

Tema 2. Aplicaciones de la IA en tareas de visión por computador

Ana Jiménez Pastor
anjipas@gmail.com

Visión por computador

La visión por computador es una campo de la inteligencia artificial que permite a los ordenadores obtener información de utilidad de las **imágenes, videos y otras fuentes de entrada visuales**, y, tomar decisiones en función de esta información.

De igual manera que la IA permite a los ordenadores pensar, la visión por computador permite a los ordenadores **ver, observar y entender**.

Gracias a la digitalización de las imágenes y le crecimiento de los **datos** disponibles, las mejoras del **hardware** y los **avances en la IA**, en los últimos años, la visión por computador ha sido capaz de ganar al humano en determinadas tareas de detección y etiquetado de objetos.



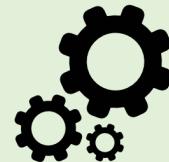
Visión por computador

Antes de la aparición del Deep Learning, las tareas de visión por computador que se podían realizar eran muy limitadas y requerían de mucha programación y esfuerzo por parte de los desarrolladores.



Manual

- Generación de una base de datos anotada de manera manual con rasgos principales (e.g. distancia entre ojo, tamaño boca, etc.)
- Con imagen nueva capturada, realizar las mismas medidas y comparar con base de datos anotada para detectar similitudes.
- Grandes márgenes de error.



Machine Learning tradicional

- Desarrollo de algoritmos para extracción de múltiples características.
- Entrenamiento de un modelo de aprendizaje máquina.
- Permite resolver nuevos problemas (e.g., predicción supervivencia en cáncer de mama).
- Se requiere conocimiento experto.



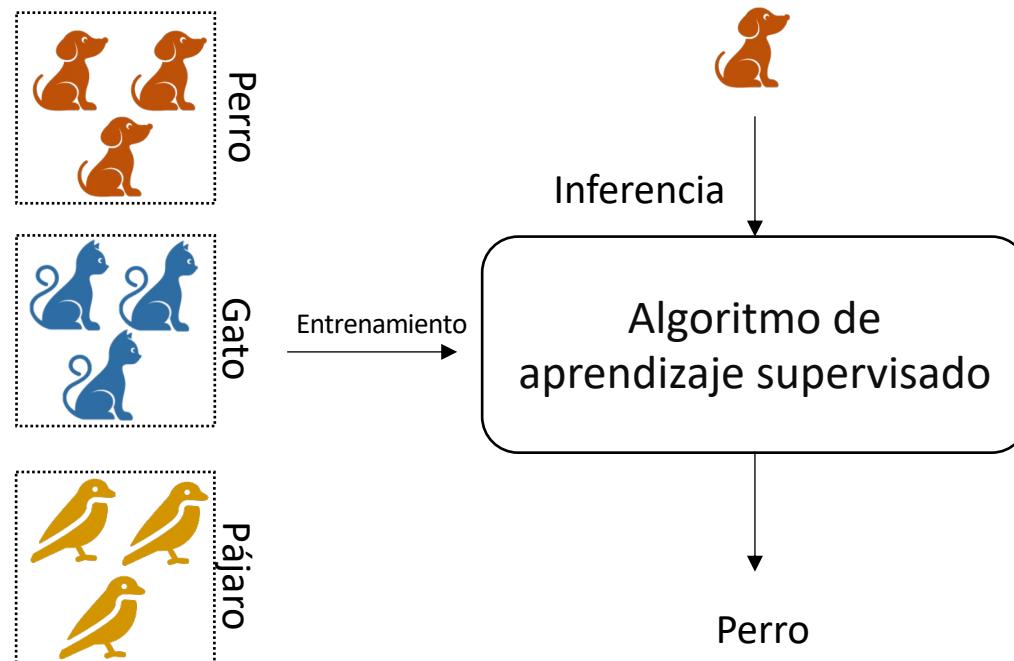
Deep Learning

- La propia red neuronal extrae las características apropiadas para resolver el problema de interés.
- Con la base de datos apropiada, ofrece grandes resultados en problemas que antes no se podían resolver.
- Requiere menos conocimiento experto en el campo.

Tipos de aprendizaje

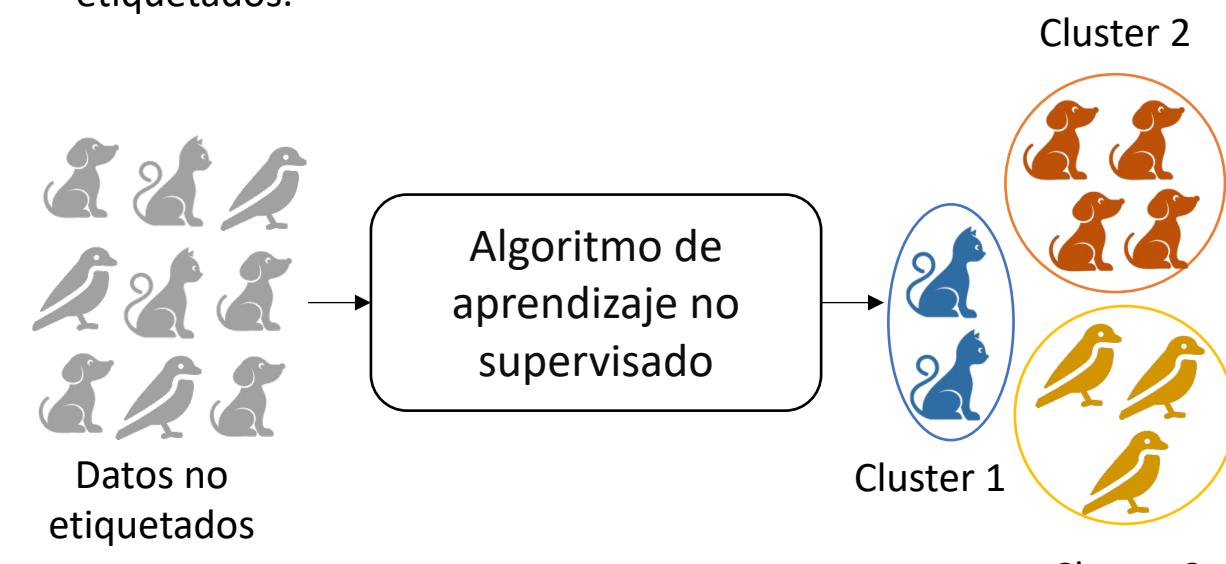
Aprendizaje supervisado

Modelos predictivos haciendo uso de datos etiquetados, es decir, para un registro de entrada tenemos su etiqueta correspondiente.

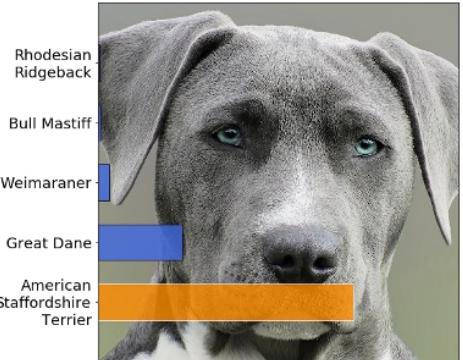


Aprendizaje no supervisado

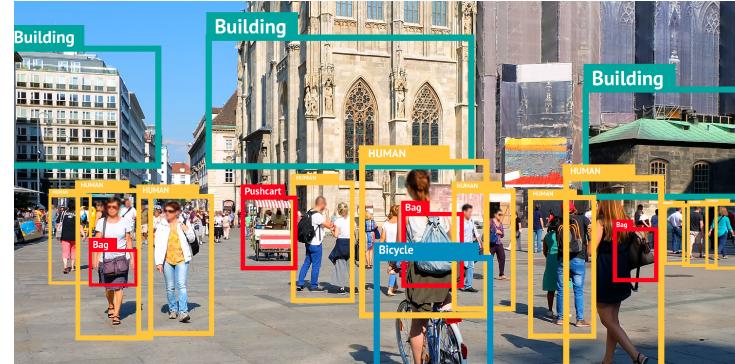
Modelos encargados de detectar patrones en los datos de entrada sin que estos se encuentran etiquetados.



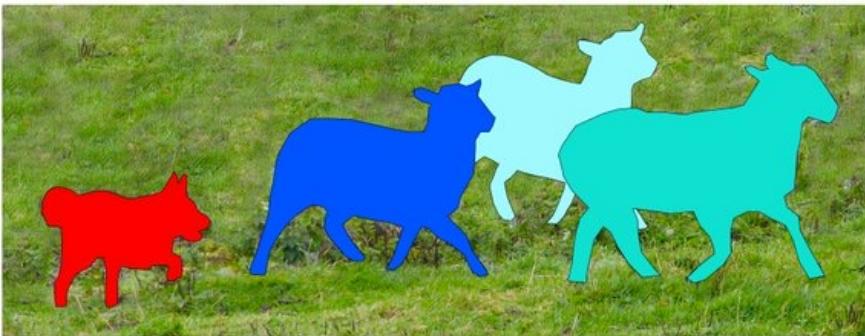
Aprendizaje supervisado



Clasificación / Regresión



Detección de objetos



Segmentación de instancias



Segmentación semántica

Aprendizaje no supervisado

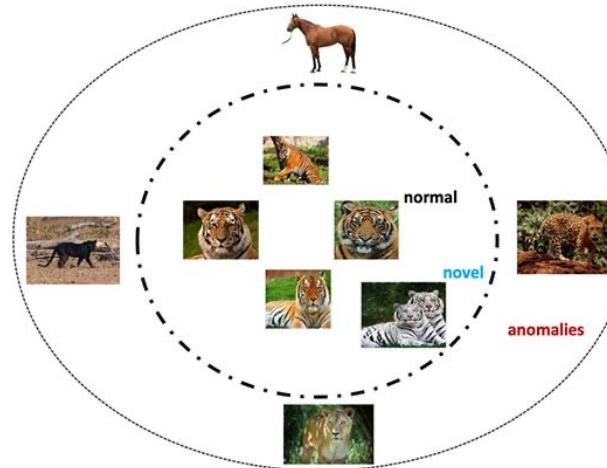
A new offline handwritten database for language, which contains full Spanish sentence, has been developed: the Spartacus database. Spanish Restricted-domain Task of Cursive were two main reasons for creating this collection. Most databases do not contain Spanish sentence. Spanish is a widespread major language. A reason was to create a corpus from semantic. These tasks are commonly used in practice of linguistic knowledge beyond the lexicon recognition process.

A new offline handwritten database for language, which contains full Spanish sentence, has been developed: the Spartacus database. Spanish Restricted-domain Task of Cursive were two main reasons for creating this collection. Most databases do not contain Spanish sentence. Spanish is a widespread major language. A reason was to create a corpus from semantic. These tasks are commonly used in practice of linguistic knowledge beyond the lexicon recognition process.

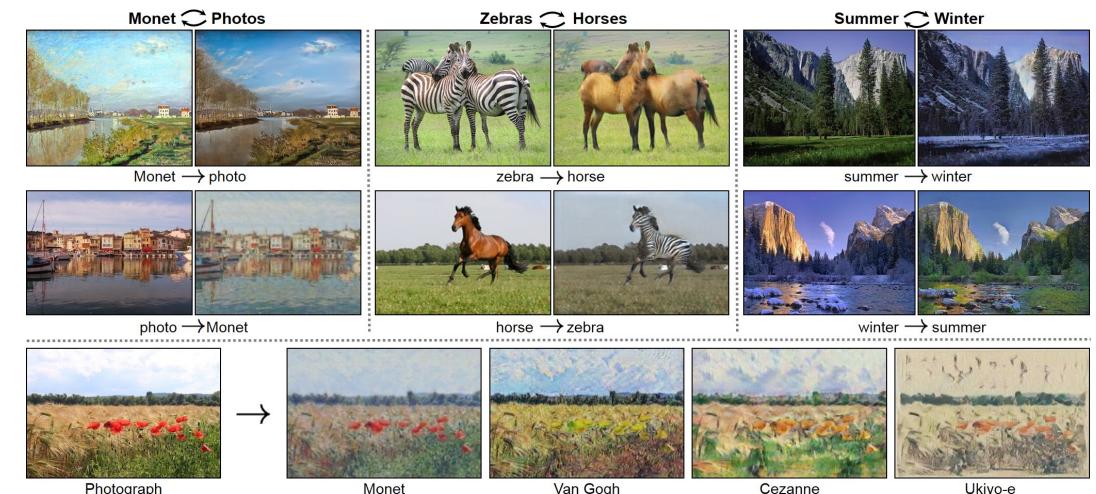
Reducción de ruido



Generación de imágenes sintéticas



Detección de anomalías

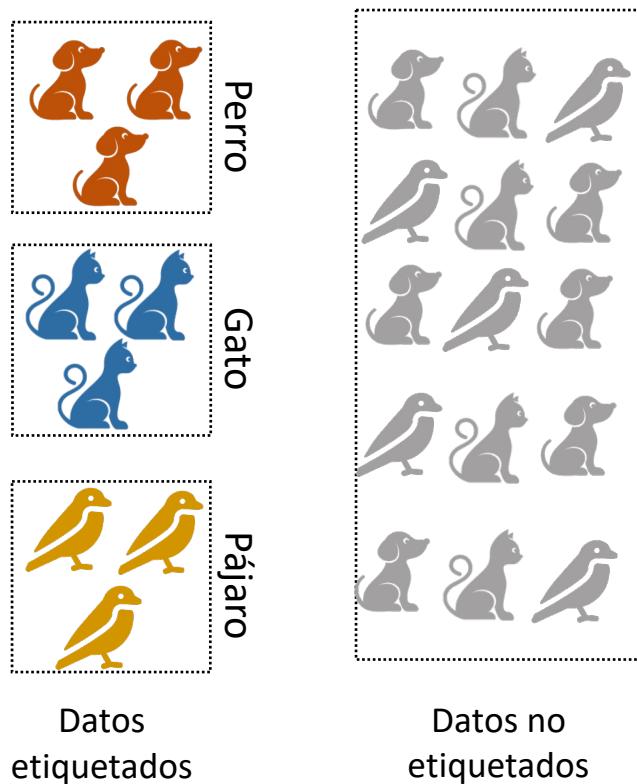


Transformación de imágenes

Tipos de aprendizaje

Aprendizaje semi-supervisado

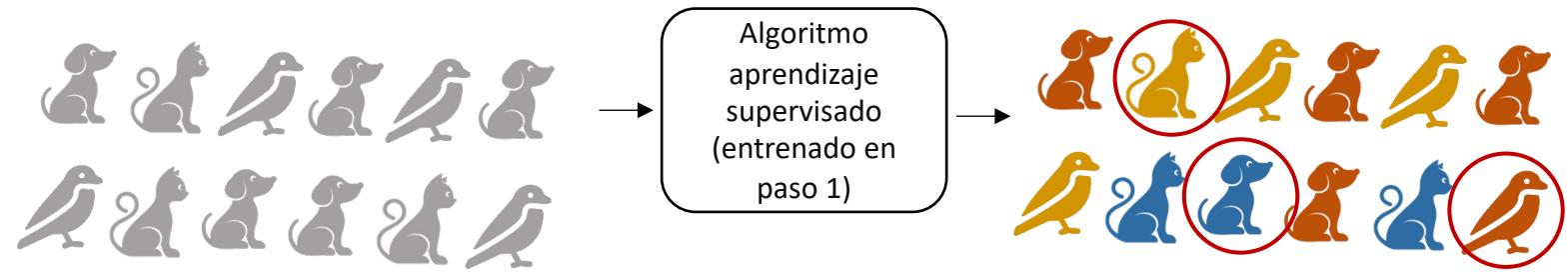
El modelo se entrena con datos tanto etiquetado como no etiquetados.



1) Entrenamos un modelo (basado en aprendizaje supervisado) empleando los datos etiquetados.



2) Aplicamos el modelo entrenado para clasificar los datos no etiquetados, generando pseudo-etiquetas.



3) Entrenamos el modelo final con **pseudo-etiquetas**



Tipos de aprendizaje

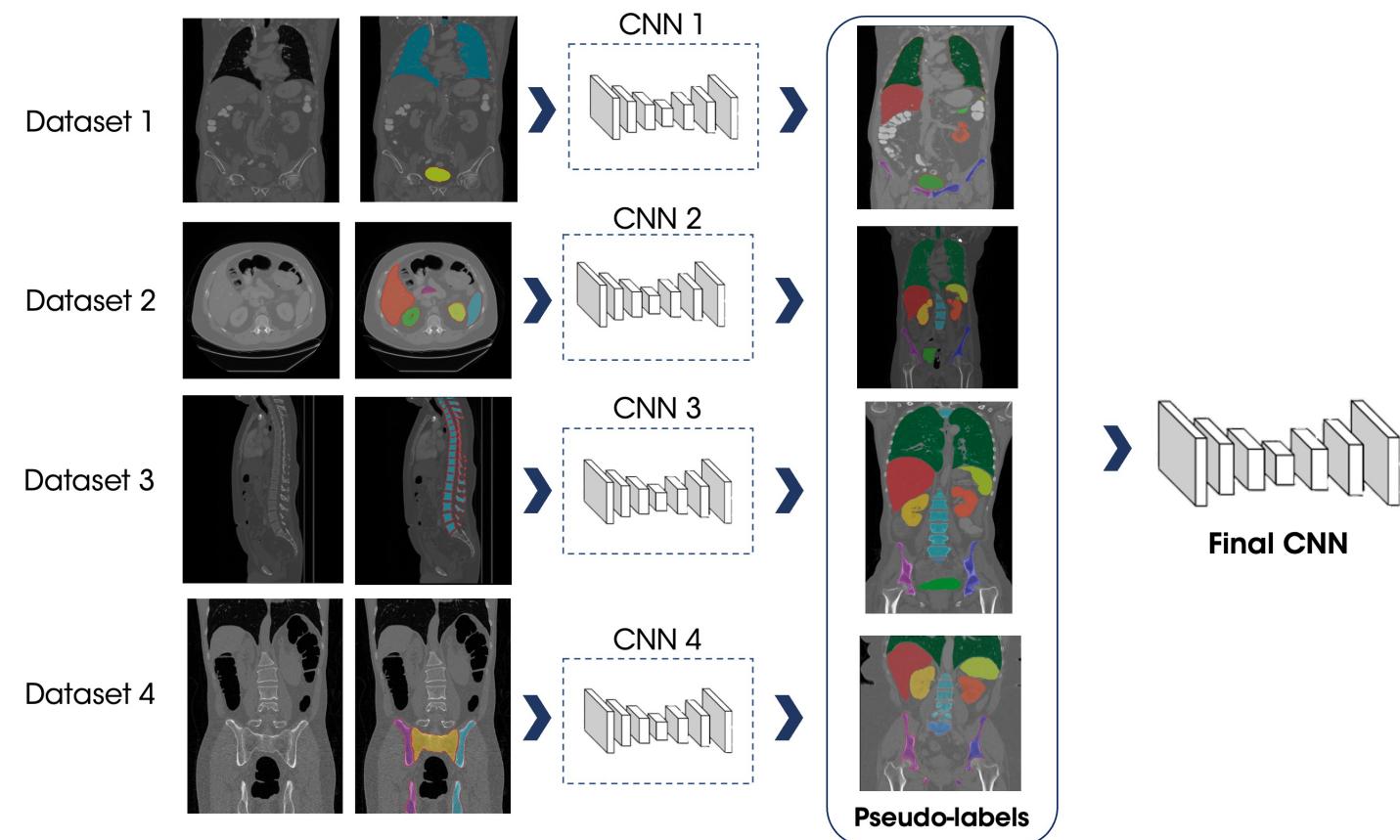
Aprendizaje semi-supervisado

Datos de 4 fuentes diferentes, cada uno de ellos con diferentes regiones segmentadas manualmente (etiqueta):

1. CT-Org: 140 casos (pulmón y vejiga)
2. AbdomenCT1k: 1.000 casos (hígado, riñones izq. Y dcho., bazo, páncreas)
3. VerSe: 136 casos (columna vertebral)
4. CTPelvic1k: 938 casos (sacro y cadera izq. y dcha.)

Metodología:

1. 4 modelos diferentes, entrenados con los datos y etiquetas de cada fuente.
2. Cada modelo se empleó para segmentar las estructuras restantes (pseudo-etiquetas) en el resto de fuentes. (p.ej., el modelo obtenido con el dataset 1 se empleó para segmentar los pulmones y la vejiga en los datasets 2, 3 y 4).
3. El modelo final se entrenó haciendo uso de las etiquetas originales y las pseudo-etiquetas.



Tipos de aprendizaje

Auto-aprendizaje

Es un tipo de aprendizaje **no supervisado** donde el modelo se crea su propia “supervisión” de los propios datos. El modelo aprende mediante la creación de tareas auxiliares que no requieren de etiquetas externas. Una vez el modelo se ha entrenado en estas tareas auxiliares, se puede transferir el conocimiento aprendido a tareas secundarias, como clasificación o regresión (*transfer learning*).

Generación de texto



ChatGPT

ChatGPT se basa en GPT, el cual se entrenó en una cantidad masiva de texto disponible en internet. Durante el entrenamiento, el modelo aprendió a predecir la siguiente palabra de una frase.

Tipos de aprendizaje

Auto-aprendizaje

Es un tipo de aprendizaje **no supervisado** donde el modelo se crea su propia “supervisión” de los propios datos. El modelo aprende mediante la creación de tareas auxiliares que no requieren de etiquetas externas. Una vez el modelo se ha entrenado en estas tareas auxiliares, se puede transferir el conocimiento aprendido a tareas secundarias, como clasificación o regresión (*transfer learning*).

Impainting



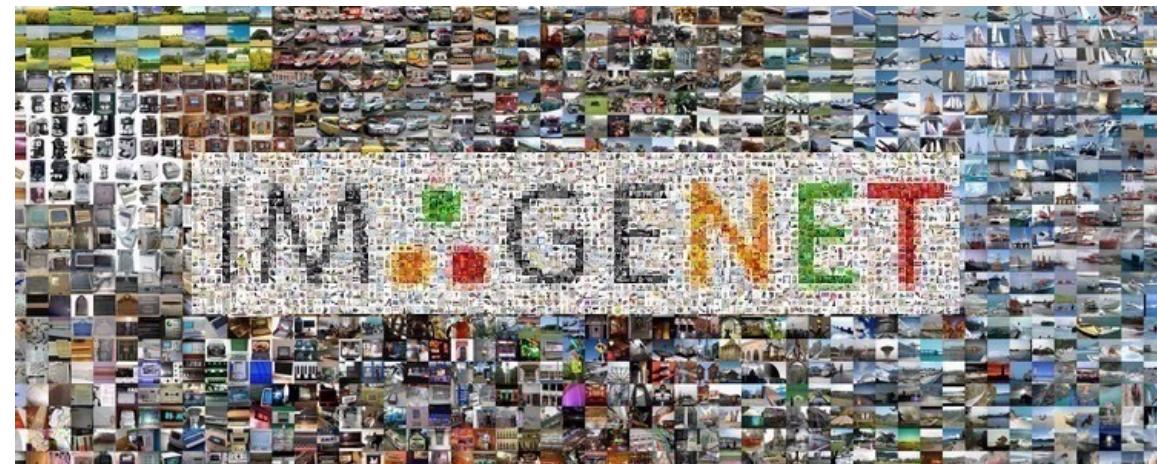
Clasificación

Asignar a una imagen una etiqueta dada.

Estas son dos de las bases de datos más conocidas para la resolución de problema de clasificación:



MNIST → Clasificación de dígitos escritos a mano formada por 70.000 imágenes etiquetadas



ImageNet → Clasificación de imágenes de la vida cotidiana formada por 14M de imágenes pertenecientes a 22k clases diferentes.

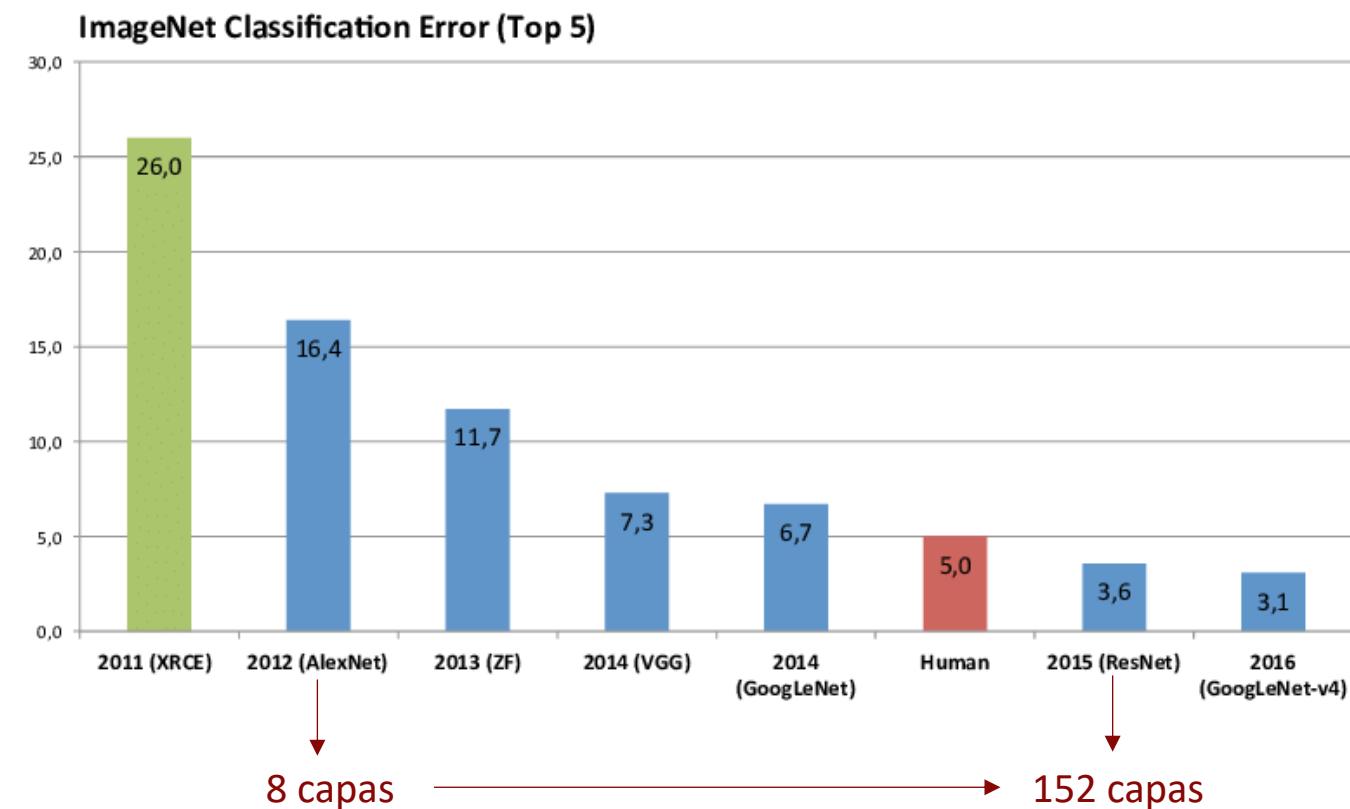
Clasificación



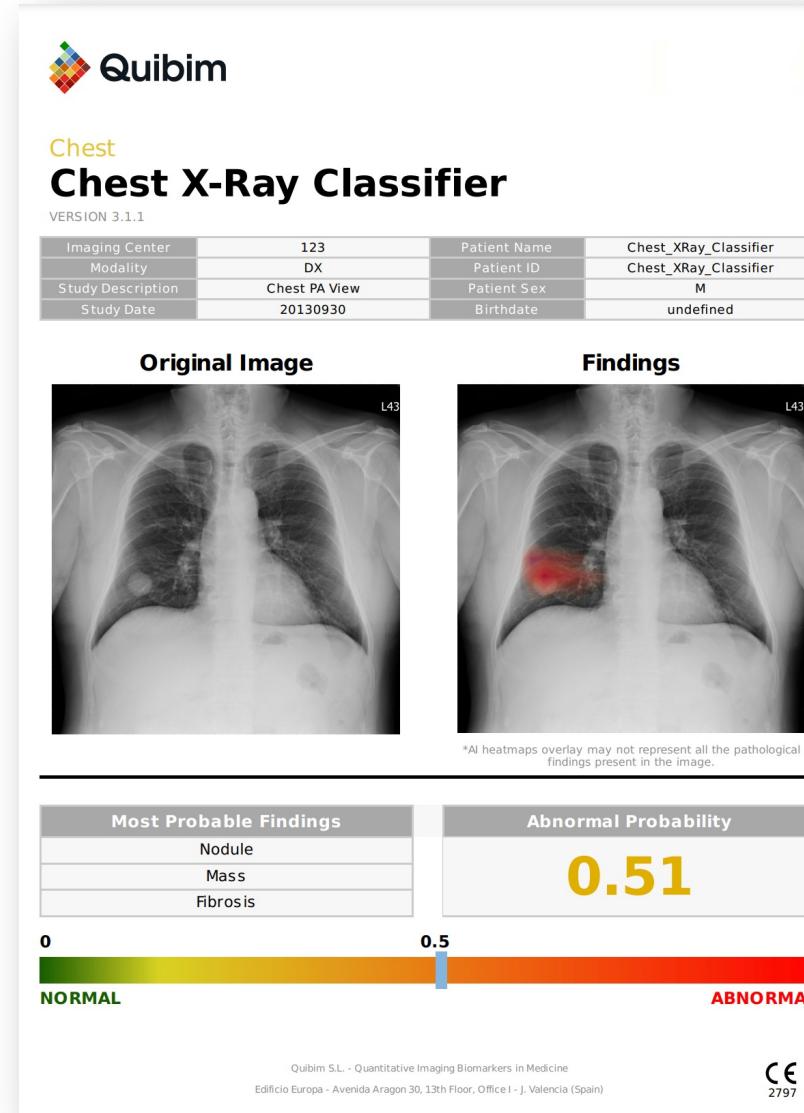
Large Scale Visual Recognition Challenge



- Reto anual (desde 2009 a 2017) de clasificación de imágenes y detección de objetos.
- 1.28M de imágenes de entrenamiento.
- 50k imágenes de validación.
- 1k clases.
- Una imagen se da como correctamente clasificada si el algoritmo predice la etiqueta correcta entre las 5 de mayor probabilidad.



Clasificación

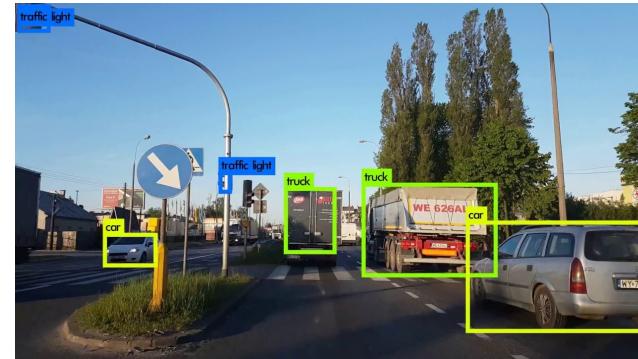


En medicina los modelos de IA han tomado gran importancia en la ayuda a la toma de decisiones.

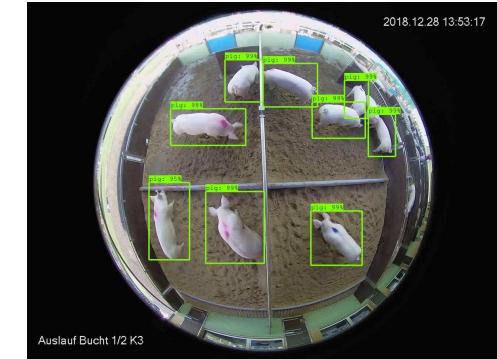
Detección de objetos



Venta al público



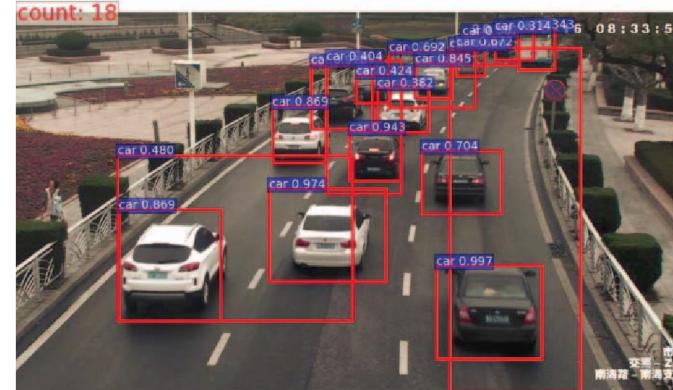
Coches autónomos



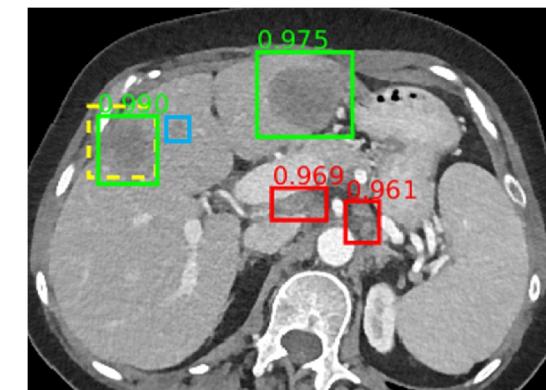
Monitorización de animales



Detección de personas en peligro



Detección de coches para control de tráfico



Detección de lesiones en imágenes médicas

Detección de objetos

La base de datos pública más conocida es COCO para la resolución de problemas de detección de objetos, segmentación y estimación de la postura.



<https://cocodataset.org/#home>

Segmentación

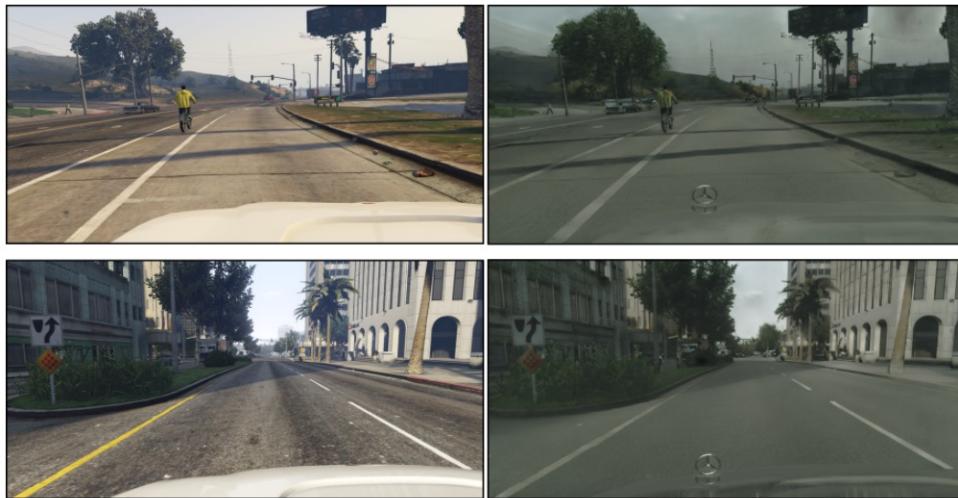


Coches autónomos

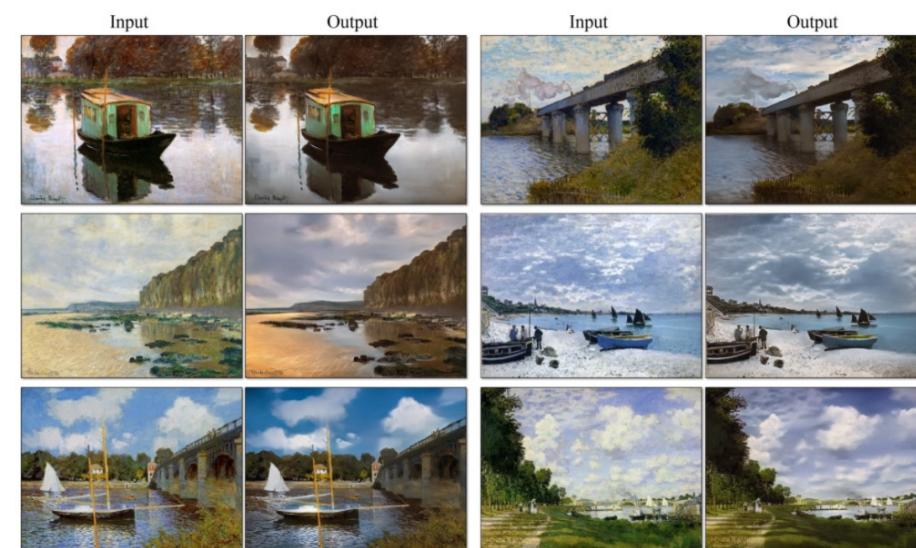
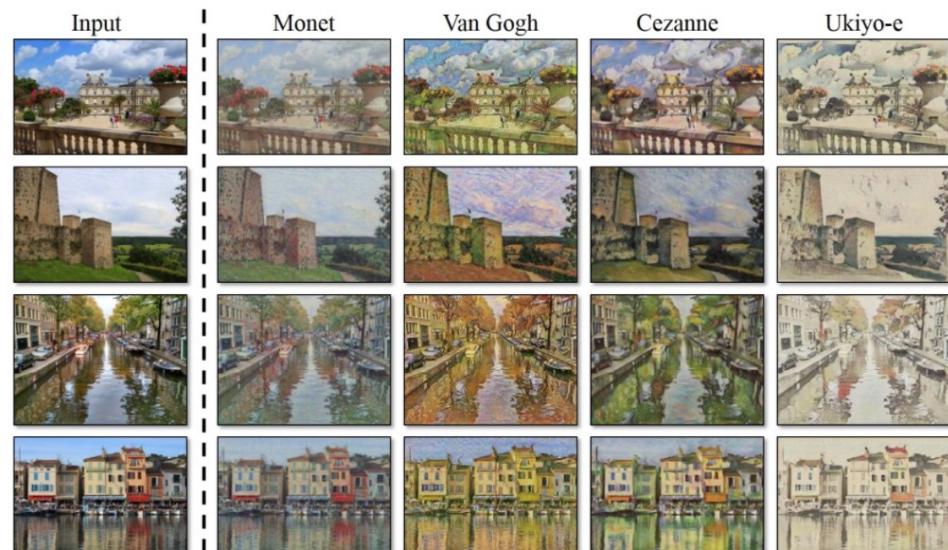
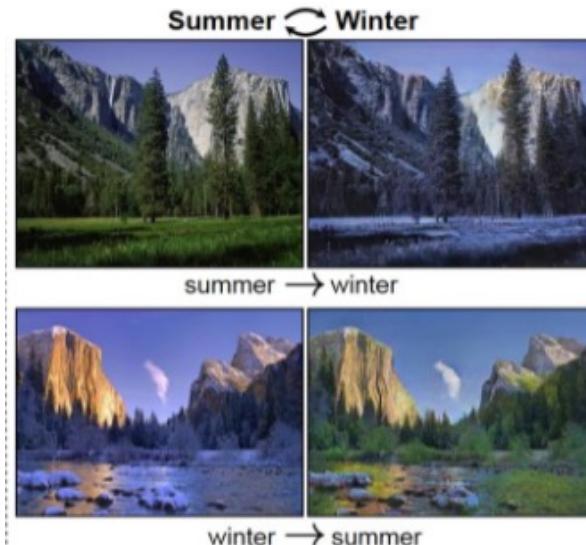
Aprendizaje no supervisado



Aprendizaje no supervisado

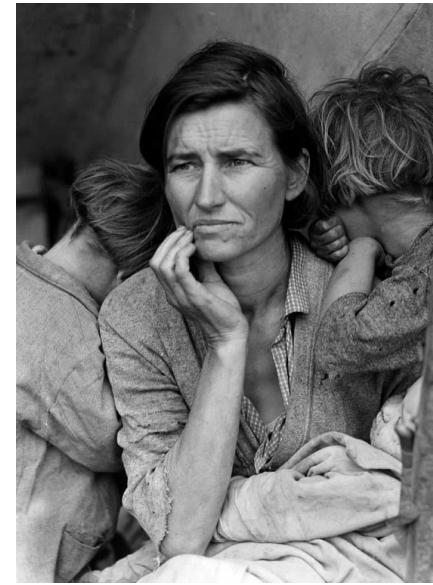


GTA → Cityscapes



Aprendizaje no supervisado

DeOldify



Aprendizaje no supervisado

<https://www.nvidia.com/en-us/research/ai-playground/>

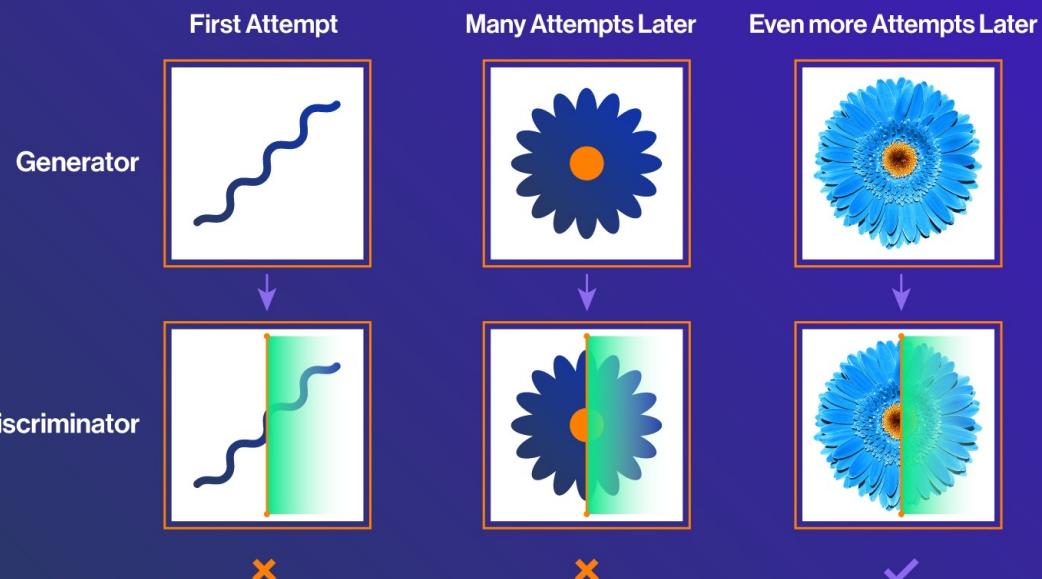
<https://www.nvidia.com/research/inpainting/index.html>



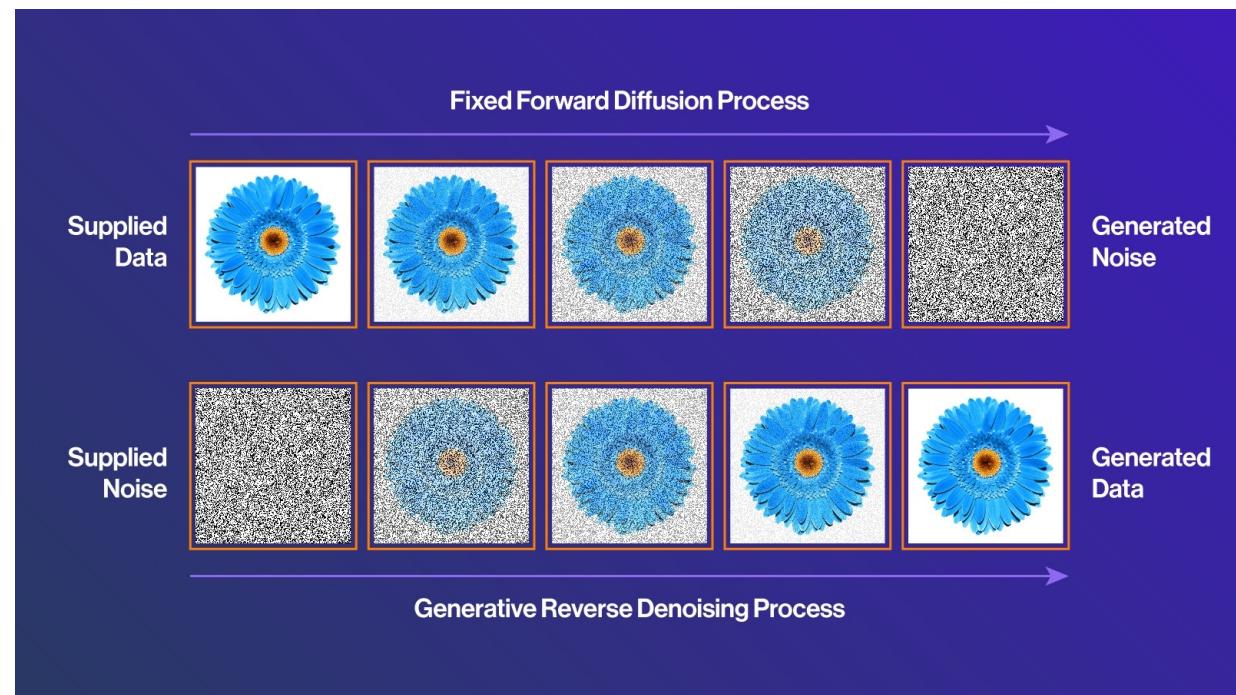
Aprendizaje no supervisado

IA generativa

Redes Generativas Adversarias (GANs)



Modelos de difusión



Aprendizaje no supervisado

¿Para qué podríamos emplear estos modelos?

Sesgos en las bases de datos

Sesgos en las bases de datos

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



TEMPE | SELF-DRIVING VEHICLE HITS BICYCLIST abc 15 ARIZONA

WIRED Uber's Self-Driving Car Didn't Know Pedestrians Could Jaywalk

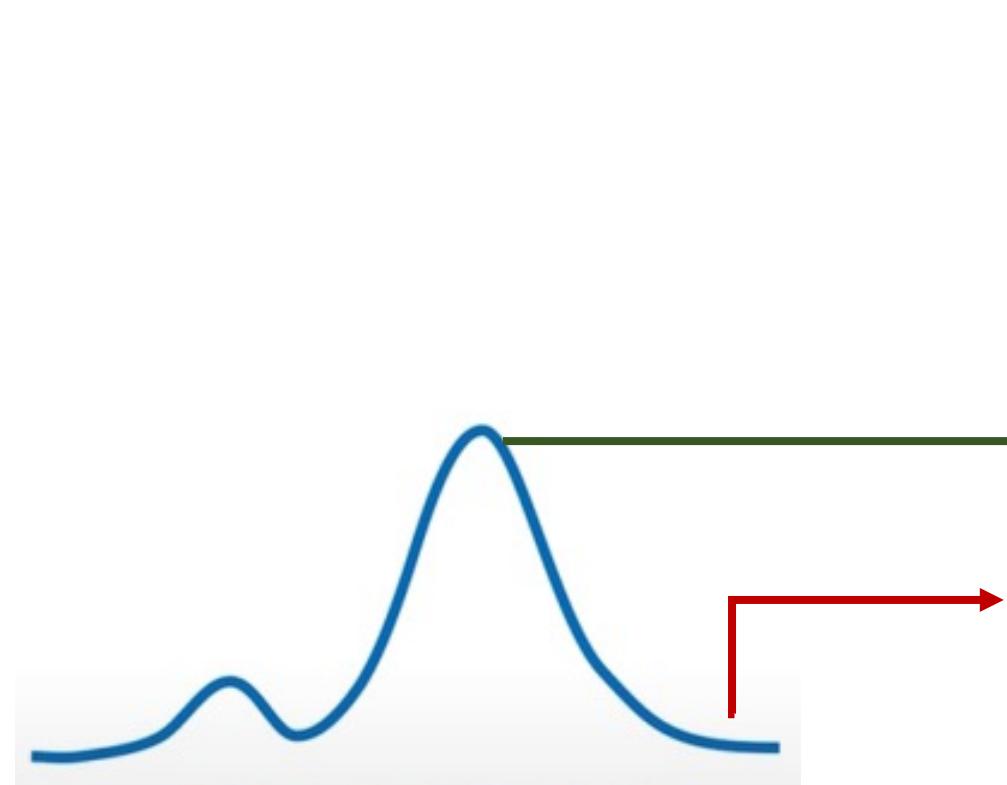
THE SOFTWARE INSIDE the [Uber](#) self-driving SUV that [killed an Arizona woman last year](#) was not designed to detect pedestrians outside of a crosswalk, according to new documents released as part of a federal investigation into the incident.

AAA NewsRoom

New research from AAA reveals that automatic emergency braking systems with pedestrian detection perform inconsistently, and proved to be completely ineffective at night.

<http://gendershades.org/>

Sesgos en las bases de datos



95% de los casos: soleado, carretera, calle recta



Detectar valores atípicos para evitar un comportamiento impredecible durante el entrenamiento.



Práctica

En esta práctica vamos a ver cómo resolver tres problemas diferentes de visión por computador haciendo uso de modelos existentes. Actualmente, la plataforma donde más modelos en abierto existe es [HuggingFace](#).

The image shows the Hugging Face website. On the left, there's a dark sidebar with a yellow emoji at the top, followed by the text "The AI community building the future." and a subtitle "The platform where the machine learning community collaborates on models, datasets, and applications." In the center, there's a search bar with placeholder text "Search models, datasets, users...". Above the search bar are navigation links for "Models", "Datasets", "Spaces", "Docs", "Solutions", "Pricing", "Log In", and "Sign Up". Below the search bar is a navigation bar with tabs for "Tasks", "Libraries", "Datasets", "Languages", "Licenses", and "Other". Under "Tasks", there are sections for "Multimodal" (Text-to-Image, Image-to-Text, Text-to-Video, Visual Question Answering, Document Question Answering, Graph Machine Learning), "Computer Vision" (Depth Estimation, Image Classification, Object Detection, Image Segmentation, Image-to-Image, Unconditional Image Generation, Video Classification, Zero-Shot Image Classification), "Natural Language Processing" (Text Classification, Token Classification, Table Question Answering, Question Answering, Zero-Shot Classification, Translation, Summarization, Conversational, Text Generation, Text2Text Generation, Sentence Similarity), "Audio" (Text-to-Speech, Automatic Speech Recognition, Audio-to-Audio, Audio Classification, Voice Activity Detection), and "Tabular". On the right, there's a large list of "Models" with 469,541 entries. Each entry includes the name of the model, its type (e.g., Text Generation, Image-to-Text), its last update time, its size, and its number of stars (e.g., 64, 393, 136, 334, 64, 899, 322, 5.72k, 2.81k).

Práctica

En esta práctica vamos a ver cómo resolver tres problemas diferentes de visión por computador haciendo uso de modelos existentes. Actualmente, la plataforma donde más modelos en abierto existe es [HuggingFace](#).

The screenshot shows the Hugging Face platform interface. On the left, there's a sidebar with navigation links like 'Tasks', 'Libraries', 'Datasets', 'Languages', 'Licenses', and 'Other'. Below these are sections for 'Multimodal' tasks: 'Feature Extraction', 'Text-to-Image', 'Image-to-Text', 'Text-to-Video', 'Visual Question Answering', 'Document Question Answering', and 'Graph Machine Learning'. A red box highlights the 'Computer Vision' section, which includes 'Depth Estimation', 'Image Classification', 'Object Detection', 'Image Segmentation', 'Image-to-Image', 'Unconditional Image Generation', 'Video Classification', and 'Zero-Shot Image Classification'. The main area displays a list of 391,370 models. Each model entry includes the name, type (e.g., 'Text Generation'), last update time, number of downloads, and likes. Some examples shown are '01-ai/Yi-34B', 'openchat/openchat_3.5', '01-ai/Yi-6B', 'openai/whisper-large-v3', 'HuggingFaceH4/zephyr-7b-beta', 'NousResearch/Yarn-Mistral-7b-128k', and '01-ai/Yi-34B-200K'. At the top right, there are buttons for 'Log In' and 'Sign Up'.

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

Encontrarás aplicaciones para resolver problemas de clasificación, segmentación y detección de objetos.

Los modelos los cogeremos de la API de HuggingFace y haremos directamente inferencia sobre ellos.

```
classifier = pipeline("image-classification", model="nateraw/food")

# Cargamos la imagen
url = 'https://www.annarecetasfáciles.com/files/miniatura-58-1536x862.jpg'
image = Image.open(requests.get(url, stream=True).raw)

# Clasificamos imagen con modelo pre-entrenado
result = classifier(image)
print(result)
```

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

model="nateraw/food"

grilled_salmon: 0.99



model="trpakov/vit-face-expression"

angry: 0.88



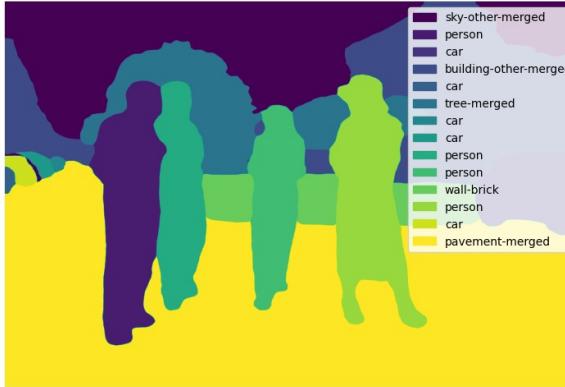
```
[{'score': 0.9936506152153015, 'label': 'grilled_salmon'},  
 {'score': 0.00047507102135568857, 'label': 'sushi'},  
 {'score': 0.00037883027107454836, 'label': 'sashimi'},  
 {'score': 0.0003085696662310511, 'label': 'fish_and_chips'},  
 {'score': 0.00024178232706617564, 'label': 'pork_chop'}]
```

```
[{'score': 0.882022500038147, 'label': 'angry'},  
 {'score': 0.054023776203393936, 'label': 'happy'},  
 {'score': 0.04549011215567589, 'label': 'fear'},  
 {'score': 0.013408415019512177, 'label': 'surprise'},  
 {'score': 0.002085126470774412, 'label': 'neutral'}]
```

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

model="facebook/mask2former-swin-base-coco-panoptic"



```
[{"score": 0.999196, "label": "sky-other-merged", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.99982, "label": "person", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.998283, "label": "car", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.985628, "label": "building-other-merged", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.957127, "label": "car", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.996432, "label": "tree-merged", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.966579, "label": "car", "mask": <PIL.Image.Image image mode=L size=6000x4000>}, {"score": 0.998756, "label": "cat", "mask": <PIL.Image.Image image mode=L size=640x480 at 0x7EE38AD64250>}, {"score": 0.997873, "label": "cat", "mask": <PIL.Image.Image image mode=L size=640x480 at 0x7EE391146590>}, {"score": 0.997116, "label": "remote", "mask": <PIL.Image.Image image mode=L size=640x480 at 0x7EE389F13490>}, {"score": 0.974273, "label": "remote", "mask": <PIL.Image.Image image mode=L size=640x480 at 0x7EE389F13CD0>}, {"score": 0.956032, "label": "couch", "mask": <PIL.Image.Image image mode=L size=640x480 at 0x7EE389F13E20>}]
```

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

model="microsoft/table-transformer-detection"

The screenshot shows a PDF document titled "PDF Parser Data Test.pdf". The content includes a list of animals with their counts: Lizard 25, Dog 250, Cat 115, Bird 50, Guinea pig 30, Hamster 45, Fish 75, Ferret 10. Below this is an example section about electromagnets:

Example 2: Electromagnets—Increasing Coils (GR 3–5)
The following data were collected using an electromagnet with a 1.5 volt battery, a switch, a piece of #20 insulated wire, and a nail. Three trials were run. *Safety precautions in repeating this experiment include using safety goggles or safety spectacles and avoiding short circuits.*

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Below the table is another example section about pH values:

Example 3: pH of Substances (GR 5–10)
The following are pH values of common household substances taken by three different teams using pH probes. *Safety precautions in repeating this experiment include hooded ventilation, chemical-splash safety goggles, gloves, and apron. Do not use bleach, ammonia, or strong acids with children.*

Lemon juice 2.4, 2.0, 2.2; Baking soda (1 Tbsp) in Water (1 cup) 8.4, 8.3, 8.7;
Orange juice 3.5, 4.0, 3.4; Battery acid 1.0, 0.7, 0.5; Apples 3.0, 3.2, 3.5;
Tomatoes 4.5, 4.0, 4.0; Bottled water 5.7, 7.0, 7.2; Milk of magnesia 10.5, 10.2

```
[{'score': 0.9995410442352295, 'label': 'table', 'box': {'xmin': 231, 'ymin': 336, 'xmax': 653, 'ymax': 484}}]
```

Table name : daily historical stock prices & volumes

Date	Open	High	Low	Close / Last	Volume
01/04/2017	62.48	62.75	62.12	62.3	21,325,140
01/03/2017	62.79	62.84	62.125	62.58	20,655,190
2/30/2016	62.96	62.99	62.03	62.14	25,575,720
2/29/2016	62.86	63.2	62.73	62.9	10,248,460
2/28/2016	63.4	63.4	62.83	62.99	14,348,340
2/27/2016	63.21	64.07	63.21	63.28	11,743,650
2/23/2016	63.45	63.54	62.8	63.24	12,399,540
2/22/2016	63.84	64.1	63.405	63.55	22,175,270
2/21/2016	63.43	63.7	63.12	63.54	17,084,370
2/20/2016	63.69	63.8	63.025	63.54	26,017,470
2/19/2016	62.56	63.77	62.42	63.62	34,318,500
2/16/2016	62.95	62.95	62.115	62.3	42,452,660

http://www.nasdaq.com/symbol/tis/historical

Date	Open	High	Low	Close / Last	Volume
01/04/2017	117.55	119.66	117.29	118.69	19,594,560
01/03/2017	116.03	117.84	115.51	116.86	20,635,600
2/30/2016	116.595	116.83	114.7739	115.05	18,668,290
2/29/2016	117	117.531	116.06	116.35	9,925,082
2/28/2016	118.19	118.25	116.65	116.92	11,985,740
2/27/2016	116.96	118.68	116.864	118.01	12,034,590
2/23/2016	117	117.56	116.3	117.27	10,885,030
2/22/2016	118.86	118.99	116.93	117.4	16,226,770
2/21/2016	118.92	119.2	118.48	119.04	10,747,610
2/20/2016	119.5	119.77	118.8	119.09	13,673,570
2/19/2016	119.85	120.36	118.51	119.24	15,871,360
2/16/2016	120.9	121.5	119.27	119.87	25,316,220

```
[{'score': 0.9955059289932251, 'label': 'table', 'box': {'xmin': 65, 'ymin': 52, 'xmax': 505, 'ymax': 270}}, {'score': 0.9984880685806274, 'label': 'table', 'box': {'xmin': 65, 'ymin': 315, 'xmax': 504, 'ymax': 533}}]
```

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

model="microsoft/table-transformer-structure-recognition"

Lizard 25, Dog 250, Cat 115, Bird 50, Guinea pig 30, Hamster 45, Fish 75, Ferret 10

Example 2: Electromagnets—Increasing Coils (GR 3–5)
The following data were collected using an electromagnet with a 1.5 volt battery, a switch, a piece of #20 insulated wire, and a nail. Three trials were run. *Safety precautions in repeating this experiment include using safety goggles or safety spectacles and avoiding short circuits.*

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Example 3: pH of Substances (GR 5–10)
The following are pH values of common household substances taken by three different teams using pH probes. *Safety precautions in repeating this experiment include hooded ventilation, chemical-splash safety goggles, gloves, and apron. Do not use bleach, ammonia, or strong acids with children.*

Lemon juice 2.4, 2.0, 2.2; Baking soda (1 Tbsp) in Water (1 cup) 8.4, 8.3, 8.7; Orange juice 3.5, 4.0, 3.4; Battery acid 1.0, 0.7, 0.5; Apples 3.0, 3.2, 3.5; Tomatoes 4.5, 4.3, 4.0; Bottled water 5.7, 7.0, 7.2; Milk of magnesia 10.5, 10.2

Lizard 25, Dog 250, Cat 115, Bird 50, Guinea pig 30, Hamster 45, Fish 75, Ferret 10

Example 2: Electromagnets—Increasing Coils (GR 3–5)
The following data were collected using an electromagnet with a 1.5 volt battery, a switch, a piece of #20 insulated wire, and a nail. Three trials were run. *Safety precautions in repeating this experiment include using safety goggles or safety spectacles and avoiding short circuits.*

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Example 3: pH of Substances (GR 5–10)
The following are pH values of common household substances taken by three different teams using pH probes. *Safety precautions in repeating this experiment include hooded ventilation, chemical-splash safety goggles, gloves, and apron. Do not use bleach, ammonia, or strong acids with children.*

Lemon juice 2.4, 2.0, 2.2; Baking soda (1 Tbsp) in Water (1 cup) 8.4, 8.3, 8.7; Orange juice 3.5, 4.0, 3.4; Battery acid 1.0, 0.7, 0.5; Apples 3.0, 3.2, 3.5; Tomatoes 4.5, 4.3, 4.0; Bottled water 5.7, 7.0, 7.2; Milk of magnesia 10.5, 10.2

Error en detección de filas y columnas.

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

model="microsoft/table-transformer-structure-recognition"

Lizard 25, Dog 250, Cat 115, Bird 50, Guinea pig 30, Hamster 45, Fish 75,
Ferret 10

Example 2: Electromagnets—Increasing Coils (GR 3–5)
The following data were collected using an electromagnet with a 1.5 volt battery, a switch, a piece of #20 insulated wire, and a nail. Three trials were run. *Safety precautions in repeating this experiment include using safety goggles or safety spectacles and avoiding short circuits.*

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Example 3: pH of Substances (GR 5–10)
The following are pH values of common household substances taken by three different teams using pH probes. *Safety precautions in repeating this experiment include hooded ventilation, chemical-splash safety goggles, gloves, and apron. Do not use bleach, ammonia, or strong acids with children.*

Lemon juice 2.4, 2.0, 2.2; Baking soda (1 Tbsp) in Water (1 cup) 8.4, 8.3, 8.7;
Orange juice 3.5, 4.0, 3.4; Battery acid 1.0, 0.7, 0.5; Apples 3.0, 3.2, 3.5;
Tomato juice 4.5, 4.7, 4.0; Distilled water 6.7, 7.0, 7.2; Milk of magnesia 10.5, 10.2

RESULTS.

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

model="microsoft/table-transformer-structure-recognition"

:u11s.

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

:u11s.

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

:u11s.

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

:u11s.

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

Finalmente, encontrarás un ejemplo de IA generativa. En el que podremos cambiar una imagen de entrada mediante un prompt.

"turn him into cyborg"



Imagen entrada

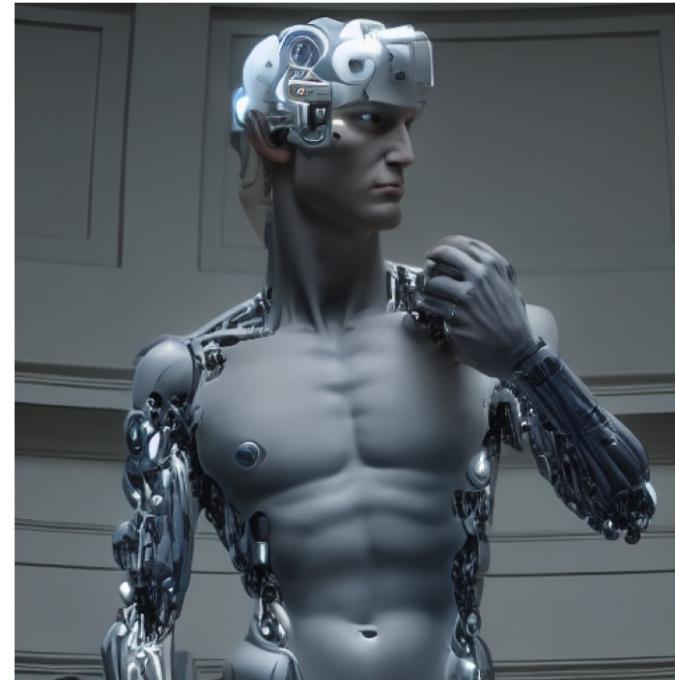


Imagen salida

Práctica

<https://colab.research.google.com/drive/1wxXUWTwOaNa-HuMk-BC9WGbzpSbHwEky?usp=sharing>

Ejercicio 1

En cada aplicación encontrarás diferentes imágenes. Ejecuta el código con cada una de ellas para ver los resultados. Busca imágenes adicionales en internet para comprobar el funcionamiento.

Ejercicio 2

Implementa una pipeline de segmentación haciendo uso del modelo "mattmdjaga/segformer_b2_clothes" (https://huggingface.co/mattmdjaga/segformer_b2_clothes). Este, dada una imagen de entrada, segmenta y categoriza la ropa presente en la imagen.

Usa tus propias imágenes o algunas de internet para ver los resultados.

Tema 2. Aplicaciones de la IA en tareas de visión por computador

Ana Jiménez Pastor
anjipas@gmail.com