Reconocimiento de objetos en imágenes Ichnaea Lens v0.6

Albornoz Martín Nahuel

Hoy en día, el reconocimiento de imágenes ha adquirido gran relevancia debido a la popularidad y avances en inteligencia artificial y redes convolucionales. En este paper, se presenta una investigación sobre la aplicación de estas tecnologías en el reconocimiento de imágenes, utilizando redes neuronales convolucionales (CNN)

Al finalizar la investigación se logra comprobar la viabilidad de una aplicación para el reconocimiento de objetos en imágenes mediante la utilización de inteligencia artificial con redes neuronales convolucionales. Las imágenes son procesadas y analizadas por la CNN entrenada y devuelven los objetos identificados. Los resultados obtenidos le serán mostrados al usuario según la disponibilidad en tiendas cercanas que cuenten con el producto a la venta

I. NOMENCLATURA

AI: Inteligencia Artificial
CNN: Redes Neuronales Convolucionales
API: interfaz de programación de aplicaciones

II. INTRODUCCIÓN

En los últimos años se ha notado un incremento en el uso de inteligencias artificiales las cuales cuentan con uso cada vez mayor del usuario promedio con interfaces cada vez mas sencillas de usar para el publico en general y ya no solo para expertos en áreas especificas.

Dentro de este incremento de AIs y nuevas tecnologías el reconocimiento de objetos en imágenes, aunque aun no perfecto, invita a un futuro de grandes posibilidades para las inteligencias artificiales. Con grandes beneficios para los usuarios.

Teniendo esto en cuenta con el aumento de las compras en línea, muchas personas buscan formas más convenientes de encontrar los productos que desean comprar. La aplicación de reconocimiento de objetos puede ayudar a los usuarios a encontrar productos en línea a partir de imágenes tomadas con sus dispositivos móviles. En este trabajo, se presenta una aplicación que utiliza redes neuronales convolucionales para el reconocimiento de objetos y una API de tiendas en línea para mostrar los resultados de búsqueda.

Esta aplicación permitirá el reconocimiento de objetos y prendas en fotografías y permitirá al usuario buscar los mismos en comercios cercanos.

III. CASO DE ESTUDIO

Para las personas que al ver prendas o artículos en alguna fotografía y desconocen donde conseguirlo o talvez siquiera como buscarlo. Para solucionar esto se plantea Ichnaea Lens. Una aplicación que analiza los objetos en una fotografía. Ya sea subida desde el dispositivo o de una foto tomada en el momento. Los resultados de la foto le muestran al usuario los objetos destacables de la fotografía que puede encontrar en tiendas cercas para adquirirlos.
[12] [2]

IV. MARCO TEORICO

A. La Inteligencia Artificial

La IA se ha convertido en un termino que representa a las aplicaciones que realizan tareas complejas que anteriormente requieran aportes humanos. Como ha llegado a ser la creación de ilustraciones, el procesamiento de textos, etc.
Es decir, la inteligencia artificial es la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear. La IA permite que los sistemas tecnológicos perciban su entorno, se relacionen con él, resuelvan problemas y actúen con un fin específico. La máquina recibe datos (ya preparados o recopilados a través de sus propios sensores, por ejemplo, una cámara), los procesa y responde a ellos.
En estos últimos años se ha visto un gran impulso de las AI llegando al mercado en una amplia variedad de sectores modificando radicalmente muchas industrias. Para obtener el valor completo de la IA, muchas empresas están haciendo inversiones significativas en equipos de ciencia de datos para mantenerse a la vanguardia de la tecnología

La inteligencia artificial se puede organizar de varias maneras, según las etapas de desarrollo o las acciones que se están realizando.
Se suelen reconocer cuatro etapas de desarrollo de la IA.

- Máquinas reactivas: IA limitada que solo reacciona a diferentes tipos de estímulos basados en reglas preprogramadas. No usa memoria y, por lo tanto, no puede aprender con datos nuevos. Deep Blue de IBM, que venció al campeón de ajedrez Garry Kasparov en 1997, fue un ejemplo de una máquina reactiva.
- Memoria limitada: Se considera que la mayor parte de la IA moderna es de memoria limitada. Puede usar la

memoria para mejorar con el tiempo mediante el entrenamiento con datos nuevos, por lo general, a través de una red neuronal artificial o algún otro modelo de entrenamiento. El aprendizaje profundo, un subconjunto del aprendizaje automático, se considera inteligencia artificial con memoria limitada.

- Teoría de la mente: Actualmente no existe IA con teoría de la mente, pero se están investigando distintas posibilidades. El término hace referencia a IA que puede emular la mente humana y tiene capacidades de toma de decisiones similares a las de un ser humano, lo cual incluye reconocer y recordar emociones, y reaccionar en situaciones sociales como lo haría un ser humano.
- Autoconocimiento: Un paso más allá de la IA con teoría de la mente, el concepto de IA con autoconocimiento describe una máquina mítica que tiene conocimiento de su propia existencia y tiene las capacidades intelectuales y emocionales de un ser humano. Al igual que la IA con teoría de la mente, la IA con autoconciencia no existe en la actualidad.

[4][13]

B. Aprendizaje automático

El aprendizaje automático (en inglés, machine learning) es uno de los enfoques principales de la inteligencia artificial. En pocas palabras, se trata un aspecto de la informática en el que los ordenadores o las máquinas tienen la capacidad de aprender sin estar programados para ello. Un resultado típico serían las sugerencias o predicciones en una situación particular.

El aprendizaje automático usa algoritmos para aprender de los patrones de datos. Un ejemplo de estos son los filtros de spam de los correos electrónicos. Los cuales utilizan este tipo de aprendizaje para poder detectar qué mensajes son correo basura y separarlos de aquellos que no lo son. A medida que va filtrando y recibe feedbacks del usuario aprende a diferenciarlos.

Existen tres subconjuntos del aprendizaje automático que pueden utilizarse. El automático, el supervisado, el no supervisado y el de refuerzo. (Fig. 1)

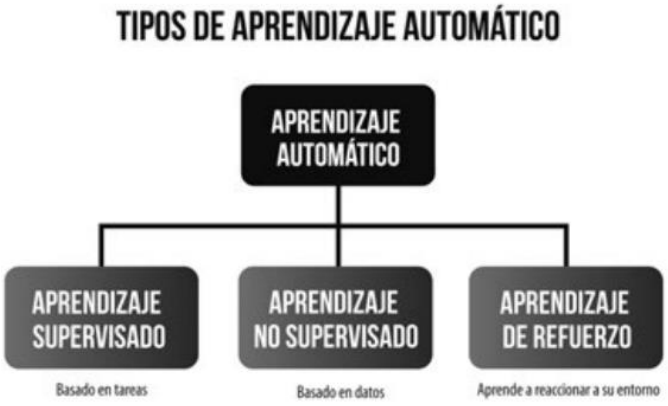


Fig. 1 Tipos de aprendizaje automático

El aprendizaje supervisado es un modelo de aprendizaje automático que asigna una entrada específica a un resultado mediante datos de entrenamiento etiquetados (datos estructurados). En términos simples, para entrenar un algoritmo que reconozca imágenes de gatos, se lo alimenta con imágenes etiquetadas como gatos.

El aprendizaje no supervisado es un modelo de aprendizaje automático que aprende patrones en función de datos no etiquetados (datos no estructurados). A diferencia del aprendizaje supervisado, el resultado final no se conoce con anticipación. En cambio, el algoritmo aprende de los datos y los clasifica en grupos en función de diversos atributos. Por lo tanto, este método no requiere la intervención humana. En el ejemplo anterior, los algoritmos tendrían que clasificar ellos mismos todas las fotos en las que apareciera un gato en una categoría. Este tipo de aprendizaje no supervisado es bueno para identificar patrones y realizar modelado descriptivo.

El aprendizaje por refuerzo es un modelo de aprendizaje automático que se puede describir en términos generales como “aprender haciendo”. Un "agente" aprende a realizar una tarea definida mediante prueba y error (un ciclo de reacción) hasta que su rendimiento está dentro de un rango deseado. El agente recibe un refuerzo positivo cuando realiza la tarea de forma correcta y un refuerzo negativo cuando tiene bajo rendimiento. Un ejemplo de aprendizaje por refuerzo sería enseñarle a una mano robótica a recoger una pelota.

[13]

C. Machine Learning

El término AI es a menudo usado de manera intercambiable con sus subcampos, el aprendizaje automático (Machine Learning) y el aprendizaje profundo (Deep Learning). Sin embargo, hay ciertas diferencias. Por ejemplo, el machine learning se centra en la creación de sistemas que aprenden o mejoran su rendimiento en función de los datos que consumen. Es importante tener en cuenta que, aunque todo machine learning es IA, no toda la IA es machine learning.

El Machine Learning en su uso más básico es la práctica de usar algoritmos para parsear datos, aprender de ellos y luego ser capaces de hacer una predicción o sugerencia sobre algo. Los programadores deben perfeccionar algoritmos que especifiquen un conjunto de variables para ser lo más precisos posibles en una tarea en concreto. La máquina es entrenada utilizando una gran cantidad de datos dando la oportunidad a los algoritmos a ser perfeccionados.(Fig.2)

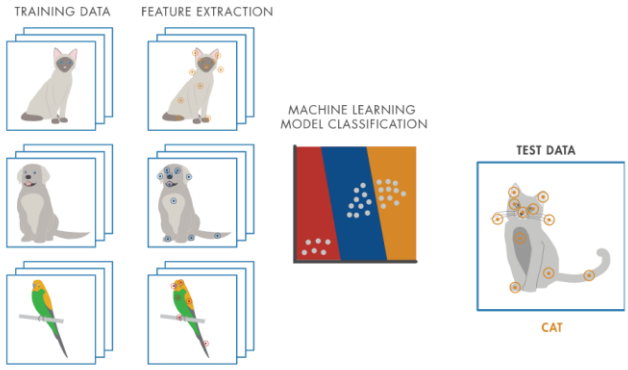


Fig. 2 Flujo de trabajo de Machine Learning para reconocimiento de imágenes.

[5][14]

D. Deep Learning (Aprendizaje profundo)

Siguiendo la evolución del Machine Learning en la última década se ha propagado con más fuerza una técnica concreta de Machine Learning conocida como Deep Learning.

Por definición, Deep Learning es un subconjunto dentro del campo del Machine Learning, el cual predica con la idea del aprendizaje desde el ejemplo. Su uso es excelente extrayendo patrones a partir de datos en bruto

En Deep Learning, en lugar de enseñarle a ordenador una lista enorme de reglas para solventar un problema, le damos un modelo que pueda evaluar ejemplos y una pequeña colección de instrucciones (Fig. 3) para modificar el modelo cuando se produzcan errores. Con el tiempo esperamos que esos modelos sean capaces de solucionar el problema de forma extremadamente precisa, gracias a que el sistema es capaz de extraer patrones.

La mayoría de los métodos de Deep learning usan arquitectura de redes neuronales o "Deep neural networks". Se le conoce como "Deep" en referencia a las capas que tienen estas redes neuronales.

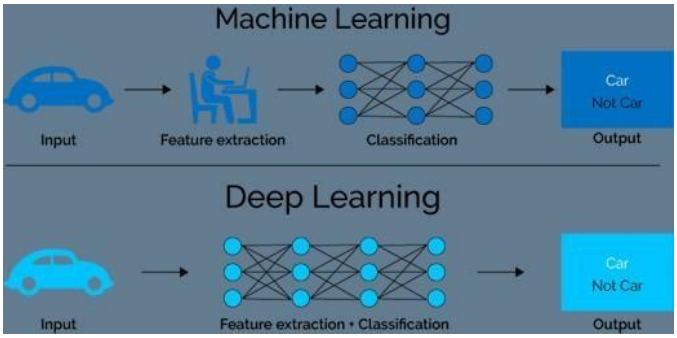


Fig. 3 Aprendizaje profundo vs Machine Learning

Para definir a grandes rasgos

• IA (Inteligencia Artificial): una máquina que es capaz de imitar el razonamiento humano.

• ML (Machine Learning): un subconjunto de Inteligencia Artificial donde las personas "entrenan" a las máquinas para reconocer patrones basados en datos y realizar sus predicciones.

• DL (Deep Learning): un subconjunto de ML en el que la máquina es capaz de razonar y sacar sus propias conclusiones, aprendiendo por sí misma

[8][17]

E. Reconocimiento de imágenes

Reconocimiento de imagen utiliza tecnología y técnicas para ayudar a las computadoras a identificar, etiquetar y clasificar elementos de interés en una imagen.

Mientras que los seres humanos procesan las imágenes y clasifican los objetos dentro de las imágenes con bastante facilidad, lo mismo es imposible para una máquina a menos que haya sido entrenada específicamente para hacerlo. El resultado del reconocimiento de imágenes es identificar y clasificar con precisión los objetos detectados en varias categorías predeterminadas con la ayuda de la tecnología de aprendizaje profundo.

[1]

F. Algoritmos Procesamiento de imágenes

Las técnicas para analizar una imagen digital se agrupan en un área llamada Procesamiento Digital de Imágenes. Estas herramientas se organizan según el nivel de procesamiento que se desea realizar para analizar la información contenida en una imagen digital.

Estos niveles forman una cadena de procesamiento (Ver Fig.4). A continuación, se describen brevemente:

1)**Preprocesamiento**: Operaciones para adaptar la información de una imagen y tener mejor análisis en pasos posteriores. Ejemplos de procesamiento son las operaciones de brillo y contraste.

2) **Segmentación**: Operaciones para hacer una partición de la imagen en varias regiones que representen la información necesaria para el problema a resolver. Dentro de este se puede desglosar en:

- Segmentación de forma (SS). Se refiere a generar una nueva imagen binarizada donde el resultado son regiones representativas de la información que se desea analizar.
- Detección de bordes (CS). Extracción de altas frecuencias en la imagen que representen contornos y límites de objetos en la imagen.
- Segmentación de color (CrS). Partición de la imagen en regiones que agrupan colores similares.
- Segmentación de Textura (TS). Partición de una imagen con base a los patrones de variación espacial de las intensidades de grises y color.
- Segmentación de movimiento (VS). Detección de objetos dinámicos en una secuencia de imágenes

3) *Detección de objetos y clasificación*. Determinación y clasificación de los objetos contenidos en la imagen.

4) *Análisis de imagen*. Obtener información de alto nivel acerca de lo que la imagen muestra

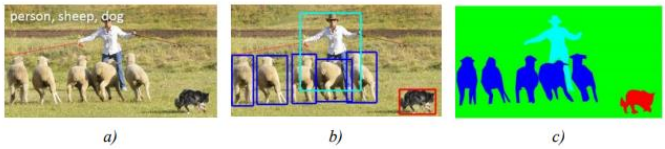


Fig. 4: a) reconocimiento de imágenes, b) detección y reconocimiento c) segmentación semántica

[7][11]

G. Redes neuronales Convolucionales

Las redes neuronales (Fig. 5) son muy útiles para, a partir de un dato de entrada, hacer una clasificación o regresión a partir de un modelo entrenado. Esta misma idea se puede aplicar a una imagen, para obtener, por ejemplo, qué representa la imagen (clasificación) o detectar un objeto que forme parte de ella (mediante algoritmos de segmentación, detección, identificación, etc.).

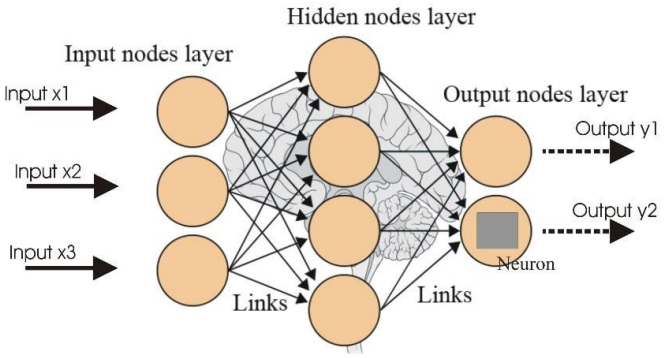


Fig. 5 Estructura básica de una red neuronal

A partir de esta surgen las redes neuronales convolucionales (CNN), las cuales han supuesto una revolución en el sector del reconocimiento de imágenes, ya que, a diferencia de las redes neuronales convencionales y otros algoritmos de clasificación de imágenes, usan un procesamiento relativamente pequeño.

La red es entrenada a partir de imágenes utilizando solo píxeles y etiquetas como entradas. Aprenden usando filtros y aplicándolos a las imágenes. El algoritmo toma un cuadrado pequeño o ventana y comienza a aplicarlo sobre la imagen. Cada filtro permite que la CNN identifique ciertos patrones en la imagen. La CNN busca partes de la imagen donde un filtro coincida con el contenido de la imagen.

En este caso las CNN toman como entrada tensores de forma, que incluyen alto y ancho de la imagen, en pixeles, y el numero de canales de color (1, si la imagen es en blanco y negro, o 3, si es en color RGB. Como salida se obtendría la clase o la probabilidad de que la imagen pertenezca a una clase particular. Como muestra la Fig. 6.

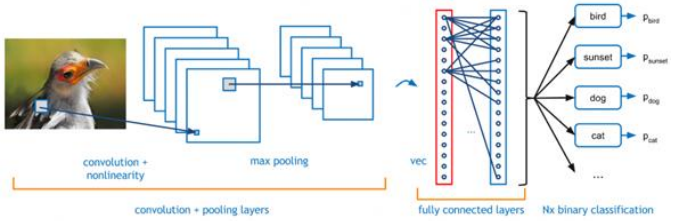


Fig.6 Estructura de una red neuronal convolucional para la clasificación de una imagen

Una operación de convolución por sí sola no puede detectar cosas más complejas que bordes y patrones muy simples, pero volver a hacer detecciones sobre las detecciones anteriores permite componer cada vez patrones más complejos.

El diseño de una arquitectura convolucional se suele representar como una especie de “embudo” (Fig. 7) donde la imagen inicial se va comprimiendo espacialmente, es decir, su resolución va disminuyendo al mismo tiempo que su grosor va aumentando; el número de mapas de características que vamos detectando va en aumento. Es al final de este “embudo convolucional” llegaremos a un punto en el que habremos detectado todos los patrones necesarios, contaremos con mapas de características que podremos introducir como inputs independientes dentro de una red neuronal multicapa que acabará por tomar la decisión de averiguar qué es o que hay en la imagen.

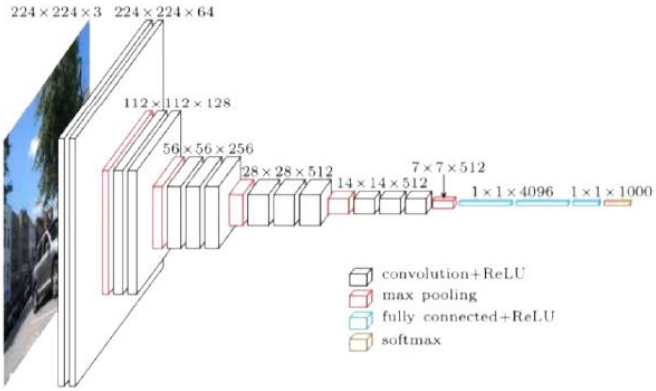


Fig. 7. Embudo convolucional
[6][10][15]

H. Python

Dentro de los lenguajes utilizados en la inteligencia artificial destacan por ser los mas utilizados Python y R. Siendo Python el mas útil a la hora de trabajar con reconocimiento de imágenes, en parte gracias al amplio numero de bibliotecas y frameworks disponibles como PyTorch, Caffe, Tensor Flow, Keras etc.

Es por esto por lo que se opta por Python 3.6 para este proyecto
[16]

I. Tensor Flow

Tensor Flow es una librería de código libre para Machine Learning (ML). Fue desarrollado por Google para satisfacer las necesidades a partir de redes neuronales artificiales. TensorFlow te permite construir y entrenar redes neuronales para detectar patrones y razonamientos usados por los humanos.

Además de trabajar con redes neuronales, TensorFlow es multiplataforma. Trabaja con GPUs y CPUs e incluso con las unidades de procesamiento de tensores (TPUs).

TensorFlow es una plataforma que facilita la creación e implementación de modelos de aprendizaje automático. Lo que se busca es automatizar cientos de procesos y TensorFlow nos permite llegar a esta automatización. Nos permite crear y entrenar modelos de Aprendizaje Automático con facilidad mediante APIs intuitivas.
[9][10]

J. Keras

Keras es una biblioteca de código abierto (con licencia MIT) El objetivo de la biblioteca es acelerar la creación de redes neuronales: para ello, Keras no funciona como un framework independiente, sino como una interfaz de programación de aplicaciones de uso intuitivo (API) que permite acceder a varios frameworks de aprendizaje automático y desarrollarlos. Esta interfaz es una excelente manera de comenzar a implementar el aprendizaje automático y, específicamente, las redes neuronales profundas ya que permite una experimentación rápida y se enfoca en su sencillez en la

utilización. Es una opción popular tanto para principiantes o aficionados como para desarrolladores profesionales y proporciona módulos incorporados para todos los cálculos de redes neuronales. Al mismo tiempo, los cálculos que implican tensores, gráficos de cálculo o sesiones, pueden hacerse a medida utilizando la API del núcleo de TensorFlow, lo que provoca una flexibilidad y un control total sobre nuestra aplicación y nos permite implementar nuestras ideas en un tiempo relativamente corto.
[9][10]

H. API de busqueda

Para la búsqueda del producto en la web se hará uso de una API de búsqueda. Si bien existen múltiples APIS disponibles o buscadoes de este tipo. La mas destacada de estas es Content API for Shopping de Google esta permite administrar mediante programación los elementos de datos estructurados que carga en Google para usarlos en las propiedades de Google, como los anuncios de Shopping y la Búsqueda de Google. Puede ser usado para consultar artículos en el mercado.
[3] [5]

H. ImageNet

ImageNet es un proyecto que proporciona una gran base de datos de imágenes con sus correspondientes anotaciones indicando el contenido de las imágenes. Una base de datos de este tipo es imprescindible para la investigación en visión artificial.
[18]

V. DESARROLLO

A. Prueba de concepto

En esta prueba de concepto, hemos desarrollado una aplicación que utiliza inteligencia artificial y una red neuronal convolucional (CNN) para reconocer objetos en imágenes y proporcionar resultados de tiendas cercanas que venden esos productos. La POC demuestra la viabilidad técnica y funcional de la aplicación al incluir el preprocesamiento de imágenes y el entrenamiento de la CNN utilizando TensorFlow y Keras.

Objetivos:

- Recopilar un conjunto de datos de imágenes etiquetadas que contengan una variedad de productos.
- Realizar el preprocesamiento de imágenes
- Entrenar una red neuronal convolucional (CNN) utilizando TensorFlow y Keras, utilizando el conjunto de datos recopilado y preprocesado.
- Implementar un sistema que permita a los usuarios cargar imágenes y obtener resultados de reconocimiento de objetos y resultados de tiendas cercanas.
- Evaluar la precisión del modelo y la relevancia de los resultados de búsqueda obtenidos.

Metodología:

- Recopilación de un conjunto de datos de imágenes etiquetadas de productos.
- Preprocesamiento de imágenes utilizando la clase ImageDataGenerator de Keras, que incluye escalado, rotación, desplazamiento y otras transformaciones.
- Entrenamiento de una red neuronal convolucional (CNN) utilizando TensorFlow y Keras, utilizando el conjunto de datos preprocesado.
- Implementación de un sistema que permite a los usuarios cargar imágenes y obtener resultados de reconocimiento de objetos y resultados de tiendas cercanas utilizando la CNN entrenada.
- Pruebas de integración y evaluación de la relevancia de los resultados de búsqueda obtenidos.

Resultados esperados:

- Una CNN entrenada en el reconocimiento de objetos.
- Un sistema funcional que permite a los usuarios cargar imágenes y obtener resultados de reconocimiento de objetos y resultados de tiendas cercanas.

B. Lista de requerimientos

1. La aplicación presentara un menú para que el usuario elija entre la opción de sacar una foto con la cámara o utilizar una foto de la galería
2. La aplicación analizara la imagen mediante el uso de redes convolucionales. Y detectando en la imagen el objeto destacable.
3. El usuario podrá moverse en la imagen para enfocar otro objeto si el indicado no es el deseado. La aplicación hará el proceso para reconocerlo de nuevo.
4. La aplicación buscara el objeto identificado en la web usando la locación del dispositivo para traer los resultados mas cercanos primero.
4. La app presentara los resultados al usuario de las tiendas cercanas a su locación que encuentre en la web. Que vendan el producto deseado o lo mas similares posible.
5. La app necesitara una conexión a internet durante todo el proceso para poder realizar la búsqueda de este.

C. Hardware

Android:

- OS: Android 6+
- CPU: Dual core 1.2GHz
- Camara
- GPS

También será necesario el equipo conectado a internet para el funcionamiento de la aplicación y hacer el procesamiento de la imagen.

D. Funcionamiento

Con la aplicación descargada el usuario puede abrirla y en la intuitiva interfaz tendrá las 2 opciones disponles. Sacar fotos o buscar una de su galería. En el grafico se puede ver el procesamiento de la imagen. Donde al capturar la imagen la misma es procesada. Si el usuario decide hacer modificaciones la foto es recortada entonces procesada. La misma puede ser

anulada y comenzar el proceso de nuevo insertando la imagen como se ve en el diagrama de la figura 8.

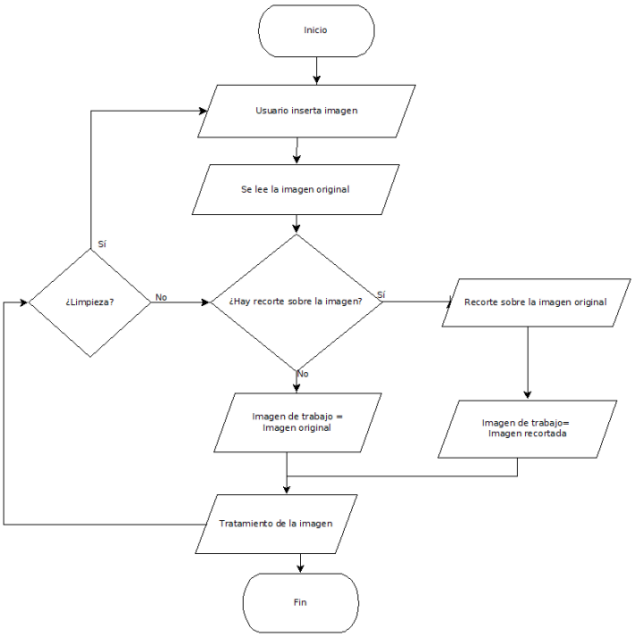


Fig. 8 Diagrama de flujo del procesamiento de la imagen

E. Modelado del Diseño para la AI

- **Recolección de datos:** Se recopilan imágenes y etiquetas de objetos de diversas fuentes de datos.
- **Preprocesamiento de datos:** Las imágenes recopiladas se procesan para prepararlas para su uso en el modelo de aprendizaje automático. Se realizan tareas como el cambio de tamaño, la normalización y el aumento de datos para aumentar la variedad de imágenes disponibles.
- **Construcción del modelo de aprendizaje automático:** Se utiliza una red neuronal convolucional (CNN) para entrenar al modelo de reconocimiento de objetos. Este modelo utiliza TensorFlow y Keras para construir una red de capas que pueden aprender y extraer características importantes de las imágenes de entrada.
- **Entrenamiento del modelo:** El modelo de aprendizaje automático se entrena utilizando un conjunto de datos de imágenes etiquetadas. El objetivo es ajustar los parámetros del modelo para que pueda predecir la clase de un objeto en una imagen de entrada.
- **Validación y ajuste del modelo:** Una vez que se entrena el modelo, se evalúa su precisión utilizando un conjunto de datos de prueba separado. Si el modelo no funciona correctamente, se pueden ajustar sus parámetros y volver a entrenarlo hasta que se logre un rendimiento aceptable.
- **Integración del modelo en la aplicación:** Una vez que se logra un rendimiento satisfactorio del modelo, se integra en la aplicación de reconocimiento de objetos. El modelo se utiliza para predecir la clase de un objeto en una imagen de entrada.
- **Integración con tiendas online:** Se realiza una integración con las tiendas en línea para que la aplicación pueda buscar en línea la tienda que venda el objeto reconocido en la imagen.

F. Configuración de la IA

La aplicación consta de dos componentes principales: un modelo de aprendizaje profundo de redes neuronales convolucionales (CNN) para el reconocimiento de objetos y un sistema de búsqueda de tiendas cercanas.
Para la configuración se utilizó la biblioteca TensorFlow (Fig 19)y su interfaz de alto nivel Keras para construir una red convolucional profunda.

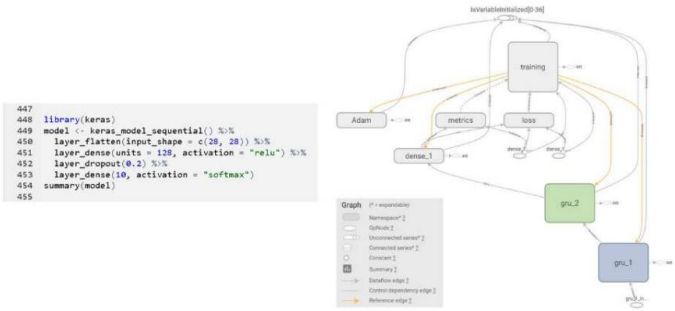


Fig. 19 Grafo generado a partir de código R

La red convolucional se entrenó utilizando un conjunto de imágenes etiquetadas como en el conjunto de datos ImageNet, que consta de más de un millón de imágenes etiquetadas en más de mil categorías. La entrada utilizada para el modelo es una imagen de un producto capturada por la cámara de un dispositivo móvil. El modelo procesa la imagen y devuelve una lista de productos coincidentes.

Se eligió una arquitectura de red convolucional basada en la red VGG16 y se realizaron varias modificaciones para adaptarla a la tarea de reconocimiento de objetos en imágenes tomadas por los usuarios.
El modelo CNN se entrena utilizando un conjunto de datos de imágenes etiquetadas y se ajusta para identificar una amplia gama de productos.

El componente básico de una red neuronal es la capa. Las capas extraen representaciones de los datos que se introducen en ellas.
La mayor parte del aprendizaje profundo consiste en encadenar capas simples. La mayoría de las capas tienen parámetros que se aprenden durante el entrenamiento. Ejemplo de esto se ve en el código siguiente:

```
# Crear una red neuronal secuencial
model = tf.keras.Sequential()

# Agregar capas a la red neuronal
model.add(layers.Dense(64,activation = 'relu', input_shape
= (input_size,)))  # Capa de entrada
model.add(layers.Dense(128, activation='relu')) #Capa oculta1
model.add(layers.Dense(64, activation='relu'))  # Capa oculta2
model.add(layers.Dense(10, activation='softmax'))  #Capa de salida

# Compilar el modelo
model.compile(optimizer='adam',     loss='categorical_crossentropy',
metrics=['accuracy'])
```

```
# Entrenar el modelo
model.fit(train_data, train_labels, epochs=10, validation_data=(test_data, test_labels))
```

En este ejemplo, se crea una red neuronal secuencial y se agregan capas a través del método add(). La primera capa agregada es la capa de entrada, seguida de dos capas ocultas y, finalmente, la capa de salida. Se utiliza la función de activación 'relu' para las capas ocultas y la función de activación 'softmax' para la capa de salida.

Una vez que se agregan todas las capas, se compila el modelo especificando el optimizador, la función de pérdida y las métricas a utilizar durante el entrenamiento. Luego, se entrena el modelo utilizando los datos de entrenamiento y se valida utilizando los datos de prueba.

Por otra parte para la integración con las tiendas en línea, se hace uso de una API de búsqueda de productos como ContentAPI de Google Shopping y la API de Google Maps para identificar las tiendas cercanas que venden el producto en cuestión. Se configuró la API para que devuelva los resultados de búsqueda relevantes para el objeto reconocido en la imagen y se mostraron en la interfaz de la aplicación. Los usuarios pueden hacer clic en un resultado de búsqueda para ir directamente a la tienda en línea donde se vende el producto.

G. Ejemplo código de implementación con Python

Codigo:

```
import cv2
import numpy as np
from tensorflow import keras
from tensorflow.keras.datasets import fashion_mnist

# Cargar la base de datos Fashion MNIST
(train_images, train_labels), (test_images, test_labels) = fashion_mnist.load_data()

# Mapear las etiquetas de Fashion MNIST a categorías de ropa
class_names = [
    "Camiseta/top", "Pantalón", "Jersey", "Vestido", "Abrigo",
    "Sandalia", "Camisa", "Zapatilla deportiva", "Bolso", "Bota"
]

# Cargar el modelo entrenado
model = keras.models.load_model('ruta/del/modelo_entrenado.h5')

# Cargar la imagen de prueba
image_path = 'ruta/de/la/imagen.jpg'
image = cv2.imread(image_path)
image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
image = cv2.resize(image, (28, 28))
image = np.expand_dims(image, axis=0)
image = np.expand_dims(image, axis=3)

# Realizar la predicción del modelo
prediction = model.predict(image)
predicted_class = np.argmax(prediction[0])

# Obtener la etiqueta de la clase predicha
predicted_label = class_names[predicted_class]

# Imprimir el resultado
print("La imagen contiene:", predicted_label)
```

Con este ejemplo, se utiliza la base de datos Fashion MNIST, que contiene imágenes de ropa etiquetadas con 10 categorías diferentes. El modelo entrenado se carga y se utiliza para realizar la predicción en una imagen de prueba. El resultado de la predicción imprimirá en consola el reconocimiento realizado mostrando la etiqueta correspondiente.
Ampliando en base a este código sumando etiquetas con una base da datos mas grandes se pueden obtener resultados de objetos mas complejos y precisos. Como se ve en la figura 20.



Fig. 20 Ejemplo de reconocimiento con etiqueteas con TensorFlow

H. Visión Arquitectura de la solución

En el diagrama (Fig. 18), podemos ver que la aplicación utiliza una base de datos de imágenes como entrada. Luego, estas imágenes se someten a un proceso de preprocesamiento para asegurarse de que tengan un formato adecuado antes de ser procesadas por la Red Neuronal Convolucional (CNN).
La CNN es la encargada de analizar las imágenes y extraer características importantes para su reconocimiento. Esta red neuronal utiliza varias capas convolucionales, de pooling y totalmente conectadas para procesar las imágenes y clasificarlas en diferentes categorías.

Una vez que la imagen es reconocida, se utiliza un motor de búsqueda para encontrar tiendas cercanas que vendan el objeto reconocido. Este motor de búsqueda utiliza una API de búsqueda de productos de Google Shopping o similar para realizar la búsqueda en diferentes tiendas en línea.
Por ultimo, se muestra al usuario una lista de resultados de las tiendas en línea que venden el objeto.

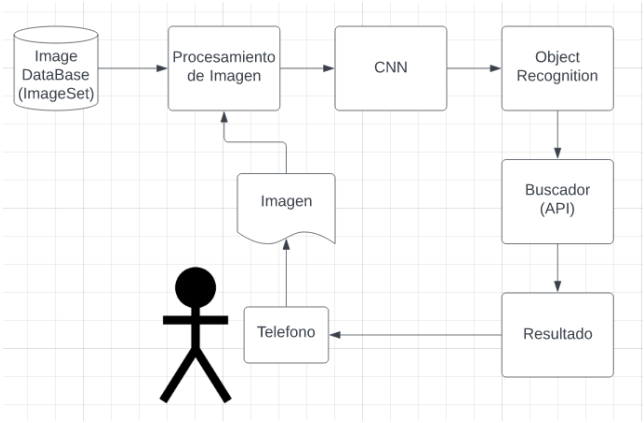


Fig. 18 Diagrama de flujo de la Arquitectura de la Solucion

F. Pantallas

Descripción grafica de ejemplo que tendrían las diferentes pantallas de la aplicación.

- Logo de la aplicación

Ichnaea el nombre de la aplicación proviene de la diosa griega de la búsqueda. Ejemplo en Fig. 9



Fig. 9 Logo de la aplicacion

- Pantalla teléfono

Ilustración de ejemplo de la aplicación instalada en el teléfono (Fig. 10)



Fig. 10 Pantalla de telelfono con la aplicacion

- Pantalla aplicación

En la Fig. 11 se puede ver un ejemplo de la ppantalla Inicial de la aplicación el usuario. En ella se tiene la opción de selección la Cámara o Galería. Según la que se elija es la foto que se analizara.



Fig. 11 Pantalla inicial de la aplicacion

- Pantalla Cámara

Pantalla de la aplicación para sacar fotos y luego aplicar el reconocimiento.
La pantalla tiene las opciones de entrar a la galería para buscar desde las carpetas del dispositivo. (Fig. 12)
Y seleccionar Flash o no.



Fig. 12 Pantalla de la aplicación con la camara seleccionada

- Pantalla búsqueda desde galería

En la opción de galería el usuario puede seleccionar una imagen para utilizar en la búsqueda de objetos como se ve en la Fig. 13



Fig. 13 Pantalla de Galeria

- Pantalla reconocimiento de objeto

Al tomar una foto o elegir una el algoritmo encuadrara los objetos destacables (Fig. 14). El usuario puede moverlo con la pantalla táctil del teléfono y seleccionar otro si así lo desea.
El algoritmo intentara reconocer el objeto por lo que debe ser lo mas claro y reconocible que sea posible. Objetos parciales o borrosos tendrán menor chance de ser reconocidos.

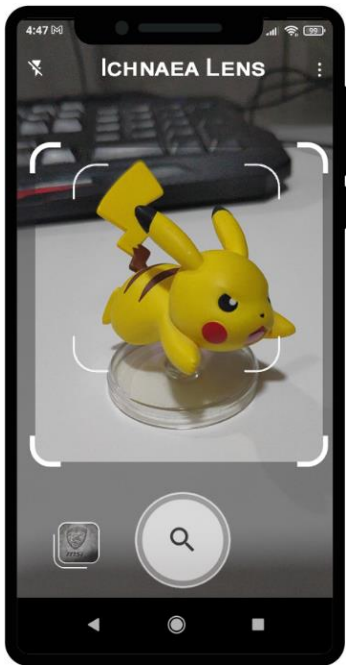


Fig. 14 Pantalla con objeto identificado

- Pantalla Mercado

El objeto seleccionado es buscado por la aplicación a travez de la web trayendo los resultados encontrados en las tiendas mas cercanas según la ubicación indicada po el gps del equipo como se ve en la Fig. 15

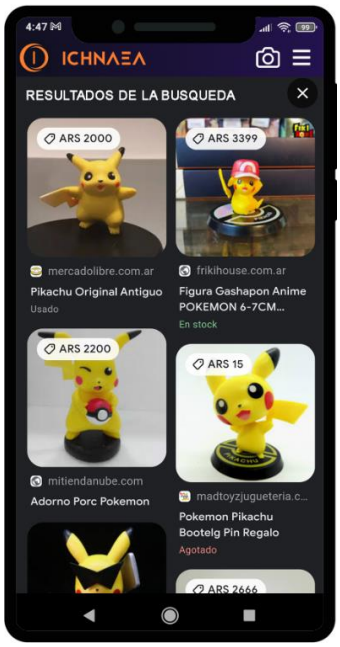


Fig. 15 Pantalla de tiendas encontradas

- Pantalla no objetos encontrados

Según la calidad de la imagen o su contendido se podra dar el caso en que el algoritmo no logre identificar ningun objeto en esta. Por lo que no le podra ser posible encontrar nada en las tiendas cercanas.(Fig. 16)



Fig. 16 Pantalla de resultados no encontrados

- Pantalla Sin Conexión

La aplicación mostrara un mensaje de error si detecta que el dispositivo no esta conectado a internet por lo que no podra realizar la busqueda (Fig. 17)



Fig. 17 Pantalla Sin acceso a Internet

- Pantalla Sin GPS

La aplicación mostrara un mensaje de error si detecta que el dispositivo no tiene el GPS acticavado (Fig. 18)

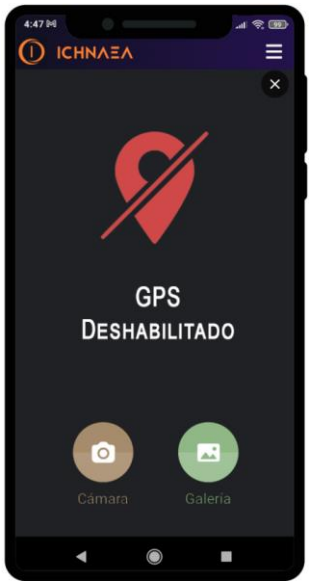


Fig. 18 Pantalla GPS no habilitado

I. Experiencia de la aplicación

¿Como funciona la aplicación?

- El usuario podrá descargar la aplicación desde el store del teléfono.
- Una vez instalada se podrá ingresar a la misma sin crear usuarios o ningún otro requisito mas que tener internet y gps conectado. De darse esta situación una pantalla de error se mostrada.
-Para usar la aplicación el usuario tendrá 2 opciones sacar una foto con la opción de “cámara” o elegir una foto que ya tenga en su galería usando la opción de “Galería” en el menú principal de la app.
- De usar la cámara el usuario podrá usar el flash o no para tomar la foto o si cambia de opinión tiene la opción de moverse a la galería desde la misma pantalla.
-Una vez seleccionada una imagen la aplicación la procesara y mostrara en un encuadre el objeto identificado. El usuario podrá mover la selección con los dedos en la pantalla del teléfono para enfocar otro objeto que desee. Tras esto la aplicación procesara el cambio y buscara el nuevo objeto indicado. De no encontrar resultados o no poder identificar el objeto una pantalla de error se mostrará.
-Al encontrar el objeto la aplicación mostrara de forma consecutiva las tiendas cercanas que cuenten con el objeto deseado o una versión similar de este.
-El usuario podrá ingresar a cualquiera de los resultados que lo llevará a la pagina del producto.

J. Resultados obtenidos con la POC

La prueba de concepto demuestra la viabilidad de la aplicación de reconocimiento de objetos y búsqueda de tiendas utilizando inteligencia artificial y redes neuronales convolucionales. El preprocesamiento de imágenes mejora la precisión del modelo y permite una mejor generalización. La implementación de la CNN utilizando TensorFlow y Keras brinda resultados prometedores en términos de precisión y relevancia de los resultados de búsqueda.

VI. CONCLUSION

Con la investigación se busco analizar la viabilidad en el desarrollo de una aplicación que permita el reconocimiento de objetos en imágenes mediante la utilización de inteligencias artificiales con el uso de redes convolucionales.

Si bien en la actualidad existen otras aplicaciones de reconocimiento de imágenes su funcionamiento es muy diverso y no muy útil para el día a día al intentar buscar algo. En especial en regiones afuera de Estados Unidos.

Como solución para esto se propuso el desarrollo de esta aplicación. únicamente para el reconocimiento de objetos en fotos o imágenes bajadas de la red y buscarlos según la localidad del usuario.

Tras lo estudiado la viabilidad de la aplicación es positiva. Si bien para su perfeccionamiento implicara un alto margen de tiempo en su realización es posible y altamente funcional.

La generación de la base de datos correcta con las etiquetas apropiadas a utilizar por la CNN abarcaría una buena cantidad de tiempo a invertir. Ya que si bien para el proceso de prueba se puede contar con bases de datos gratuitas como imageSet las mismas no podrán ser utilizadas a la hora de una aplicación orientada al mercado.

En conclusión, se comprobó que es totalmente factible realizar el efectivo reconocimiento de objetos en imágenes con el uso de AIs y redes convolucionales utilizando TensorFlow y Keras. A su vez se noto que las nuevas tecnologías permiten que este proceso sea cada vez mas sencillo.

VII. LINEAS FUTURAS DE INVESTIGACION

- Agregar funciones a la aplicación similares a otras en el mercado como el reconocimiento de texto

VIII. REFERENCIAS

Referencias:

[1] Sofia Aliaga and Ronald Cortez. Reconocimiento de objetos usados beacons. *Revista de Investigación Estudiantil Iluminate*, 9:50, 2017.

[2] Zhanna I Bilyk, Yevhenii B Shapovalov, Viktor B Shapovalov, Anna P Megalinska, Fabian Andruszkiewicz, and Agnieszka Dolhańczuk-Śródka. Assessment of mobile phone applications feasibility on plant recognition: comparison with google lens ar-app. CEUR Workshop Proceedings, 2020.

[3] Alicia Erlich. Search multiple websites at a time with all in one search app, 2013

[4] Nuria Fernández Pérez et al. La inteligencia artificial como motor de pro- greso: el dif´ıcil equilibrio entre los derechos de propiedad intelectual y la privacidad. 2021.

[5] Charles Frankel, Michael J Swain, and Vassilis Athitsos. Webseer: An image search engine for the world wide web. Technical report, Technical Report 96-14, University of Chicago, Computer Science Department, 1996.

[6] Wilmer Giraldo Loaiza. Reconocimiento de objetos usando técnicas de inteligencia artificial. 2020.

[7] W Gómez. Reconocimiento de objetos en fotografías. *CINVESTAV-LTI Top Tamaulipas*, 2015.

[8] Adam Kubany, Shimon Ben Ishay, Ruben Sacha Ohayon, Armin Shmilovici, Lior Rokach, and Tomer Doitshman. Comparison of state-of- the-art deep learning apis for image multi-label classification using semantic metrics. *Expert Systems with Applications*, 161:113656, 2020.

[9] Javier Martínez Llamas. Reconocimiento de imágenes mediante redes neuronales convolucionales. 2018.

[10] Francisco de Paula Quirós Pérez. Reconocimiento de imágenes con tensor- flow desde r. 2022.

[11] José-Luis Ramírez-Arias Ph D, Astrid Rubiano-Fonseca Ph D, and Robin- son Jiménez-Moreno Ph D. Object recognition through artificial intelligence techniques. *Revista Facultad de Ingeniería*, 29(54), 2020.

[12] Ivan Razmírez. Nuevo google lens: así puedes identificar objetos y textos directamente con la cámara de tu móvil, 2018.

[13] Lasse Rouhiainen. Inteligencia artificial. *Madrid: Alienta Editorial*, 2018.

[14] José Luis Villaluenga Morán. Openstreetcam: reconocimiento automático de objetos en imágenes mediante machine learning. 2019.

[15] Rodríguez, J. "Cómo empezar con Reconocimiento de Imagen y Redes Neuronales Convolucionales en 5 minutos." *línea]. Disponible enhttps://planetachatbot. com/c% C3% B3mo-empezar-con-reconocimiento-de-imagen-y-redes-neuronales-convolucionales-en-5-minutos-7f651054dfd7*.

[16] Moreno Fernández, Samuel. "Herramienta de Reconocimiento de Imágenes en Python." (2020).

[17] Minaya Montalvo, Marcos. Análisis comparativo de diversos modelos de Deep Neural Networks aplicados al reconocimiento de imágenes. BS thesis. 2018.

[18] Data, S.B. (2019) *ImageNet: Reconocimiento visual a Gran Escala*, *sitiobigdata.com*.Availableat:https://sitiobigdata.com/2019/05/01/imagenet-reconocimiento-visual-a-gran-escala/#

IX. CV

Albornoz Martín Nahuel (albornozmartinn@gmail.com)
CV-AlbornozMartin