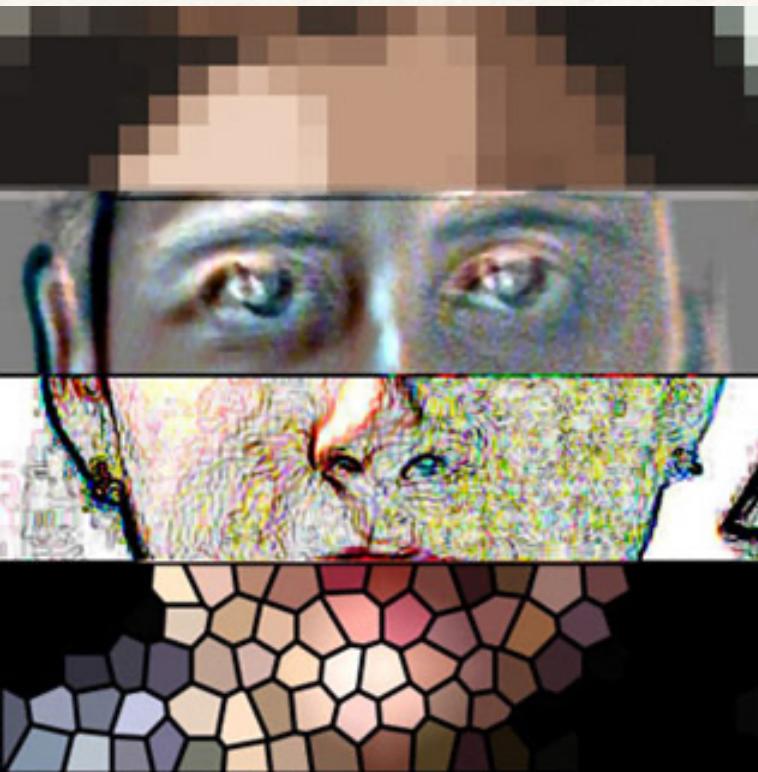


*Visión computacional*

# Detección de objetos

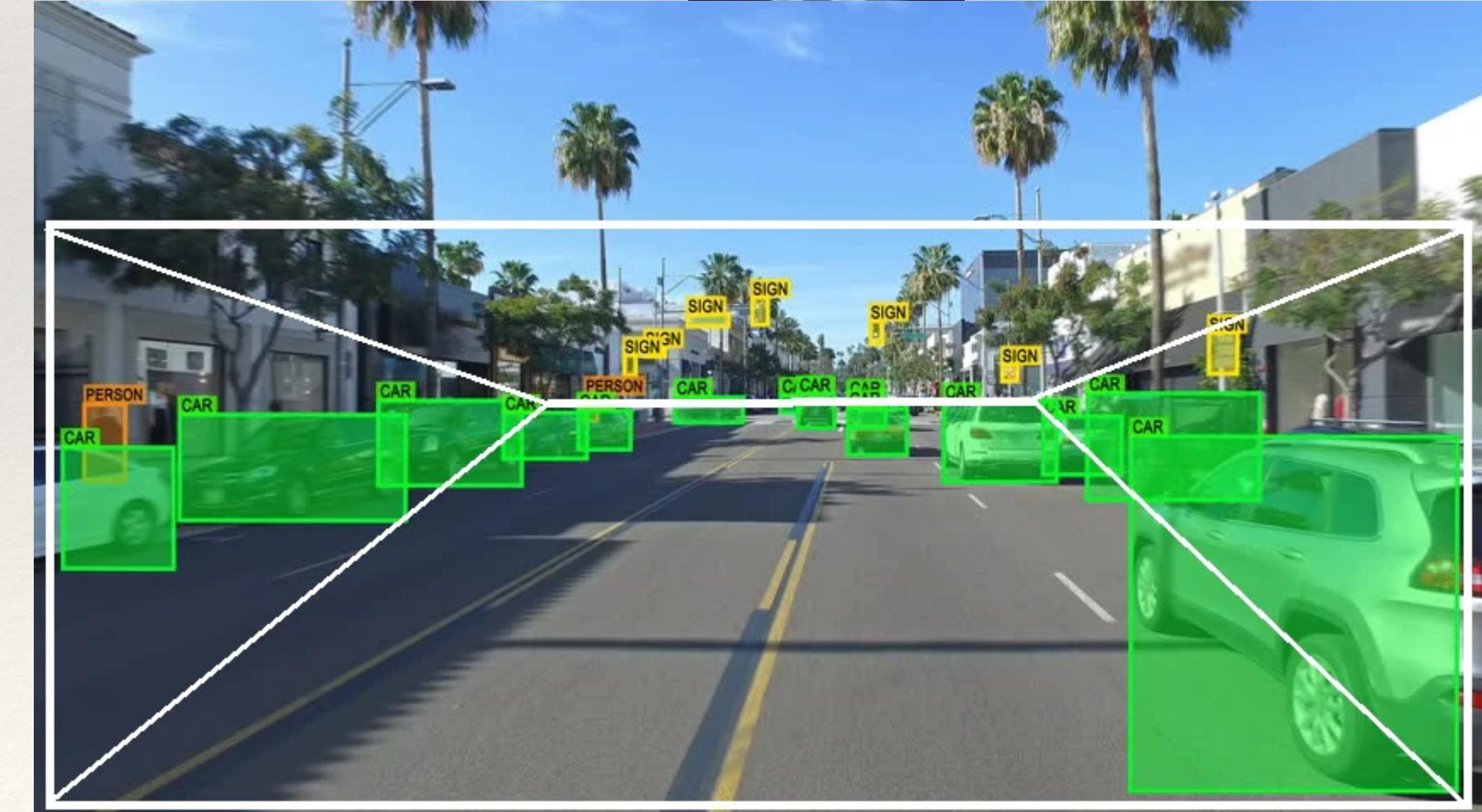
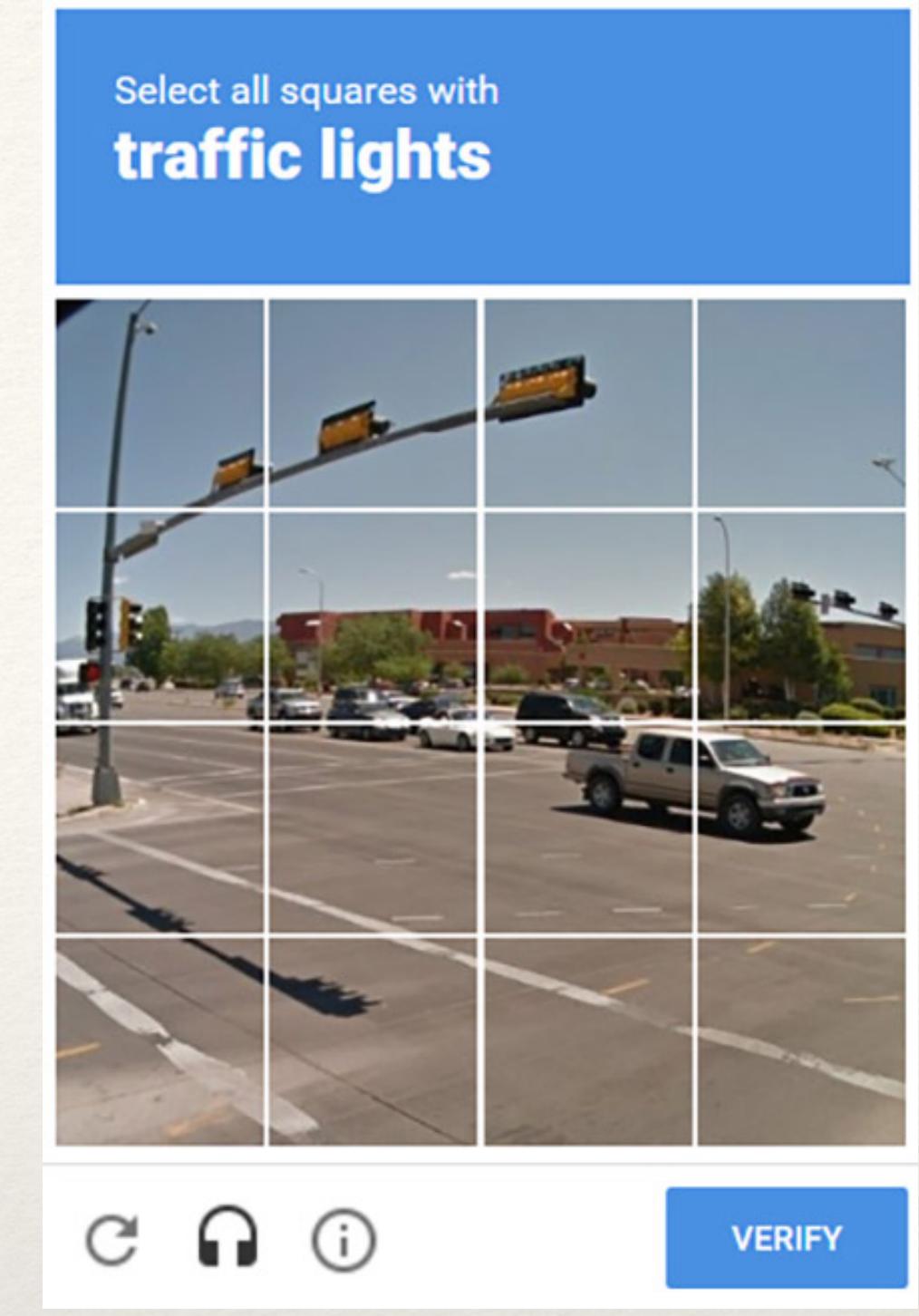
Natalia Jurado

*Ingeniera Electrónica  
Master en Ingeniería Automotriz  
ITESM - Toluca*



# Agenda

1. Visión computacional
  1. Aplicaciones
  2. Historia y estado del arte
2. Detección de objetos
  1. ¿Qué es?
  2. Importancia
  3. Aprendizaje profundo
  4. Como funciona
  5. Redes neuronales convolucionales
3. Aplicaciones de detección de objetos
4. Ejemplo de detección de objetos usando YOLOv5



# A través del tiempo@MIT

- ❖ 1996 => The summer project

Artificial Intelligence Group  
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into subproblems, which will divide a midsector picture into regions such as likely objects, likely background, and chaos.

La tarea fue particularmente seleccionada porque puede ser segmentada en subproblemas, lo que permitirá la participación individual e independiente en la construcción de un sistema lo suficientemente complejo como para dejar una marca en el desarrollo del reconocimiento de patrones

## Goals - General

El objetivo principal de este proyecto es la construcción de un sistema de programas que dividirán una imagen en regiones tales como:

likely objects

Objetos

likely background

Regiones de fondo (fondo de una imagen)

chaos.

Caos

We shall call this part of its operation FIGURE-GROUND analysis.

It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really inseparable in practice from the second goal which is REGION DESCRIPTION.

The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

# Avances en visión computacional

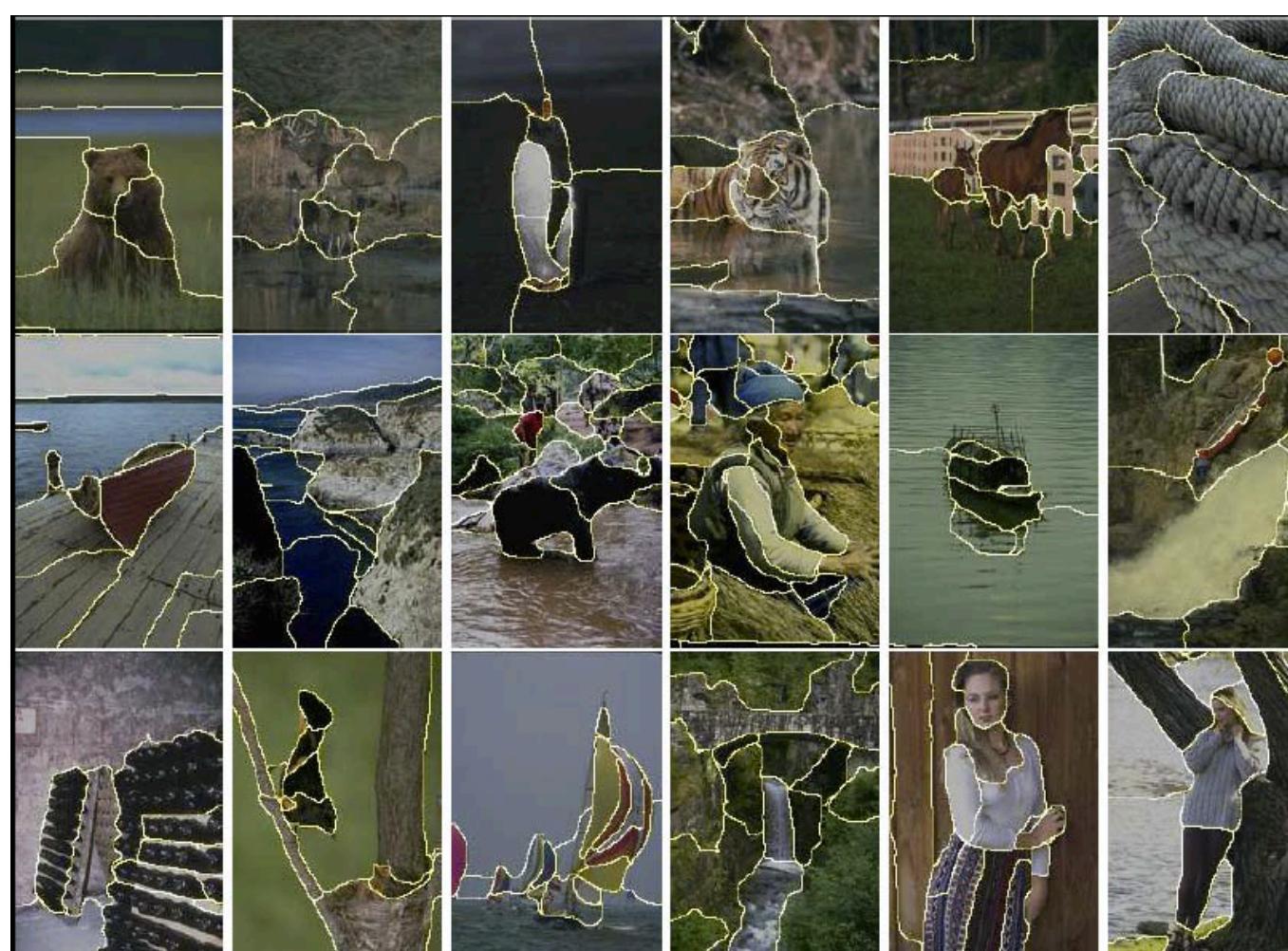
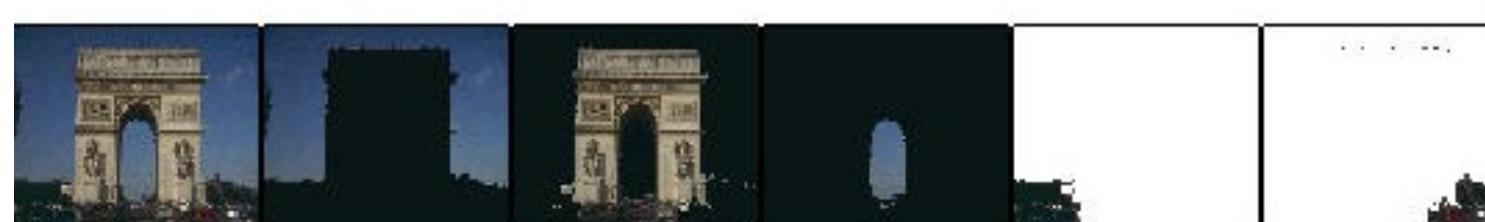
❖ 2012

## 1.2 Why is vision hard?

Es complicado puntualizar las razones exactas por las que la visión computacional es una dura tarea, debido a que no se tiene una solución todavía. Dos aspectos que tiene considerable complejidad son:

Estructura de la entrada (imagen de entrada)

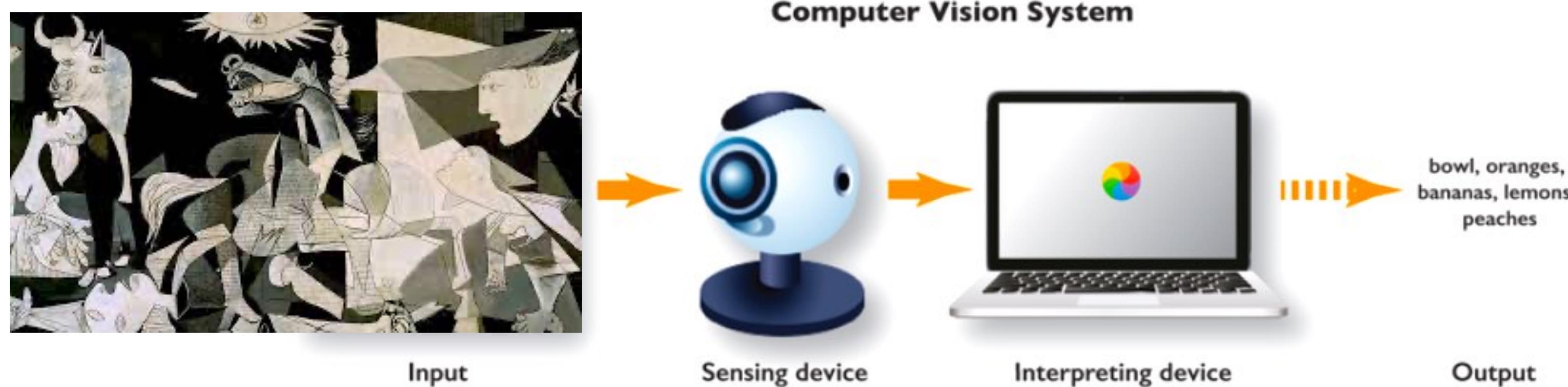
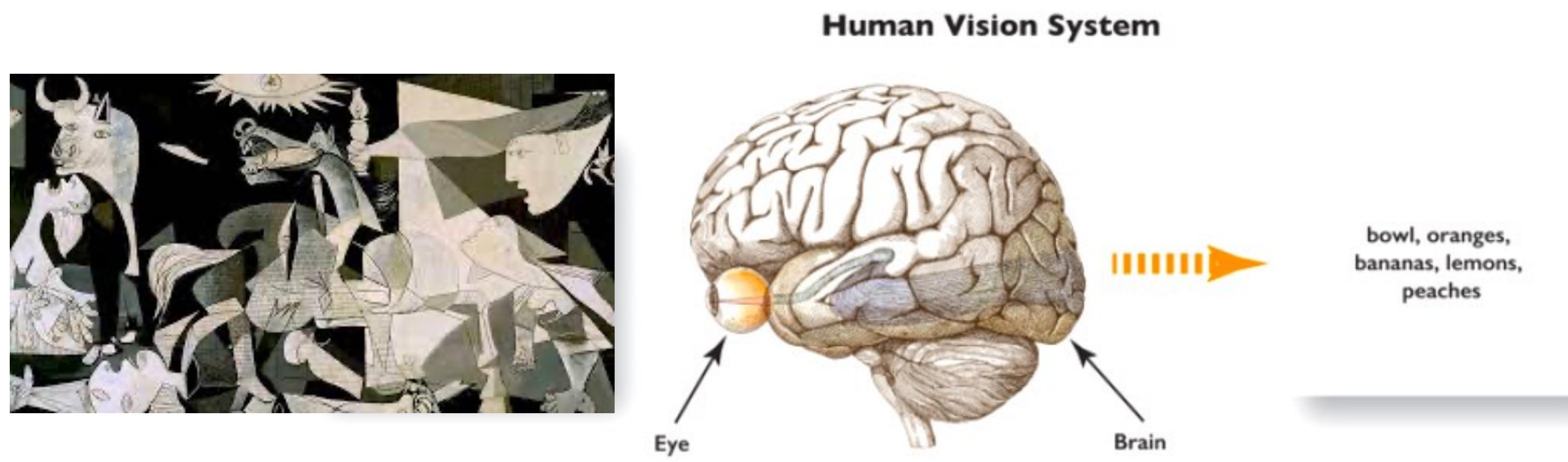
Estructura deseada para la salida (es una imagen? Un valor? Una palabra?)



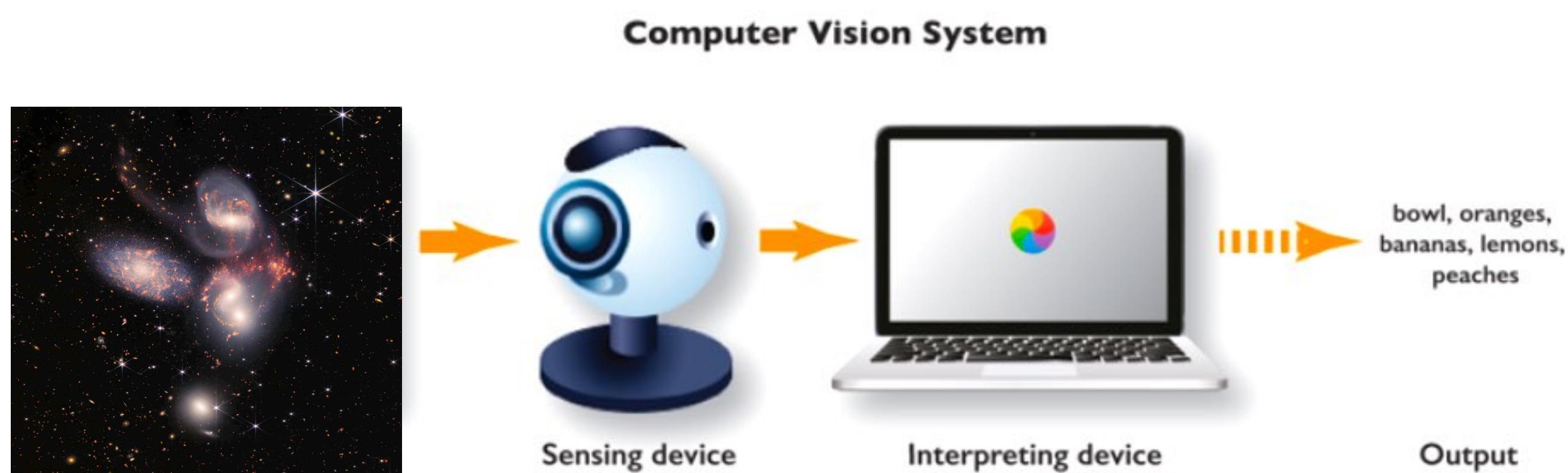
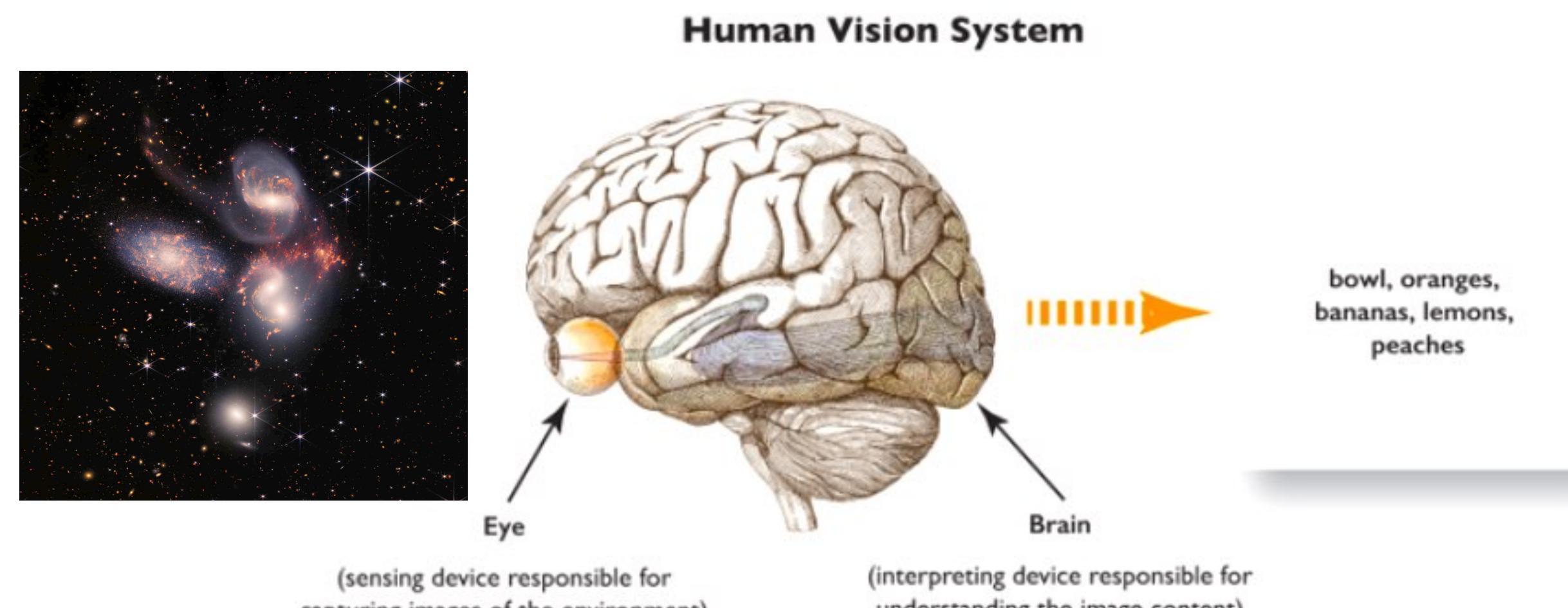
### Class Notes and Slides

Lecture 1	A Simple Vision System	<a href="#">Lecture notes</a> (Sep 6, 2012) Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 2	Linear Filters	No note. Refer to Chapter 3 of [Sz]. Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 3	Pyramids and scale space	No note. Refer to Chapter 3 of [Sz]. Slides: <a href="#">ppt</a>   <a href="#">pdf</a>   <a href="#">keynote</a>
Lecture 4	Image statistics I	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 5	Image statistics II	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 6	Color	Slides: <a href="#">keynote</a>   <a href="#">pdf</a>
Lecture 7	Texture	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 8	Edge and Segmentation	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 9	Image Formulation Models	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 10	Image Formulation Models	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 11	Shape from X	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 12	Image features	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 13	Homographies and RANSAC	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 14	Multi-view 3D Reconstruction	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 15	Markov Random Field	<a href="#">Lecture notes</a> Slides: <a href="#">pdf</a>
Lecture 16	Motion I	Slides: <a href="#">pdf</a>
Lecture 17	Motion II	Slides: <a href="#">pdf</a>
Lecture 18	Object Recognition I	Slides: <a href="#">ppt</a>
Lecture 19	Object Recognition II	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 20	Object Recognition III	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 21	Object Recognition V	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 22	Scene Understanding	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>
Lecture 23	Scene Understanding	Slides: <a href="#">ppt</a>   <a href="#">pdf</a>

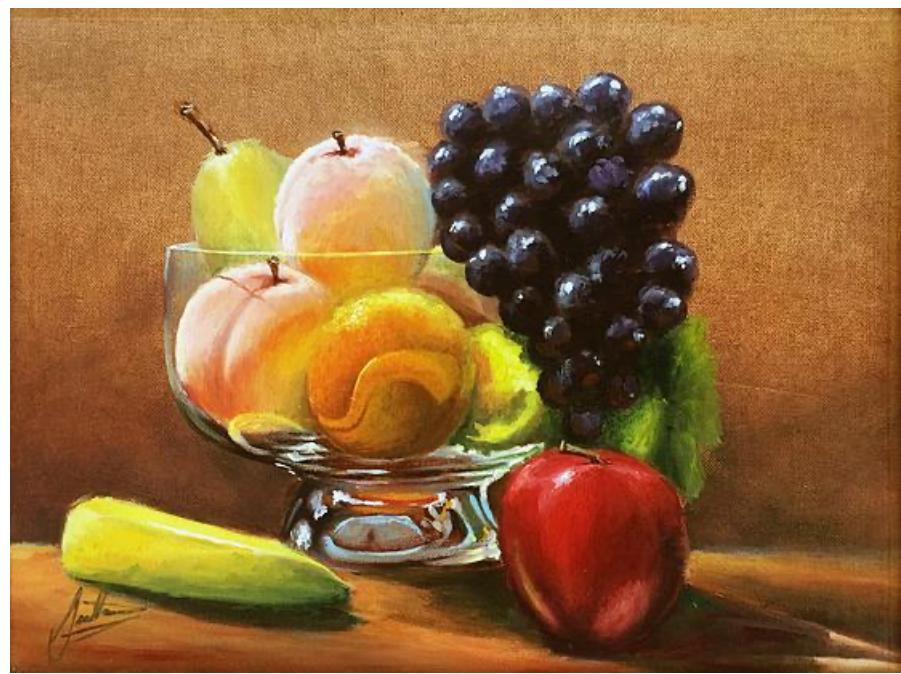
# Sistema de visión computacional



# Sistema de visión computacional



# División de tareas en la visión computacional

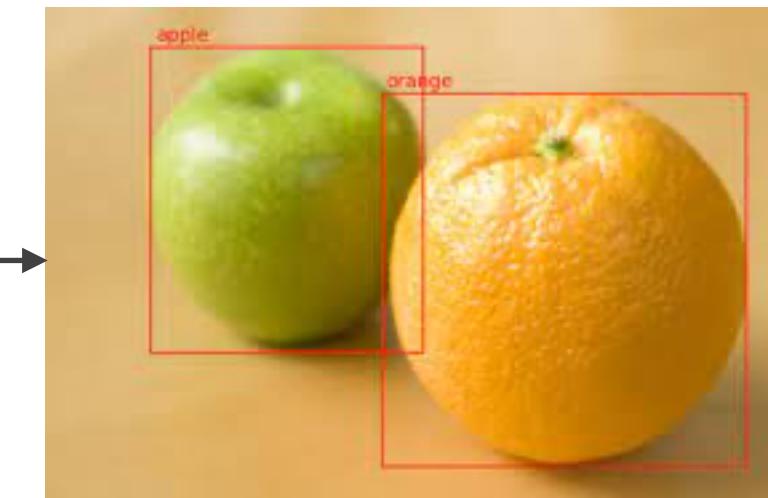


## Per-procesamiento

- remover ruido
- simplificar colores
- ...

Procesamiento

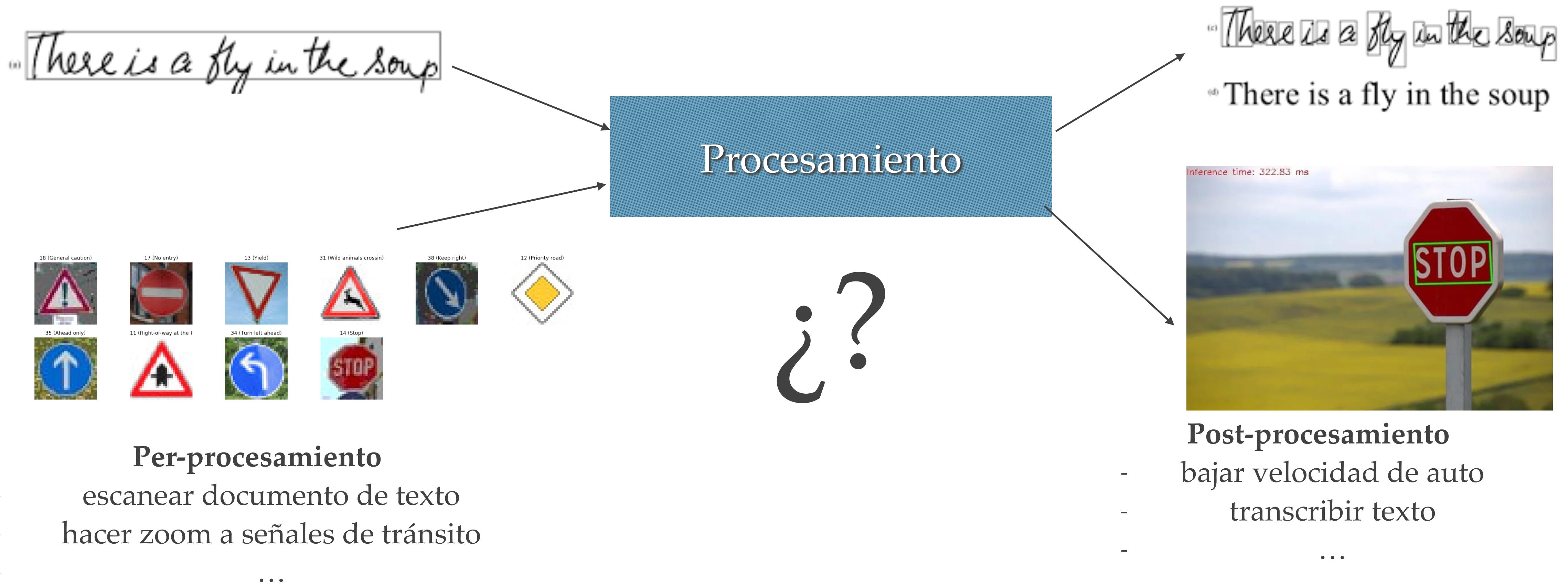
?



## Post-procesamiento

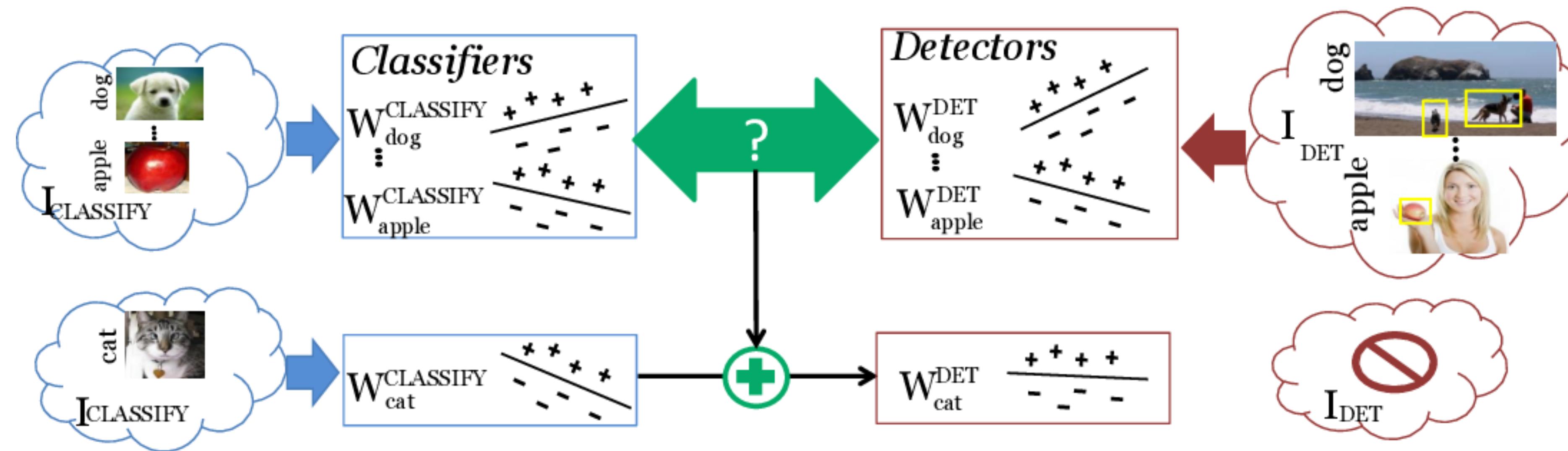
- etiquetar objetos
- encuadrar objetos
- ...

# División de tareas en la visión computacional



# En 2022...

- ❖ Bancos de imágenes: coco: CommonObjectsInCOnText, ...
- ❖ Bancos de modelos: detectores y clasificadores
- ❖ Compiladores para modelos de ML (MachineLearning)
- ❖ GPUs con librería para multithreading (CUDA)



# En 2022...

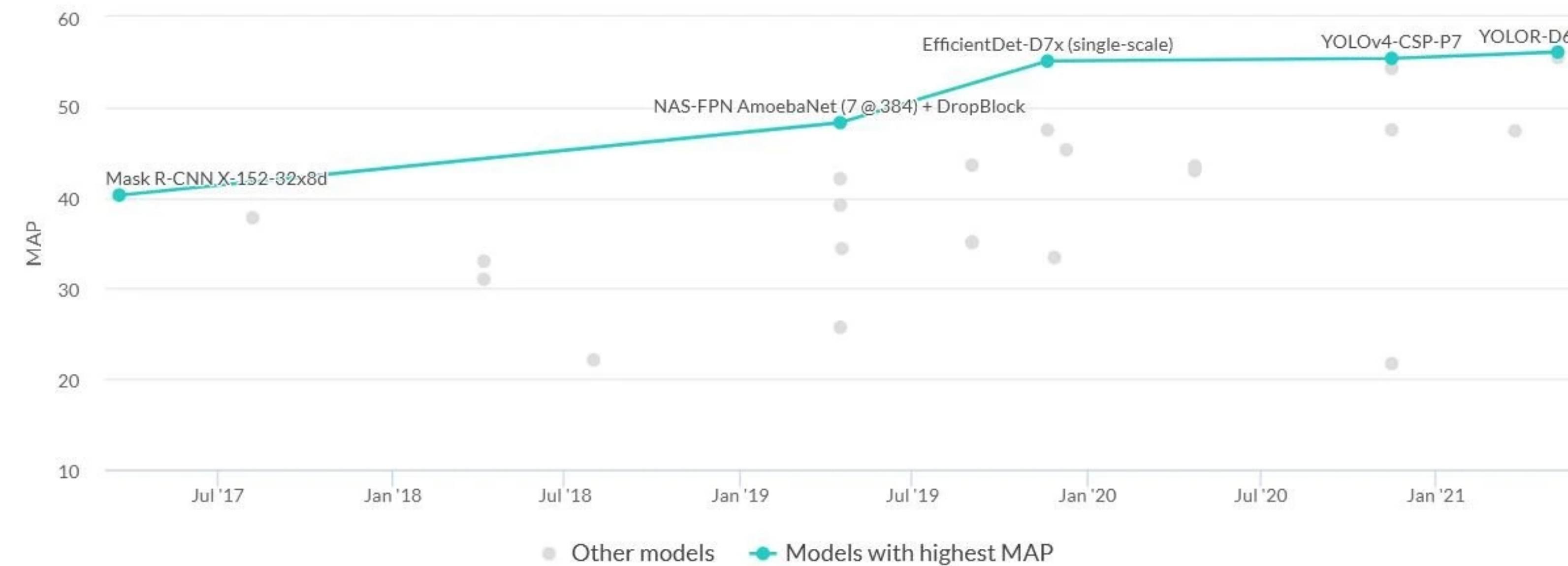
El campo de la detección de objetos no es tan nuevo como parece.

De hecho, la detección de objetos ha evolucionado en los últimos 20 años.

El progreso de la detección de objetos generalmente se divide en dos períodos históricos separados (antes y después de la introducción de Deep Learning):

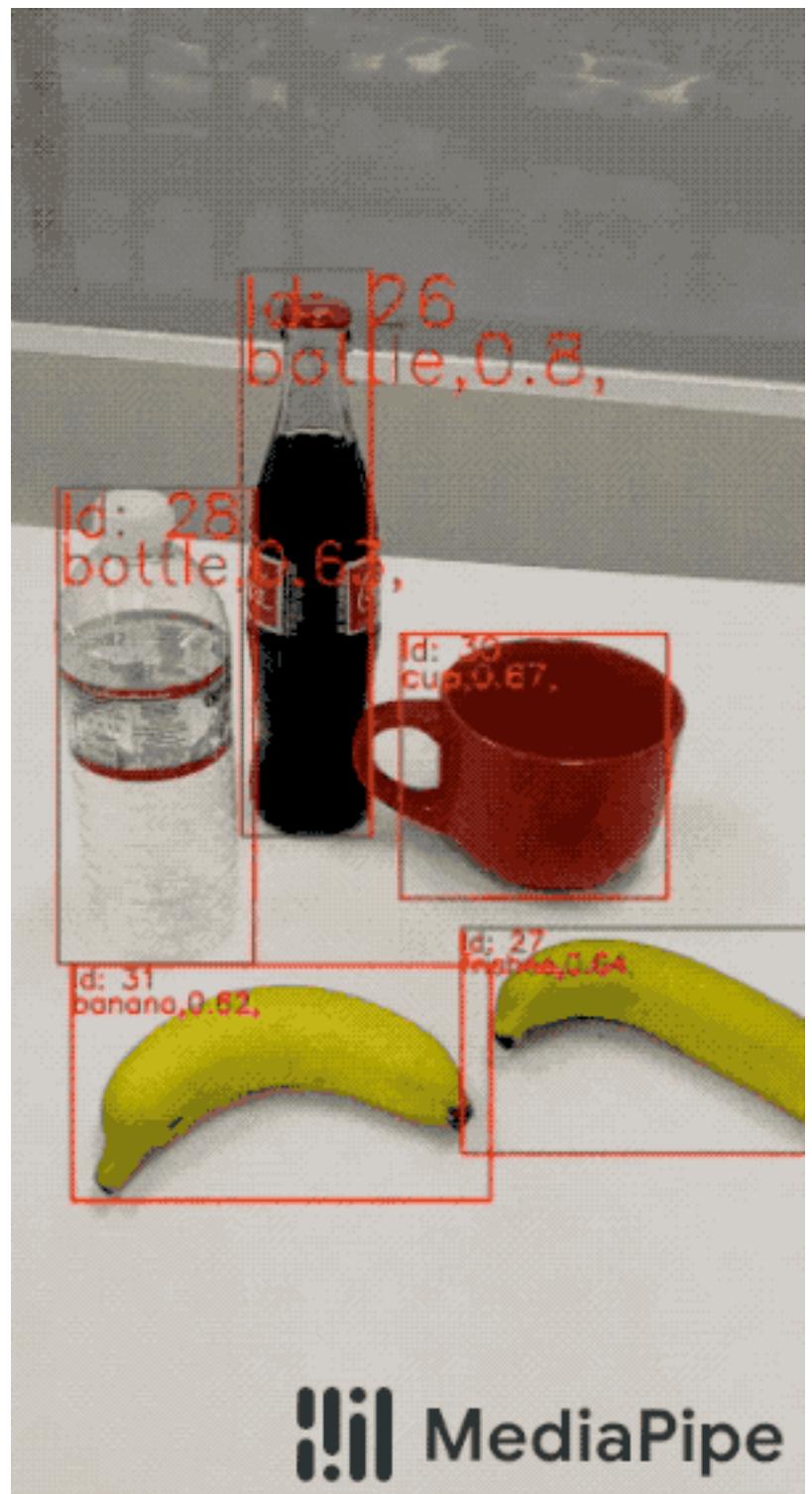
**Antes de 2014** : período tradicional de detección de objetos

**Después de 2014** : período de detección de aprendizaje profundo



# Detección y reconocimiento de objetos

# Objetivo de la detección de objetos



El objetivo de la detección de objetos es desarrollar modelos computacionales que proporcionen la información más fundamental que necesitan las aplicaciones de visión artificial : " ¿Qué objetos están dónde ?".

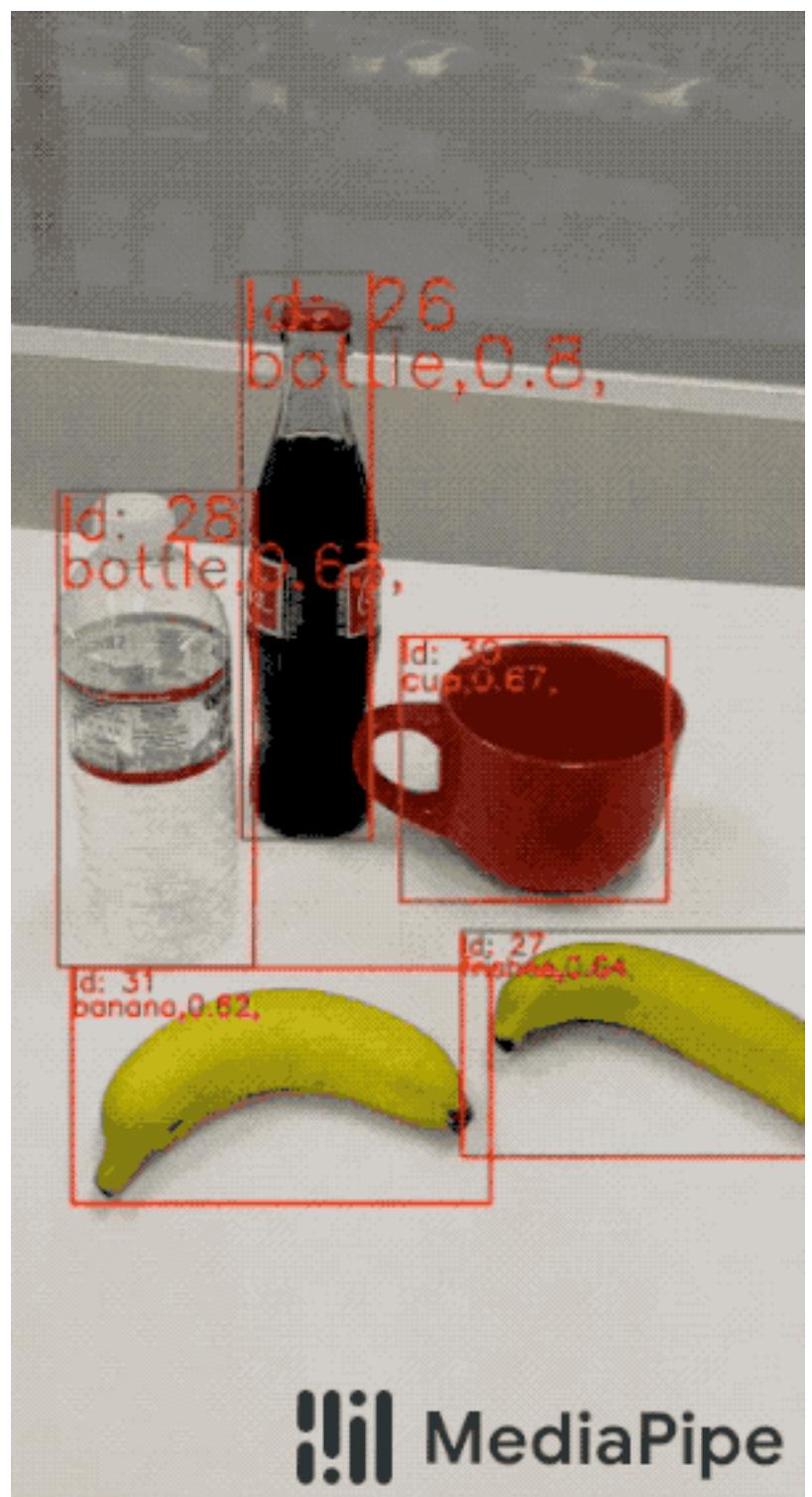
# ¿Por qué es importante la detección de objetos?

La detección de objetos es uno de los problemas fundamentales de la visión artificial.

Forma la base de muchas otras tareas posteriores de visión por computadora

- segmentación de instancias,
- subtítulos de imágenes,
- seguimiento de objetos y más.

Las aplicaciones específicas de detección de objetos incluyen detección de peatones, [conteo de personas](#) , detección de [rostros](#) , detección de texto, detección de [poses](#) o [reconocimiento de placas](#) .



## El objeto



Los tipos de objetos



El objeto



Los tipos de objetos



El objeto



La escena



# Instancias vs. Categorías

**Instancias** Encuentra estos 2 juguetes



Puede hacerlo!

**Categorías** Encuentra la botella:

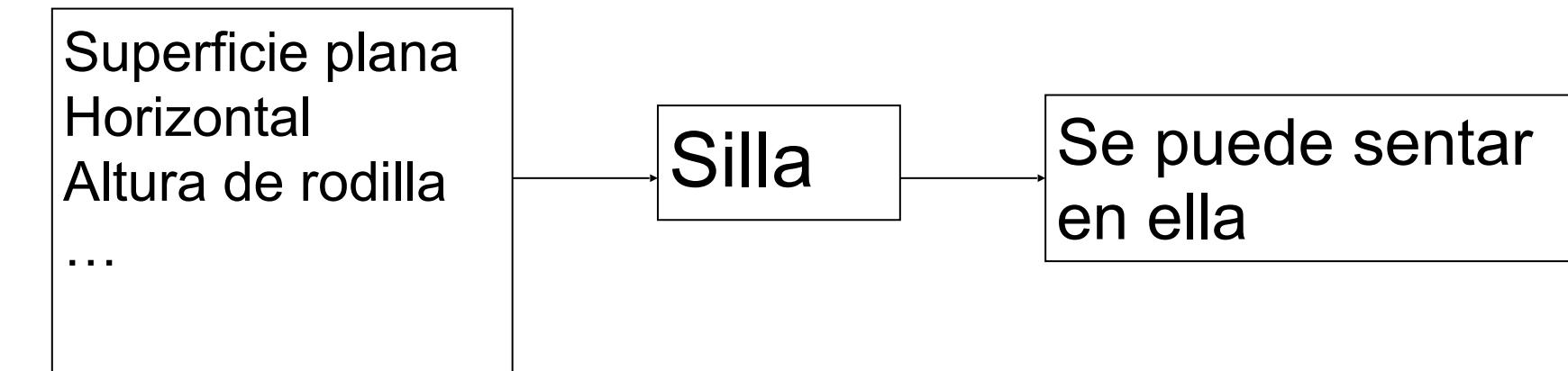


No puede hacerlo, a menos  
que no te importen ciertos  
errores...

# ¿Por qué nos importa el reconocimiento?

- ❖ Percepción de función: Nosotros podemos percibir la forma 3D, la textura, el material, propiedades del material sin conocer el objeto. **Pero, el concepto de categoría encapsula también la información acerca de lo que se puede hacer con el objeto.**

- Percepción mediata (Categorización)

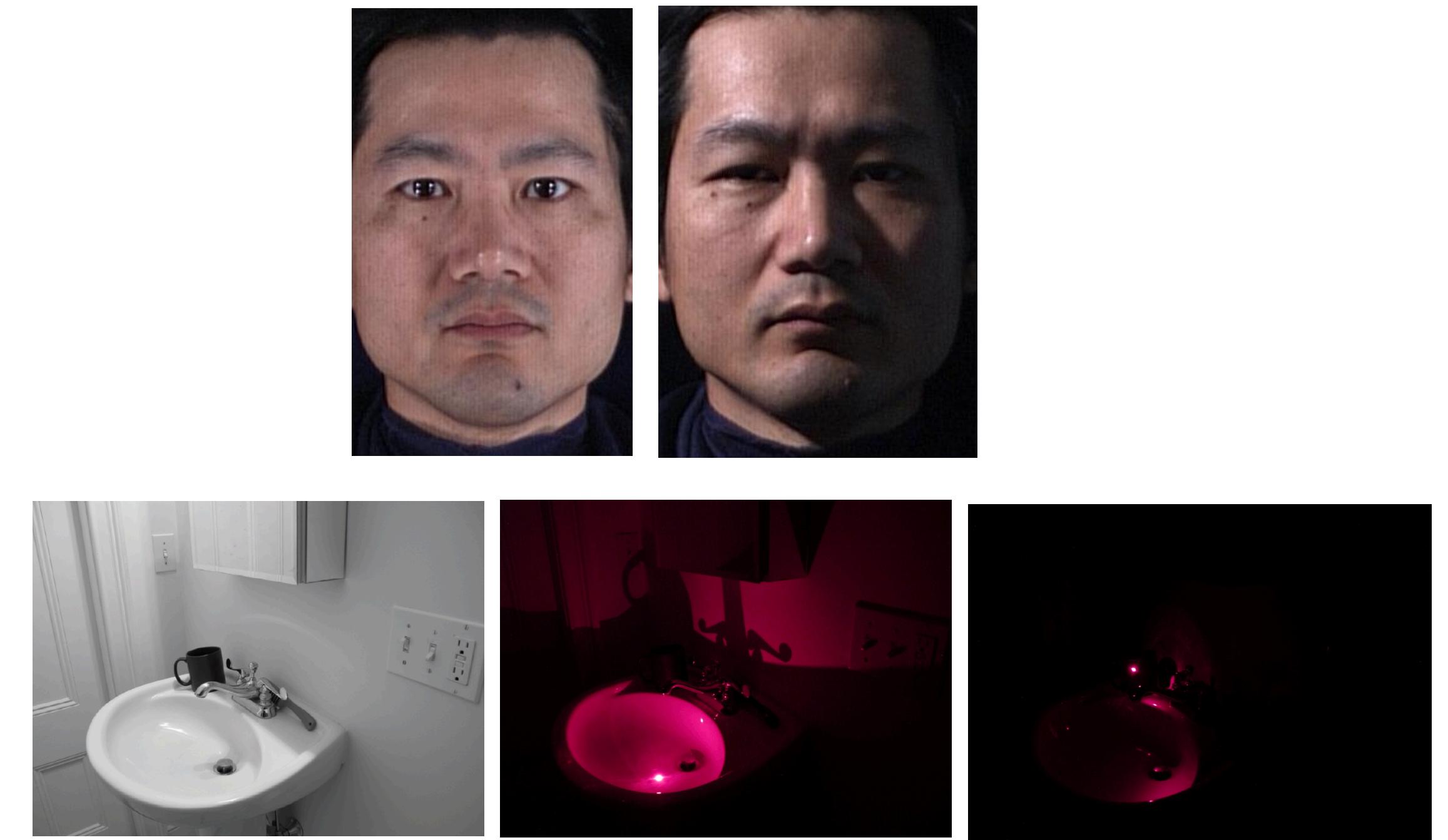


# Problemas generales en detección de objetos

❖ Variación de punto de vista



❖ Iluminación



# Problemas generales en detección de objetos

## ❖ Oclusión

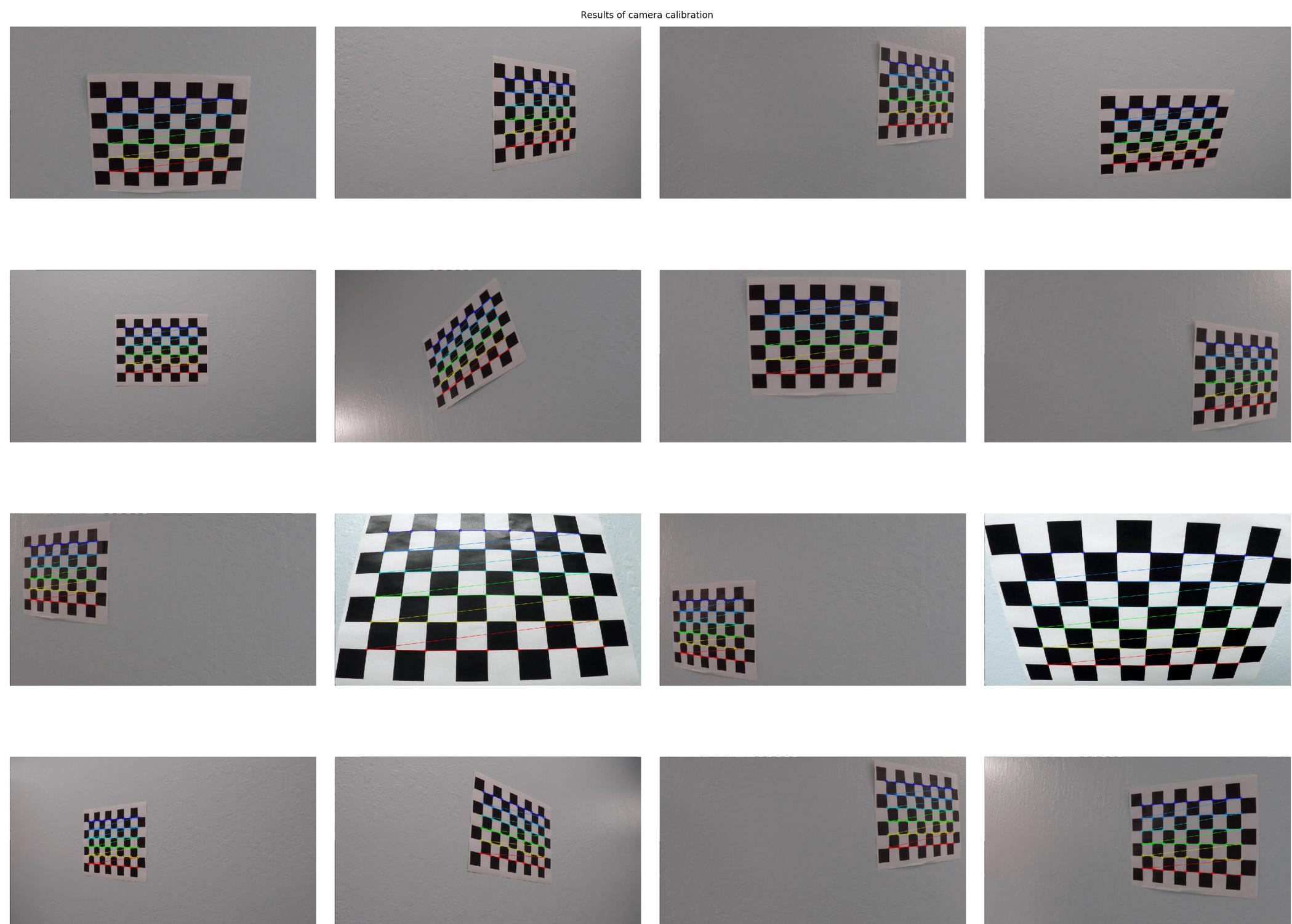


## ❖ Escala



# Problemas generales en detección de objetos

## ❖ Deformación



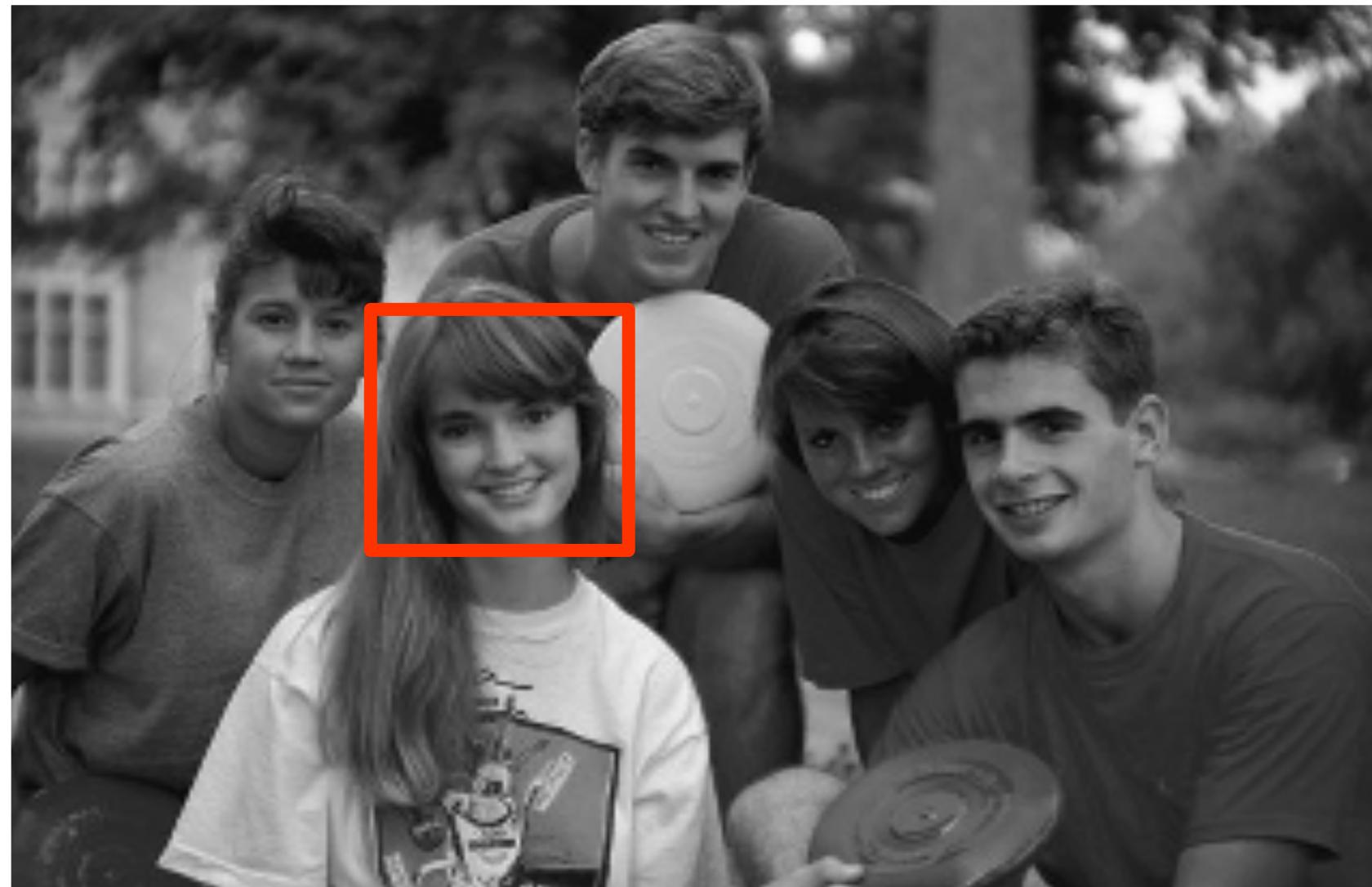
## ❖ Variación intra-class



# Reconocimiento de objetos/rostros

## ❖ Detección de objetos

¿Dónde esta Gabby?  
¿Dónde hay rostros?  
¿Hay rostros en la imagen? = Clasificación



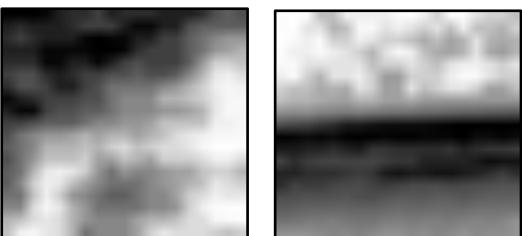
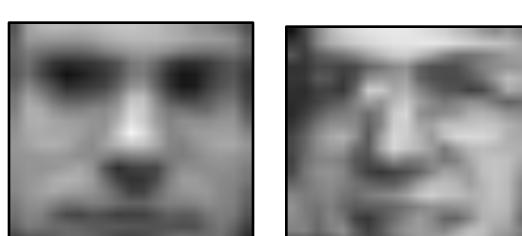
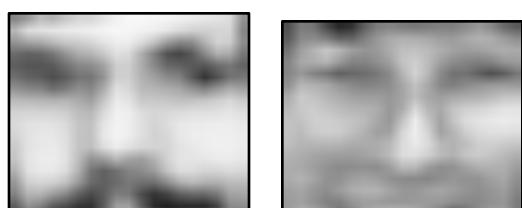
## ❖ Identificación de objetos

¿Quién es?  
¿Es Gabby o Erik?



## Detección de rostros - esquema básico

Ejemplos de rostros



Ejemplos de no-rostros

Off-line  
entrenamiento



Clasificador

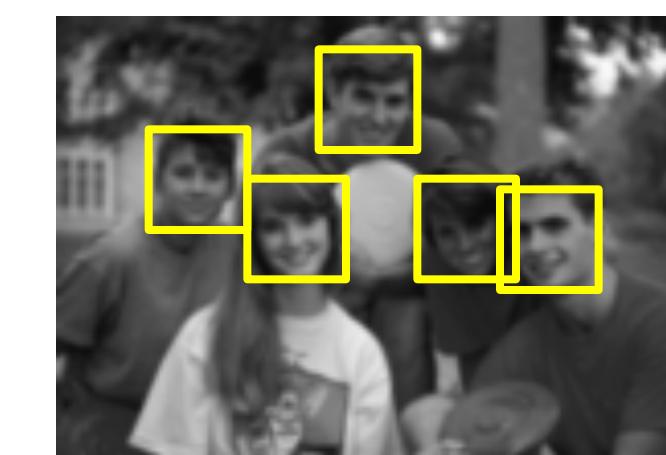
Vector de rasgos ( $x_1, x_2, \dots, x_n$ )

Extracción de rasgos

Patrón de pixeles



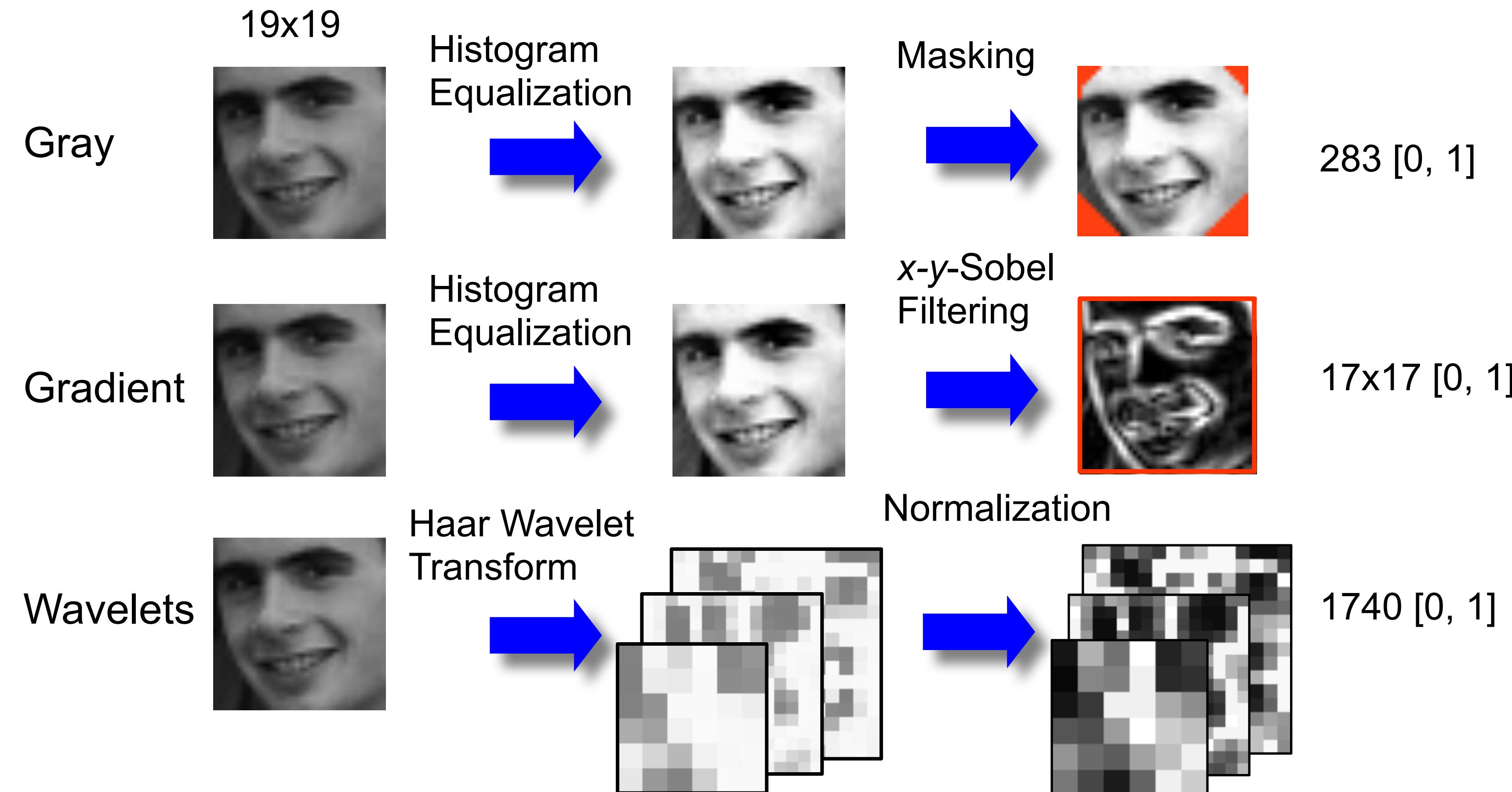
Encontrar rostros en diferentes  
ubicaciones y resoluciones



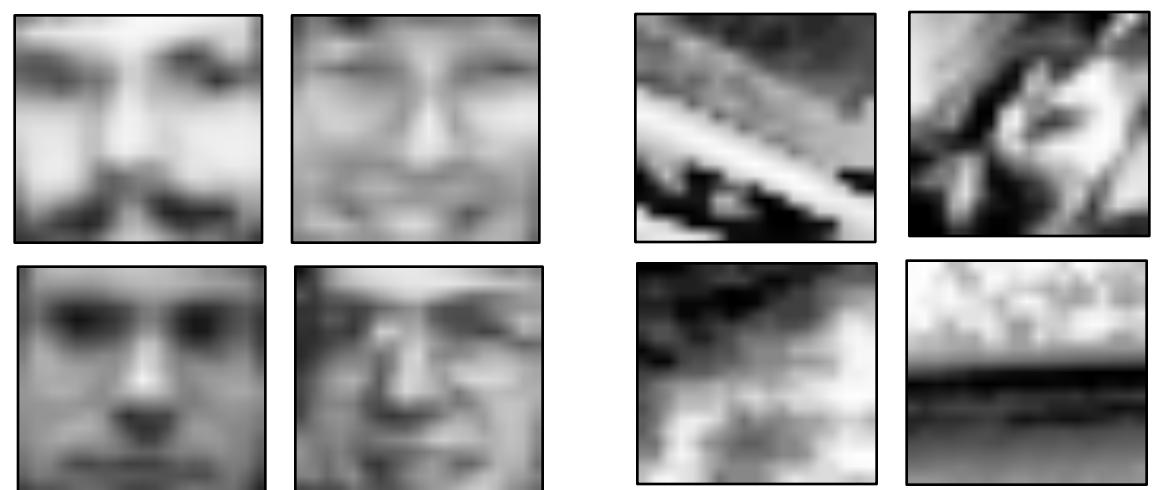
Resultado de  
la clasificación



## Rasgos de la imagen

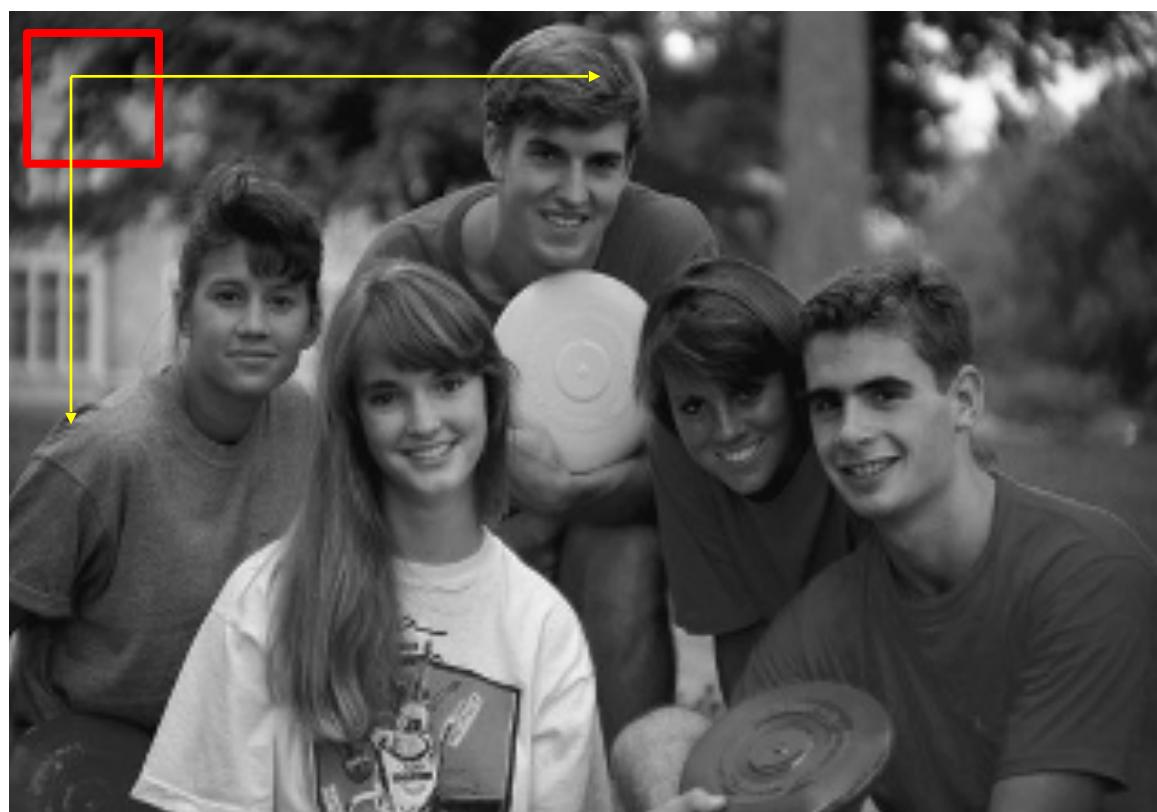


## Set de entrenamiento



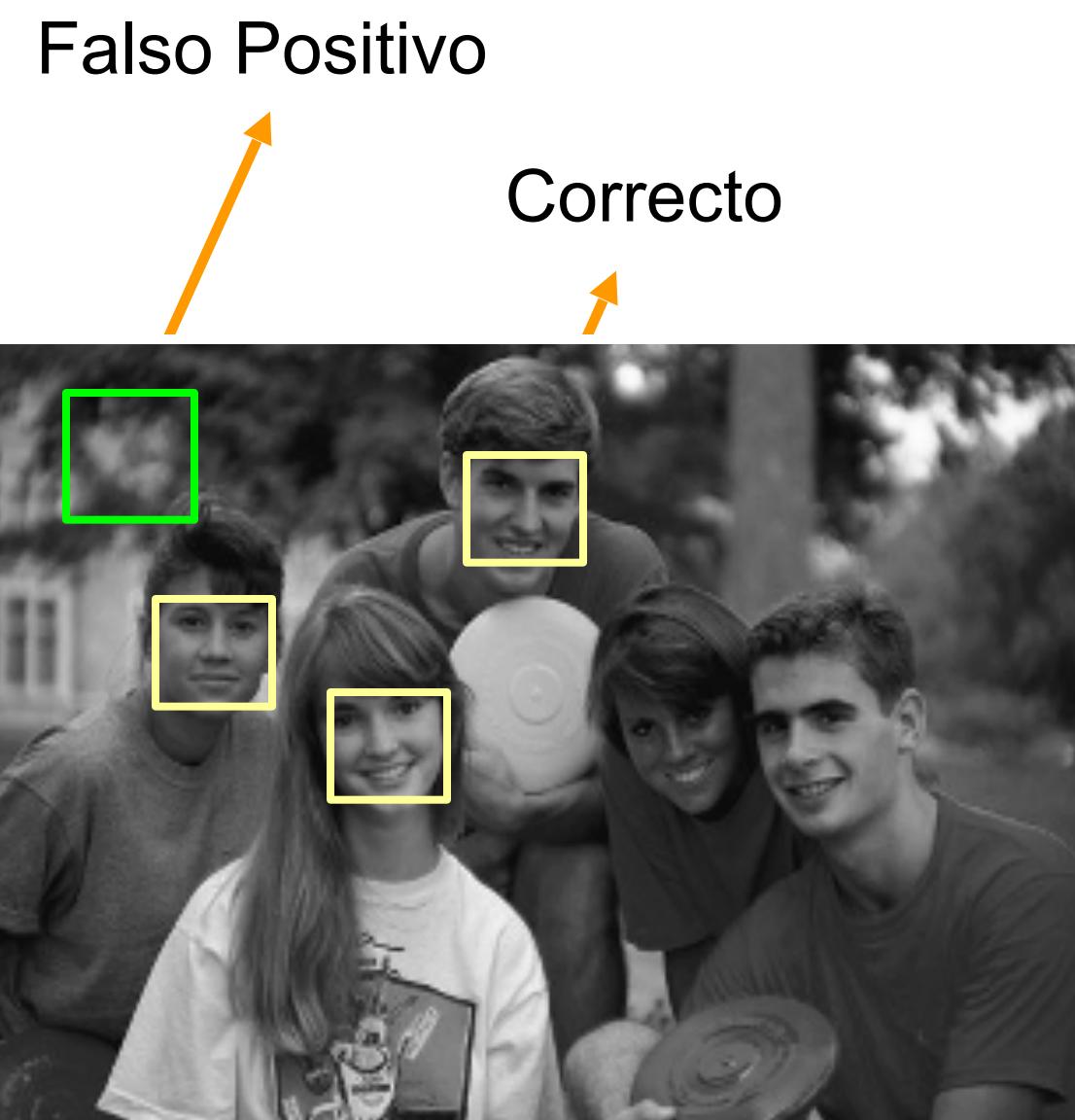
Entrenamiento  
del clasificador

## Set de pruebas



Sensitivity

Clasifica



Correcto

Falso Positivo

# En 2022...

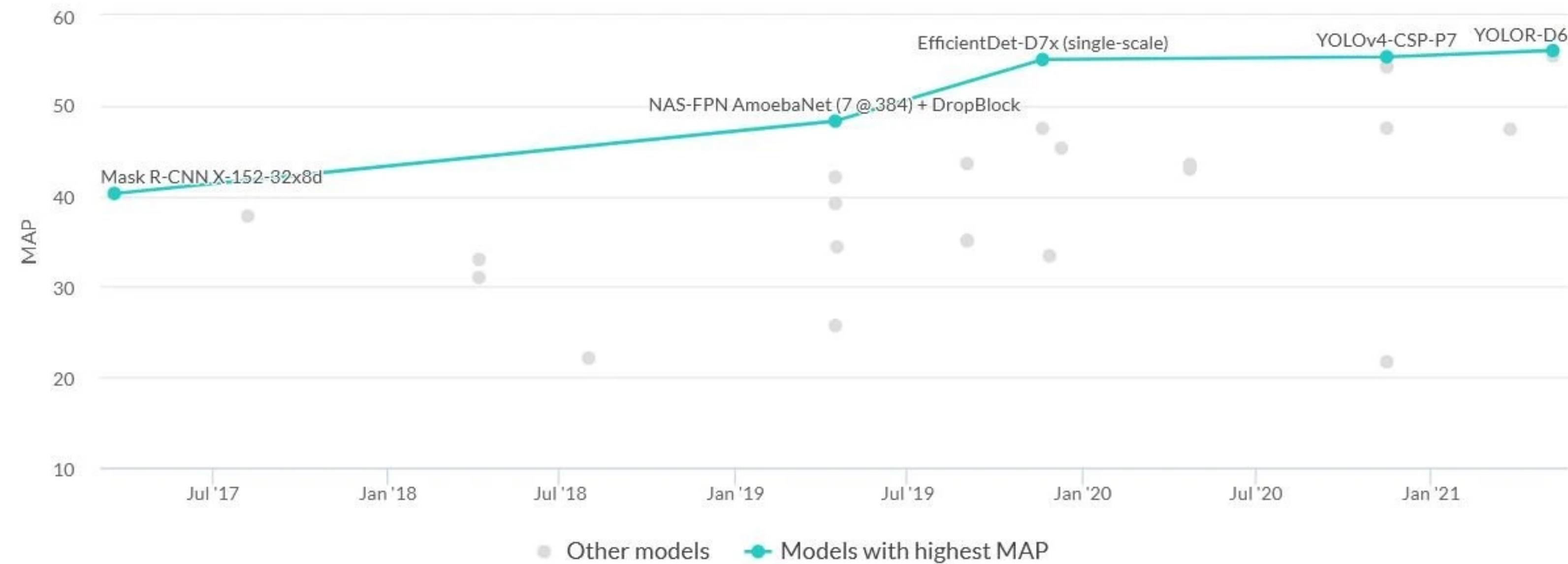
El campo de la detección de objetos no es tan nuevo como parece.

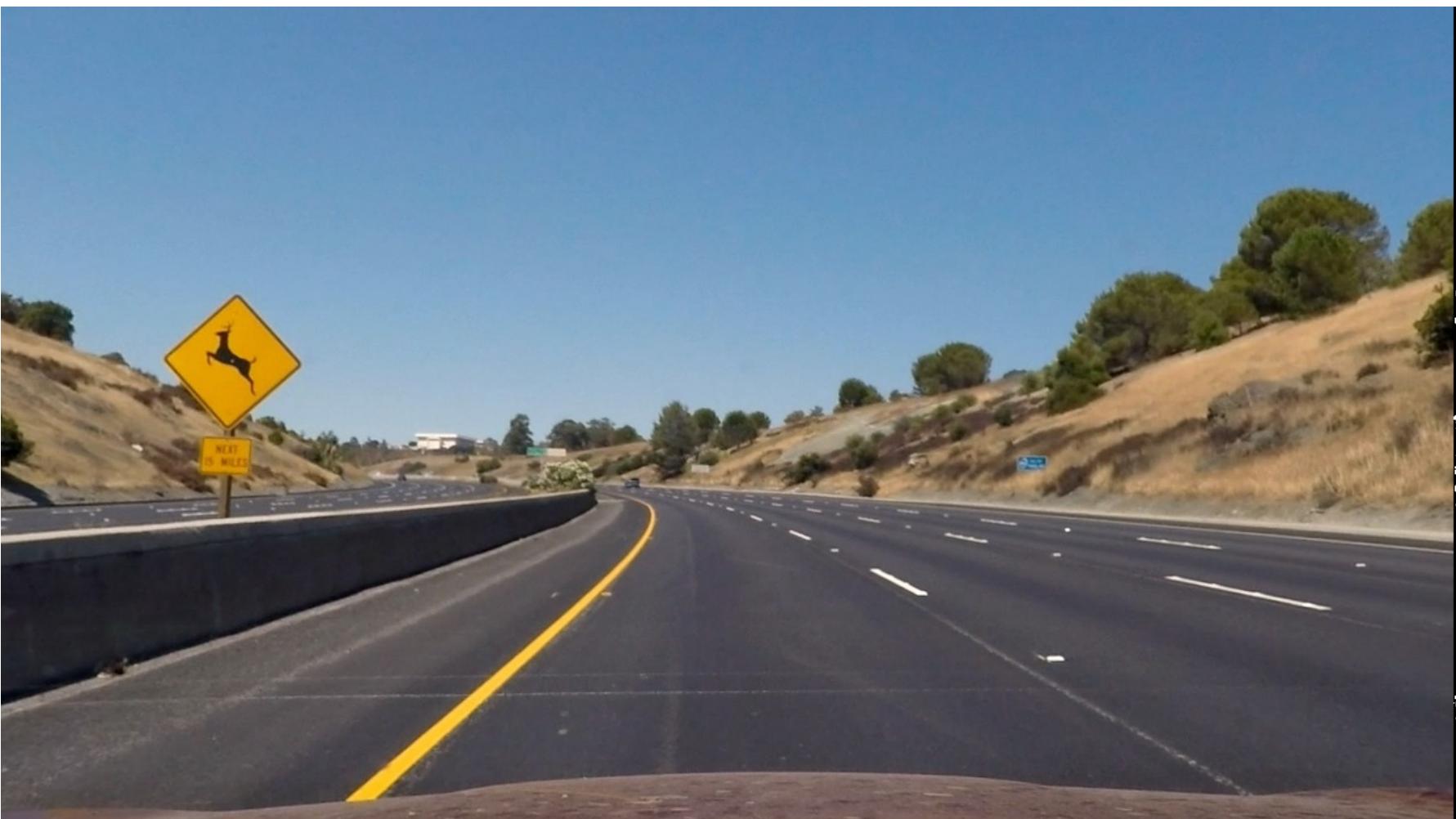
De hecho, la detección de objetos ha evolucionado en los últimos 20 años.

El progreso de la detección de objetos generalmente se divide en dos períodos históricos separados (antes y después de la introducción de Deep Learning):

**Antes de 2014** : período tradicional de detección de objetos

**Después de 2014** : período de detección de aprendizaje profundo





HLS: S Channel Binary



Edge Detection



Sobel Binary

## Área de interés y detección de objetos en esa área



# Detección de objetos y aprendizaje profundo

---

- ❖ ‘Desbloqueadores’ del aprendizaje profundo:
  - Redes de aprendizaje
  - Potencia informática de las GPUs
  - Rendimiento de los detectores
  - Rastreadores de objetos
- ❖ El **aprendizaje automático (ML)** es una rama de la inteligencia artificial (IA) que consiste en aprender patrones a partir de ejemplos o datos de muestra a medida que la máquina accede a los datos y tiene la capacidad de aprender de ellos (*Aprendizaje supervisado en imágenes anotadas/etiquetadas*). El aprendizaje profundo es una forma especializada de aprendizaje automático que implica el aprendizaje en diferentes etapas.

# ¿Cómo funciona la detección de objetos?

- ❖ La detección de objetos se puede realizar utilizando técnicas tradicionales (1) de **procesamiento de imágenes o modernas** (2) redes de aprendizaje profundo.
1. Las técnicas de procesamiento de imágenes generalmente no requieren datos históricos para el entrenamiento y son de naturaleza no supervisada.

**Pros:** por lo tanto, esas tareas no requieren **imágenes anotadas**, donde los humanos etiquetaron los datos manualmente (para entrenamiento supervisado).

**Contras:** estas técnicas están restringidas a múltiples factores, como escenarios complejos (sin fondo unicolor), occlusión (objetos parcialmente ocultos), iluminación y sombras, y efecto de desorden.

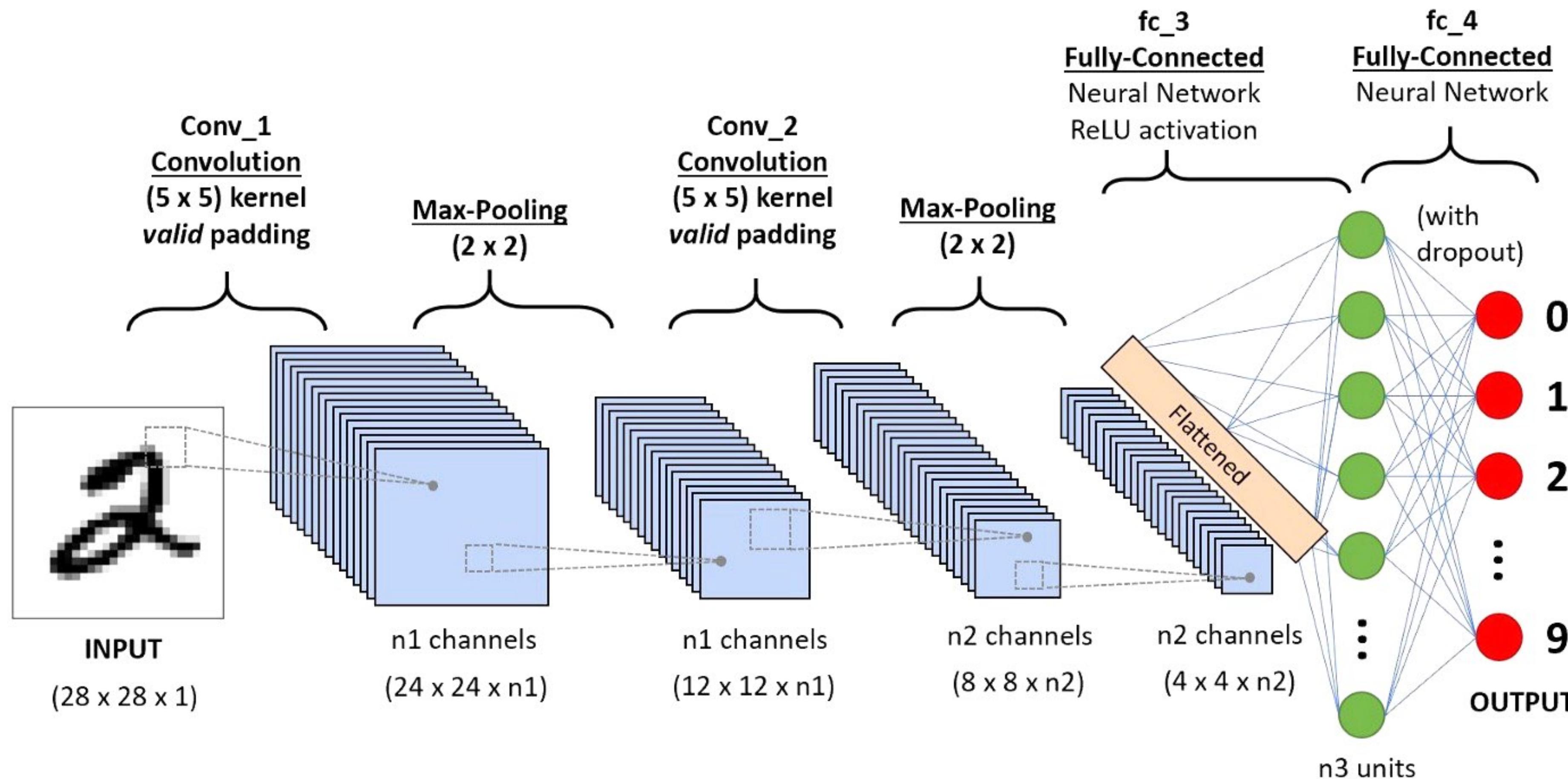
# ¿Cómo funciona la detección de objetos?

- ❖ La detección de objetos se puede realizar utilizando técnicas tradicionales (1) de procesamiento de imágenes o modernas (2) **redes de aprendizaje profundo**.
2. Los métodos de aprendizaje profundo generalmente dependen de un entrenamiento supervisado. El rendimiento está limitado por el poder de cómputo de las GPU que aumenta rápidamente año tras año.

**Pros:** la detección de objetos de aprendizaje profundo es significativamente más resistente a la oclusión, las escenas complejas y la iluminación desafiante.

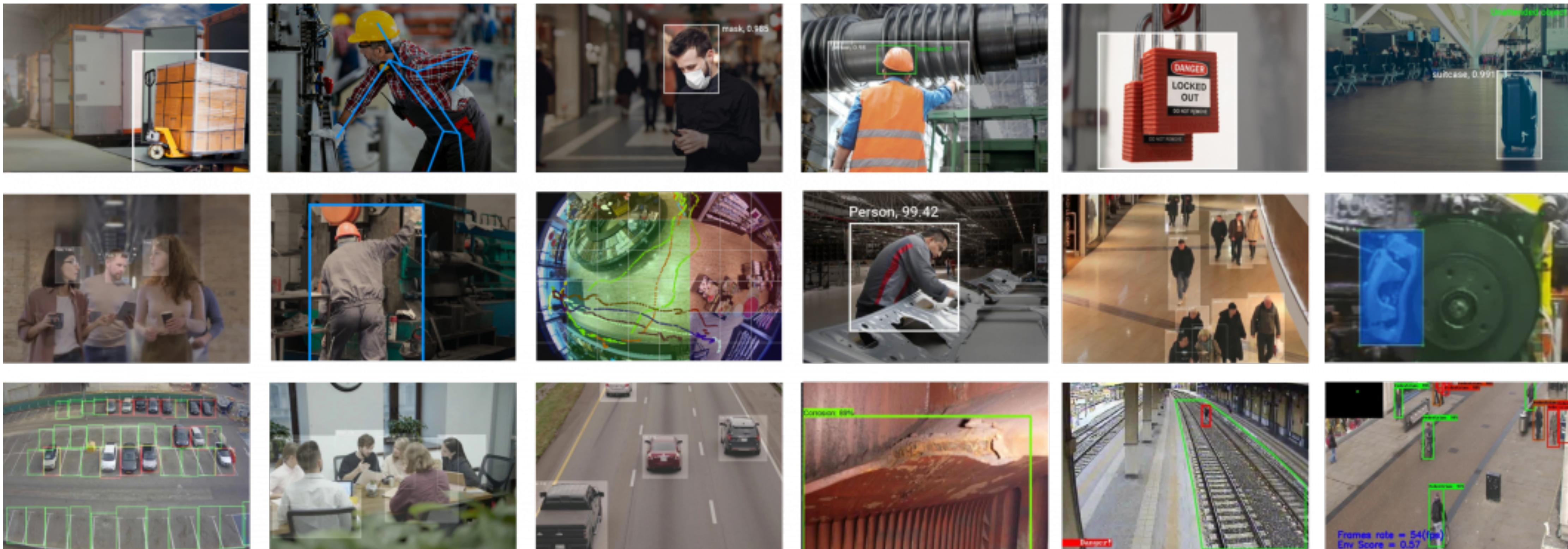
**Contras:** se requiere una gran cantidad de datos de entrenamiento; el proceso de **anotación de imágenes** es laborioso y costoso. Por ejemplo, etiquetar 500 000 imágenes para entrenar un algoritmo de detección de objetos DL personalizado se considera un conjunto de datos pequeño. Sin embargo, muchos conjuntos de datos de referencia ([MS COCO](#), Caltech, KITTI, PASCAL VOC, V5) brindan la disponibilidad de datos etiquetados.

# Modelos de visión computacional



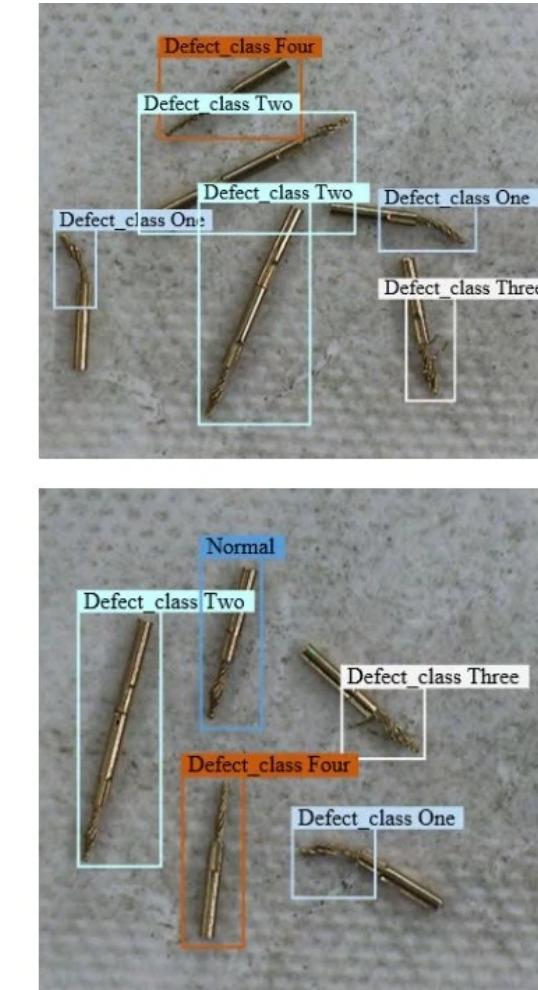
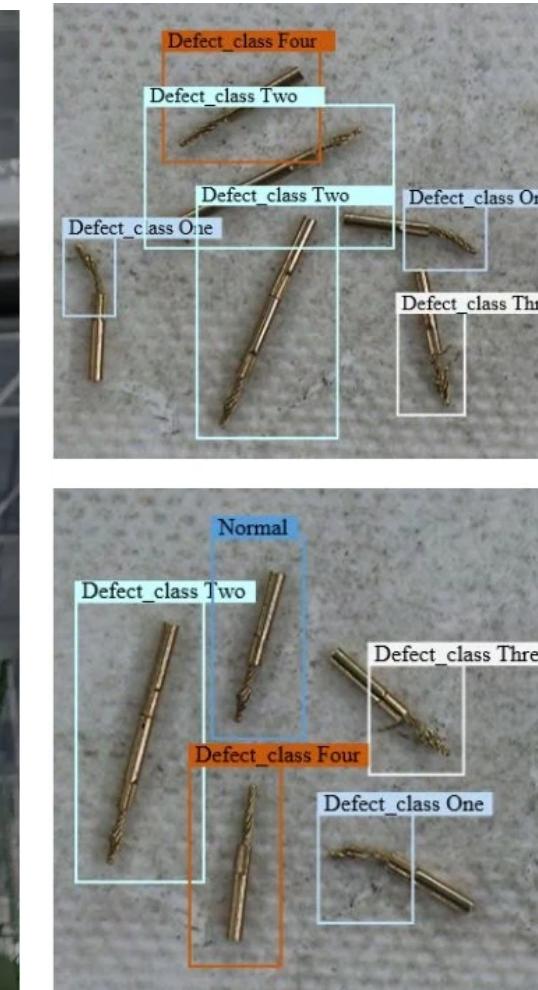
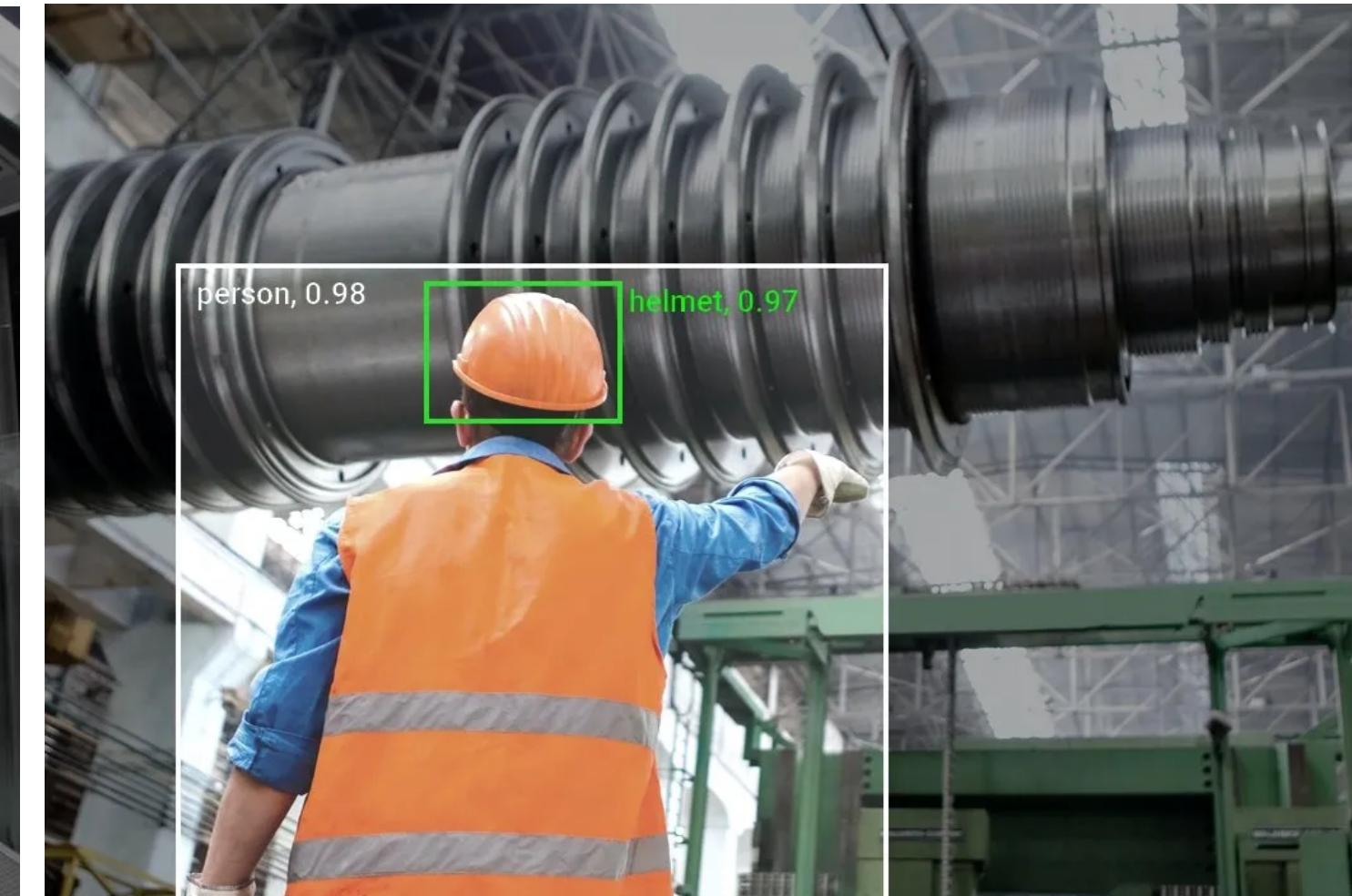
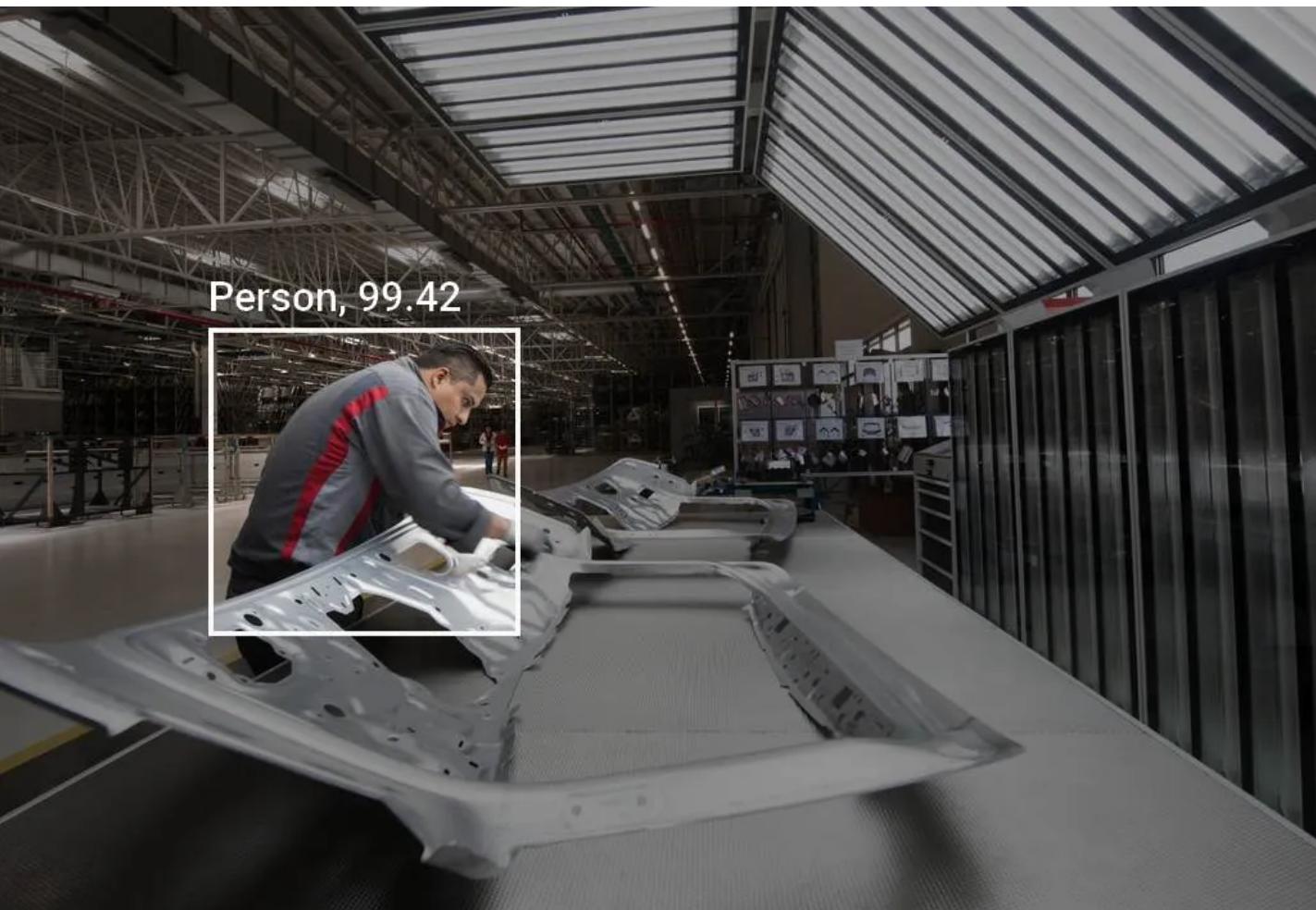
# Aplicaciones de detección de objetos

- ❖ 87 aplicaciones de visión artificial / detección de objetos



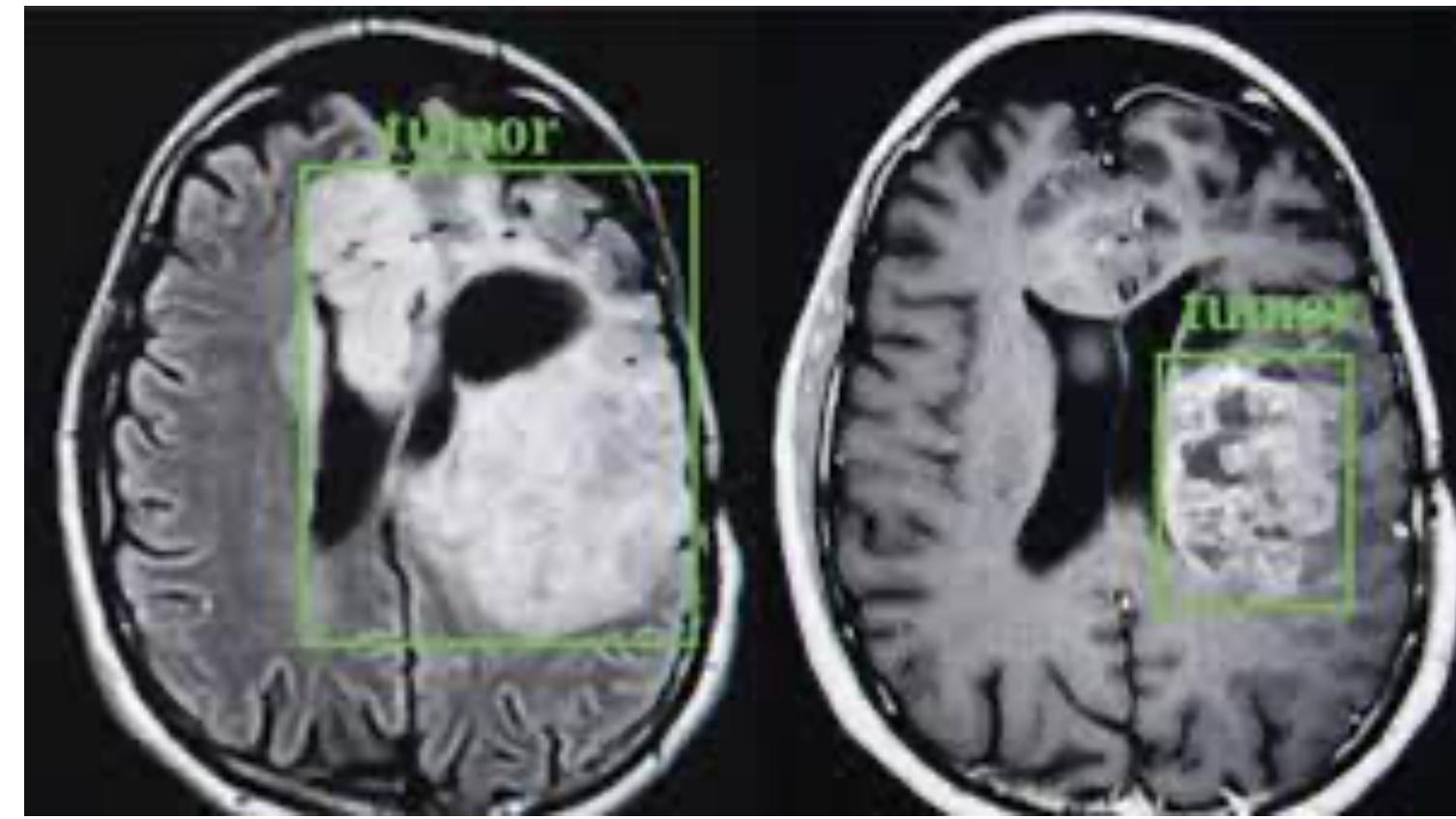
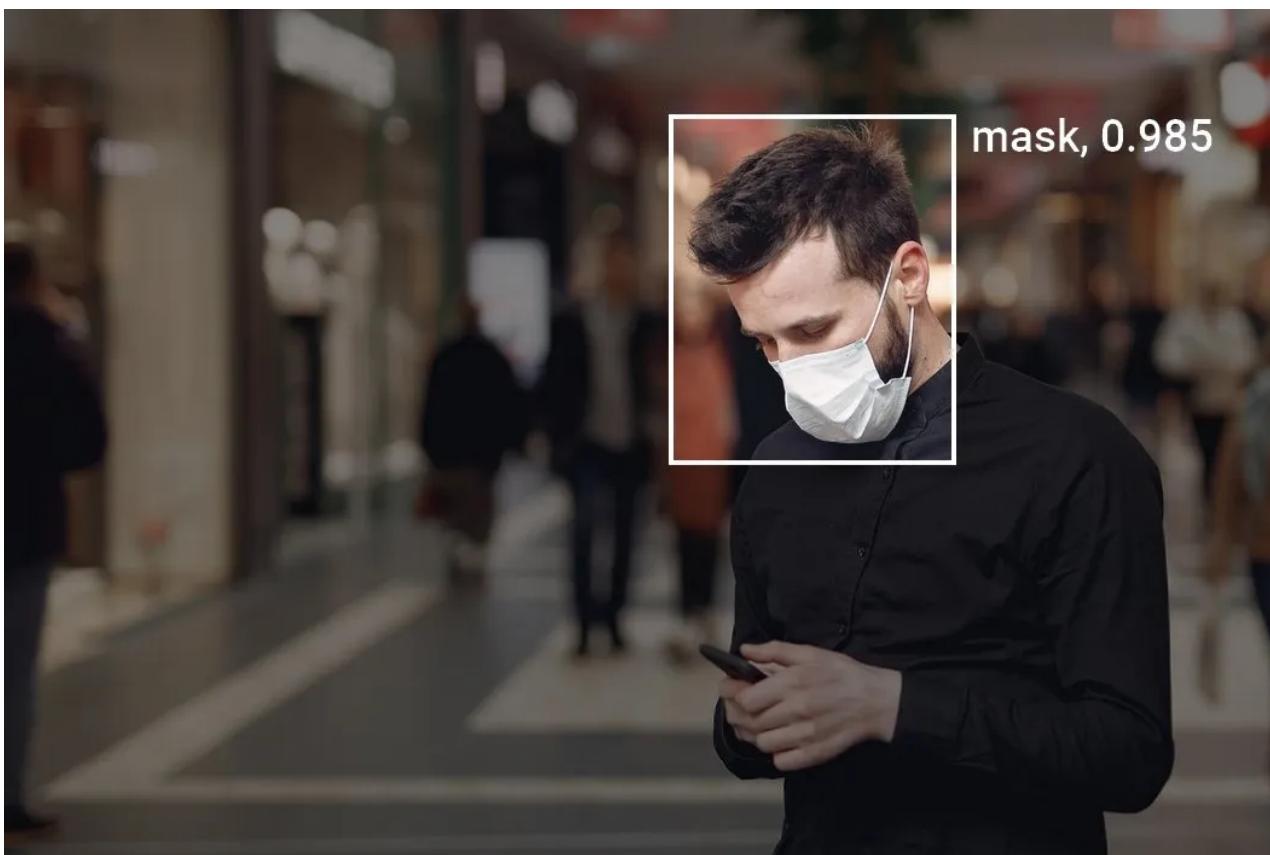
# Análisis de productividad/calidad/seguridad

- ❖ Rastrear una línea de producción, encontrar huecos de botella y analizar las posibles soluciones para un flujo constante de producto.



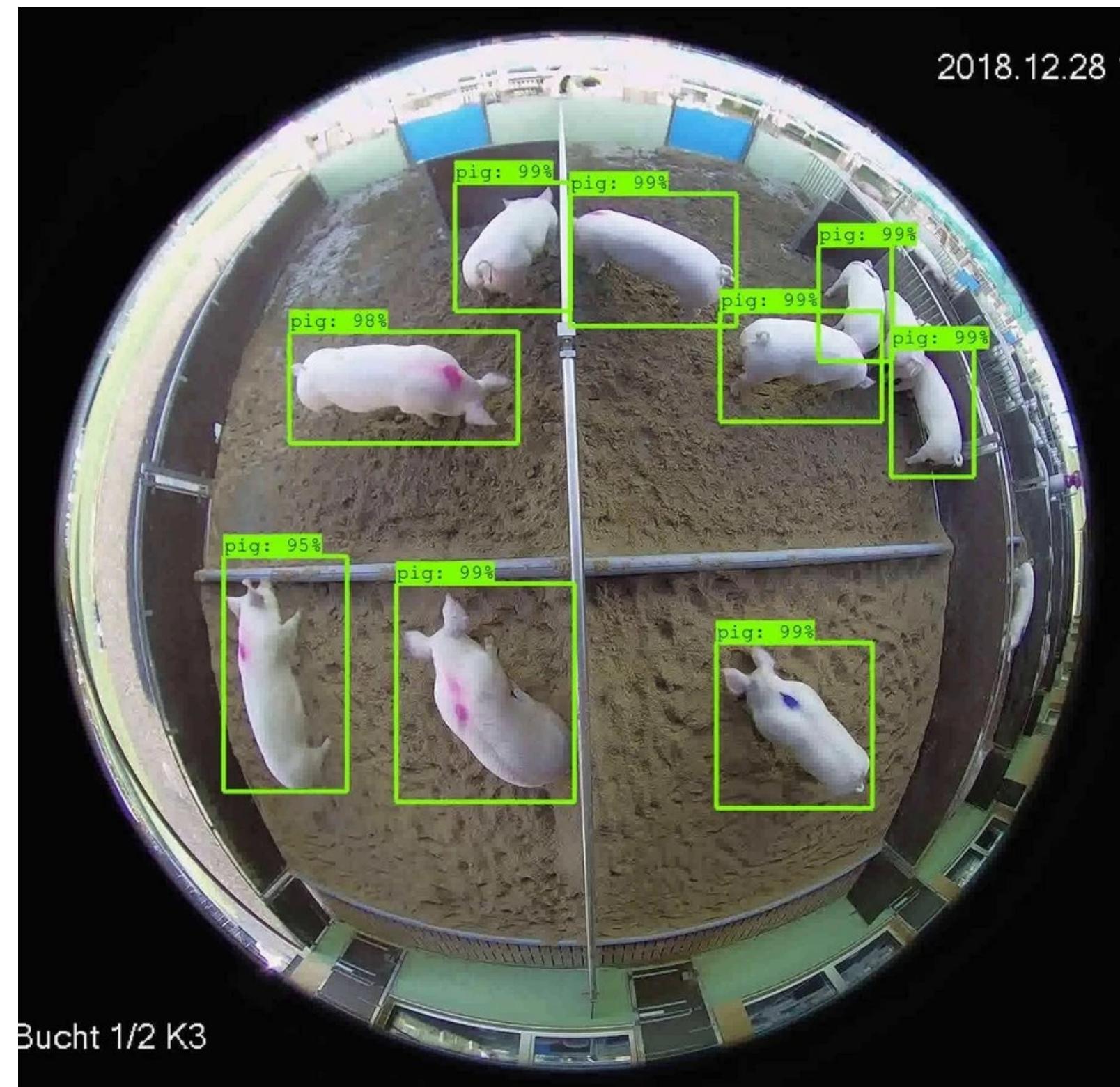
# Cuidado de la salud

- ❖ Detección de tumores
- ❖ Imágenes médicas
- ❖ Detección de cáncer
- ❖ Entrenamiento médico
- ❖ Combate a COVID-19



- ❖ Monitoreo de salud
- ❖ Diagnóstico asistido por computadora
- ❖ Detección temprana de enfermedades
- ❖ Rehabilitación y monitoreo de pacientes en casa

# En la agricultura



- ❖ Monitoreo de animales
- ❖ Automatización de granjas
- ❖ Monitoreo de cultivos
- ❖ Detección de floración
- ❖ Monitoreo de plantaciones
- ❖ Detección de insecto
- ❖ Enfermedades en las plantas
- ❖ Deshierbe automático
- ❖ Cosecha automática
- ❖ Pruebas de calidad agrícolas

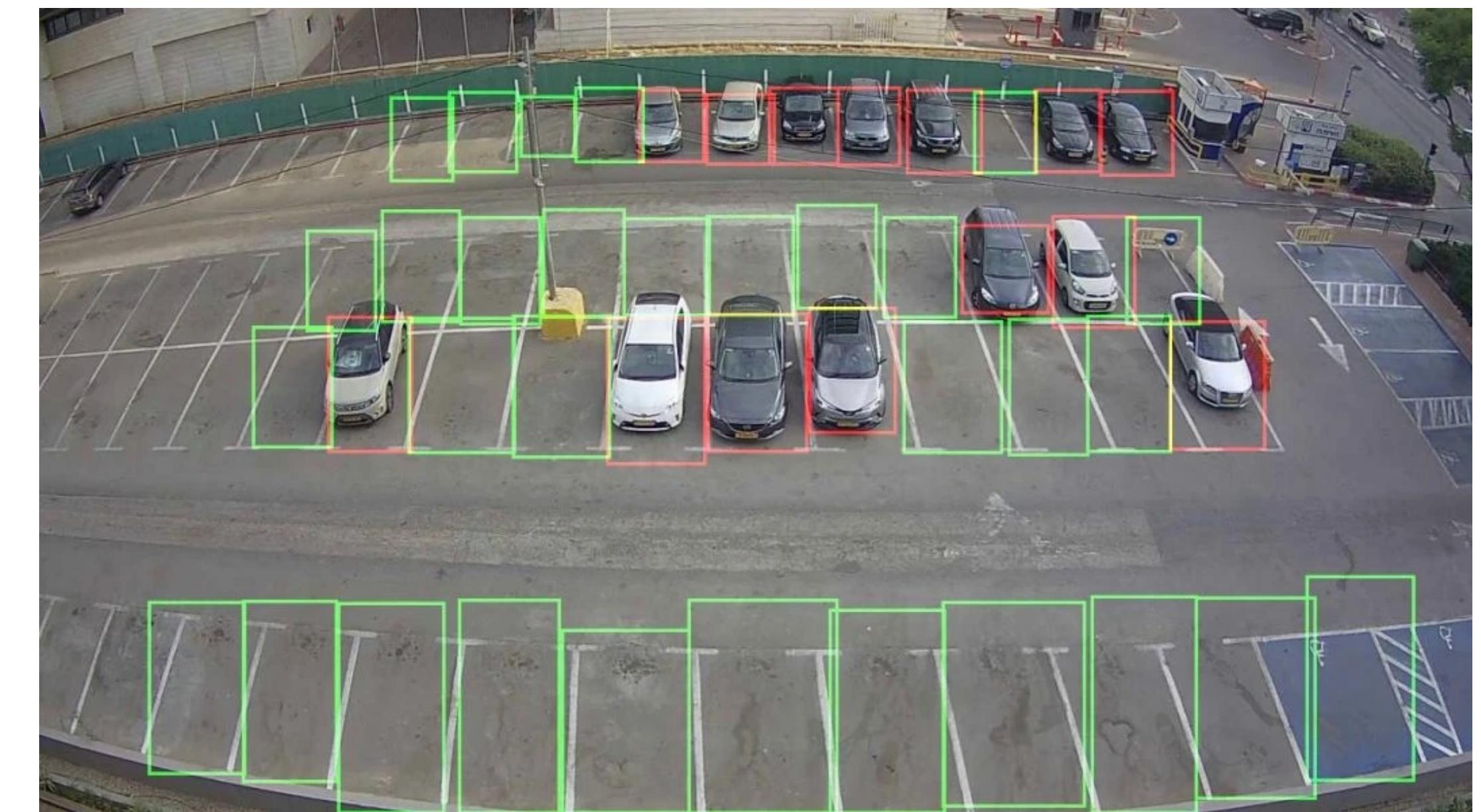
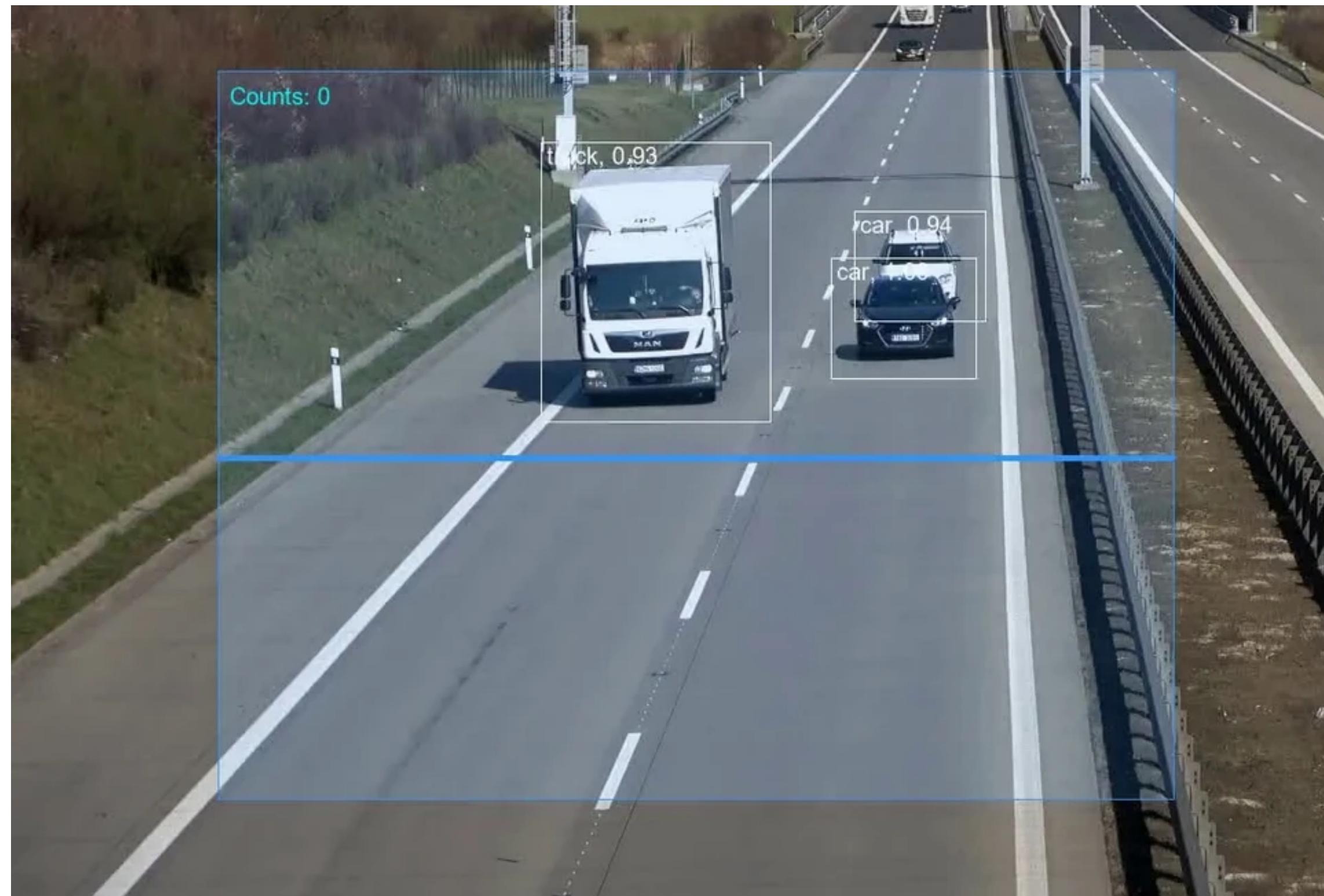


# En petróleo y gas

- ❖ Reconocimiento automático de instrumentos analógicos
- ❖ Detección de fugas de gases y corrosión
- ❖ Evaluación geológica y exploración asistida



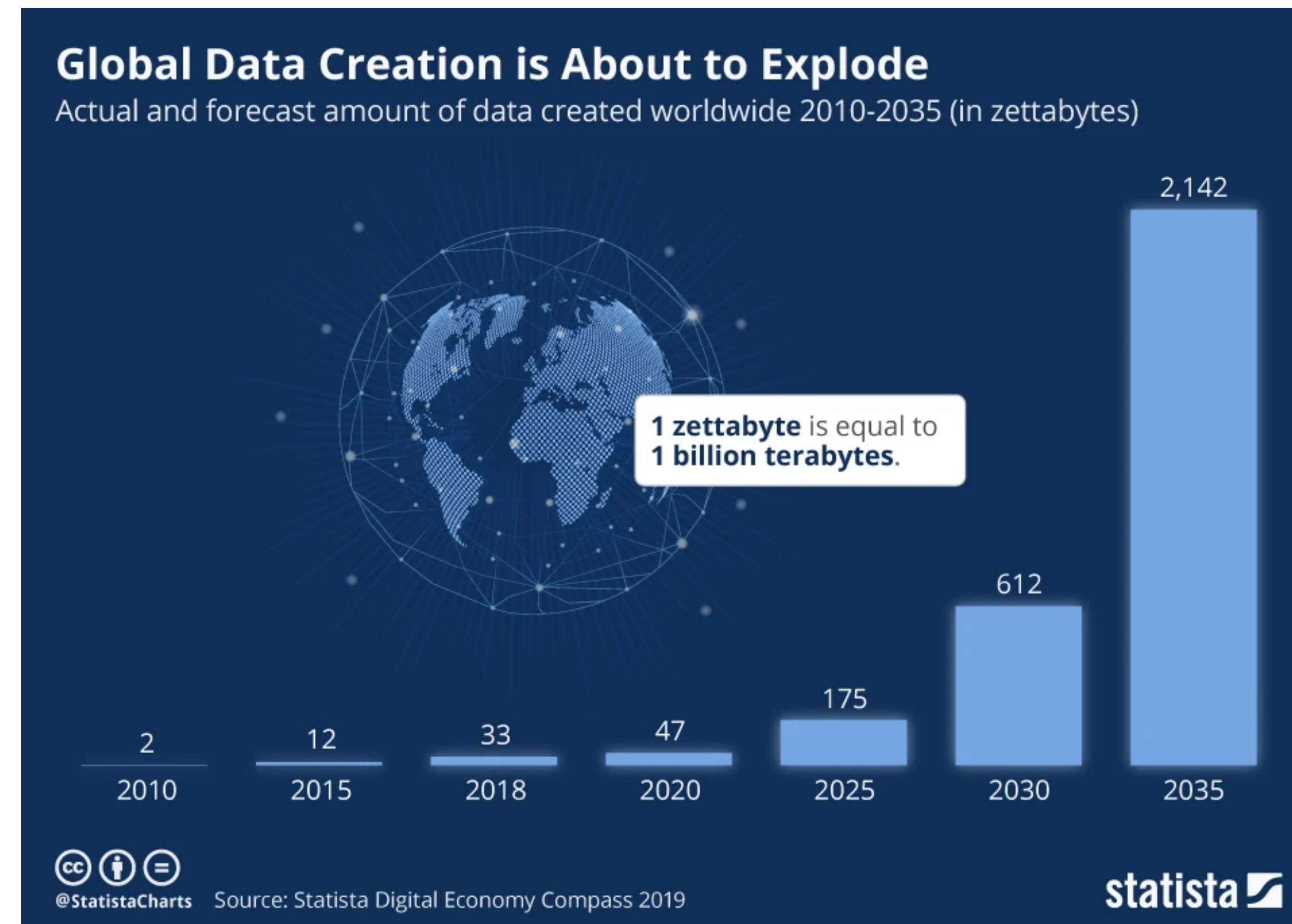
# En el manejo autónomo



# En el deporte



# Los datos globales van a explotar!

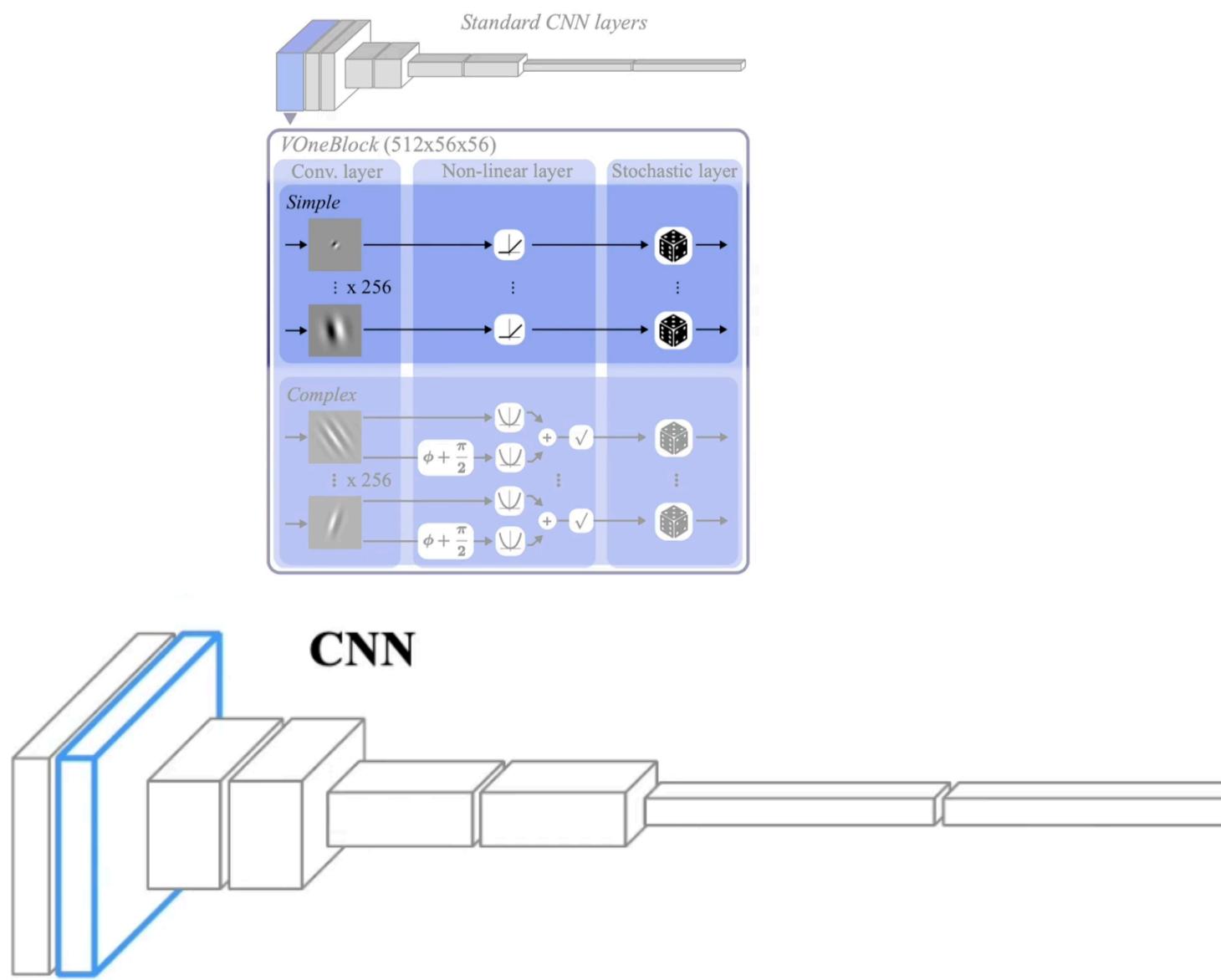


<https://www.craiyon.com>

# En neurociencia - back and forth

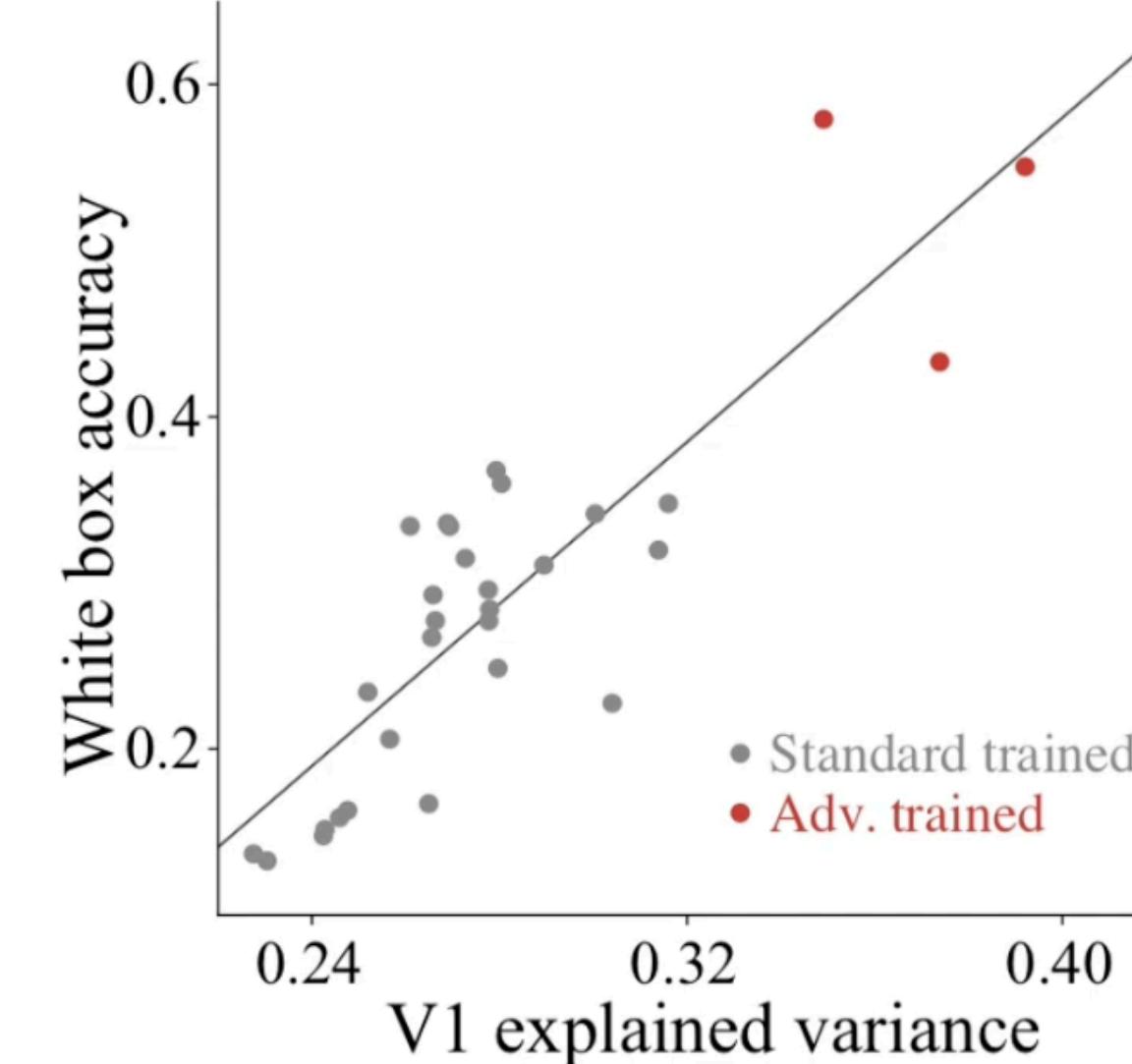
*Are models more like the primate primary visual cortex (V1) also more robust?*

VOneNet, a hybrid CNN with a V1 front-end



*White box adversarial robustness*

*vs V1 explained variance*

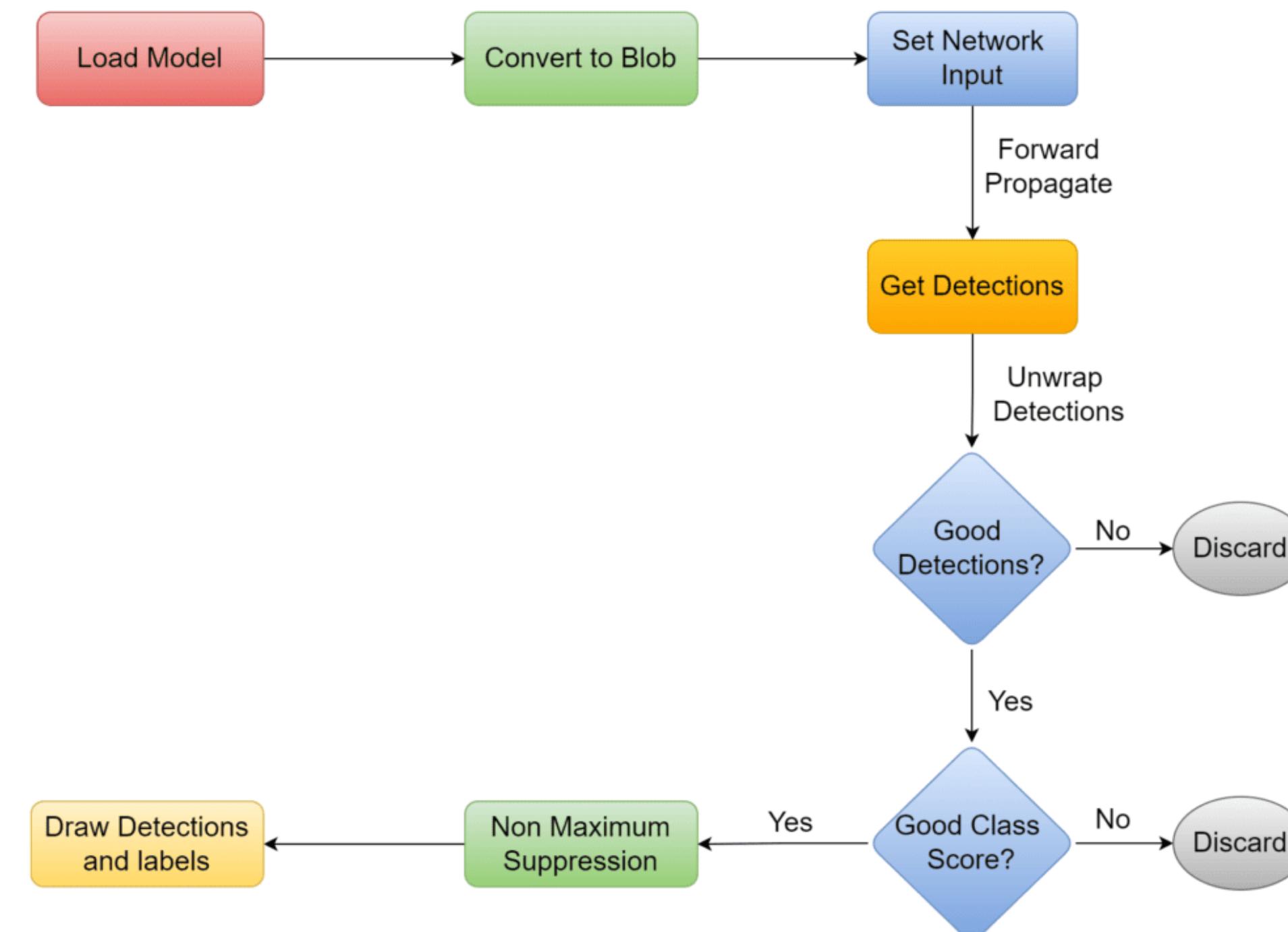


Nueva CNN basada en V1 cortex area

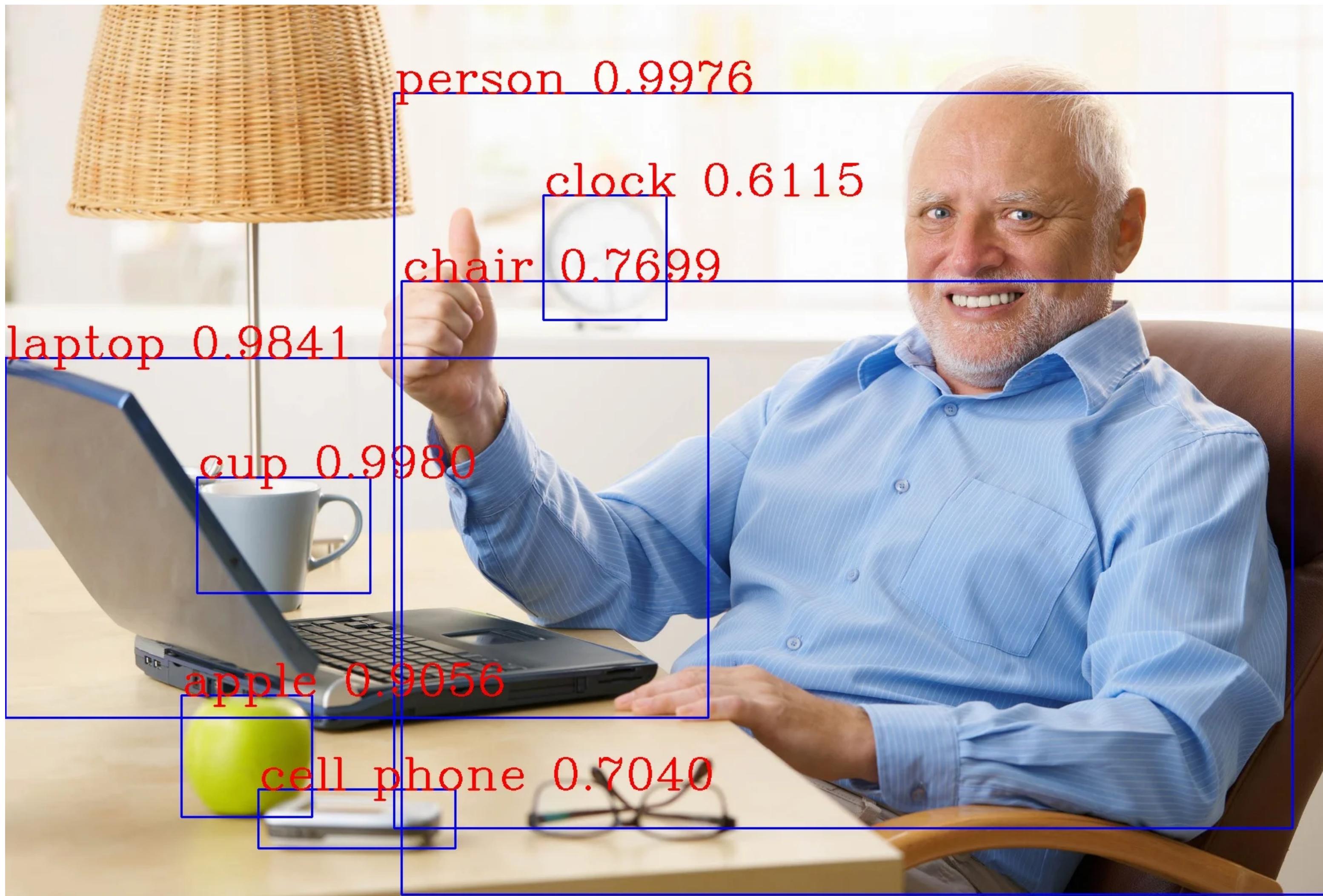
# Ejemplo: YOLOv5

Yolo: You only look once

<https://learnopencv.com/deep-learning-based-text-detection-using-opencv-c-python/>



# ¡Gracias por su atención!



# Contacto

---

Email

[natalia.juradoes@gmail.com](mailto:natalia.juradoes@gmail.com)

Github

<https://github.com/AchachayJurado>

LinkedIn

<https://www.linkedin.com/in/nataliajuradoes>