



Curso de IA Generativa Aplicada a Negocio

M3 - Clase 9: Intro a generación de imágenes - 14/oct/24 - Prof: Fran Bartolomé



Cofinanciado por
la Unión Europea



MINISTERIO
DE INDUSTRIA
Y TURISMO



Fondos Europeos





Módulo 3

Generación y edición de imagen

1. Introducción a la generación de imágenes.
2. Modelos Generativos.
3. Técnicas de entrenamiento.
4. Implicaciones en el Arte.
5. Herramientas:
 - o Dall·E 3.
 - o Stable Diffusion.
 - o Midjourney.
 - o Adobe Firefly.
 - o Ideogram.
 - o Leonardo.
 - o Flux.
 - o Krea.
6. Proyecto práctico.





Introducción a la generación de imágenes

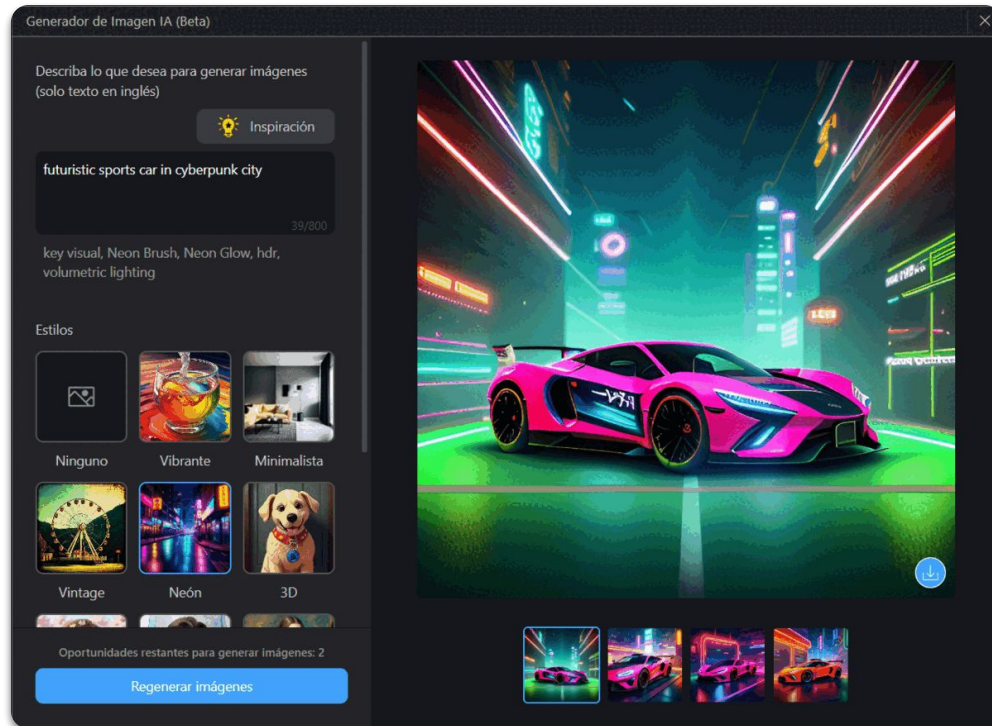




¿Qué es la generación de imágenes con IA?

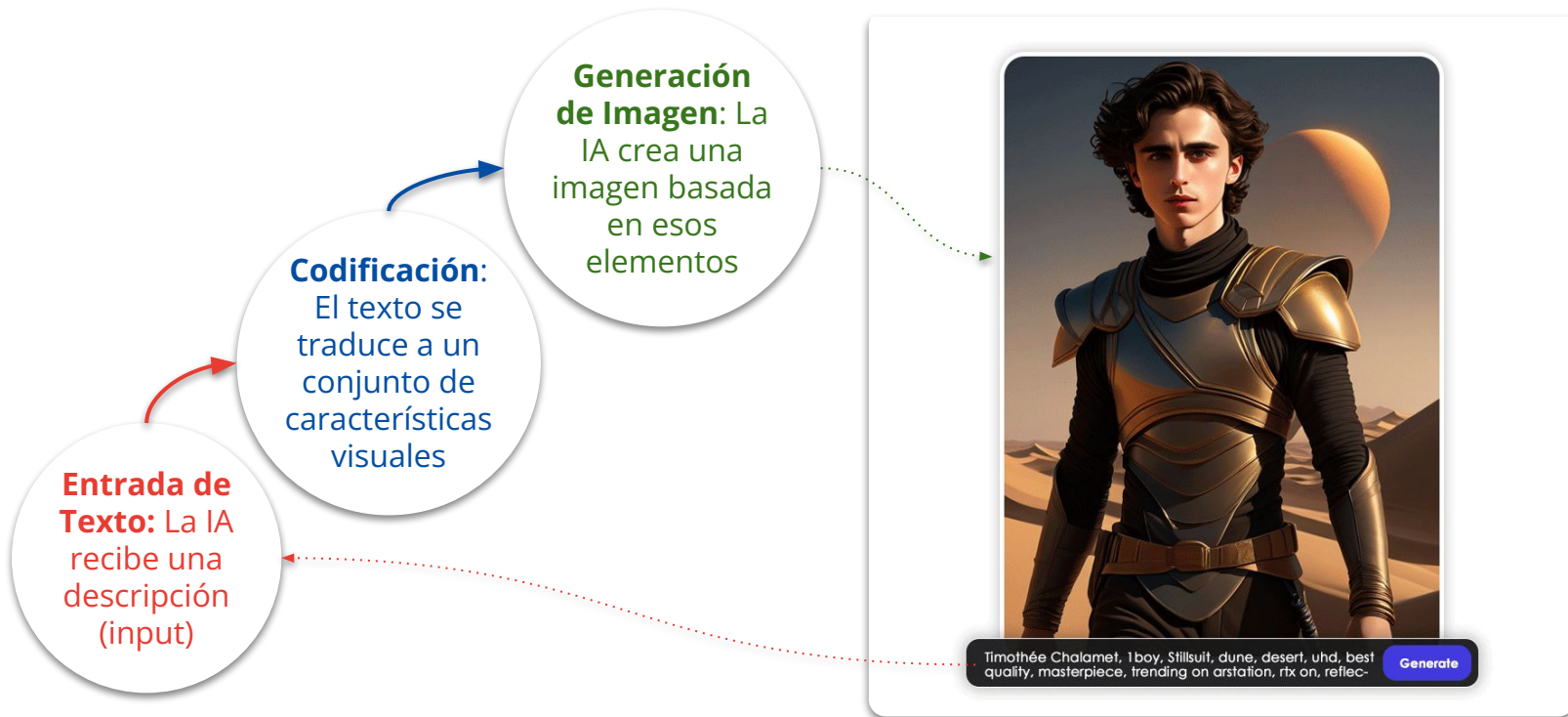
Proceso donde algoritmos de IA crean imágenes originales a partir de descripciones textuales.

Ejemplo: Coche deportivo futurista en una ciudad ciberpunk



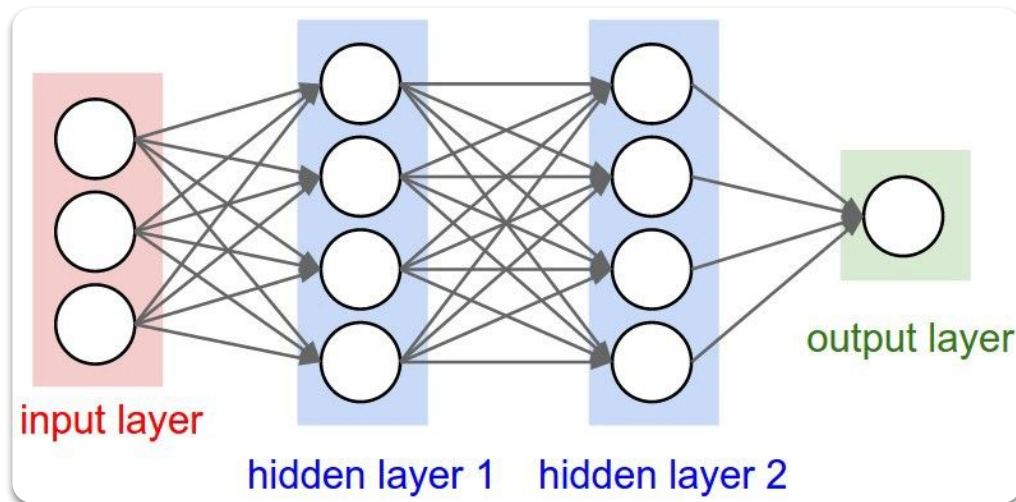


¿Cómo funciona?





Redes Neuronales

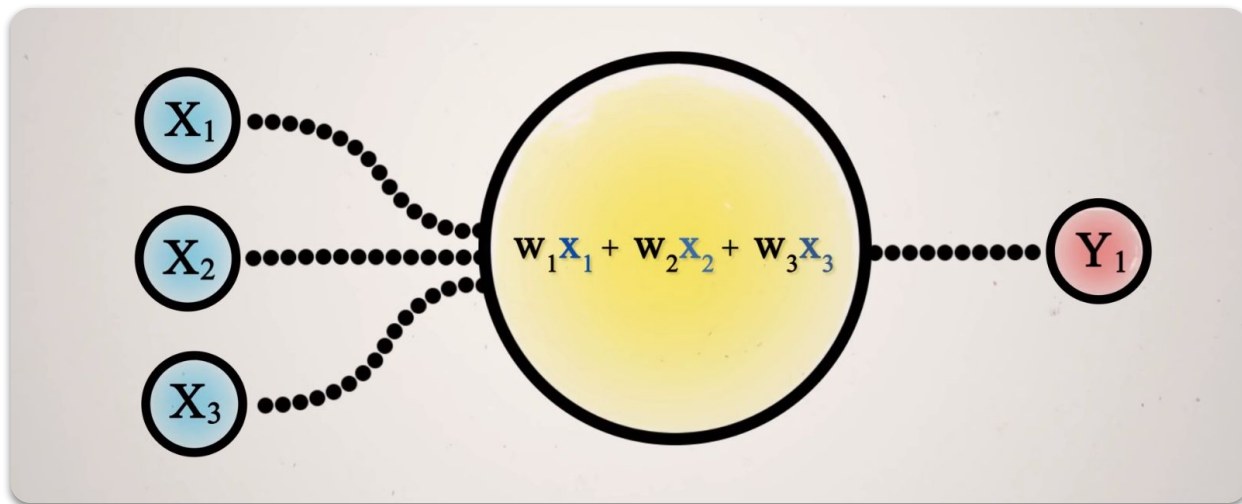


"GATO"

"PERRO"

Es un modelo de **Machine Learning**, inspirado en la biología del cerebro. La red puede *aprender* la relación entre unos **datos de entrada** determinados y unos **valores de salida**. Tras ser entrenada, la representación de los datos aprendida por la red en sus **capas intermedias** permite a la red **generalizar** su conocimiento más allá de los ejemplos proporcionados.

Redes Neuronales - ¿Cómo funciona una neurona IA?



La neurona artificial **recibe datos**: números, palabras, imágenes...

Procesa la información: realiza cálculos simples sobre estos datos (sumas, restas, medias...)

Toma una decisión: se activa y **envía información** a otras neuronas, o no se activa y no envía nada.



Redes Neuronales - Fases del entrenamiento

FORWARD PASS

1

Input → (Procesamiento info. capa previa) x N° Capas → **Output**

2

FUNCIÓN DE COSTE : F(Predicho, Real) → % ERROR

3

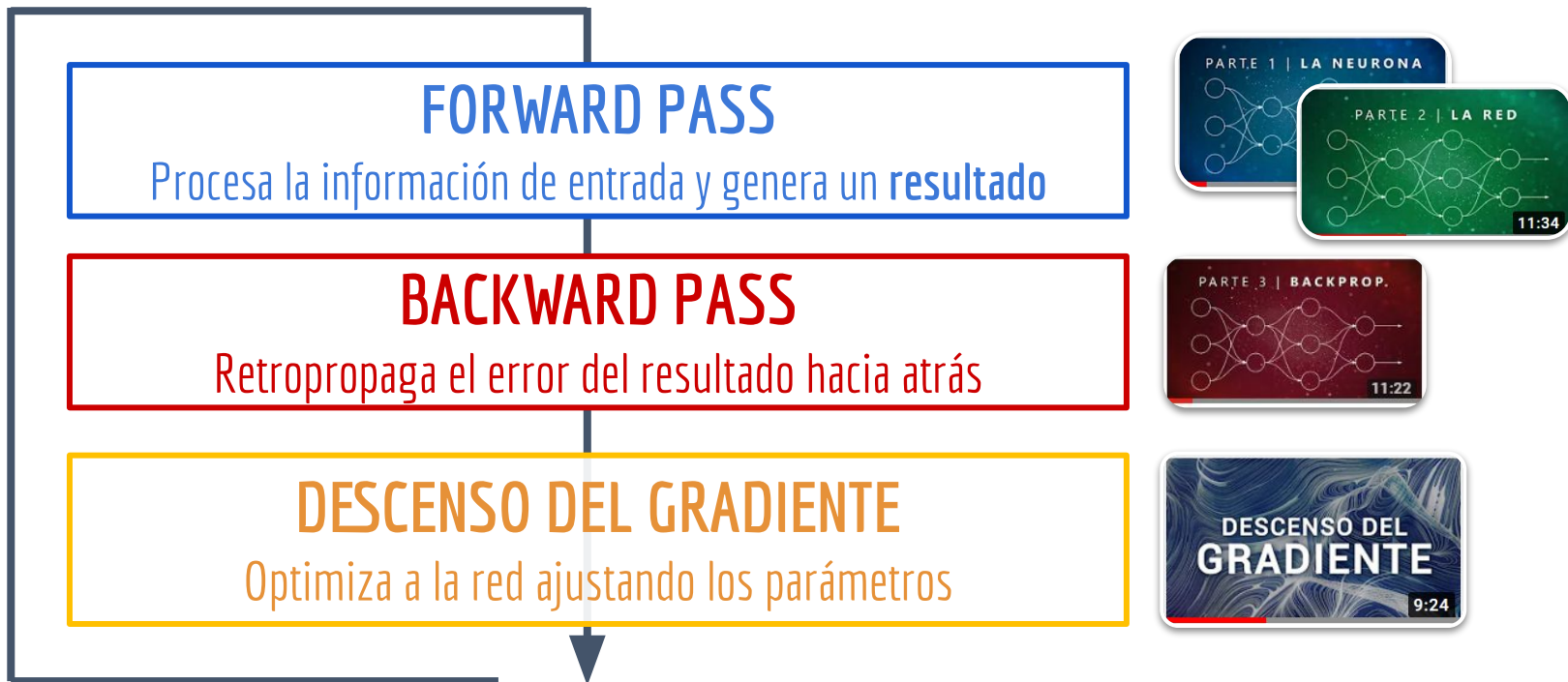
BACKWARD PASS

LOS ERRORES SON PROPAGADOS HACIA ATRÁS PARA IR AJUSTANDO LOS PARÁMETROS



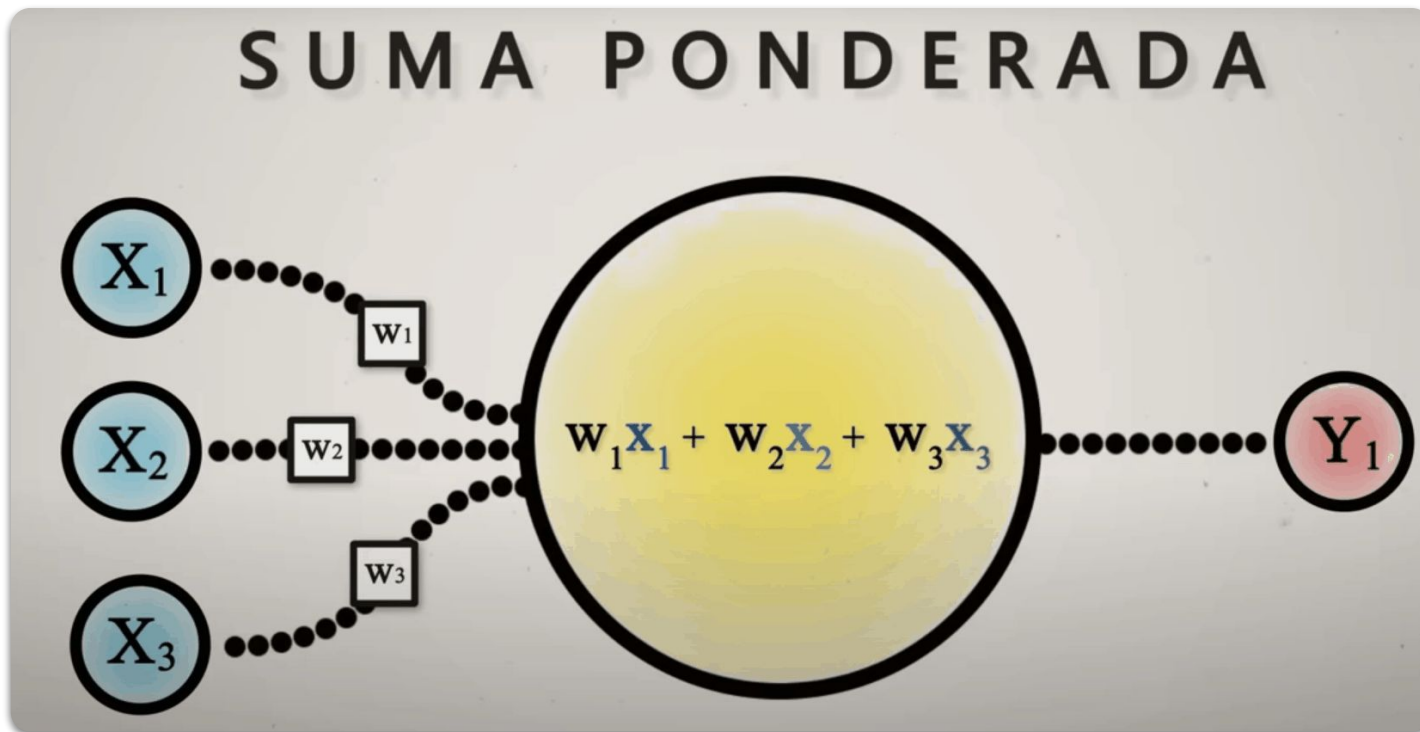


Redes Neuronales - Ciclos de entrenamiento



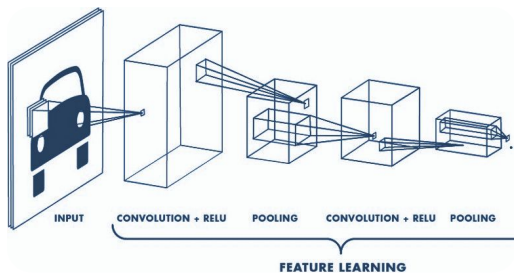


Redes Neuronales - Definición de parámetros



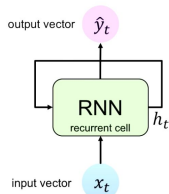


Redes Neuronales - Arquitecturas principales para imagen



RECONOCIMIENTO DE IMAGEN

Red Neuronal CONVOLUCIONAL (CNN)



Apply a recurrence relation at every time step to process a sequence:

$$h_t = f_W(h_{t-1}, x_t)$$

cell state

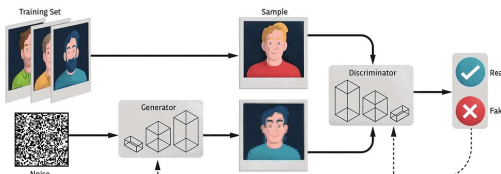
function parameterized by W

old state

input vector at time step t

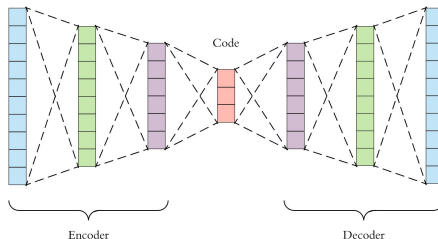
RECONOCIMIENTO DE IMAGEN

Red Neuronal RECURRENTE (RNN)



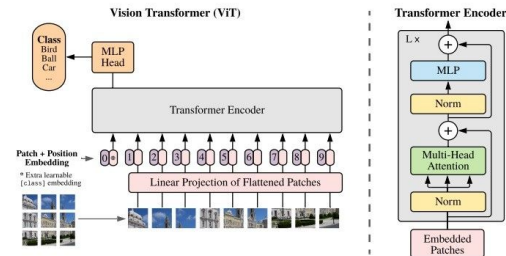
RECON. Y GENERACIÓN DE IMAGEN

Red ADVERSARIA GENERATIVA (GAN)



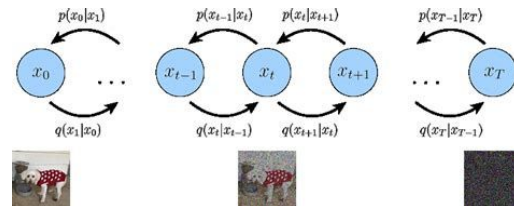
GENERACIÓN DE IMAGEN

Red Neuronal AUTOENCODER (VAE)



GENERACIÓN DE IMAGEN

Red Neuronal TRANSFORMER

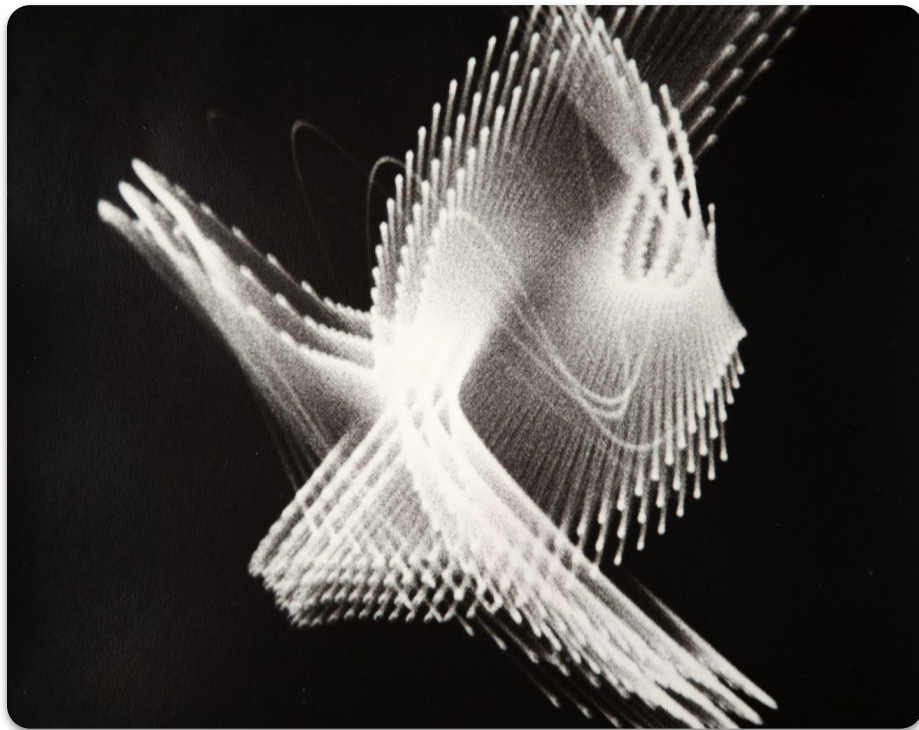


GENERACIÓN DE IMAGEN

Red Neuronal DIFFUSION MODEL



Historia y Evolución - Arte con ordenador analógico



Con un ordenadores analógicos, que utilizaban señales físicas continuas (corrientes eléctricas, voltajes...) para representar y manipular datos.

1952



Electronics Abstractions - Ben F. Laposky



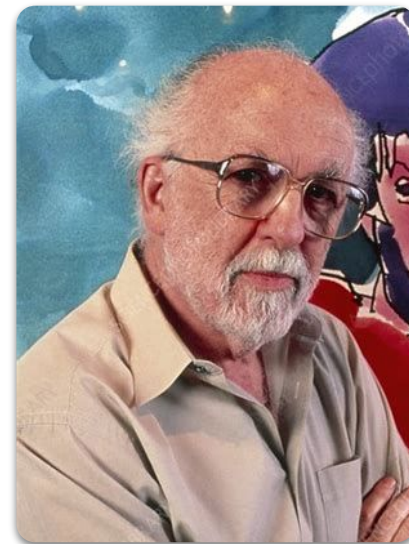


Historia y Evolución - Robot Aaron



*¿Cuál es
la condición mínima
para que un conjunto
de marcas funcione
como una imagen?*

1973



Robot AARON

-

de Harold Cohen



Cofinanciado por
la Unión Europea



GOBIERNO
DE ESPAÑA

MINISTERIO
DE INDUSTRIA
Y TURISMO



Escuela de
organización
industrial



Fondos Europeos

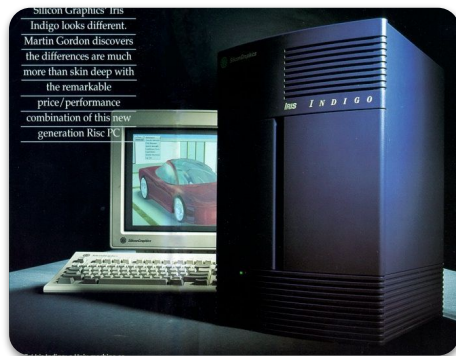




Historia y Evolución - Ordenadores personales



años 80 y 90
S.XX





Historia y Evolución - LeNet 5 - Redes Convolucionales



1989 - 1995

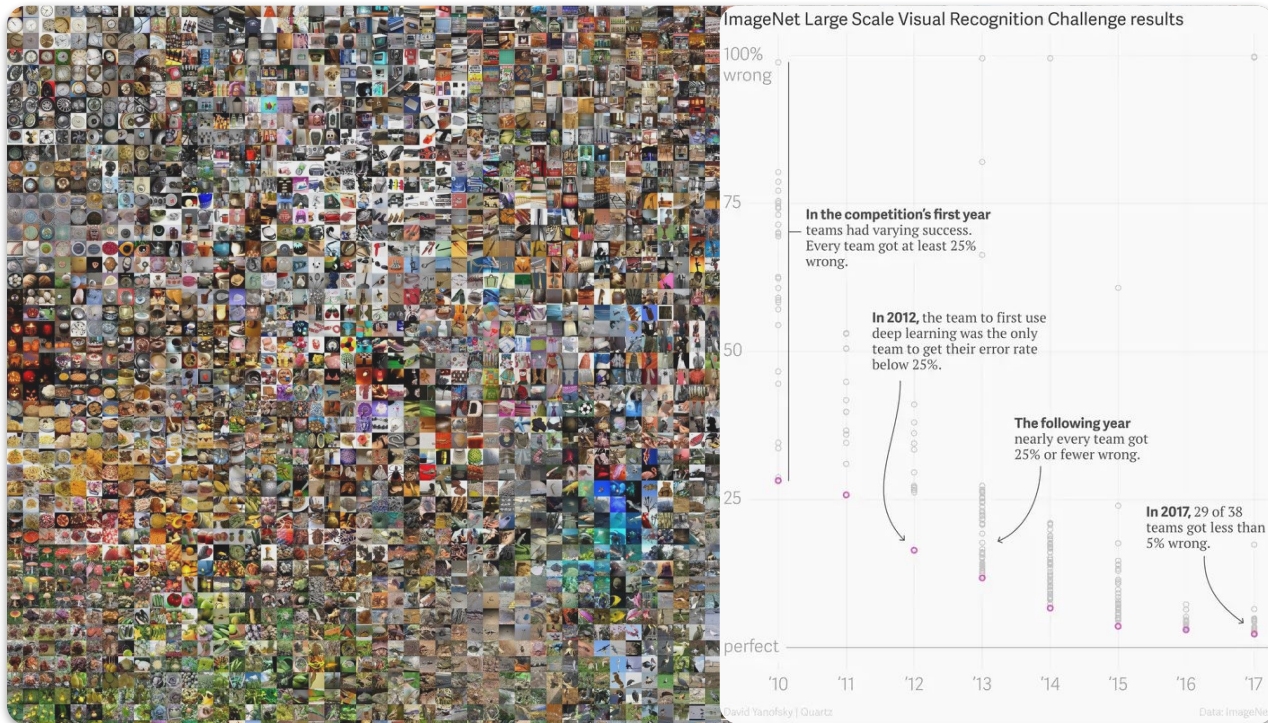


LeNet 5 - Yann LeCun





Historia y Evolución - ImageNet - Data Science y CNN



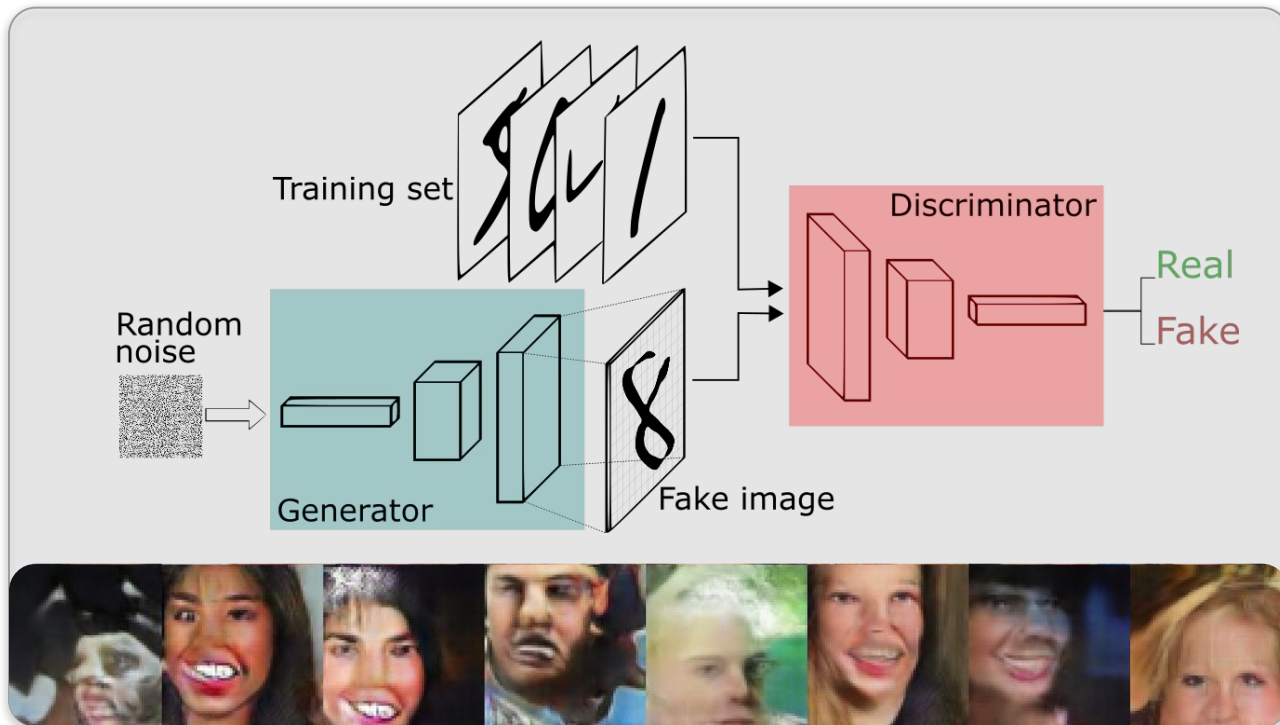
2000 - 2012 -
Actualidad



Fei-Fei Li



Historia y Evolución - Redes Generativas Adversarias (GAN)



2000 - 2012 -
Actualidad



Ian Goodfellow





Historia y Evolución - Redes Generativas Adversarias (GAN)

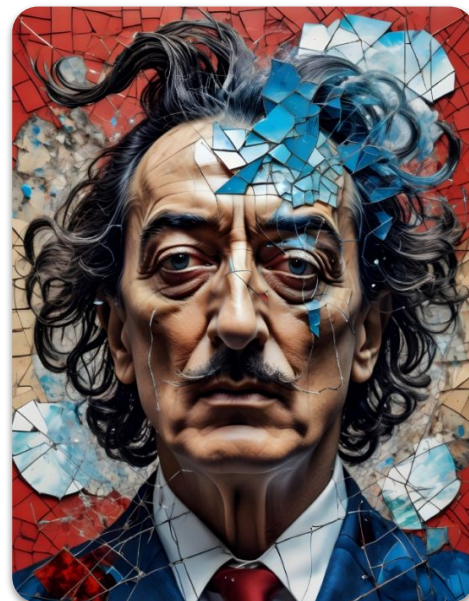




Historia y Evolución - Deep Dream de Google



2015 - Actualidad





Historia y Evolución - CycleGan - Transferencia de estilos

2017 - Actualidad

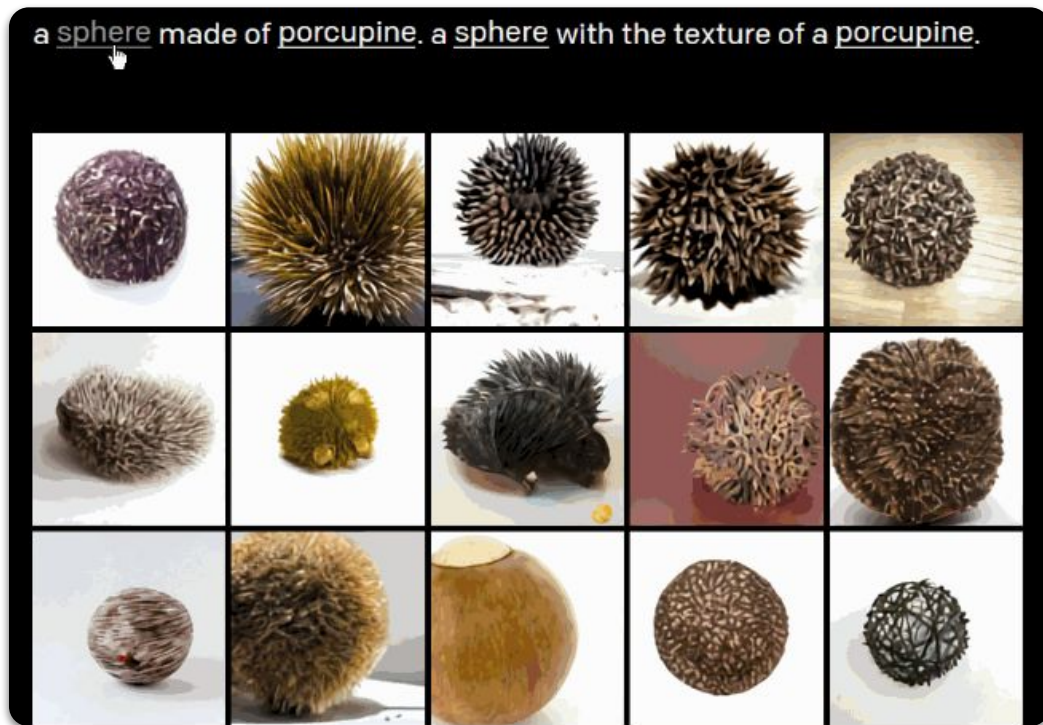




Historia y Evolución - Deepfakes - Transferencia de estilos



Historia y Evolución - Transformers - DALL·E



2019 - 2021

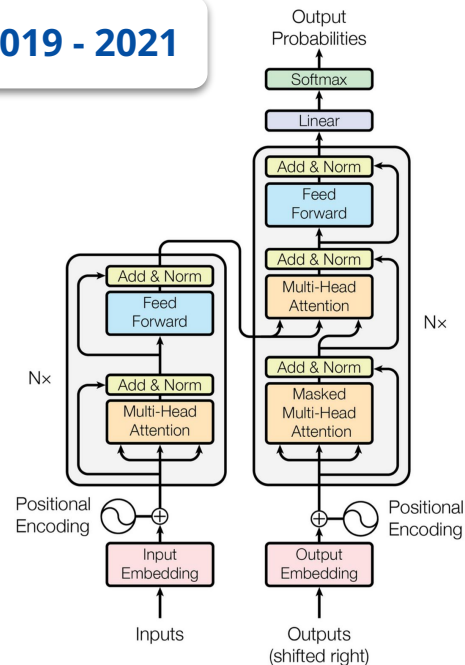


Figure 1: The Transformer - model architecture.



Aplicaciones Empresariales

1. Marketing y Publicidad:

- a. **Creación rápida de contenido visual:** Generar imágenes personalizadas para campañas publicitarias, redes sociales y materiales de marketing es más rápido y eficiente que nunca.
- b. **Personalización masiva:** Crear imágenes únicas para cada cliente o segmento de mercado, lo que aumenta la relevancia y el engagement.
- c. **Prototipado visual:** Visualizar rápidamente ideas y conceptos antes de invertir en producción.

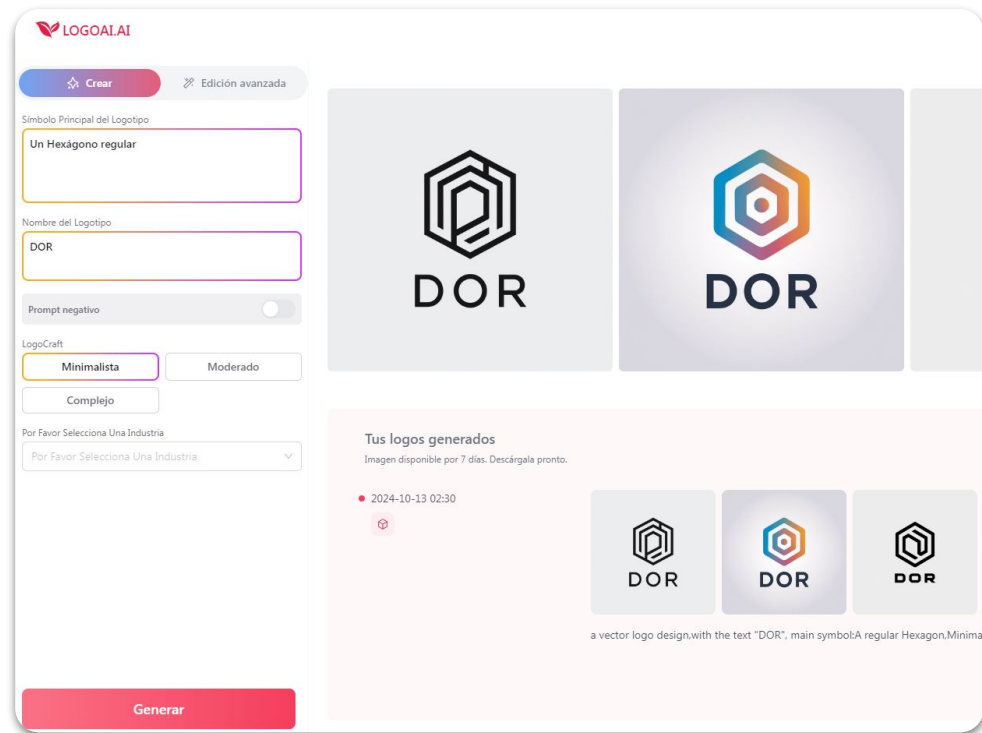




Aplicaciones Empresariales

2. Diseño Gráfico:

- Generación de ideas:** La IA puede sugerir nuevas ideas de diseño, combinaciones de colores y composiciones.
- Creación de logotipos y branding:** Desarrollar logotipos y elementos de marca personalizados.
- Diseño de productos:** Visualizar productos antes de su fabricación, lo que agiliza el proceso de diseño.





Aplicaciones Empresariales

3. Desarrollo de producto:

- a. **Simulaciones visuales:** Crear simulaciones realistas de productos en diferentes entornos, facilitando la toma de decisiones.
- b. **Diseño de interfaces de usuario:** Generar prototipos de interfaces de usuario de forma rápida y sencilla.



Aplicaciones Empresariales

4. Educación:

- a. **Material didáctico personalizado:** Crear imágenes y diagramas personalizados para explicar conceptos complejos de forma visual.
- b. **Simulaciones interactivas:** Desarrollar simulaciones interactivas para enseñar ciencias, historia y otras materias.

Results



"Greek temple on the coast"



Aplicaciones Empresariales

5. Entretenimiento:

- a. **Creación de juegos:**
Generar entornos de juego y personajes de forma más rápida y eficiente.
- b. **Producción de películas y series:** Crear efectos visuales y animaciones.



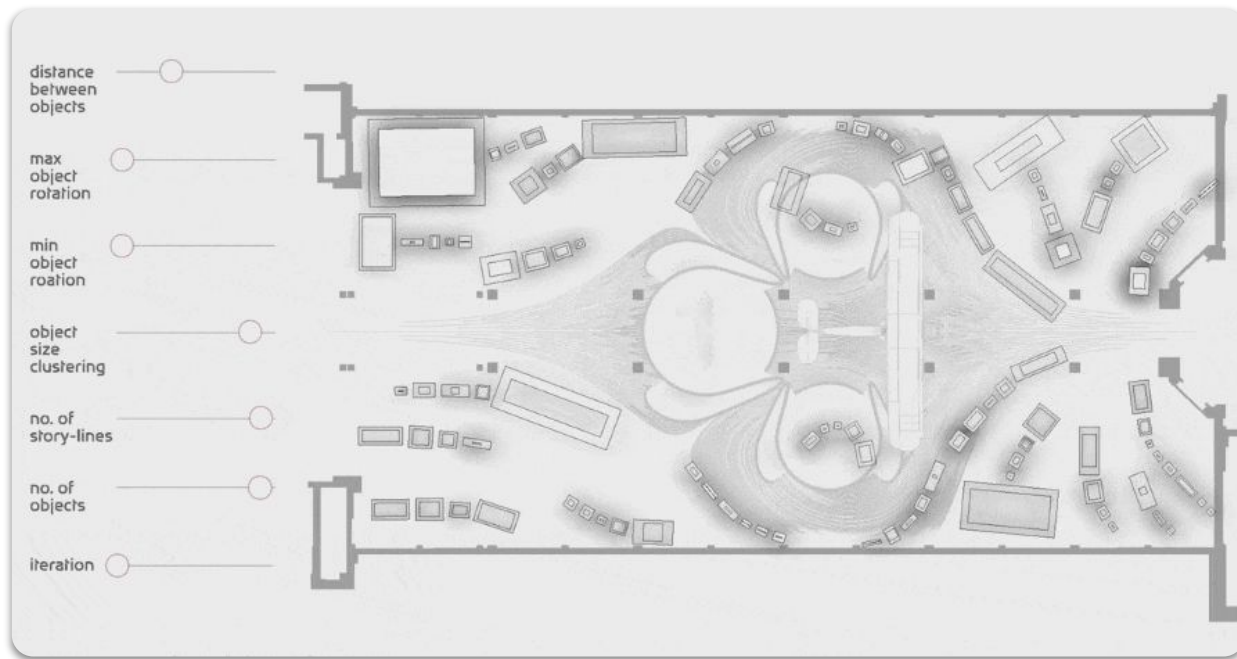


Aplicaciones Empresariales

6. Arquitectura / Interiorismo:

a. **Visualización de proyectos:** Crear renderizados realistas de proyectos arquitectónicos y de interiorismo.

b. **Diseño de espacios virtuales:** Generar entornos virtuales para experiencias inmersivas.





Aplicaciones Empresariales

7. Moda:

- a. **Diseño de prendas:**
Crear diseños de prendas únicas y personalizadas.
- b. **Catálogos de productos:** Generar imágenes de productos para catálogos en línea y tiendas virtuales.

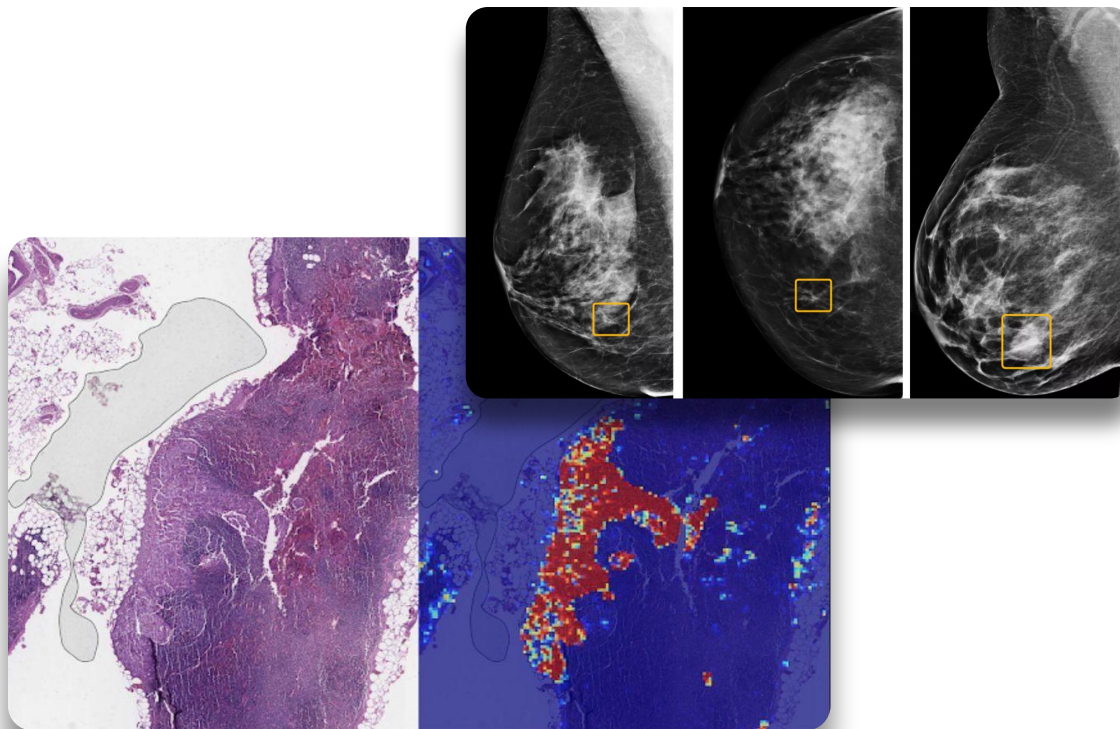




Aplicaciones Empresariales

8. Medicina:

- a. **Mejora de resolución:**
Para imágenes de resonancias magnéticas, tomografías. Permiten detectar anomalías con más precisión
- b. **Asistencia en diagnóstico:** Reconocer signos de enfermedades en imágenes, incluso detectando signos tempranos.





Modelos Generativos



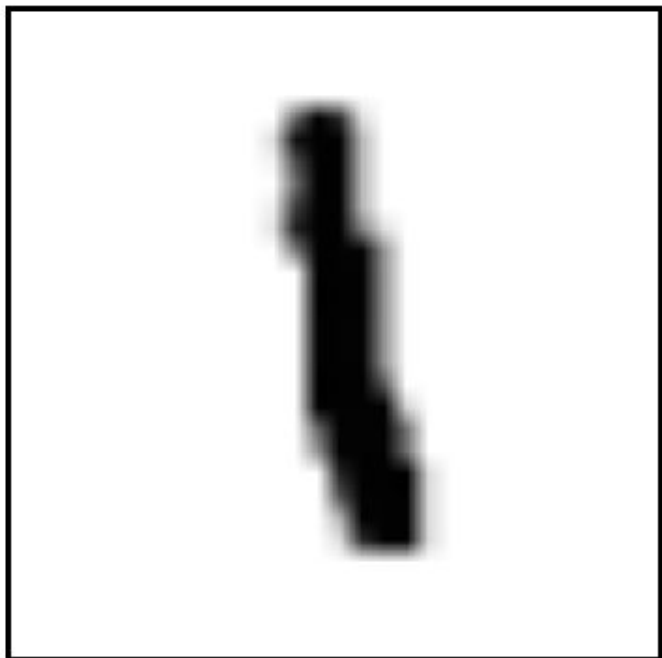


Antes de nada...





... convertimos las imágenes en datos numéricos

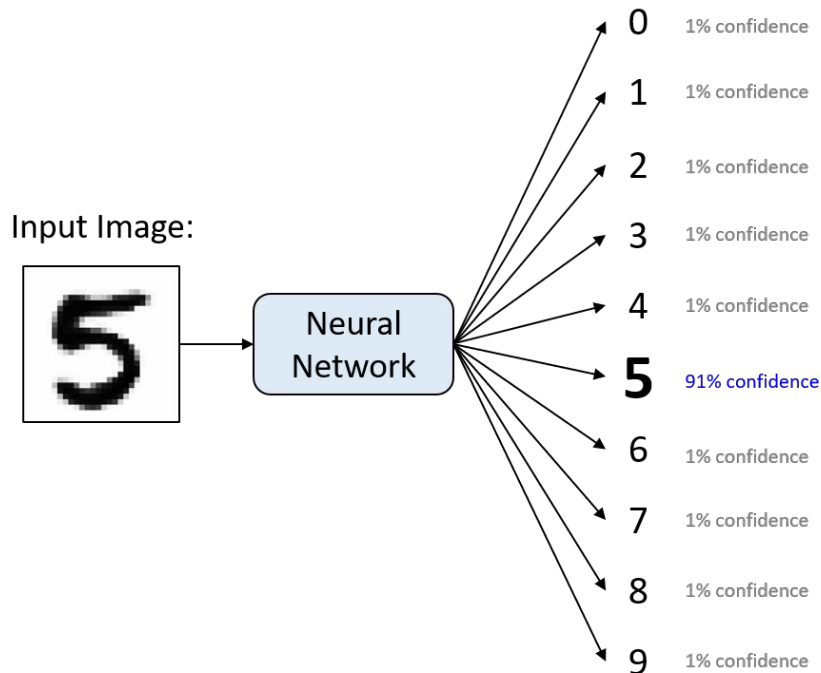


12

0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	.6	.8	0	0	0	0	0	0
0	0	0	0	0	0	.7	1	0	0	0	0	0	0
0	0	0	0	0	0	.7	1	0	0	0	0	0	0
0	0	0	0	0	0	.5	1	.4	0	0	0	0	0
0	0	0	0	0	0	0	1	.4	0	0	0	0	0
0	0	0	0	0	0	0	1	.4	0	0	0	0	0
0	0	0	0	0	0	0	1	.7	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	0	0	.9	1	.1	0	0	0	0
0	0	0	0	0	0	0	.3	1	.1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0



Y serán esos valores numéricos los que usarán cada modelo





¿Qué es una CONVOLUCIÓN?

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		

Operación por la cual **sobre una imagen dada (píxeles)** se desliza **un filtro o kernel**, que va calculando una suma ponderada de los píxeles vecinos para conformar así otra **imagen resultante (píxeles)**.



Filtros en las convoluciones



Original

•0	•0	•0
•0	•1	•0
•0	•0	•0



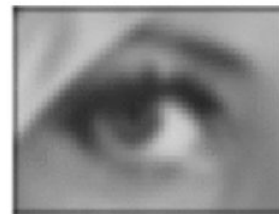
Filtered
(no change)

•0	•0	•0
•1	•0	•0
•0	•0	•0



Shifted right
By 1 pixel

$\frac{1}{9}$	•1	•1	•1
	•1	•1	•1
	•1	•1	•1



Blur (with a
box filter)

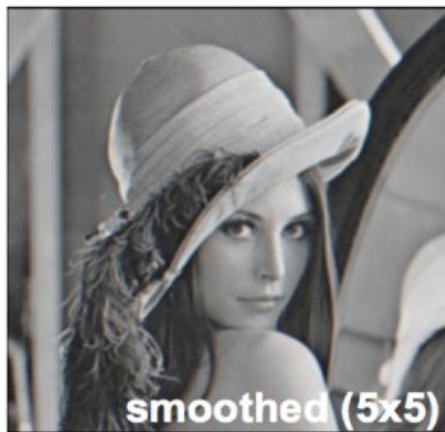


Filtros en las convoluciones



original

−



smoothed (5x5)

=



detail

•0	•0	•0
•0	•1	•0
•0	•0	•0

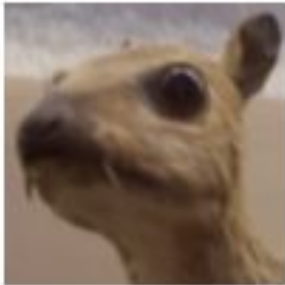
−

•1	•1	•1
•1	•1	•1
•1	•1	•1

$\frac{1}{9}$

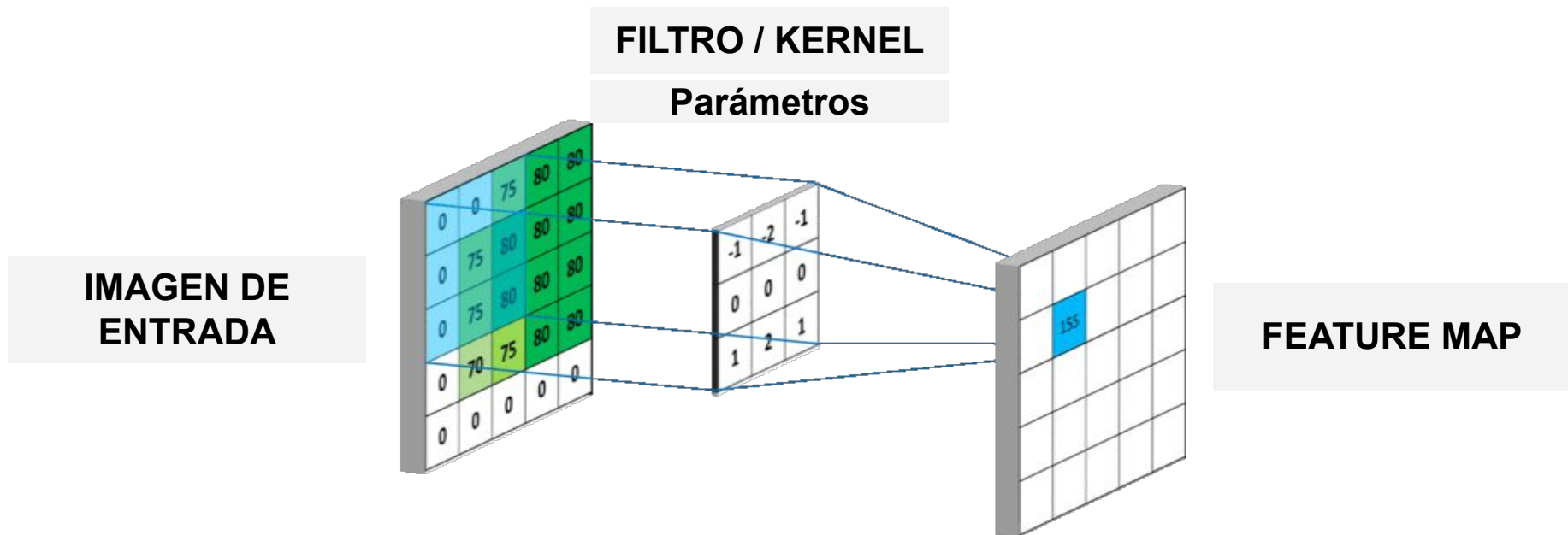


Filtros en las convoluciones

<i>Original</i>	<i>Gaussian Blur</i>	<i>Sharpen</i>	<i>Edge Detection</i>
$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$
			

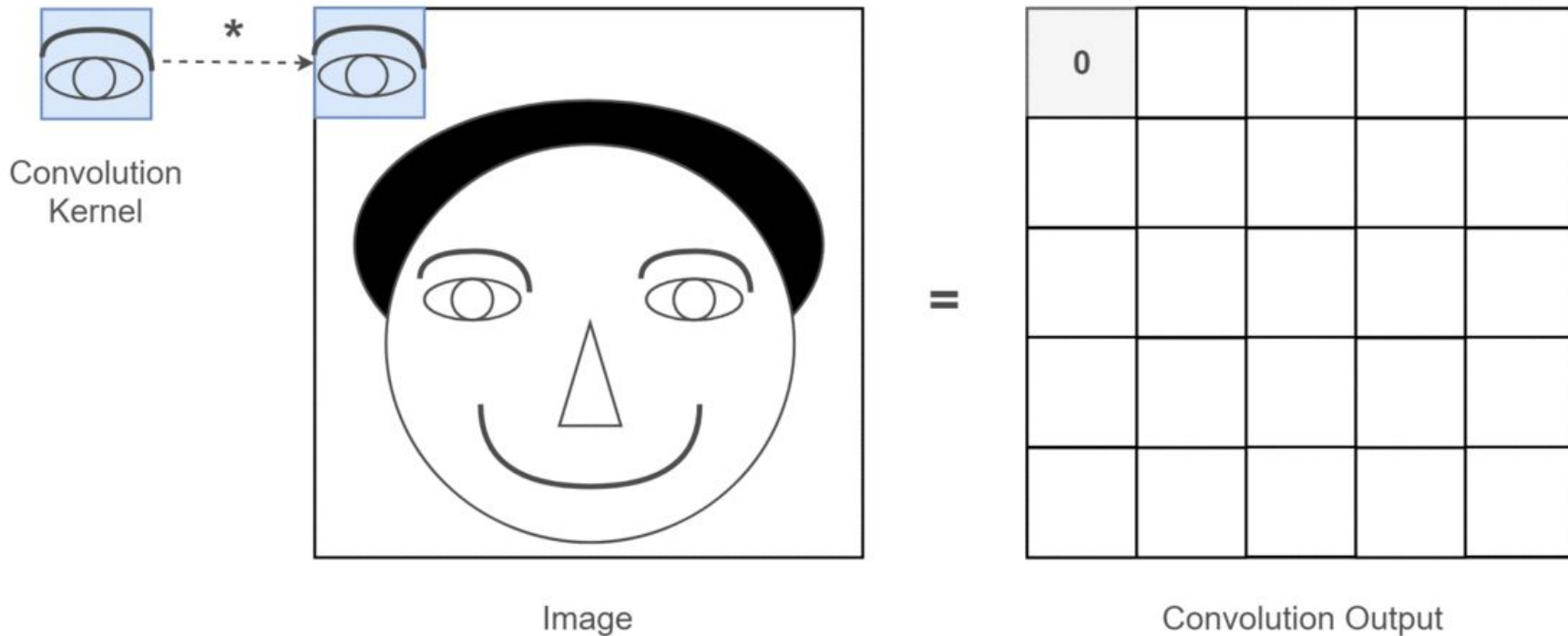


Filtros en las convoluciones





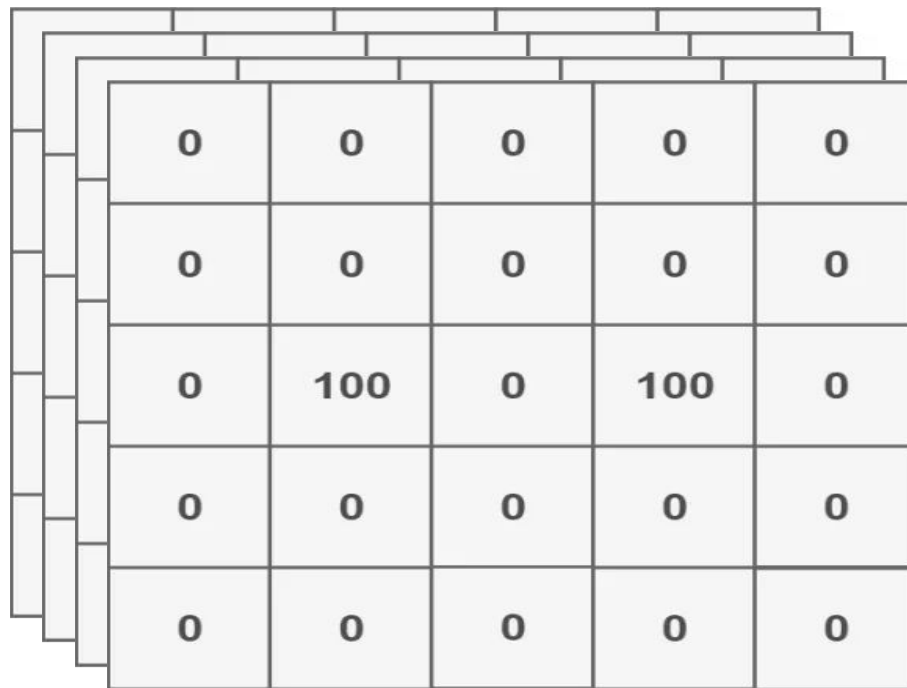
Si queremos detectar un ojo...





Feature Maps

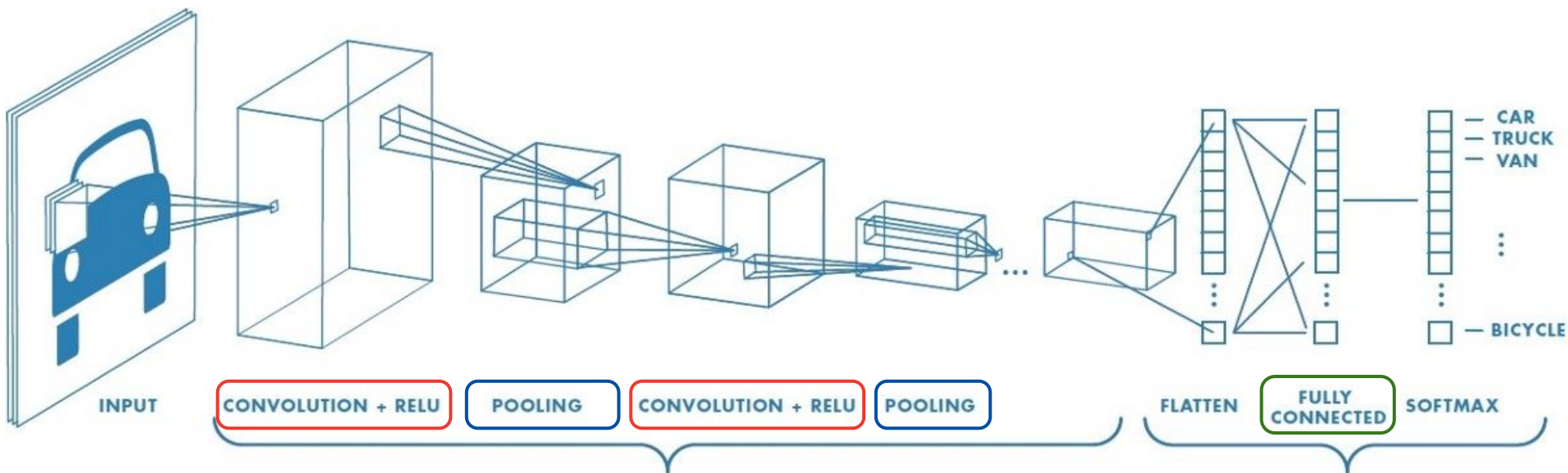
En cada modelo hay varias capas convolucionales, y cada una de ellas puede tener varios filtros, por lo que se generan múltiples Feature Maps



0	0	0	0	0
0	0	0	0	0
0	100	0	100	0
0	0	0	0	0
0	0	0	0	0



Arquitectura típica de una CNN



Capa Convolucional

Capa de Agrupamiento

Capa totalmente conectada



Cofinanciado por
la Unión Europea



GOBIERNO
DE ESPAÑA

MINISTERIO
DE INDUSTRIA
Y TURISMO



Escuela de
organización
industrial



Fondos Europeos





Capa de Agrupamiento (Pooling Layer)

Max Pooling

29	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

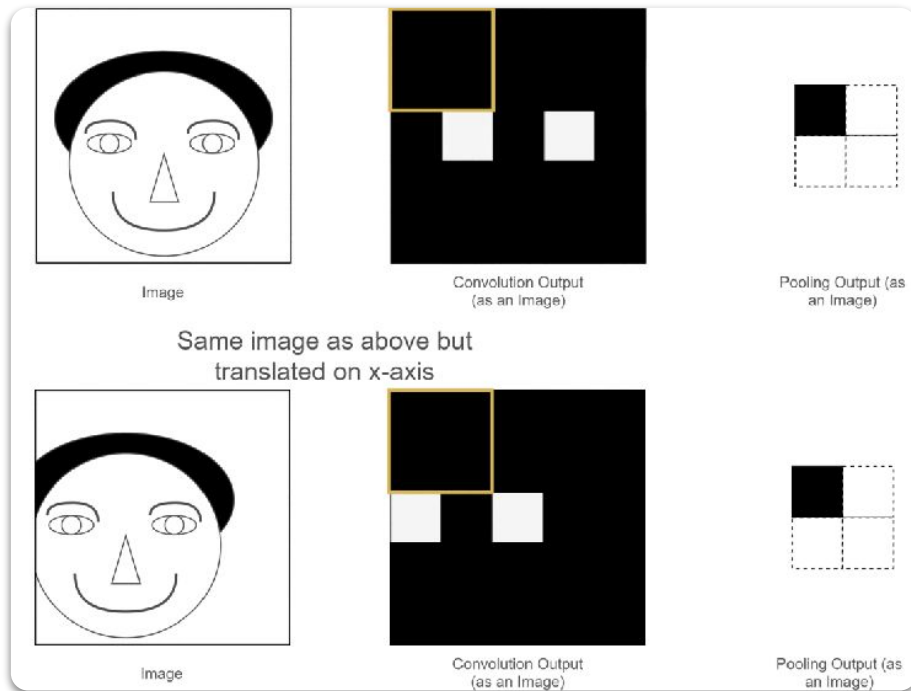
100	184
12	45

Average Pooling

31	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

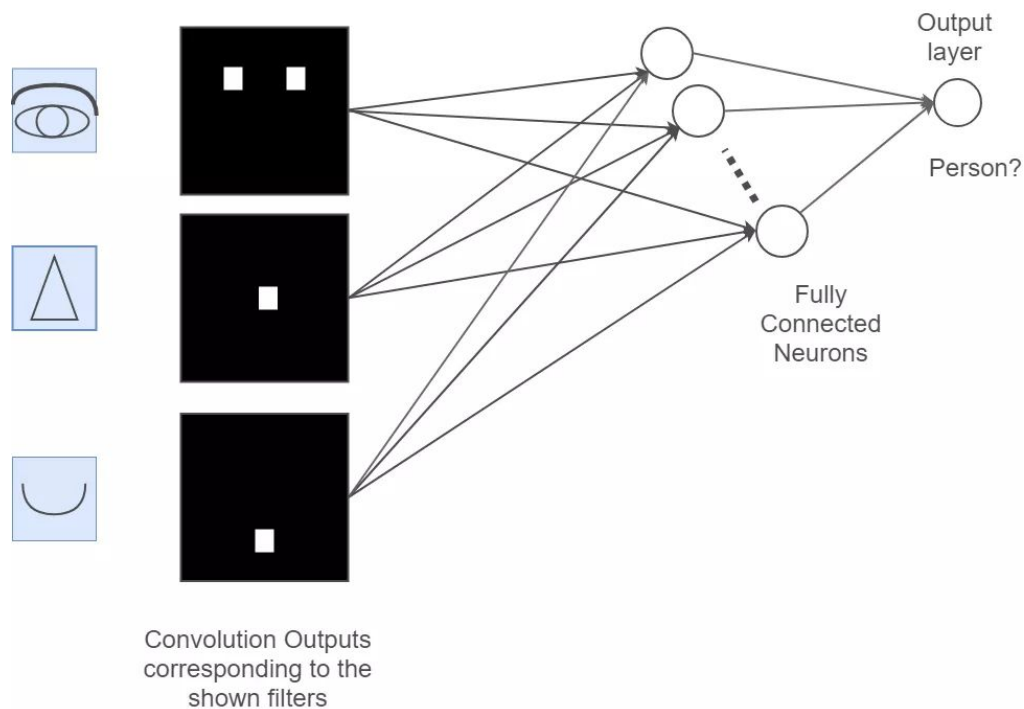
2 x 2
pool size

36	80
12	15



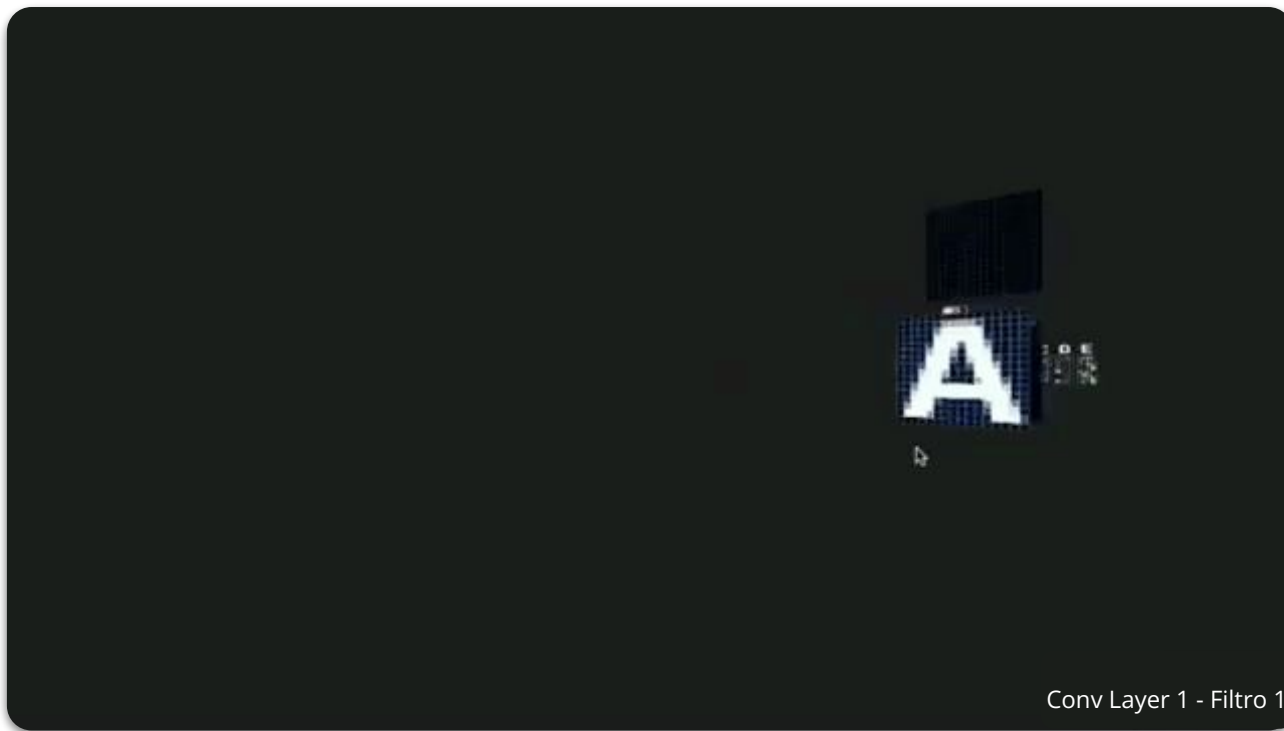


Capa totalmente conectada - Fully Connected



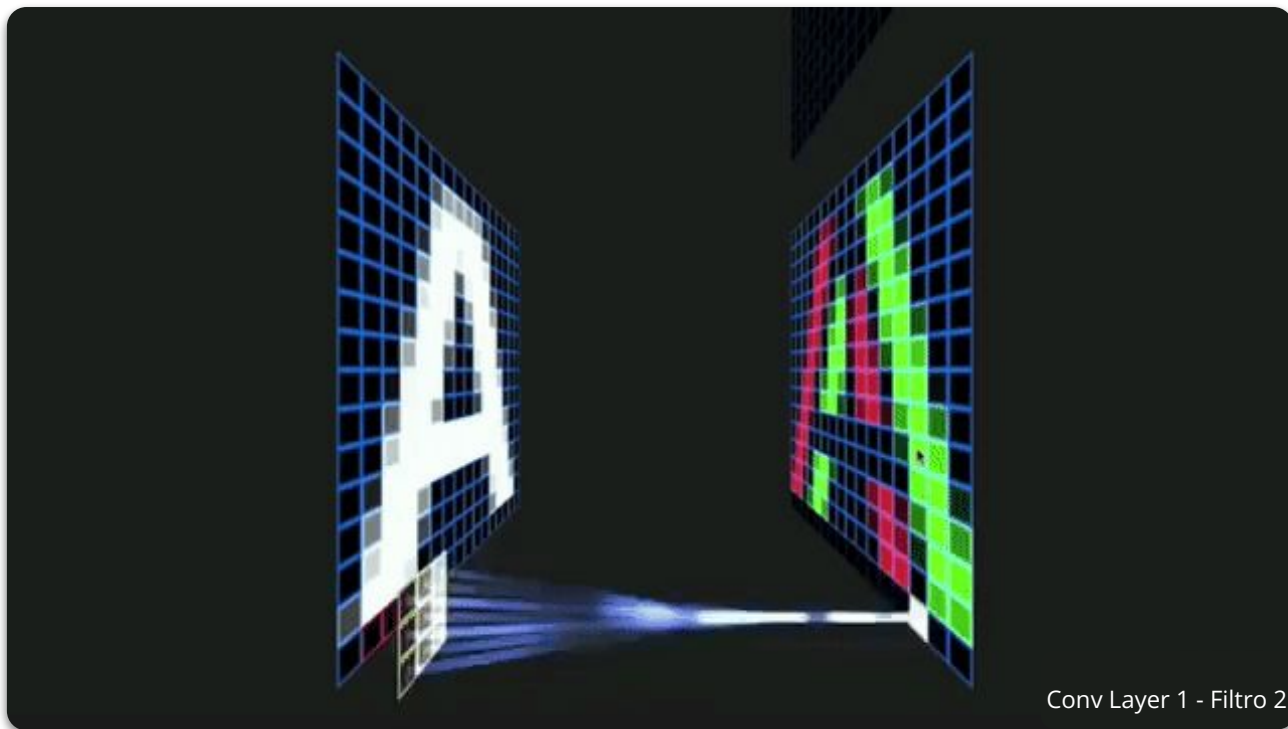


Redes Neuronales - Ejemplo Convolutional



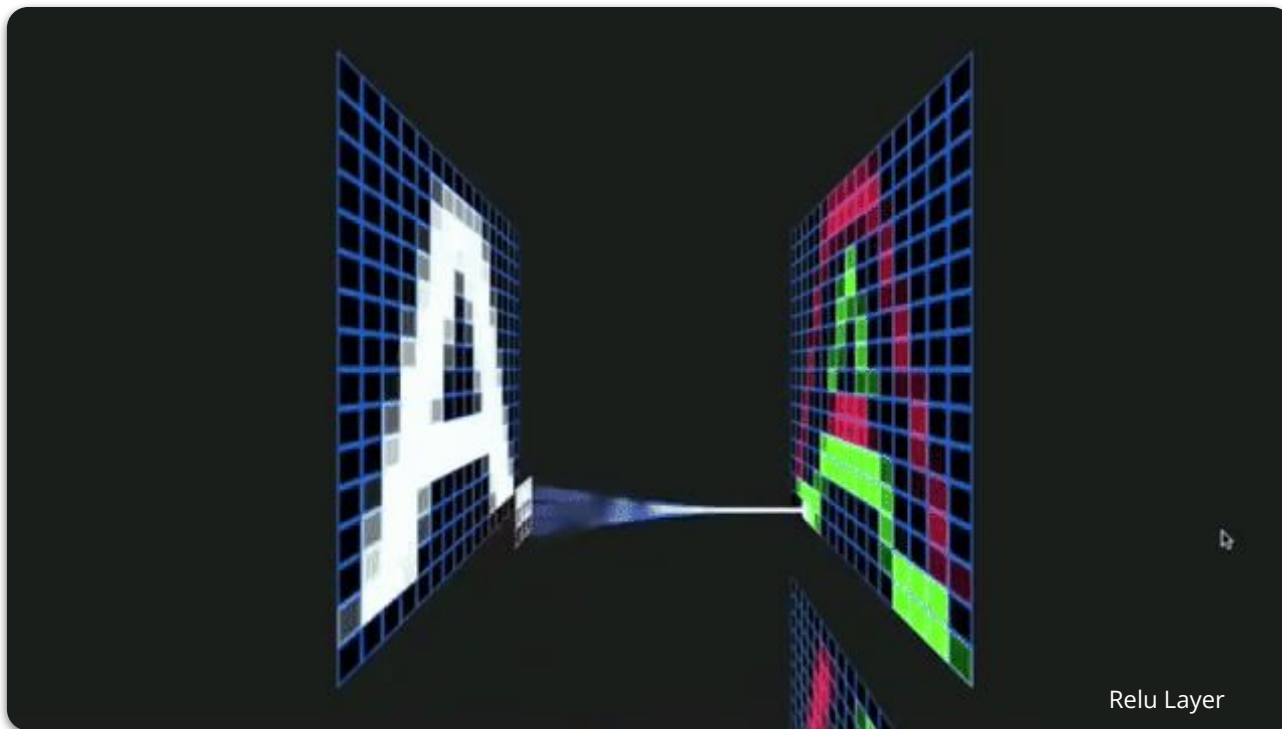


Redes Neuronales - Ejemplo Convolutional





Redes Neuronales - Ejemplo Convolutucional



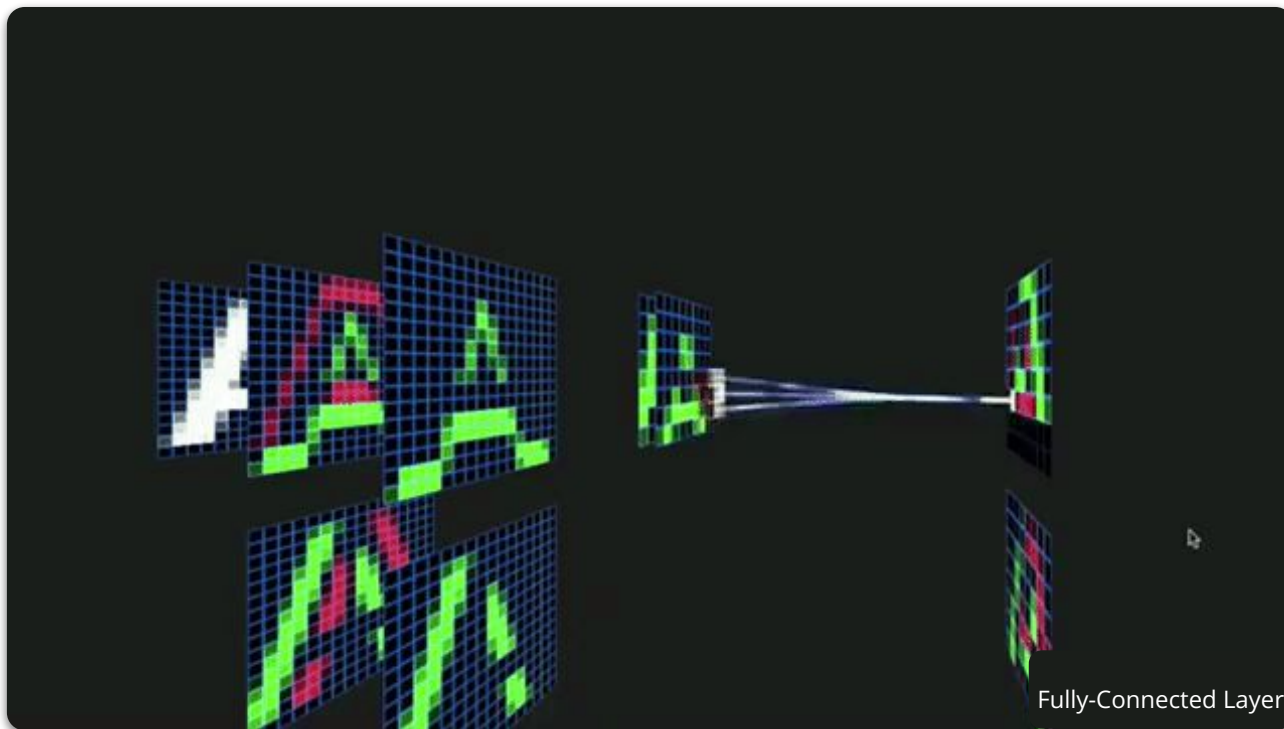


Redes Neuronales - Ejemplo Convolutional



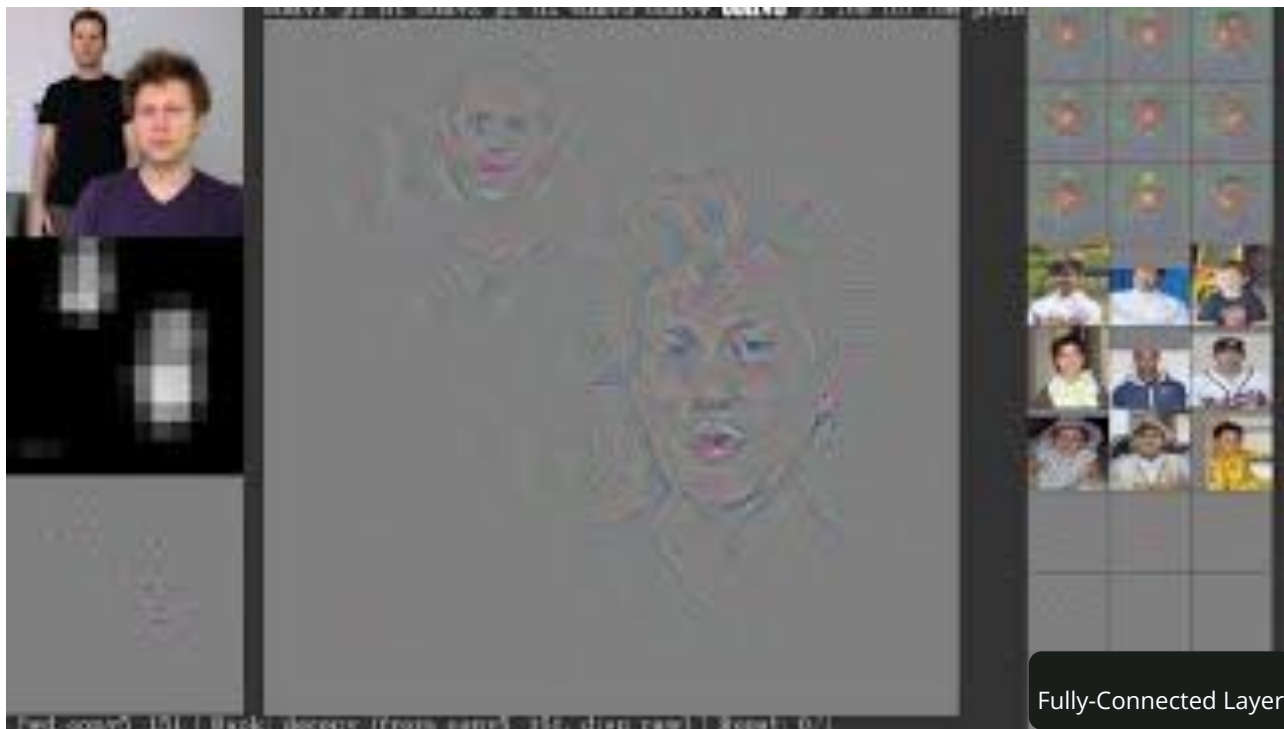


Redes Neuronales - Ejemplo Convolutucional





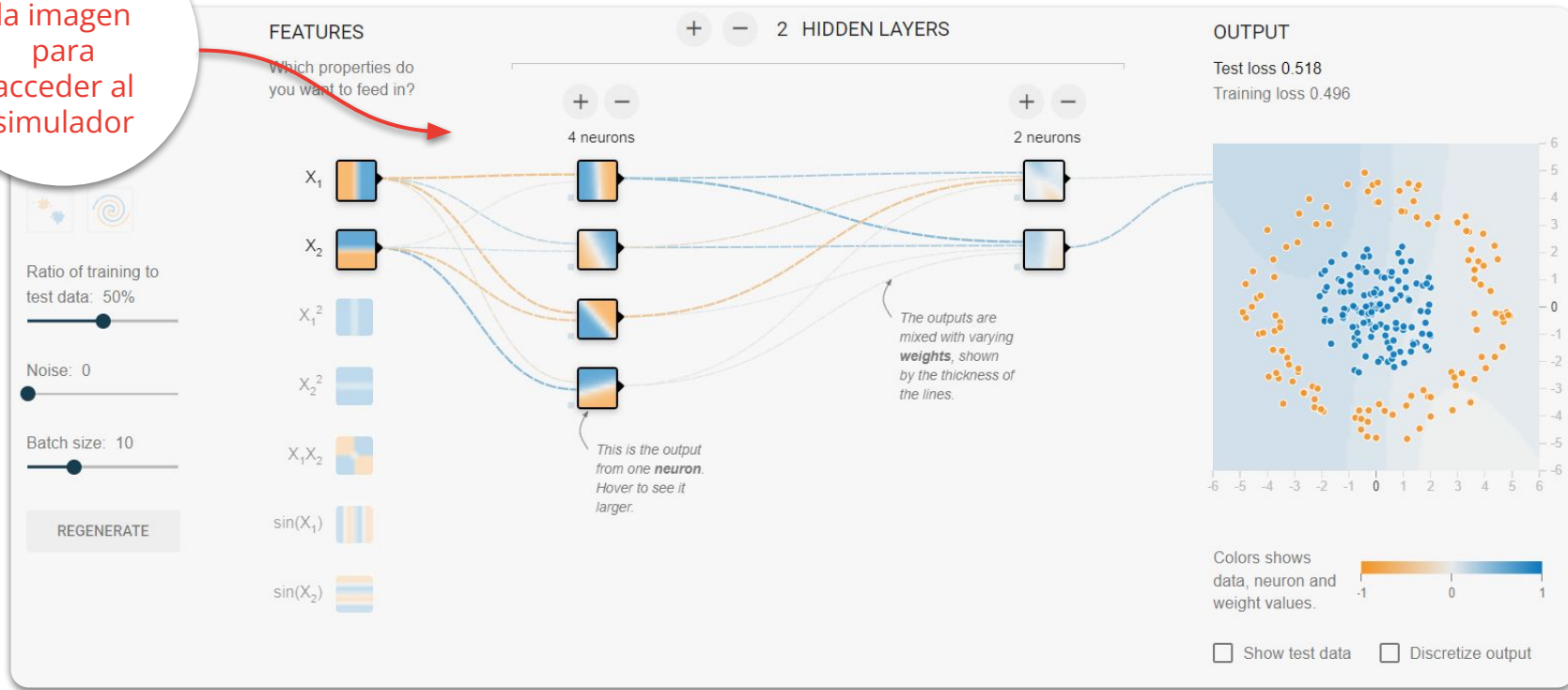
Modelo completo de CNN en acción





Redes Neuronales - Simulador interactivo

Haz clic en la imagen para acceder al simulador





Modelo GANs - Generative Adversarial Networks



Dos redes que compiten entre sí: un generador y un discriminador



Cofinanciado por
la Unión Europea



MINISTERIO
DE INDUSTRIA
Y TURISMO

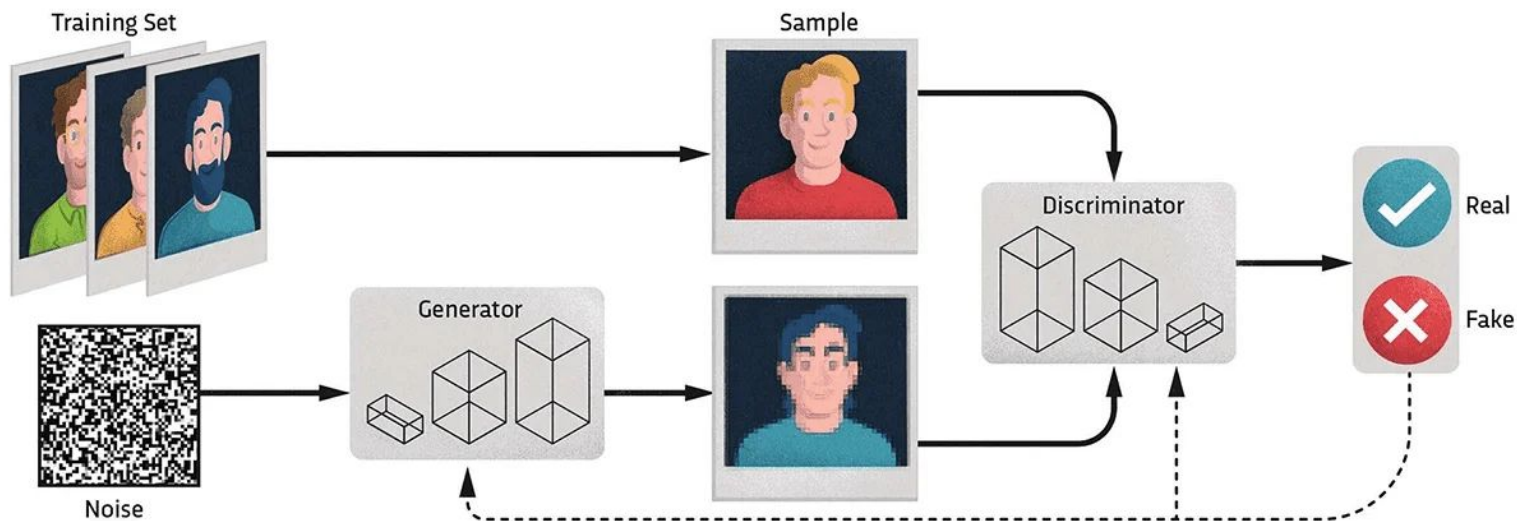


Fondos Europeos





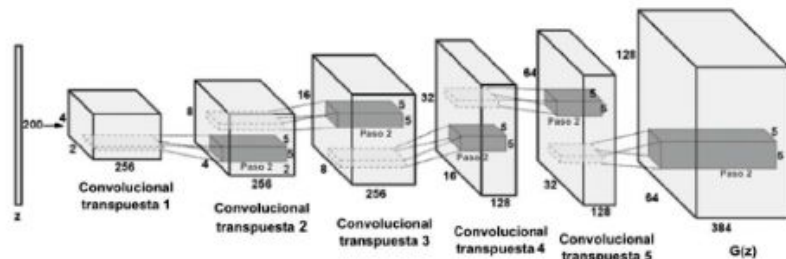
Modelo GANs - Generative Adversarial Networks



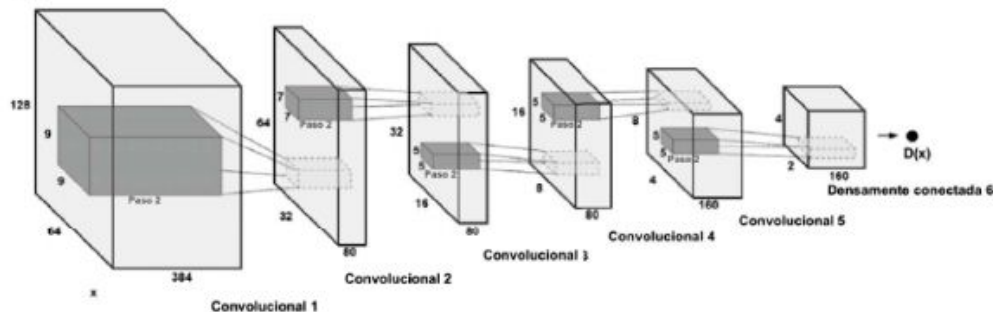


Arquitectura de las GANs

(a)

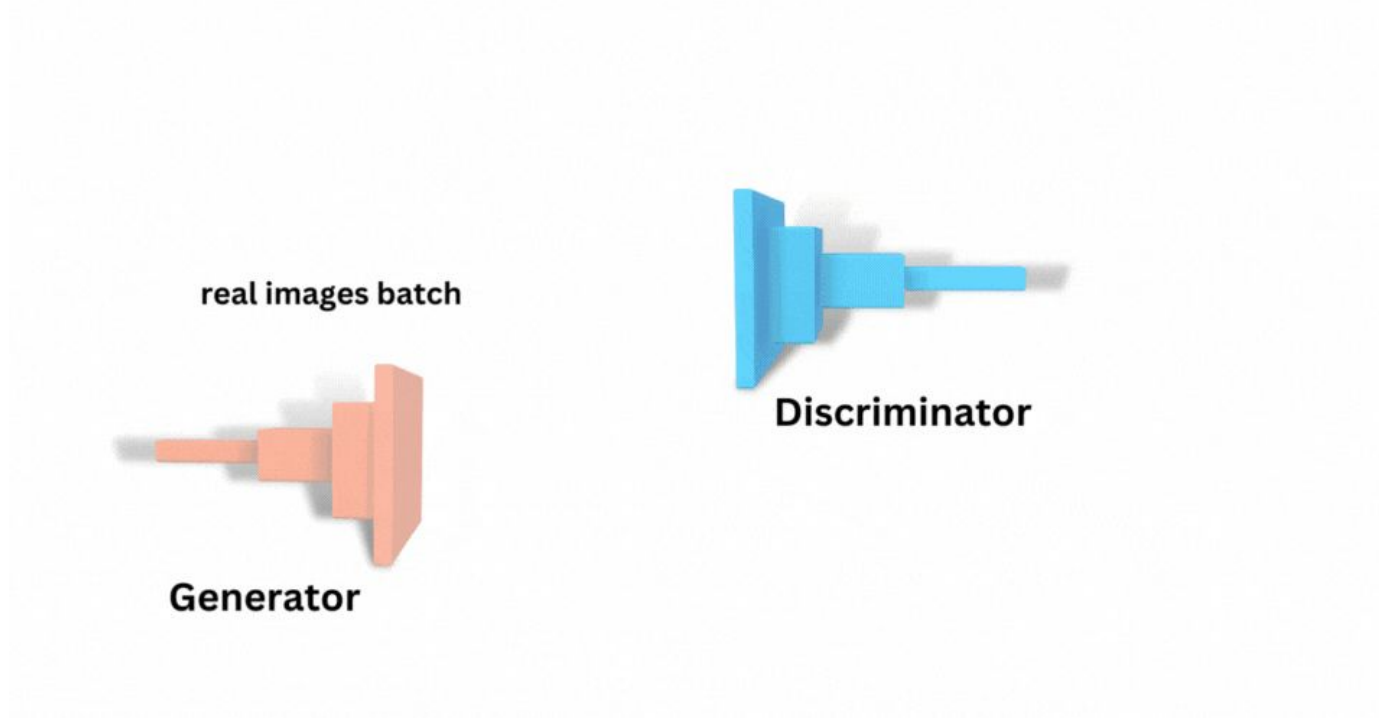


(b)





Proceso de entrenamiento de las GANs





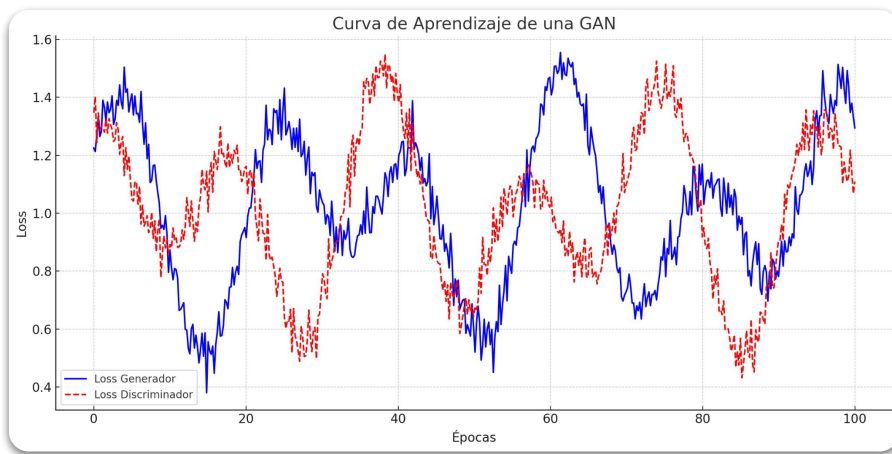
Aplicaciones de las GANs





Limitaciones y desafíos de las GANs

- **Inestabilidad en el entrenamiento:** Puede fallar si el generador o el discriminador se vuelven demasiado buenos.
- **Requerimiento de datos masivos:** Necesita grandes conjuntos de datos para entrenar.
- **Cuestiones éticas:** Uso indebido para crear contenido falso (deepfakes).





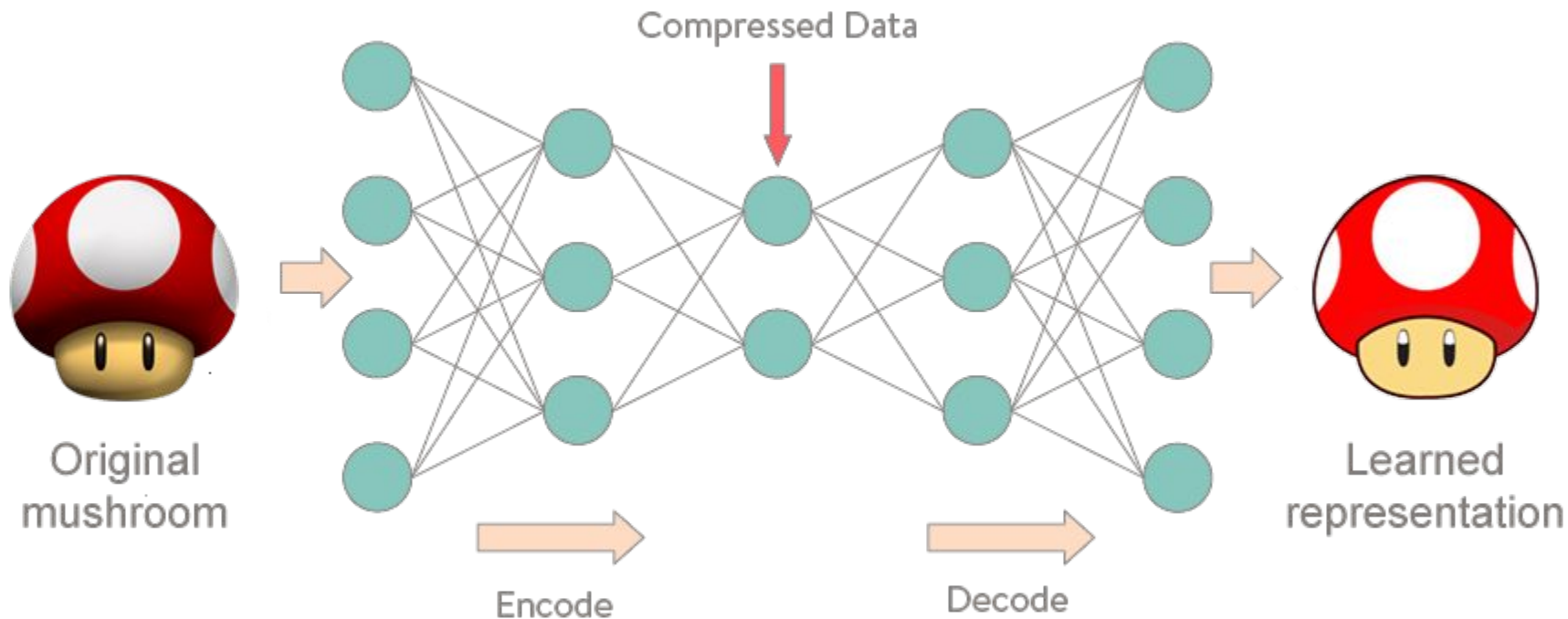
Futuro de las GANs

- Mayor estabilidad en el entrenamiento.
- Aplicaciones en la Realidad Aumentada y Virtual.
- Fusión con otras tecnologías de IA.



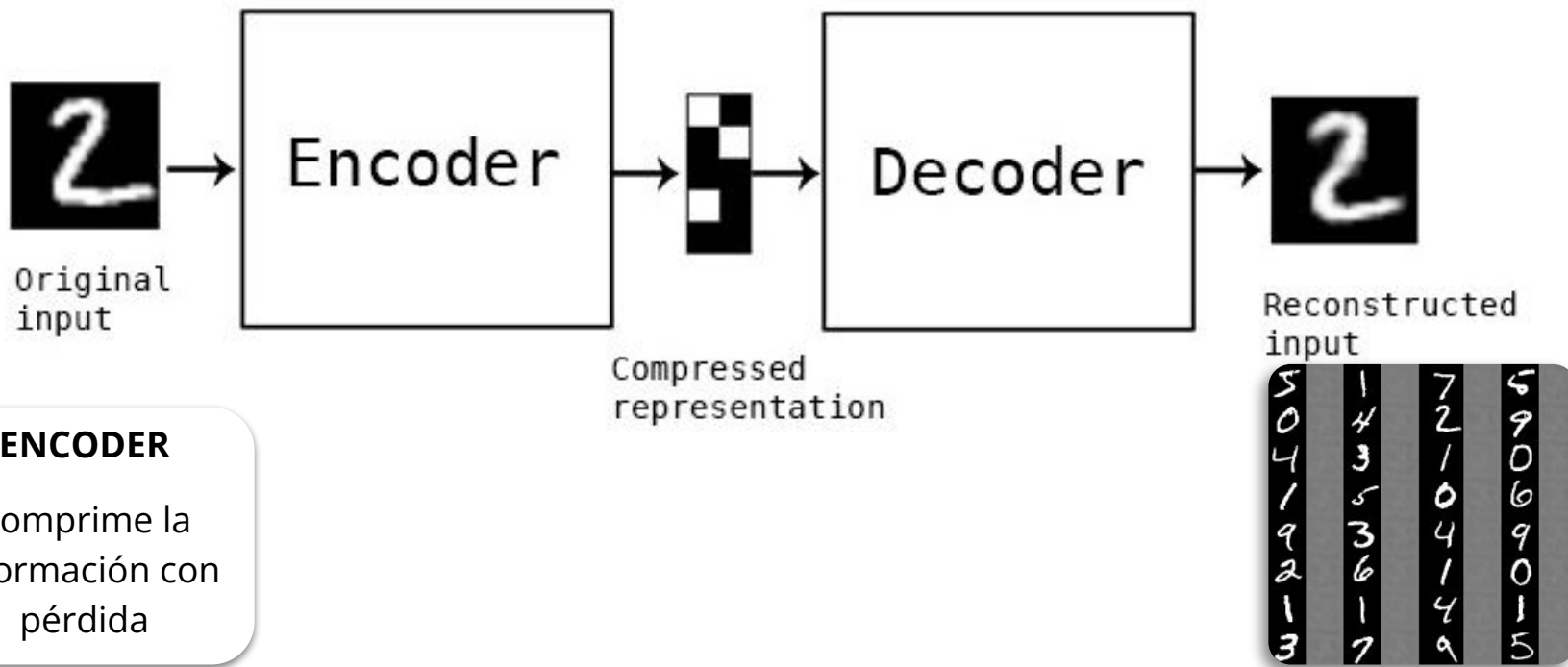


Autoencoder



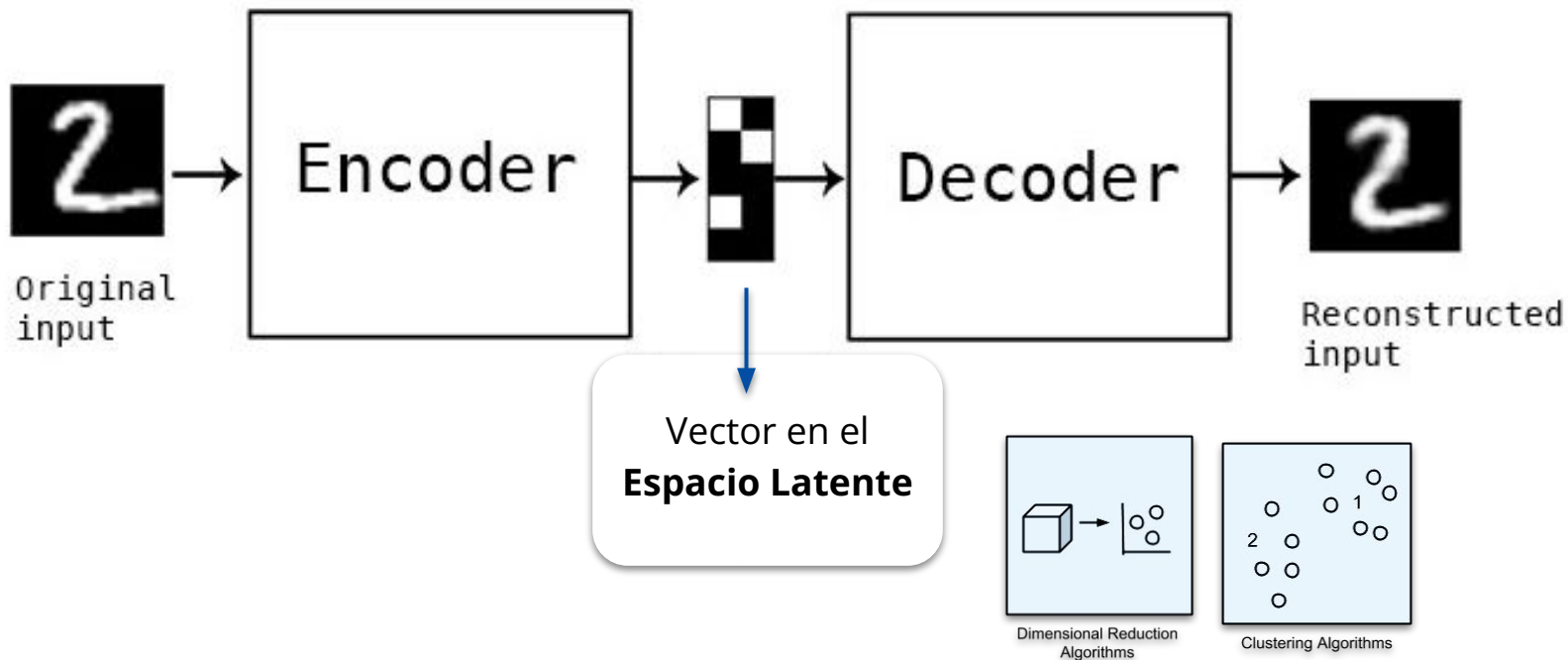


Estructura Encoder y Decoder



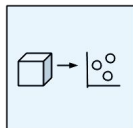


Estructura Encoder y Decoder

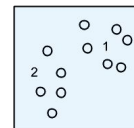




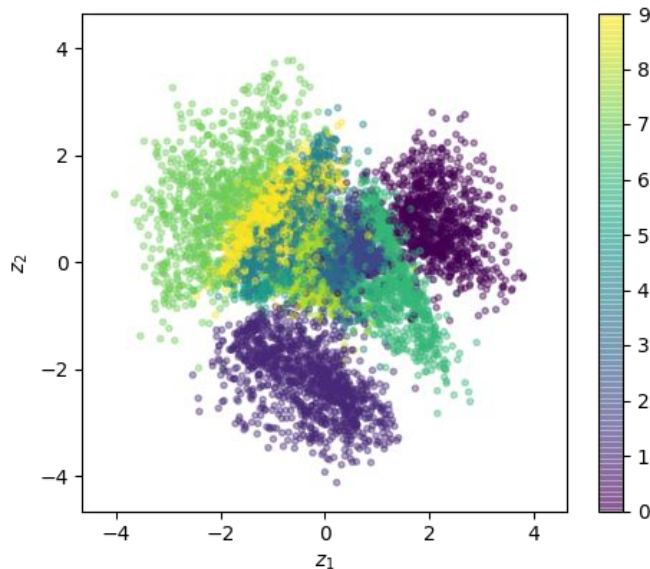
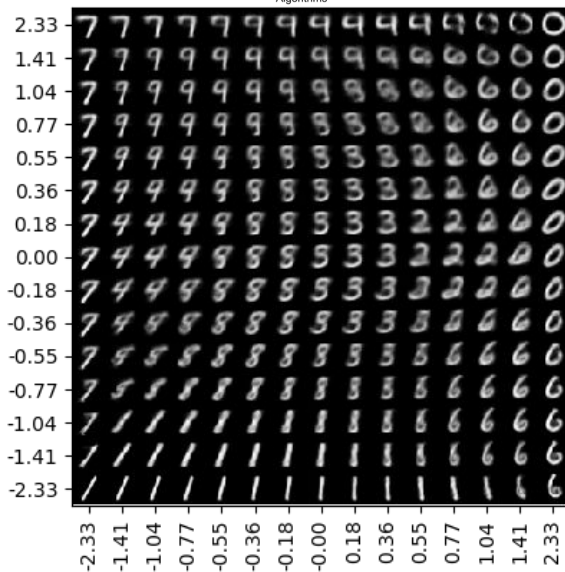
Estructura Encoder y Decoder



Dimensional Reduction Algorithms



Clustering Algorithms



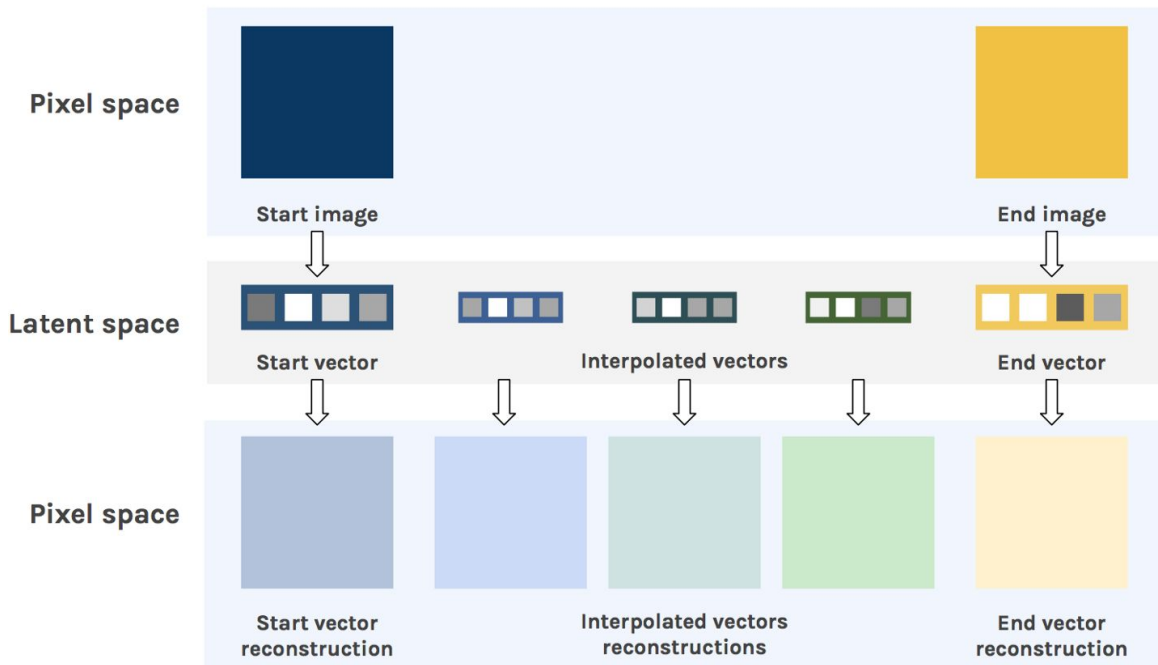


Interpolación



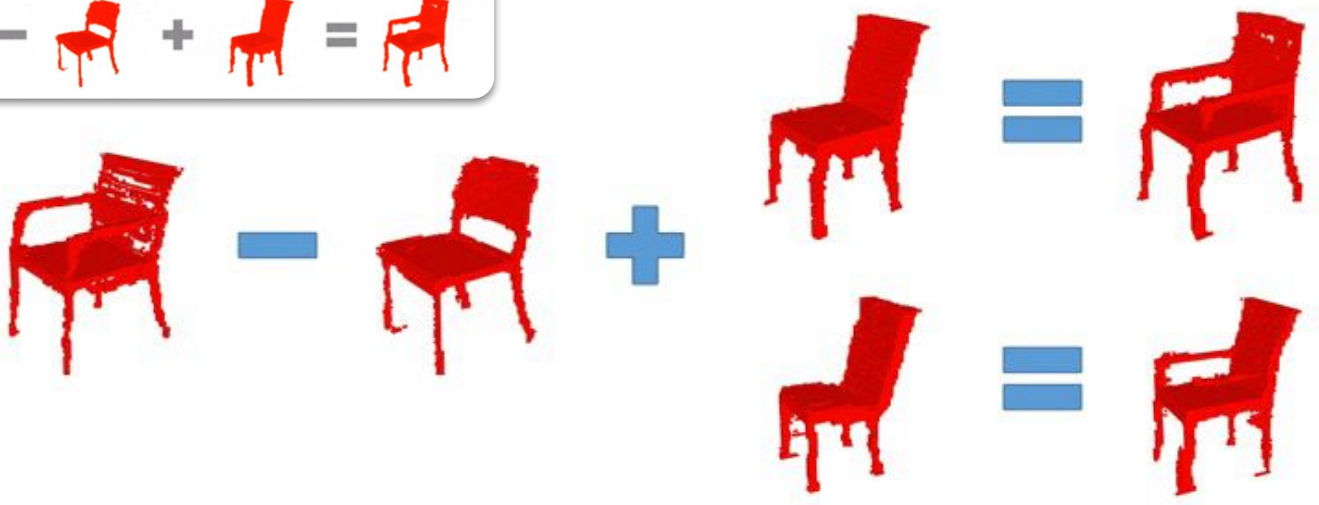
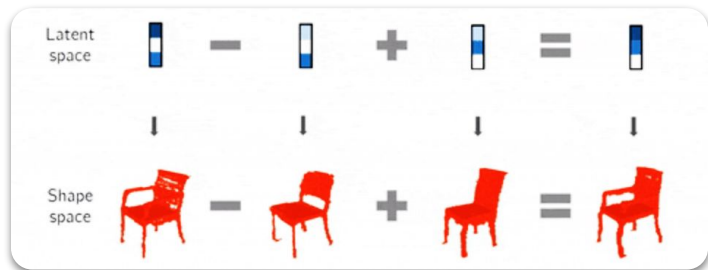


Interpolación



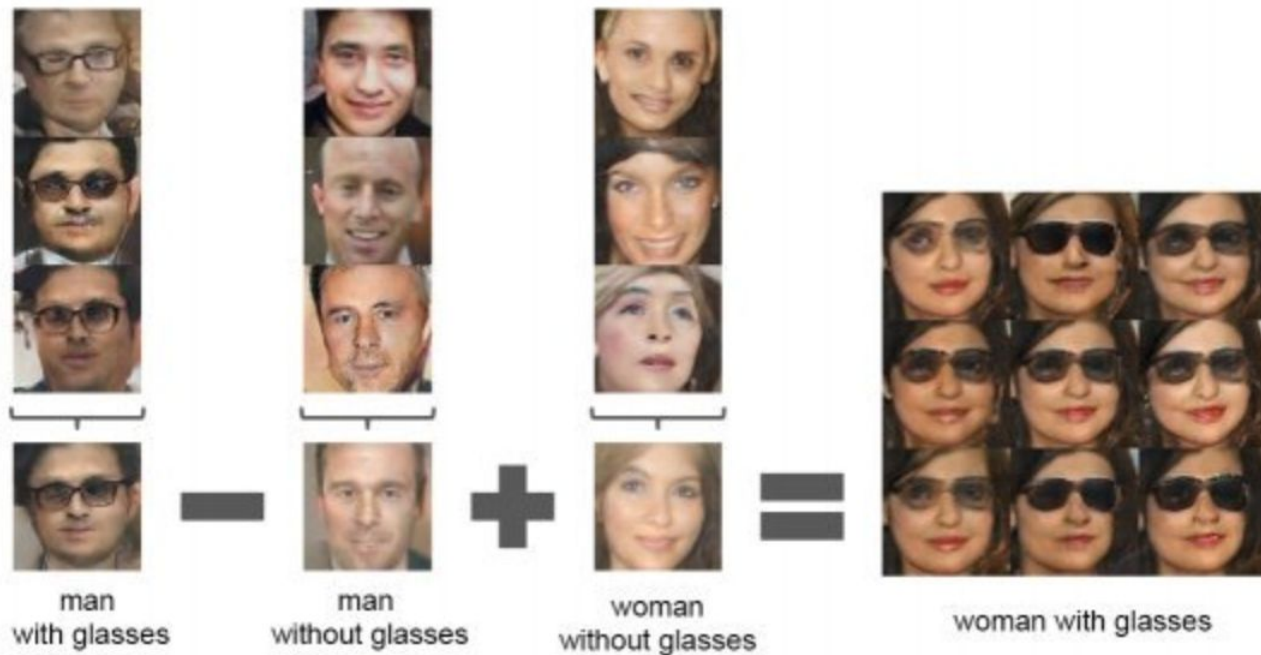


Álgebra



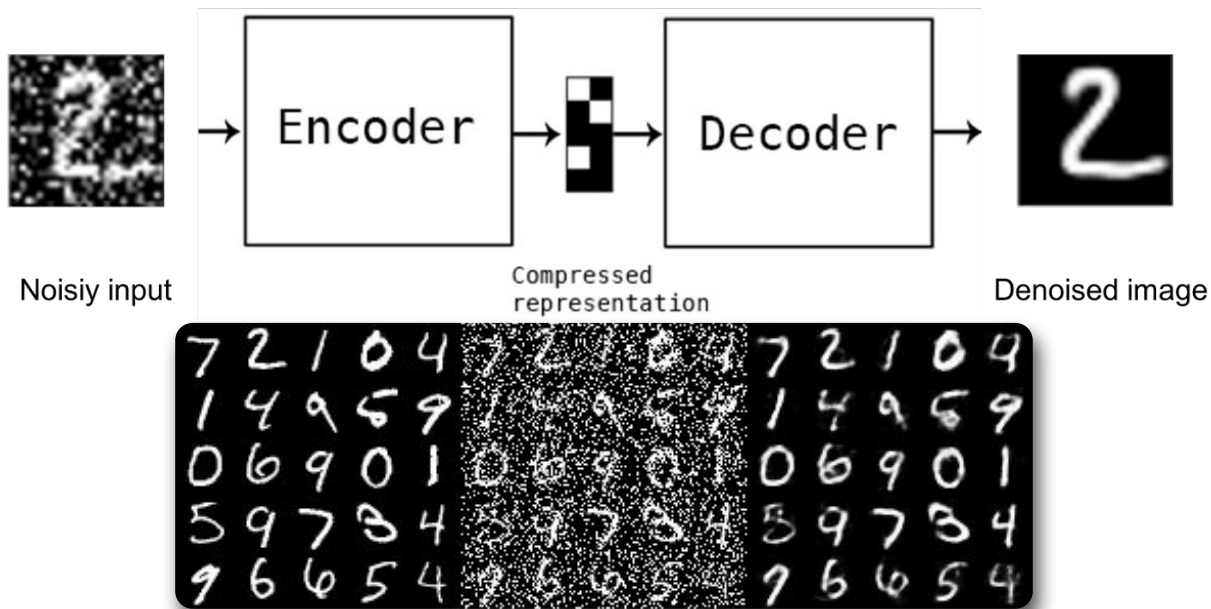


Álgebra





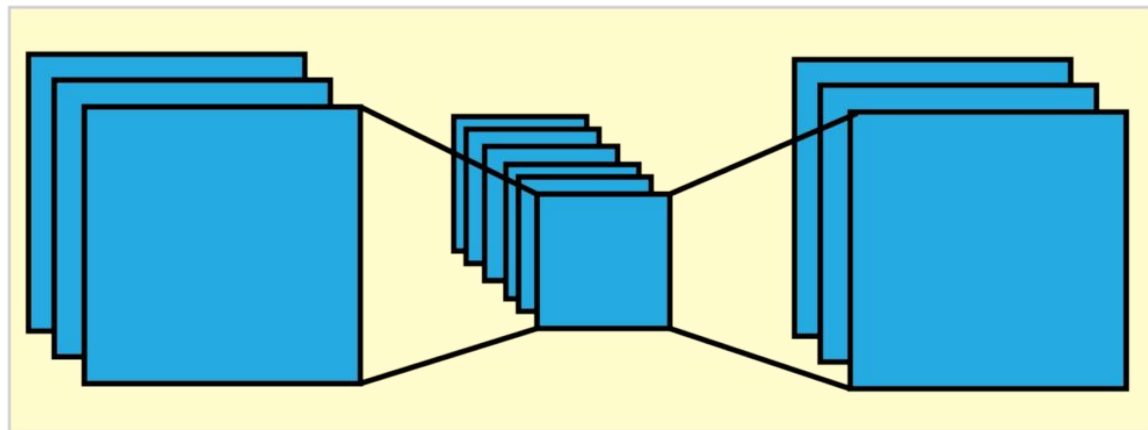
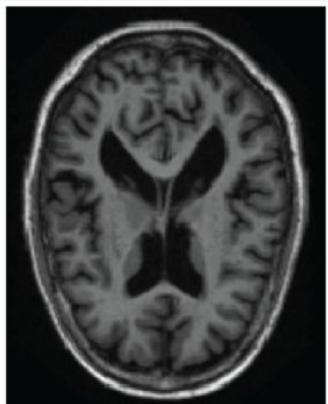
Denoiser Autoencoder





Denoiser Autoencoder

NOISY
INPUT



COMPRESSION/
CLEANING

DECOMPRESSION

DENOISED
OUTPUT



DENOISER AUTOENCODER



Original

Blender Cycles
10 Samples, 5 sec

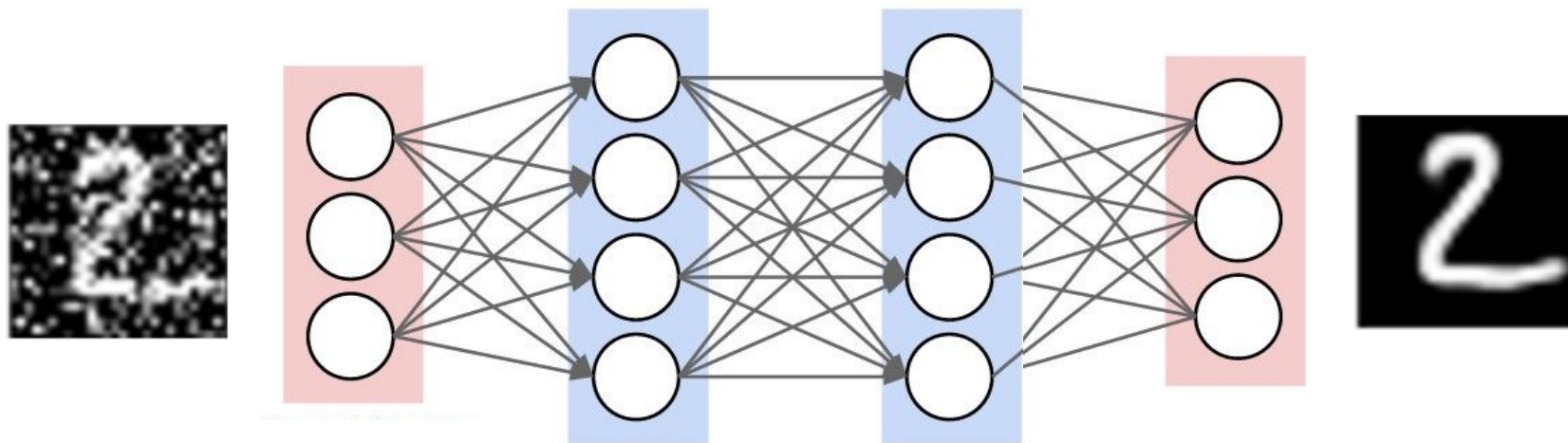
DENOISER AUTOENCODER

Raw



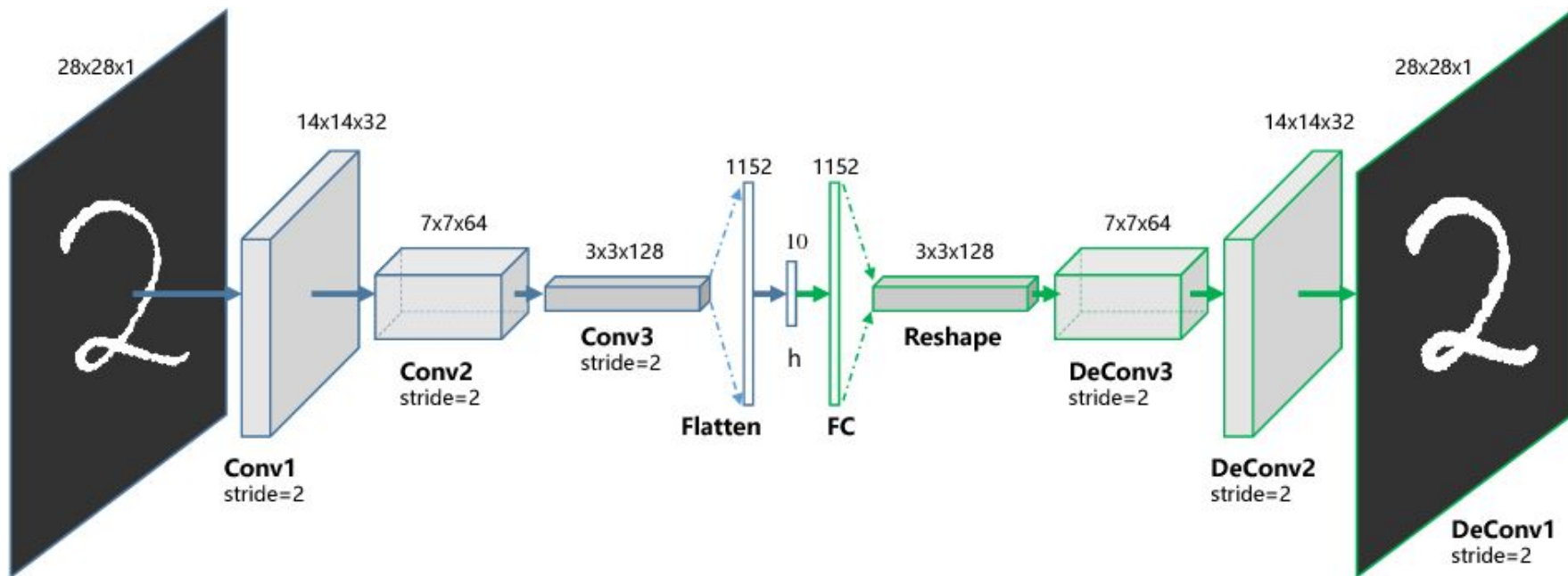


Denoiser Autoencoder



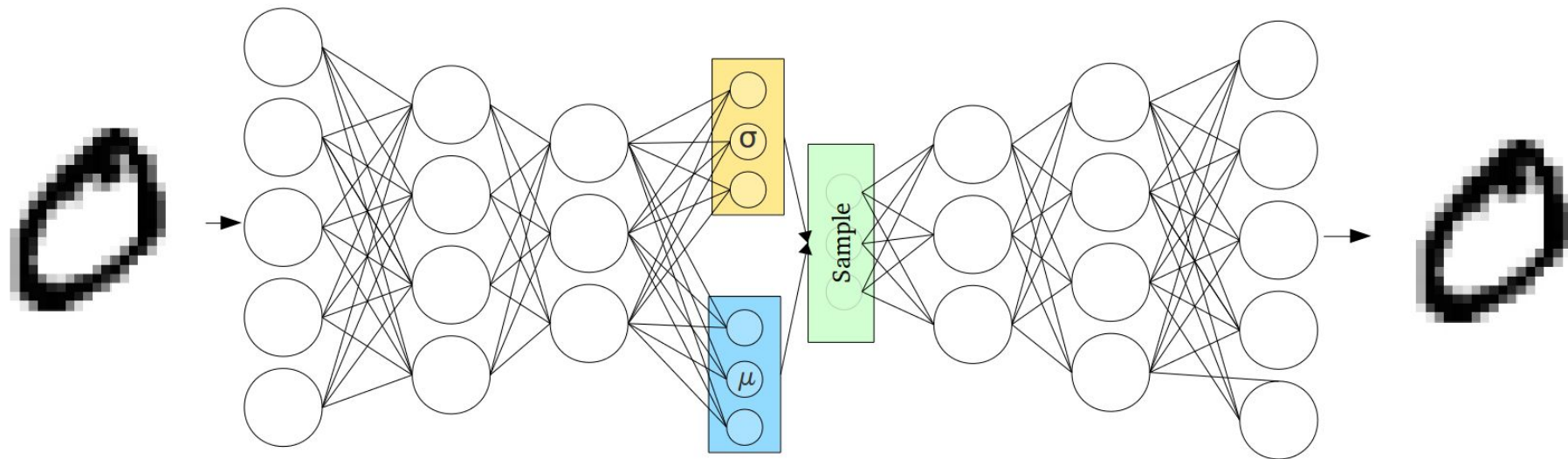


Convolutional Autoencoder





Variational Autoencoder (VAE)



Understanding Variational Autoencoders (VAEs)

<https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>



Cofinanciado por
la Unión Europea



GOBIERNO
DE ESPAÑA

MINISTERIO
DE INDUSTRIA
Y TURISMO

EOI Escuela de
organización
industrial



Fondos Europeos

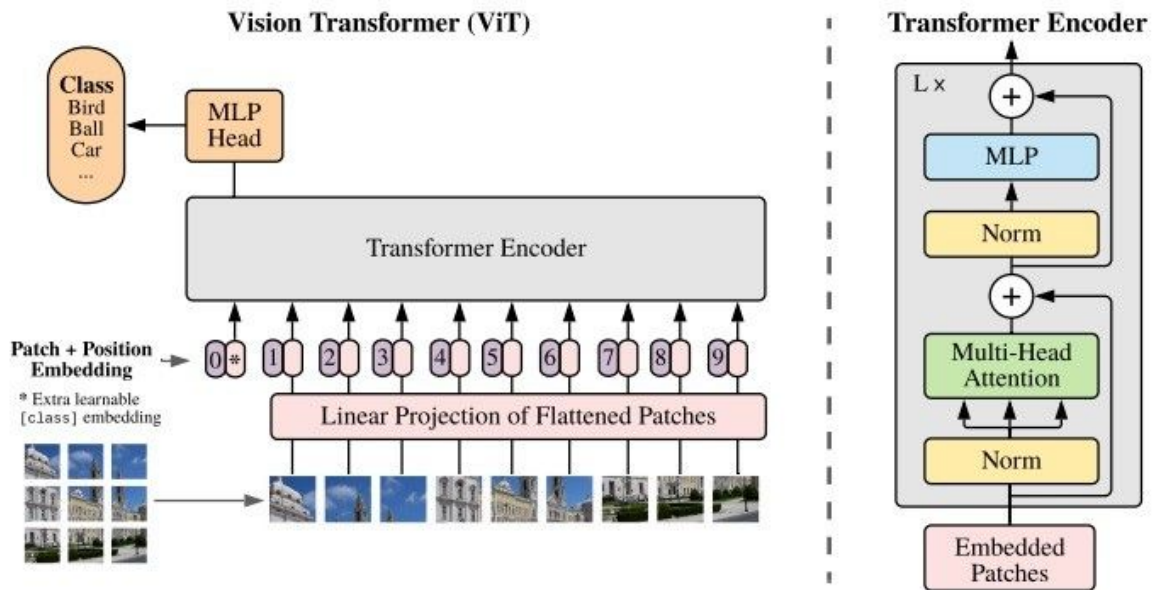




Variational Autoencoder (VAE)



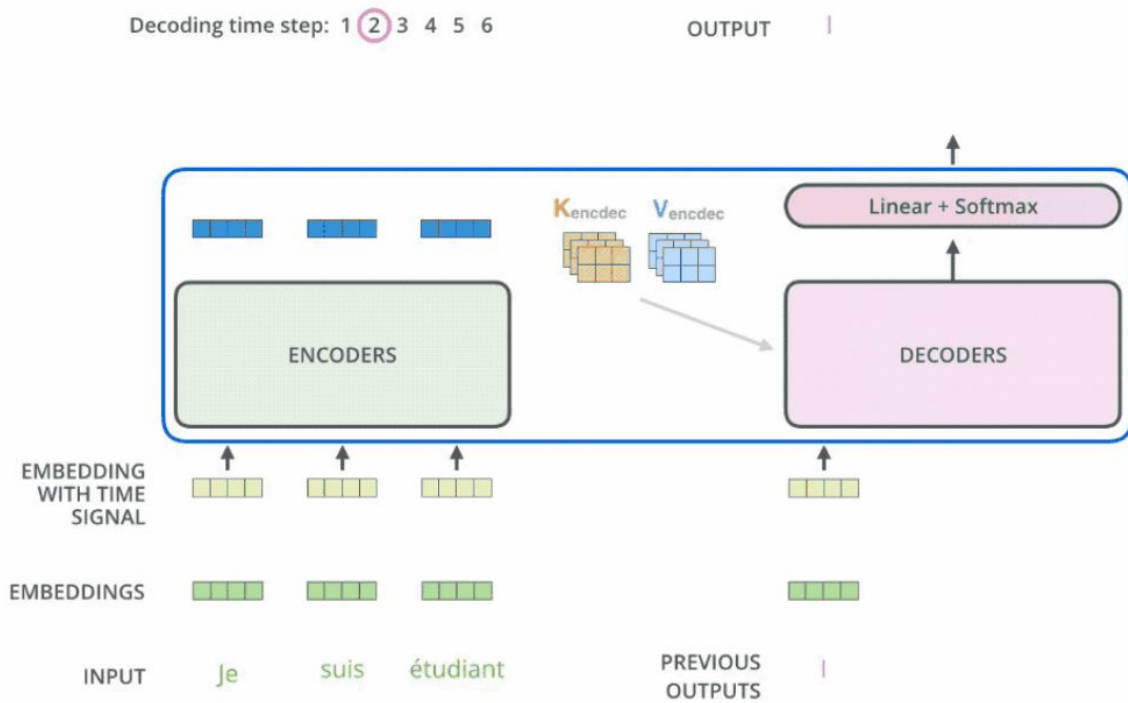
Modelo TRANSFORMERS



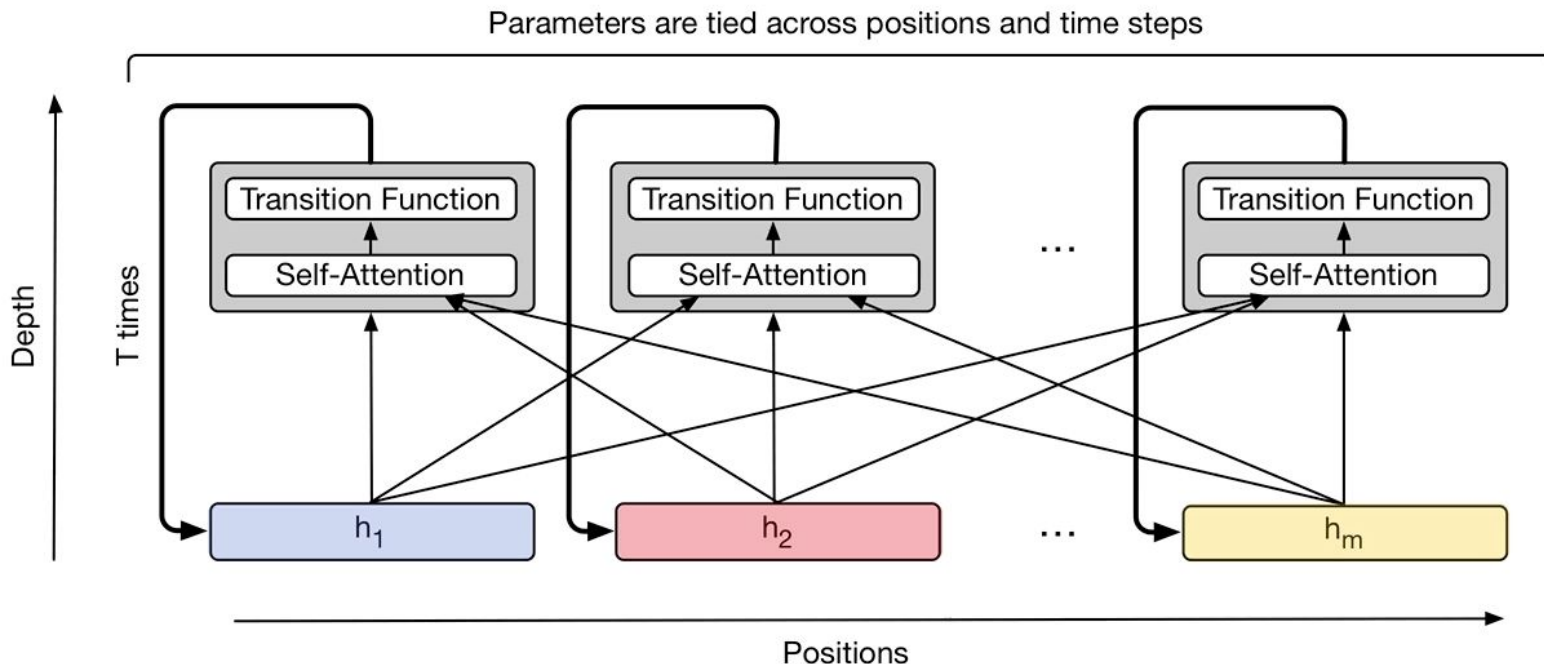
Basados en la atención que procesa secuencias de datos, como texto o imágenes



¿Cómo funcionan los Transformers?



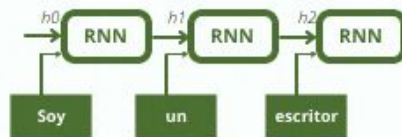
Arquitectura de los Transformers





Mecanismo de atención

$$\text{Atención} = \text{Softmax}(QK^T / \sqrt{d}) V$$



Codificador

Decodificador





Aplicación de los Transformers

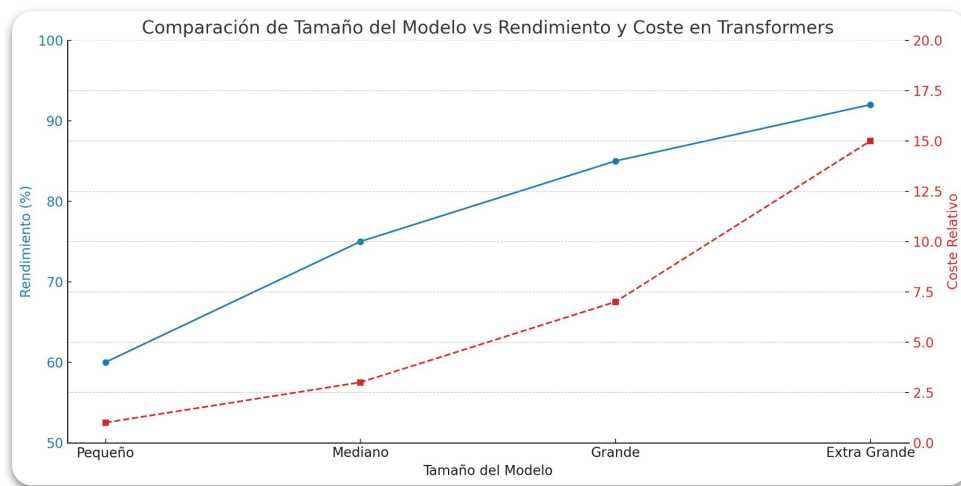
- **Traducción automática:** Convertir texto de un idioma a otro.
- **Generación de texto:** Chatbots, resúmenes automáticos.
- **Reconocimiento de voz:** Convertir voz a texto.
- **Generación de imágenes:** Creación de imágenes a partir de descripciones.
- **Predicción de estructuras de proteínas:** Uso en biología computacional.





Limitaciones y desafíos de los Transformers

- **Consumo de recursos:** Entrenar modelos grandes es costoso.
- **Dependencia de grandes conjuntos de datos:** Necesitan grandes volúmenes de datos etiquetados.
- **Cuestiones éticas:** Sesgos inherentes en los datos pueden afectar las predicciones.





El futuro de los Transformers

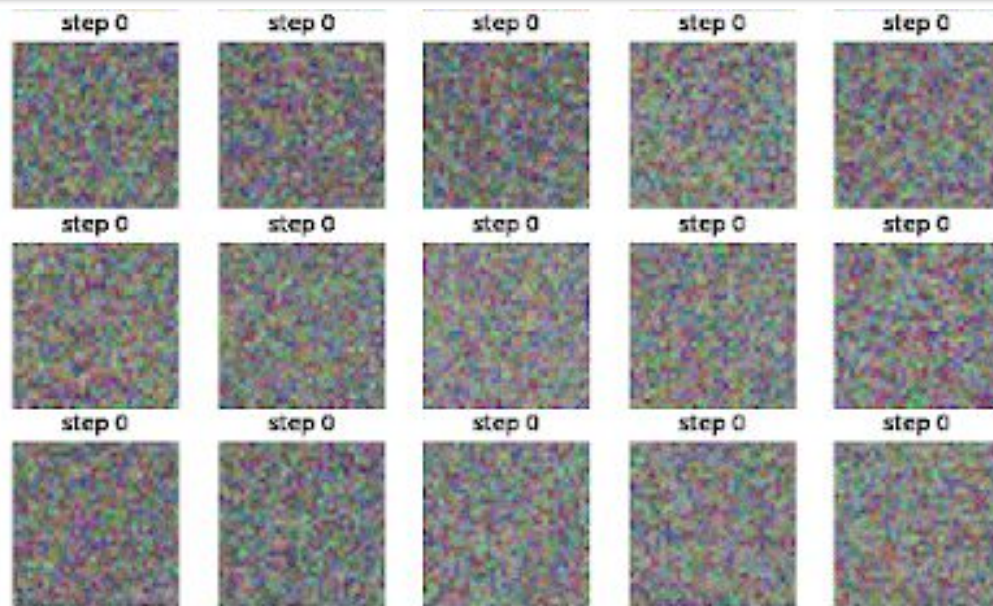
- **Modelos más pequeños y eficientes:** Uso optimizado de recursos.
- **Transformers multimodales:** Procesan texto, imágenes y otros datos simultáneamente.
- **Integración con hardware especializado:** Mejor rendimiento con aceleradores.





Modelo de DIFUSIÓN - Diffusion Models

Modelos probabilísticos que transforman una muestra de ruido aleatorio en una imagen o dato estructurado.





¿Cómo funcionan los Modelos de Difusión?

Difusión directa



$p(\mathbf{x})$
def
 $p_0(\mathbf{x})$



$\pi(\mathbf{x})$
def
 $p_1(\mathbf{x})$



Difusión inversa





Matemáticas detrás del Modelo de Difusión

Ecuación de
difusión directa:

$$q(x_t | x_{t-1})$$

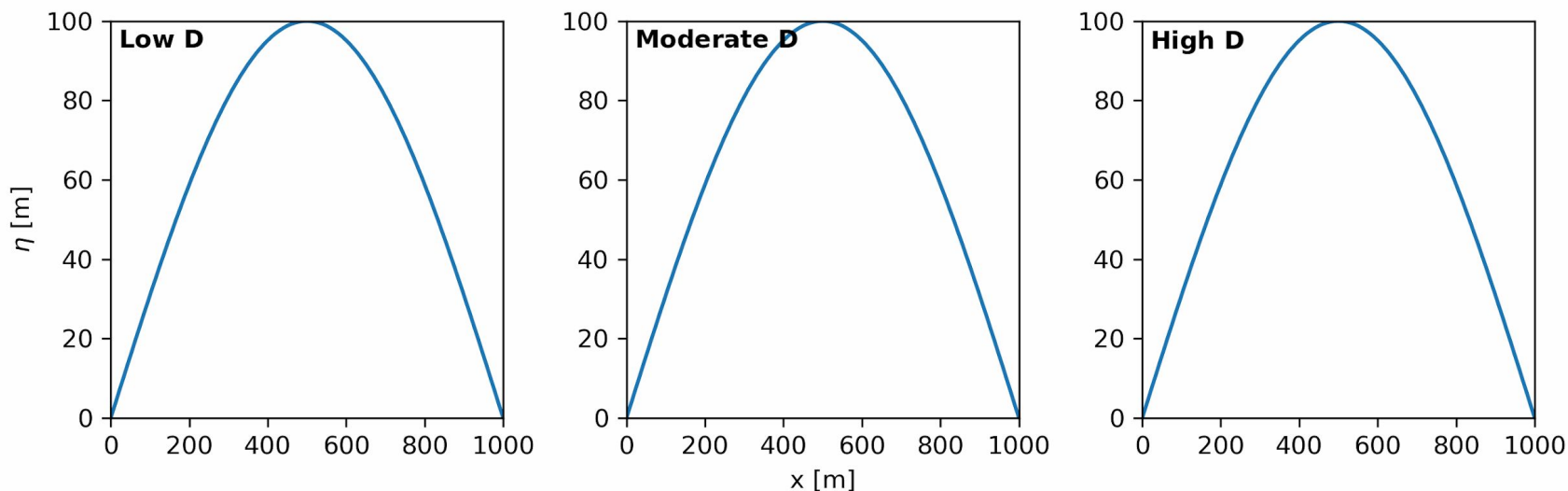
Ecuación de
difusión inversa:

$$p(x_{t-1} | x_t)$$

Variable	Símbolo	Descripción
Imagen original	x_0	La imagen inicial libre de ruido, que se desea reconstruir mediante el proceso de difusión
Imagen ruidosa	x_t	Imagen en el paso t del proceso de difusión, con un nivel específico de ruido añadido
Paso temporal	t	Índice del paso en el proceso de ruido (generalmente de 0 a T), donde T es el número total de pasos
Distribución de ruido	$(q(x_t x_{t-1}))$	
Distribución inversa	$(p(x_{t-1} x_t))$	
Tasa de ruido	β_t	Parámetro que controla la cantidad de ruido añadido en cada paso t
Parámetro de difusión	α_t	Factor acumulativo que representa la proporción de la imagen original que se conserva tras t pasos
Latente inicial	z	Muestra de ruido aleatorio que se utiliza como punto de partida para el proceso inverso de generación



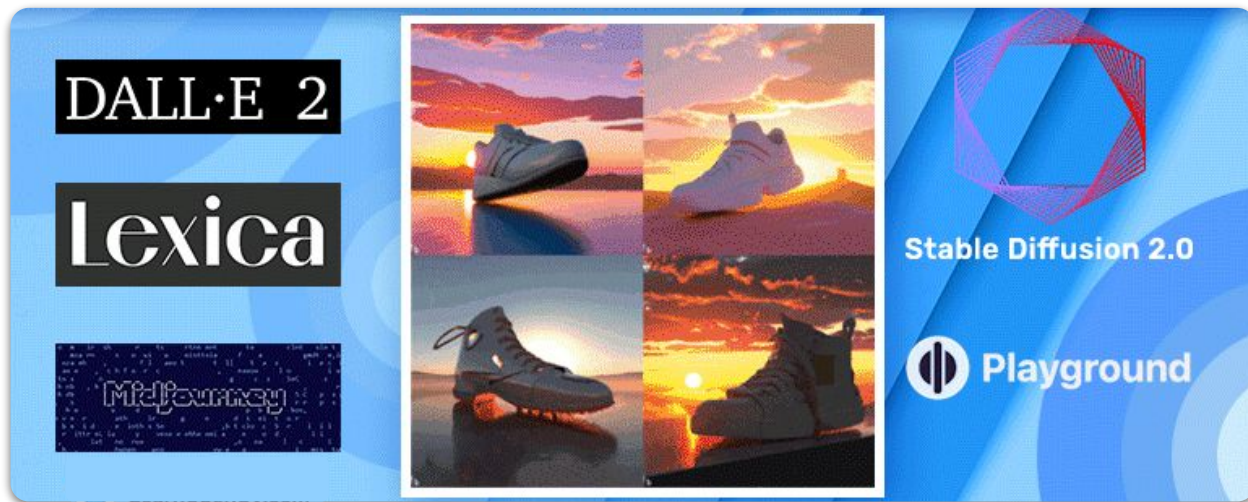
Entrenamiento de un Modelo de Difusión





Aplicaciones de los Modelos de Difusión

- **Generación de imágenes:** Creación de imágenes realistas a partir de descripciones.
- **Edición de imágenes:** Eliminar o agregar elementos a imágenes existentes.
- **Generación de audio:** Crear sonido o música realista.



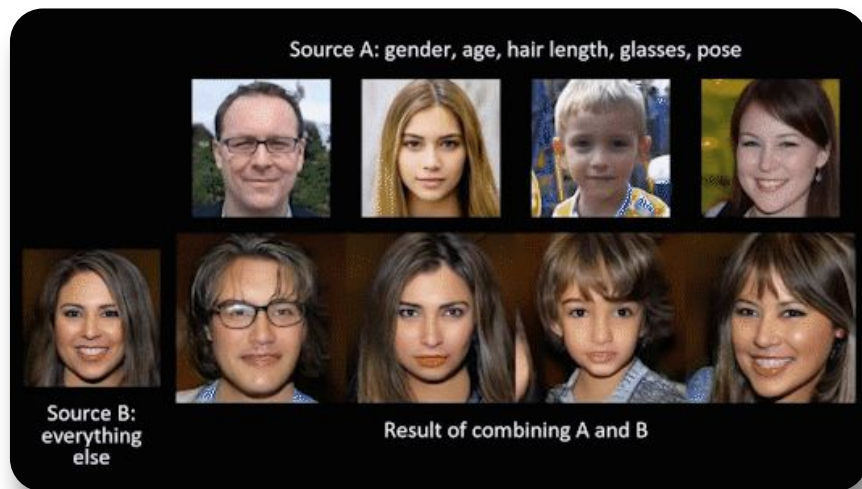
Ventajas y Desafíos de los Modelos de Difusión

Ventajas:

- Alta calidad de generación de imágenes.
- Capacidad de generar detalles finos.

Desafíos:

- Consumo de recursos: Entrenar el modelo es costoso.
- Tiempo de inferencia: Mejorar tiempos de generación.





Futuro de los Modelos de Difusión

- Modelos más eficientes.
- Aplicación en la realidad aumentada y videojuegos.
- Combinación con otros enfoques de IA.





Modelos pre entrenados

Modelos de IA que ya han sido entrenados en grandes cantidades de datos y están listos para ser utilizados o adaptados a nuevas tareas

Ventajas:

- Ahorro de tiempo y recursos: No es necesario entrenar desde cero.
- Alta calidad: Aprovechan el conocimiento de grandes conjuntos de datos.
- Uso extendido: Disponibles para diversas aplicaciones como generación de imágenes, NLP, etc.

Desventajas:

- Menos flexibilidad: Pueden no adaptarse perfectamente a tareas muy específicas.
- Sesgos inherentes: Heredan los sesgos de los datos con los que fueron entrenados.





Modelos pre entrenados vs propios

Característica	Modelos Pre entrenados	Modelos Propios
Tiempo de desarrollo	Menor	Mayor
Personalización	Limitada (necesita ajustes)	Alta
Coste computacional	Bajo a moderado	Alto
Sesgos	Puede contener sesgos de los datos previos	Depende de los datos utilizados
Calidad inicial	Generalmente alta	Variable (depende del entrenamiento)





Gracias por tu atención



¿Preguntas, dudas, inquietudes, ...?

franbvgamazo

franbvg@proton.me



Cofinanciado por
la Unión Europea



MINISTERIO
DE INDUSTRIA
Y TURISMO

EOI Escuela de
organización
industrial



Fondos Europeos

