumen.



Marco teorico




Deep Learning

El aprendizaje profundo mejora el aprendizaje automático al dar a las máquinas la capacidad de elegir entre un conjunto de algoritmos, que ofrecen una variedad de respuestas, y actuar sobre las conclusiones determinadas por una variedad de combinaciones.


Redes neuronales convolucionales

Las redes neuronales convolucionales son un algoritmo de Deep Learning que está diseñado para trabajar con imágenes, tomando estas como input, asignándole importancia (peso) a ciertos elementos en la imagen para así poder diferenciar unos de otros.

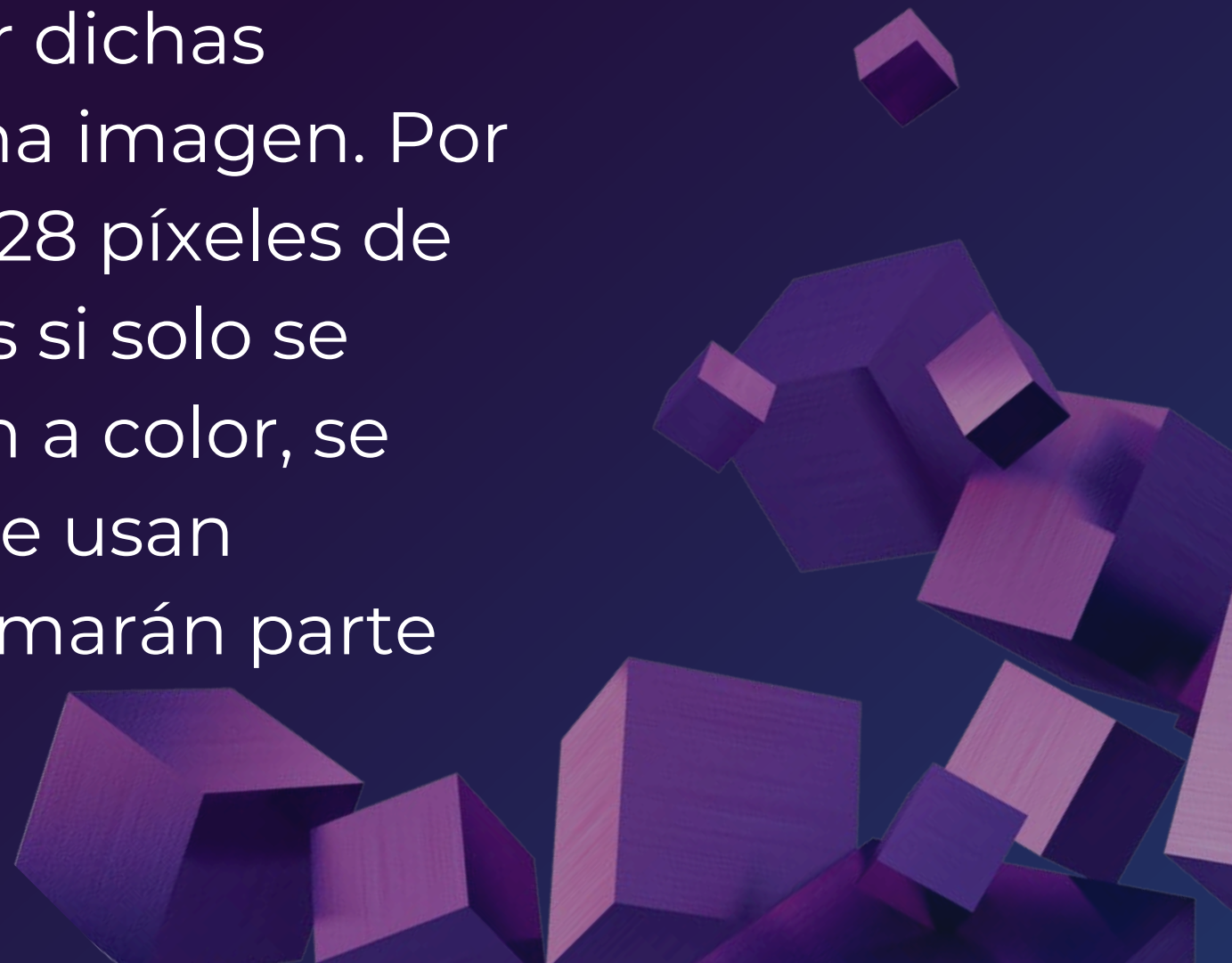




Redes neuronales convolucionales



Las tareas comunes de este tipo de redes son detección o categorización de objetos, clasificación de escenas y clasificación de imágenes en general. Para realizar dichas tareas, la red toma como entrada los píxeles de una imagen. Por ejemplo, si se tuviera una imagen con apenas 28×28 píxeles de alto y ancho, eso equivale a 784 neuronas. Y eso es si solo se tiene 1 color (escala de grises). Si fuera una imagen a color, se necesitan 3 canales (red, green, blue) y entonces se usan $28 \times 28 \times 3 = 2352$ neuronas de entrada las cuales formarán parte de la capa de entrada.



Arquitectura de las redes neuronales convolucionales

La arquitectura de una red neuronal convolucional (RNC) se puede separar en dos etapas, por un lado, la etapa de extracción de características que a su vez está compuesta por una o más capas convolucionales en donde se realizan pasos fundamentales y, por otro lado, la etapa de clasificación, también llamada capa completamente conectada, que comienza con un aplanamiento y continúa con una red neuronal tradicional totalmente conectada, en donde se realizará la clasificación final.

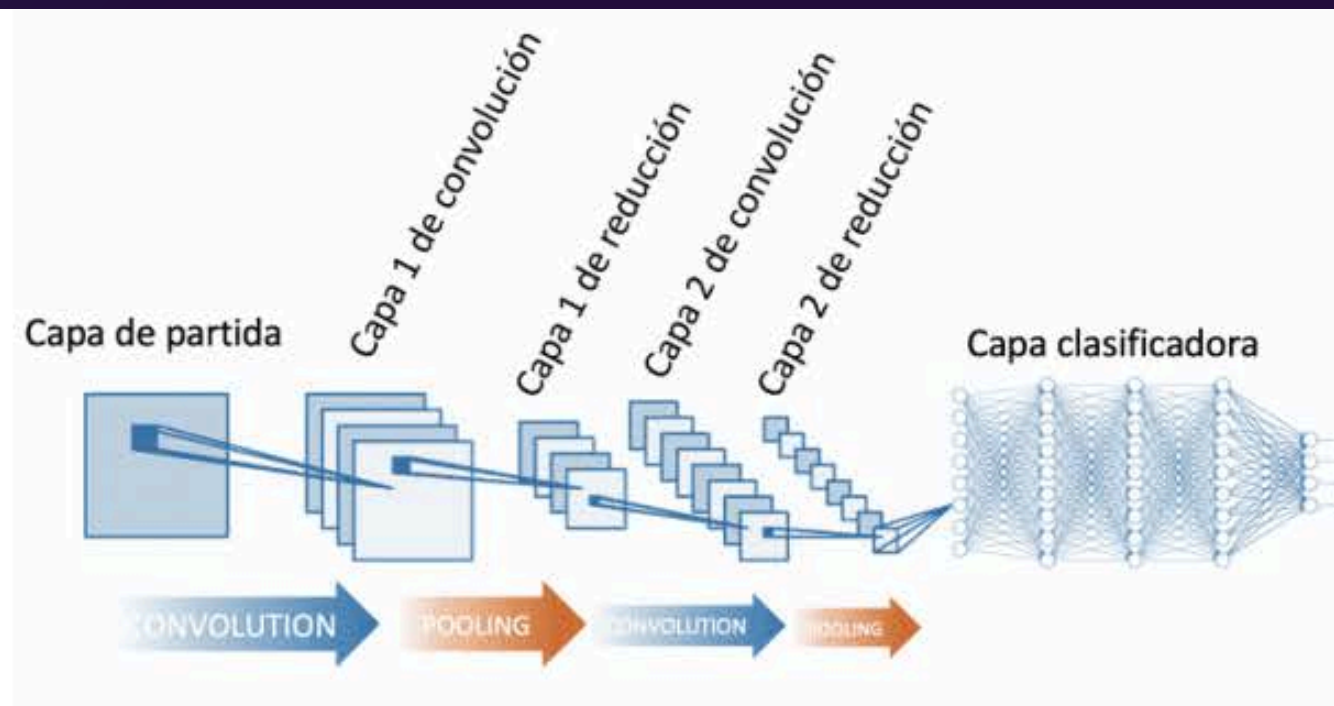
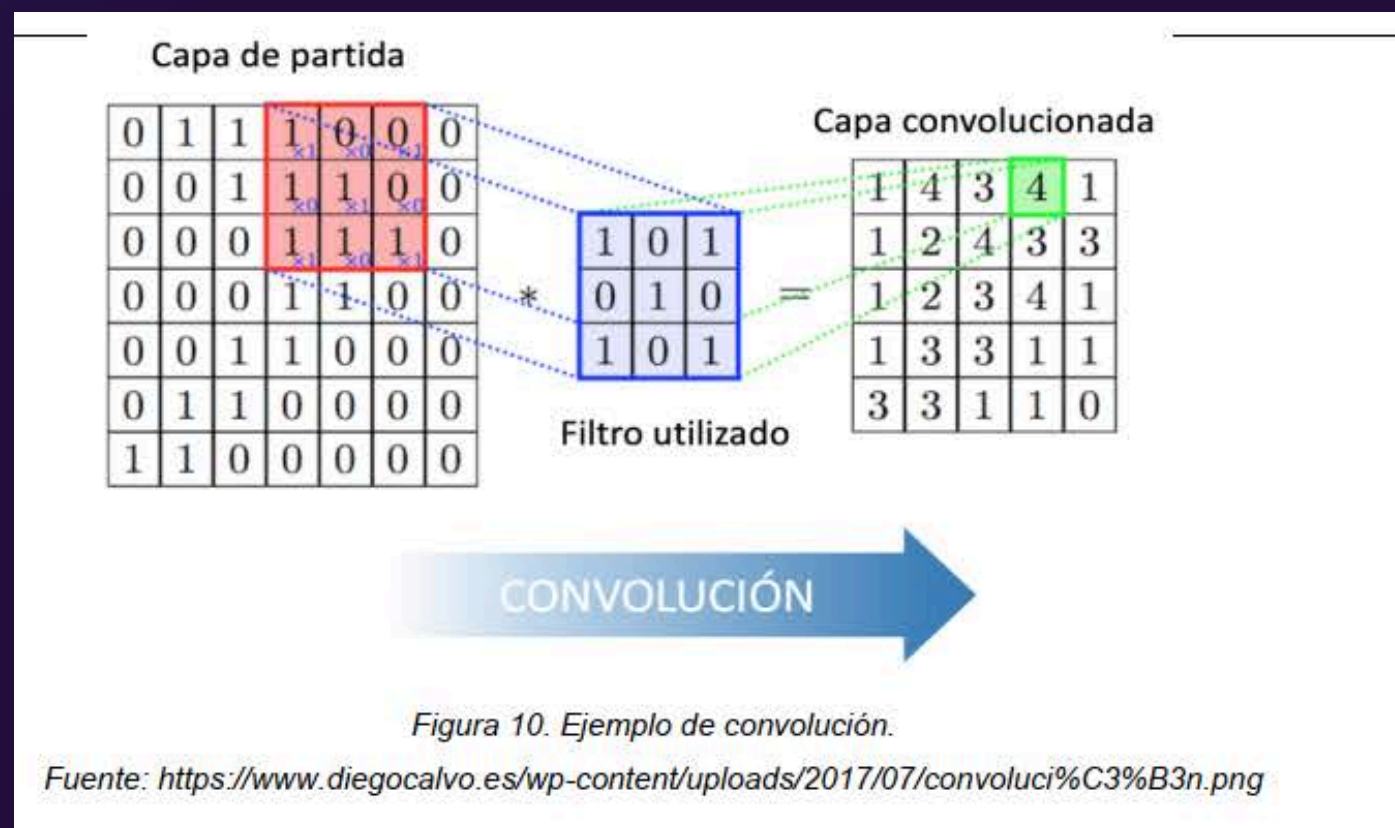


Figura 9. Arquitectura RNC.

Fuente: <https://www.diegocalvo.es/wp-content/uploads/2017/07/red-neuronal-convolucional-arquitectura.png>

Convolución

La convolución es el pilar de la arquitectura de las redes convolucionales y para llevarla a cabo, se realiza una operación matemática que describe cómo fusionar dos conjuntos de información. La diferencia fundamental entre una capa densamente conectada y una capa especializada en la operación de convolución, llamada capa de convolución, es que la capa densa aprende patrones globales tomados de todo el dato de entrada, mientras que las capas convolucionales aprenden patrones locales en pequeñas ventanas de dos dimensiones.






Capa de reducción o pooling



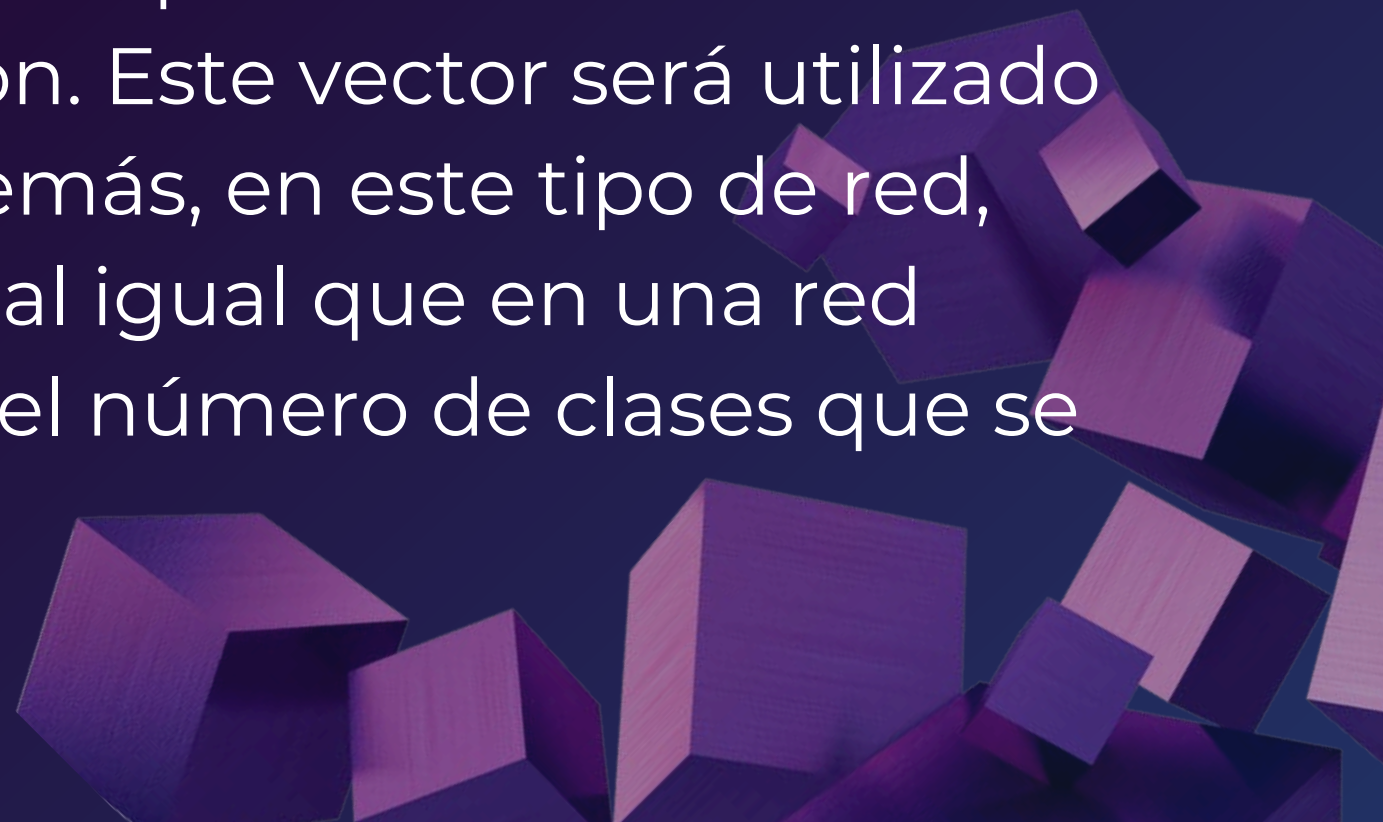
La capa de reducción o pooling se coloca generalmente después de la capa convolucional. Su utilidad principal radica en la reducción de las dimensiones espaciales (ancho x alto) del volumen de entrada para la siguiente capa convolucional.





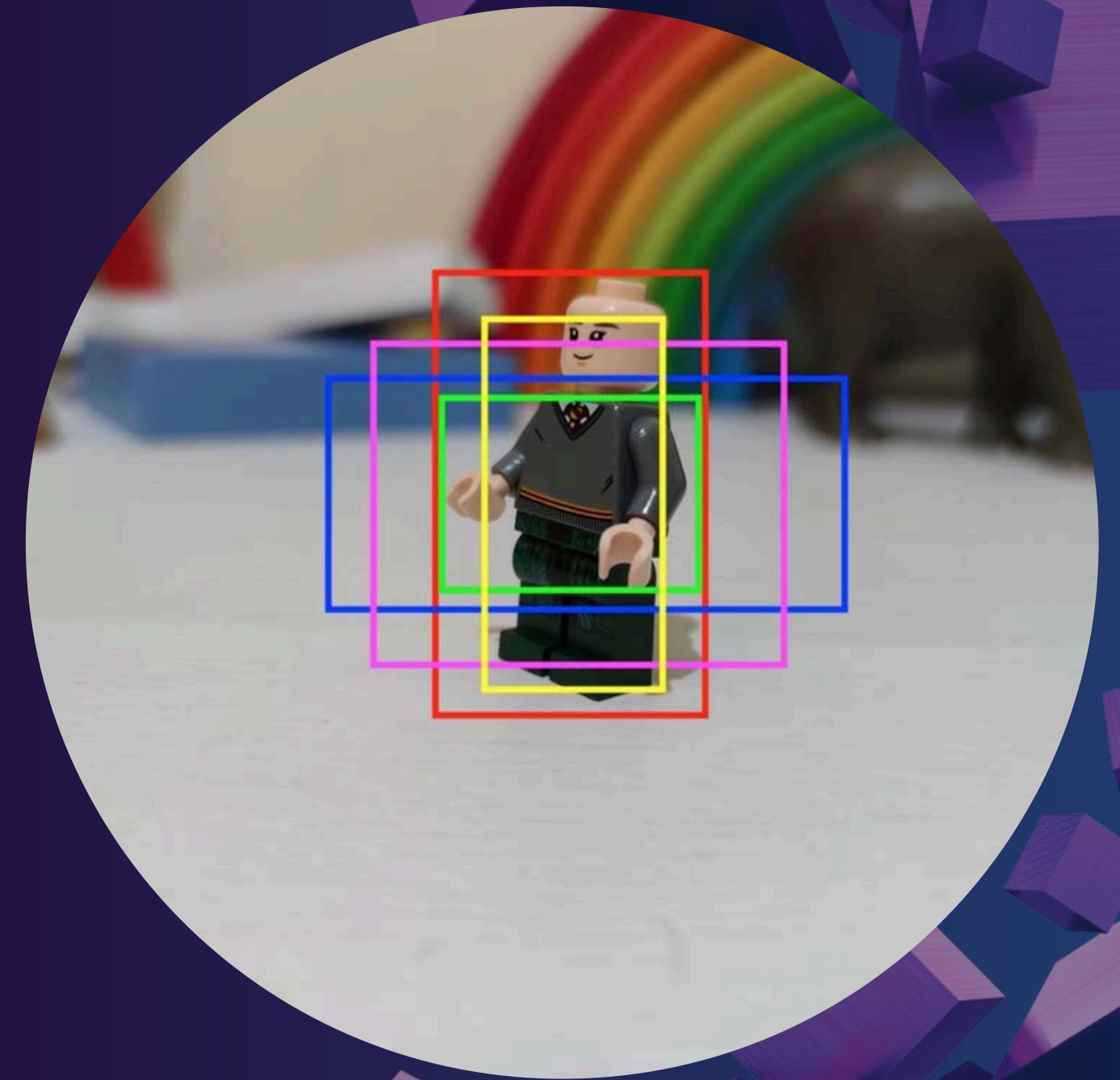
Clasificación

La etapa de clasificación es la etapa final de una CNN. Esta fase consiste, en primer lugar, en realizar un aplanamiento o “flatten” de los mapas de características y luego pasarlos por una red totalmente conectada, que se encargará de realizar la clasificación para determinar la clase a la que pertenece la imagen de entrada. El aplanamiento consiste en tomar todos los mapas de características obtenidos en la última capa convolucional y pasarlos por una función que transformará estos datos tridimensionales en un vector de una dimensión. Este vector será utilizado como entrada de una red totalmente conectada. Además, en este tipo de red, cada píxel se considera como una neurona separada al igual que en una red neuronal regular. Esta tendrá tantas neuronas como el número de clases que se debe predecir.



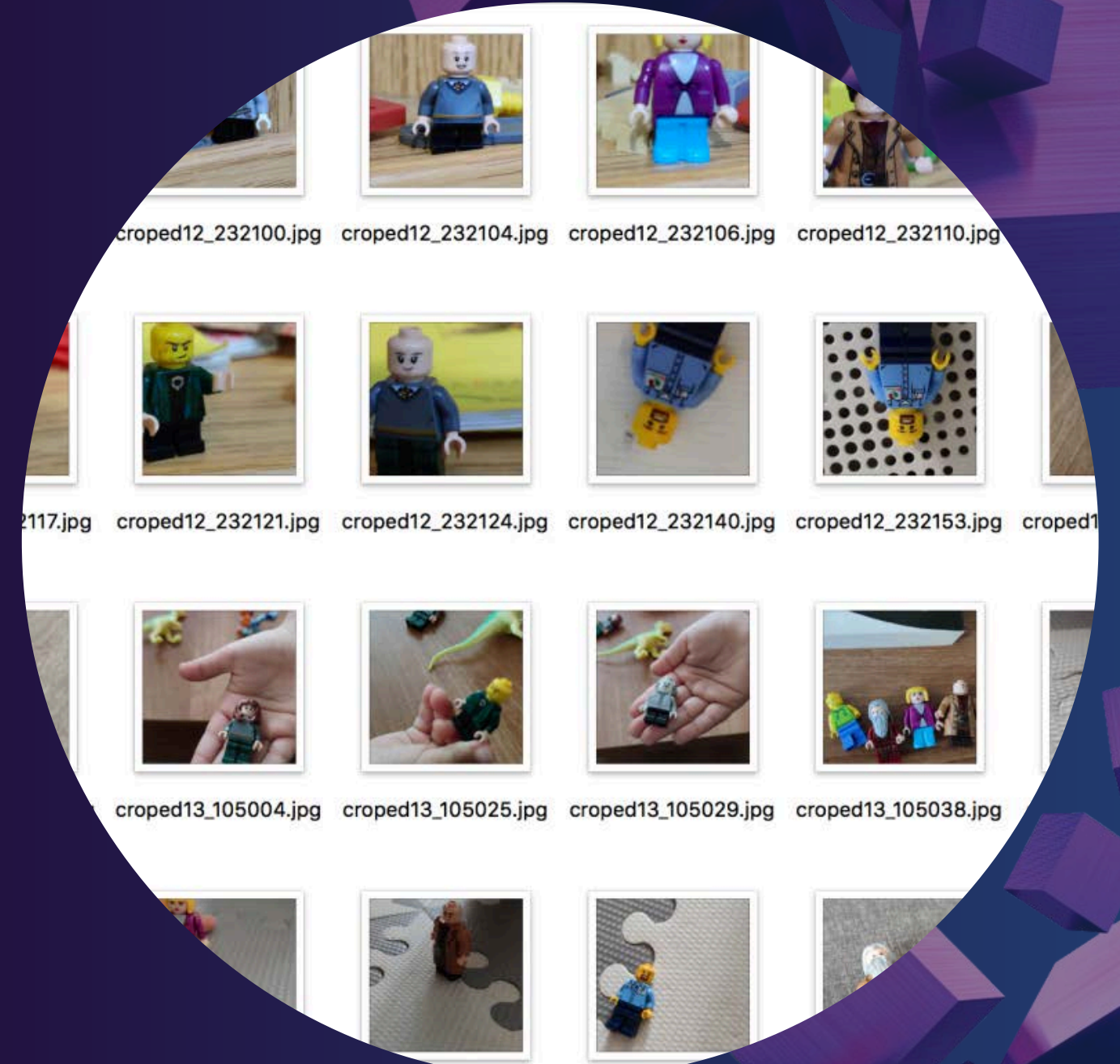
Pasos para la detección de objetos

Para dar inicio al modelado y desarrollo de una red neuronal capaz de detectar objetos es necesario primero recolectar la mayor cantidad de imágenes posibles y que estas sean lo suficientemente descriptivas para obtener resultados deseados. Cuantos más tipos de elementos (clases) se desee detectar en una imagen, mayor tendrá que ser la cantidad de imágenes. Estas imágenes no deben ser muy grandes, ya que esto podría provocar lentitud en la detección. Luego de la recolección de imágenes es necesario etiquetarlas estableciendo las coordenadas donde están los objetos que se quiere detectar.



Datos de entrenamiento o training data

Los datos de entrenamiento son los datos que se usan para entrenar un modelo, en este caso, los datos serán imágenes. La calidad del modelo de aprendizaje automático va a ser directamente proporcional a la calidad de los datos. Por eso las labores de limpieza, depuración o data wrangling (es el proceso de limpiar, transformar y organizar datos en bruto en un formato adecuado para el análisis y la toma de decisiones. Este proceso es crucial en la ciencia de datos porque asegura que los datos sean de alta calidad y estén listos para ser analizados.) consumen un porcentaje importante del tiempo de los científicos de datos.



Aumento de datos

El aumento de datos es una de las técnicas utilizadas para aumentar la cantidad de datos agregando copias ligeramente modificadas de datos ya existentes o datos sintéticos recién creados a partir de datos existentes. Actúa como regularizador y ayuda a reducir el sobreajuste al entrenar un modelo de aprendizaje automático.

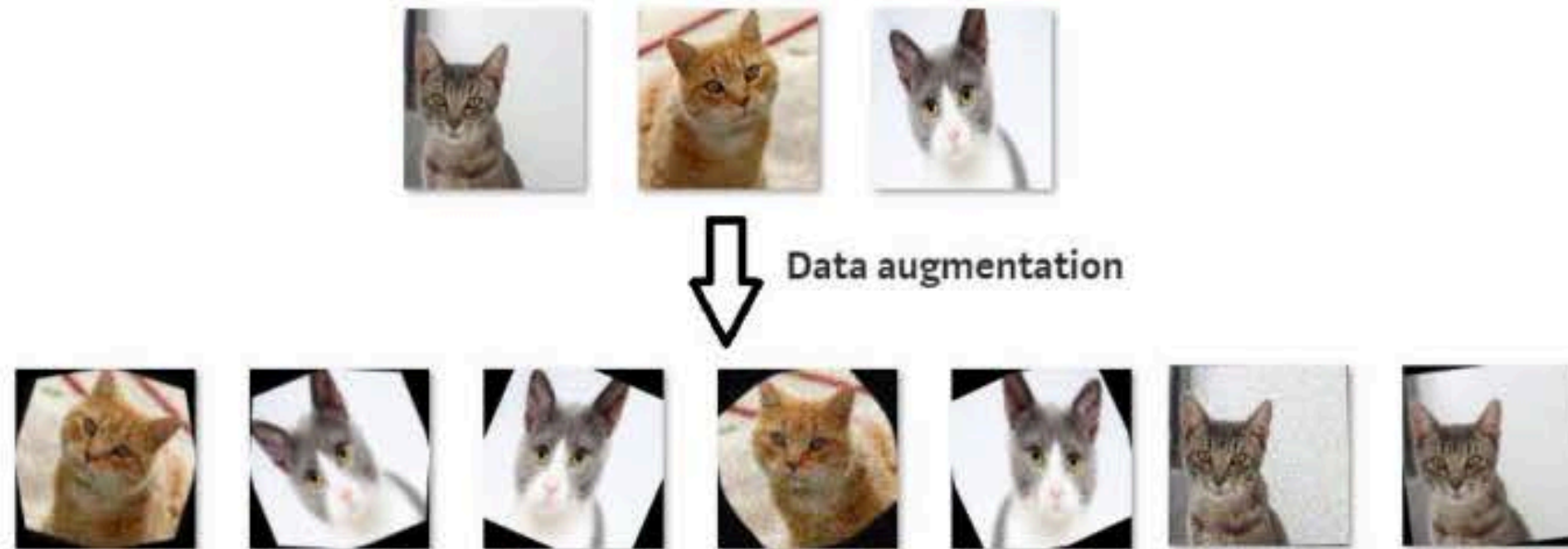


Figura 22. Ejemplo de aumento de datos.

Fuente: https://miro.medium.com/max/1400/1*G7AOG4ivciLI6y1wYFx-sQ.png

Herramientas utilizadas



LABELIMG

LabelImg es una herramienta utilizada para el etiquetado del set de imágenes. Lo que se hace con este programa es abrir una imagen, seleccionar un recuadro para marcar el objeto que se quiere que el programa aprenda a detectar y guardar un archivo XML con la información de las coordenadas.

PYTHON


Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código. Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional. Es un lenguaje interpretado, dinámico y multiplataforma.

GOOGLE COLAB

Colab es un servicio cloud, basado en los Notebooks de Jupyter, que permite el uso gratuito de las GPUs y TPUs de Google, con librerías como: Scikit-learn, PyTorch, TensorFlow, Keras y OpenCV. Todo ello con bajo Python 2.7 y 3.6

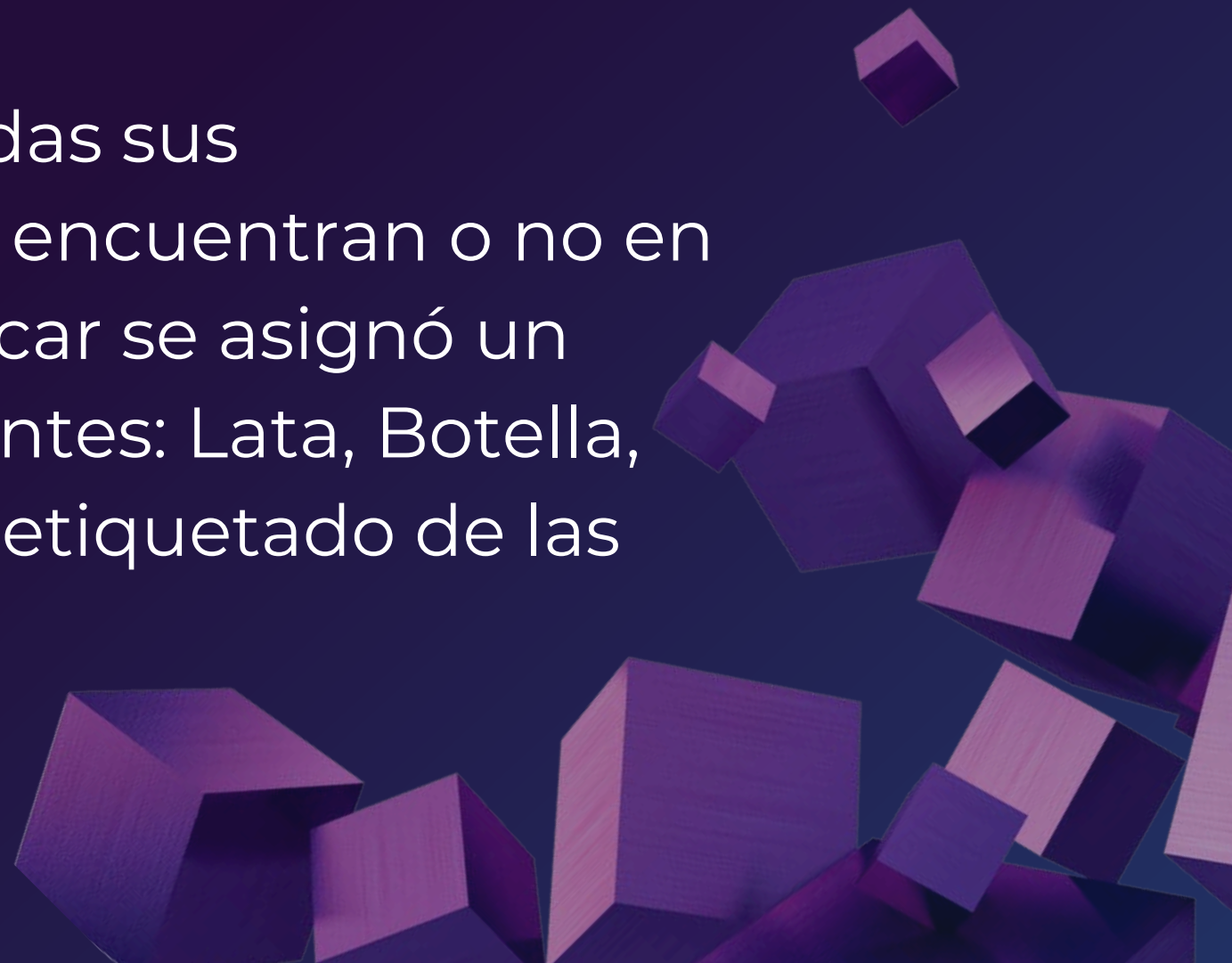


Primera aproximación



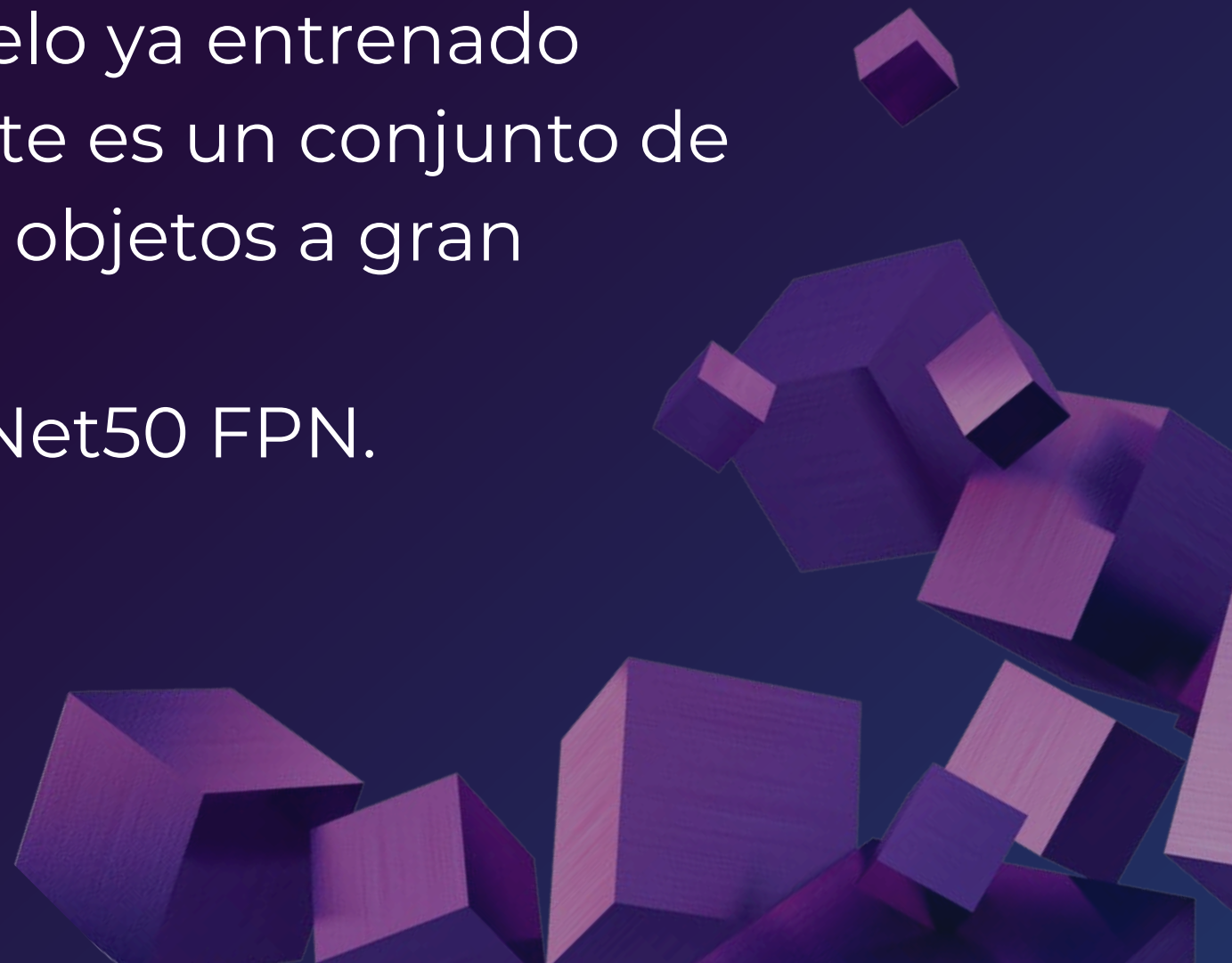
Después de la investigación llevada a cabo, se realizó un primer algoritmo de detección de objetos utilizando un set de datos de 224 imágenes

.El objetivo es la detección de latas y botellas y dadas sus características poder detectar si dichos objetos se encuentran o no en buen estado. Por lo que, por cada objeto a identificar se asignó un nombre que representará su clase y son los siguientes: Lata, Botella, Lata_defectuosa Botella_defectuosa con el fin del etiquetado de las imágenes.





Transfer Learning en conjuntos de datos personalizados



Teniendo en cuenta la poca cantidad de imágenes recolectadas en una primera prueba piloto, se decidió utilizar el método de transfer learning. Partiendo de esto se seleccionó un modelo ya entrenado originalmente con el conjunto de datos COCO (Este es un conjunto de datos de subtítulos, segmentación y detección de objetos a gran escala publicado por Microsoft.)

Finalmente se eligió el modelo Faster R-CNN ResNet50 FPN.

Resultados obtenidos

Este método muestra en una ventana las predicciones del modelo en la actual transmisión de cámara web en vivo de una computadora

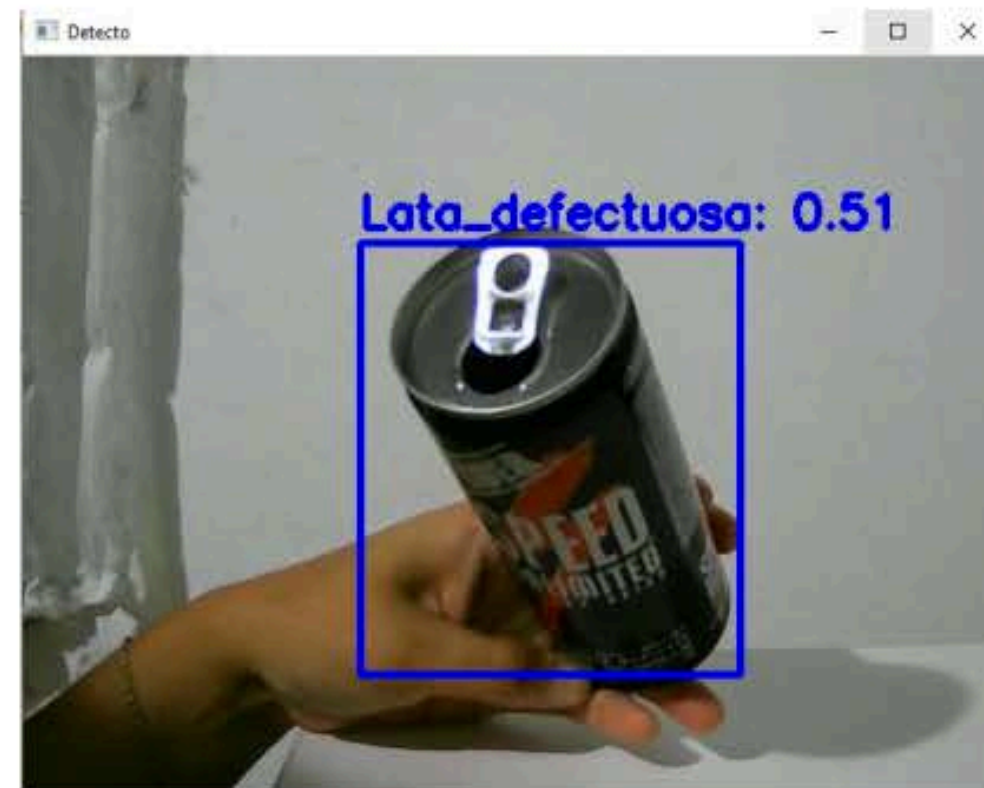


Figura 41. Detección de lata defectuosa.

Fuente: producción propia.

The background is a solid dark purple. In the top corners, there are several 3D cubes of varying sizes, some of which are open boxes. Two thin, horizontal white lines are positioned above and below the central text.

GRACIAS !