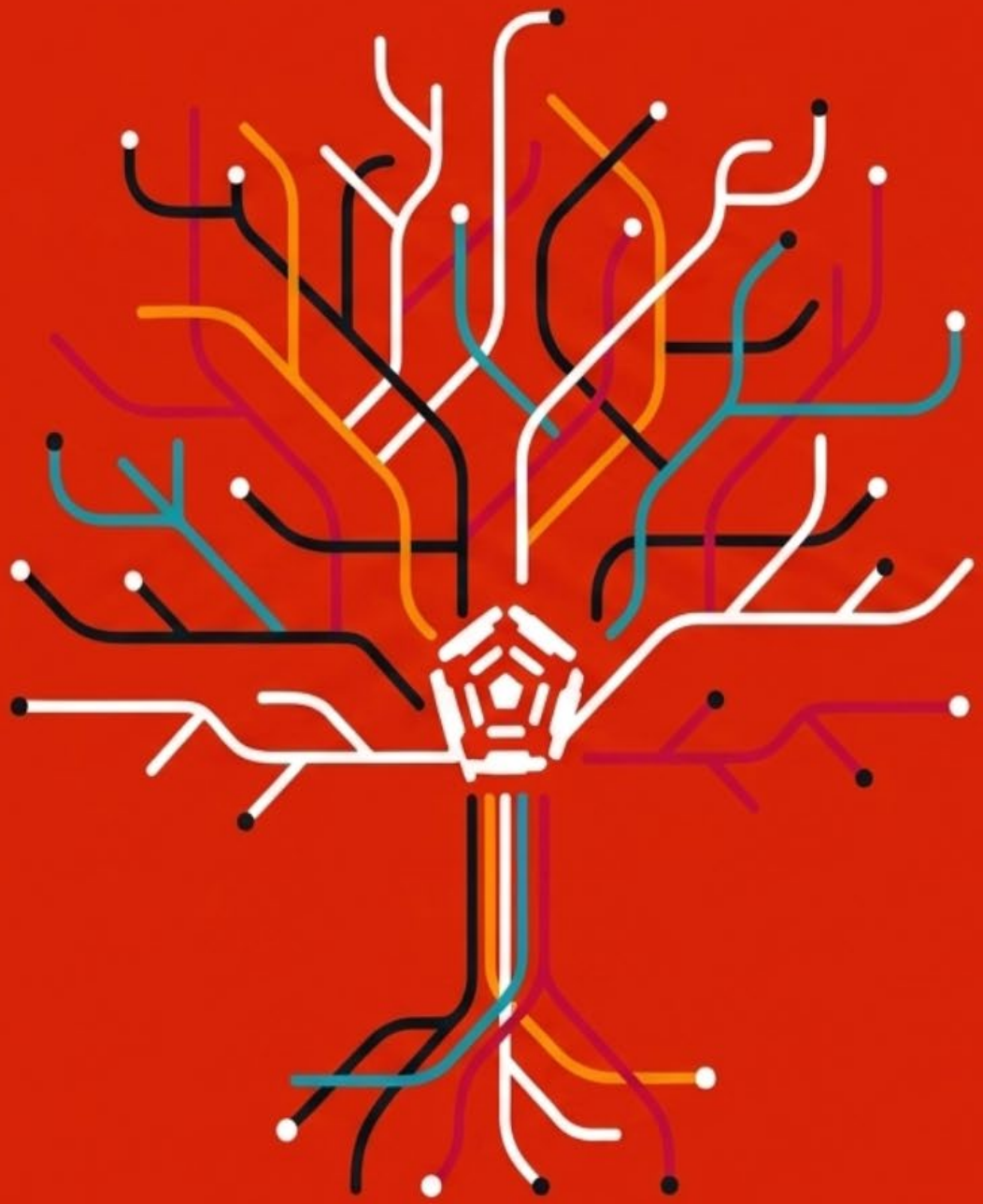
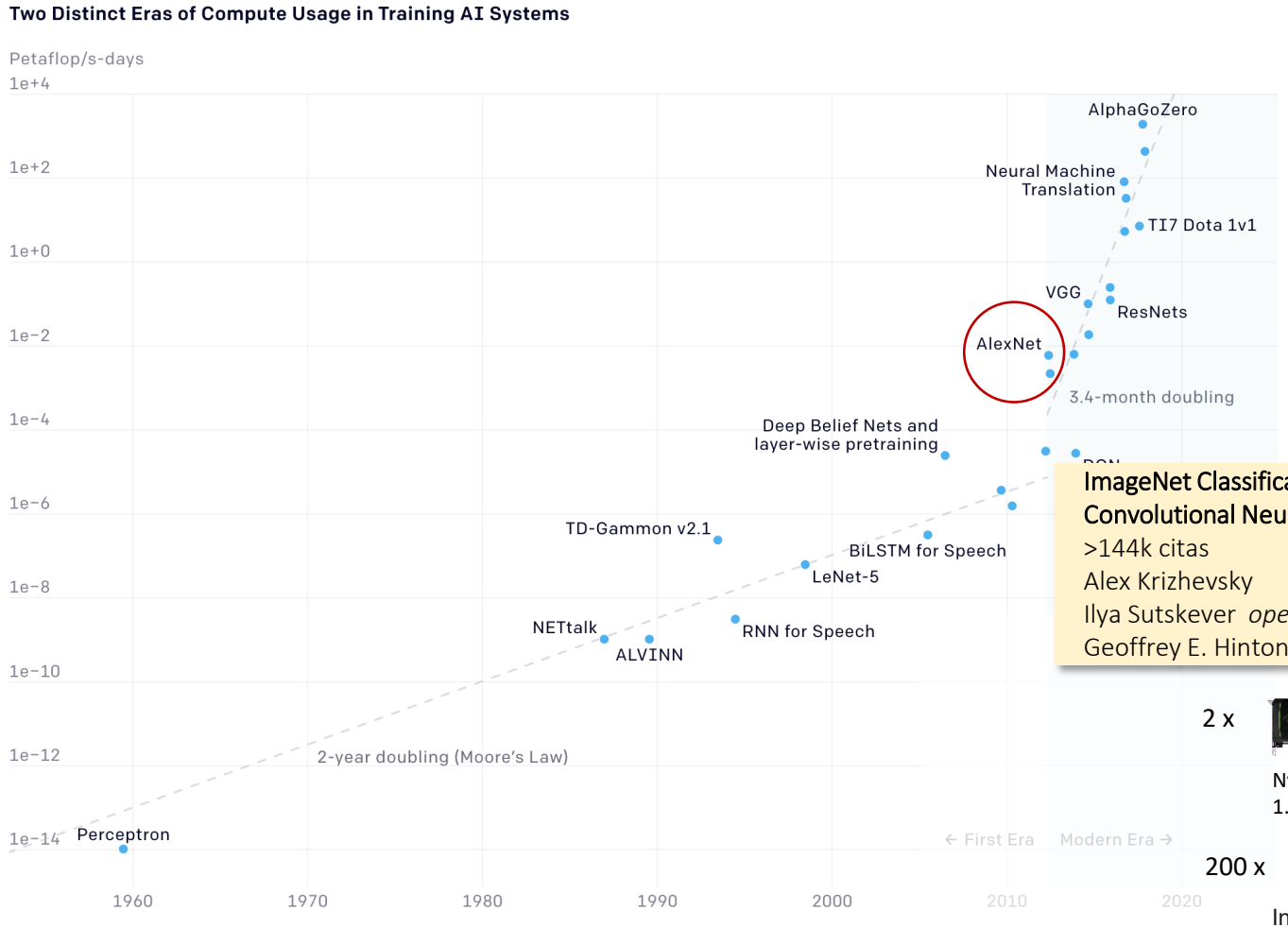


Universidad de Zaragoza | Curso 2026

¿QUÉ TIENE LA INTELIGENCIA ARTIFICIAL PARA MÍ ARCHIVO?

De la teoría a la práctica.

Antecedentes



ImageNet Classification with Deep Convolutional Neural Networks
>144k citas
Alex Krizhevsky
Ilya Sutskever *openAi Chief Scientist (GPT)*
Geoffrey E. Hinton

2 x 
Nvidia GTX 580
1.5 TFLOPS

200 x 
Intel Core i5-2550K @ 3.40GHz

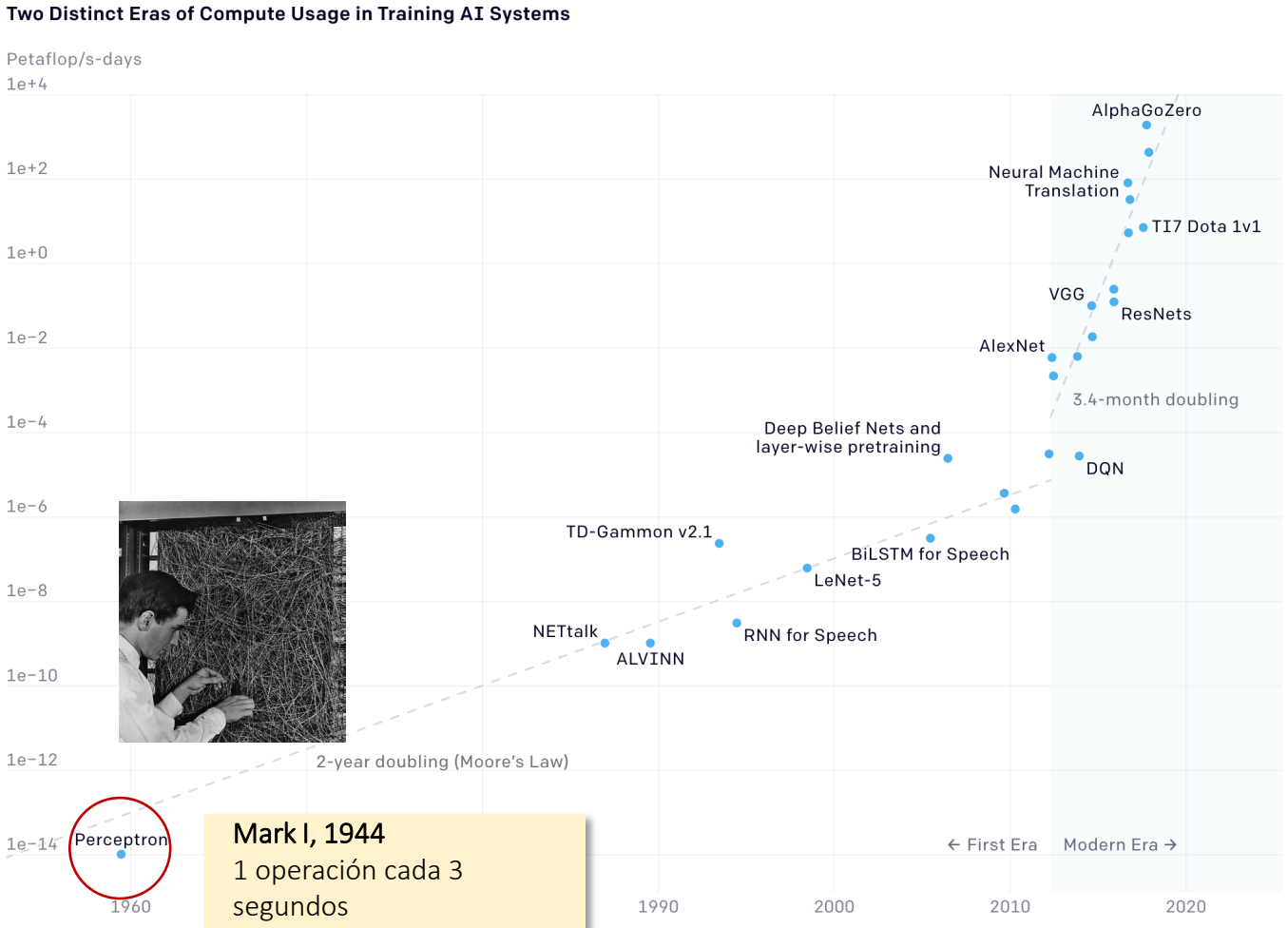

Nvidia RTX 6000
124 TFLOPS

78 x 

VIVOLAB

¿Qué tiene la IA para mi archivo?

Antecedentes



Mark I, 1944
1 operación cada 3 segundos

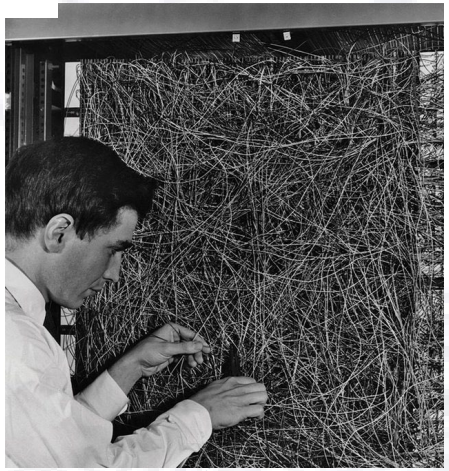
Universidad Zaragoza

Antecedentes

Máquinas electrónicas

Mark I, 1944

1 operación cada 3 segundos



Máquinas mecánicas

Mecanismo de Anticitera 200 a. C

La teoría

Frank Rosenblatt 1957

Alan Turing 1950

Antecedentes

Los microprocesadores

Intel 8086, 1978

50 mil operaciones por segundo

Intel i5, 2018

25 mil millones de operaciones por segundo

2010s La era de las GPUs

Playstation 4s, 2016

1.8 TFlops (~90 x intel i5)

Playstation 5s, 2020

10.2 TFlops (~411 x intel i5)

Nvidia RTX Titan, 2018

16 TFlops (~640 x intel i5)

Nvidia RTX 3090, 2020

35 TFlops (~1400 x intel i5)

Nvidia RTX 4090, 2023

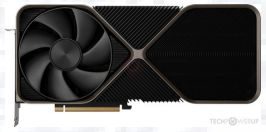
82 TFlops (~3280 x intel i5)

Nvidia RTX 5090, 2025

104 TFlops (~4160 x intel i5)

Nvidia RTX 6000, 2026

124 TFlops (~4960 x intel i5)



Antecedentes

Company/Entity	H100 GPU Count	Type
Meta	350,000	Private cloud
XAI/X	100,000	Private cloud
Tesla	35,000	Private cloud
Lambda	30,000	Public cloud
Google A3	26,000	Public cloud
Oracle Cloud	16,000	Public cloud
Poolside	10,000	Public cloud
Magic	8,000	Public cloud
Andromeda	3,632	Private cloud
Scaleway	1,016	Private cloud
Hugging Face	768	Public cloud
DeepL	544	Private cloud
Recursion	504	Private cloud
Princeton	300	National HPC
Photoroom	256	Private cloud

Año 3 AG (after GPT)

T4

FP16 ***65.13*** *TFLOPS* (~2600 x intel i5)

A100

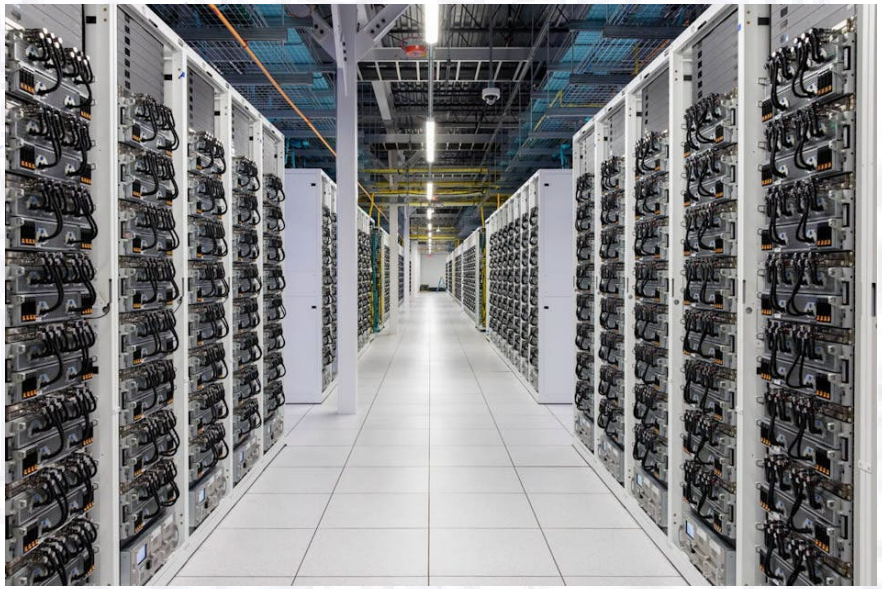
FP16 ***77.97*** *TFLOPS* (~3120 x intel i5)

H100

FP16 ***204.9*** *TFLOPS* (~8196 x intel i5)

H200

FP16 ***241.3*** *TFLOPS* (~9640 x intel i5)



xAI Colossus Supercomputer Data Center – Memphis USA

Antecedentes

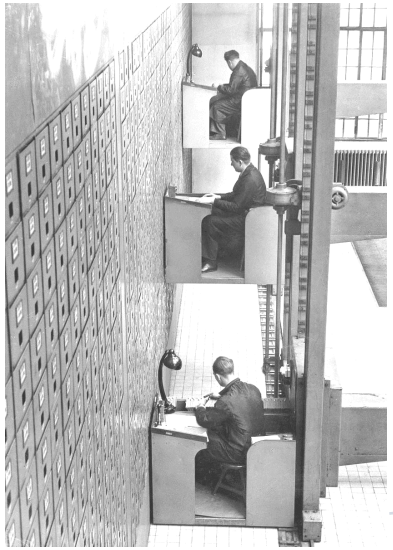
Velocidad de almacenamiento

Disco duro *2000* *18GB (48MB/s)*

HD estado sólido *2021* *1TB* (7000 MB/s)

Almacenamiento

Sistema mecánico1937 (República Checa)



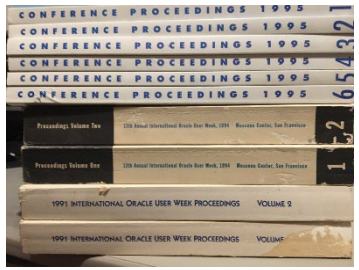
Capacidad de almacenamiento

Cinta perforada *1970* *<1 KB*

Disco 3 ½ *1987* *1.4 MB*

DVD *1995* *4.7 GB*

Antecedentes



Proceedings papel 1995
Revistas papel

Google

Buscadores internet 1998

TensorFlow

PYTORCH

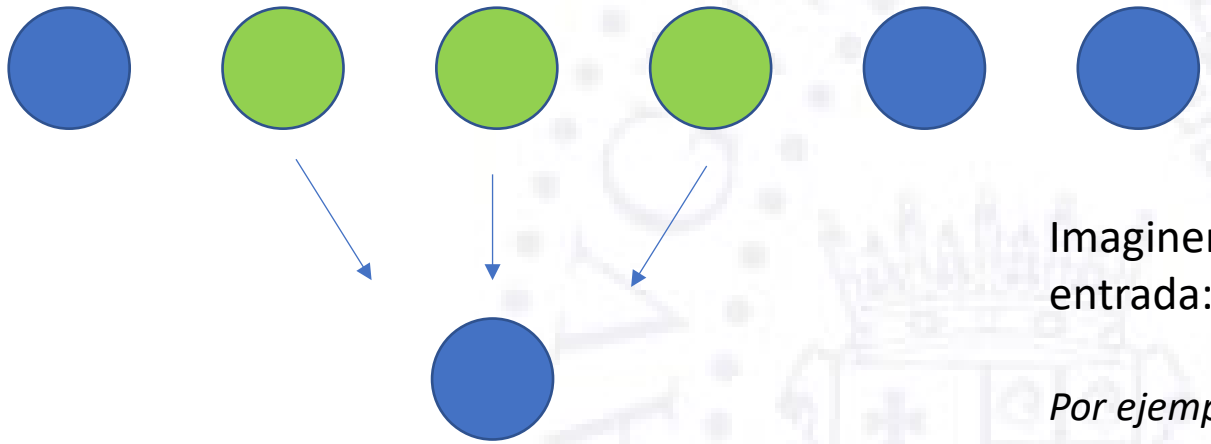
Software gratuito
Toolkits 2010

GitHub

2008 *Redes sociales /*
plataformas de desarrollo colaborativo

Fundamentos

- **Procesado en fases, capas**

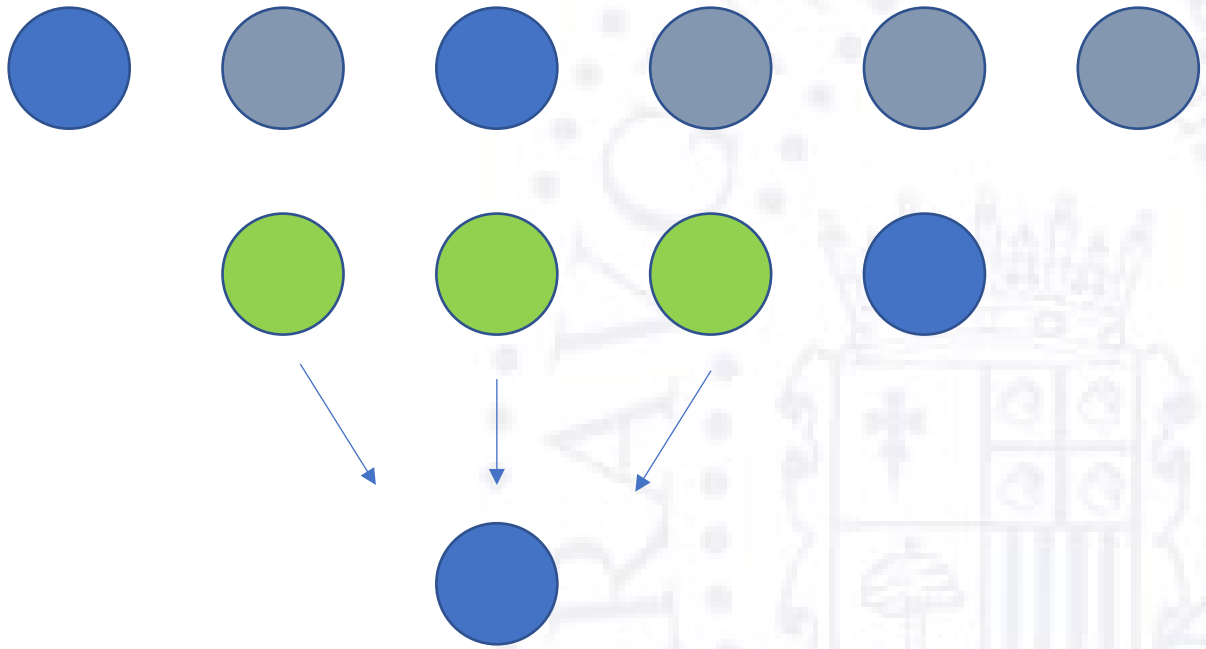


Imaginemos unos datos de entrada: altura, edad,..

Por ejemplo podríamos decir: Pregunta a los tres que tengas en la fila de delante y quédate con el máximo, mínimo, etc

Fundamentos

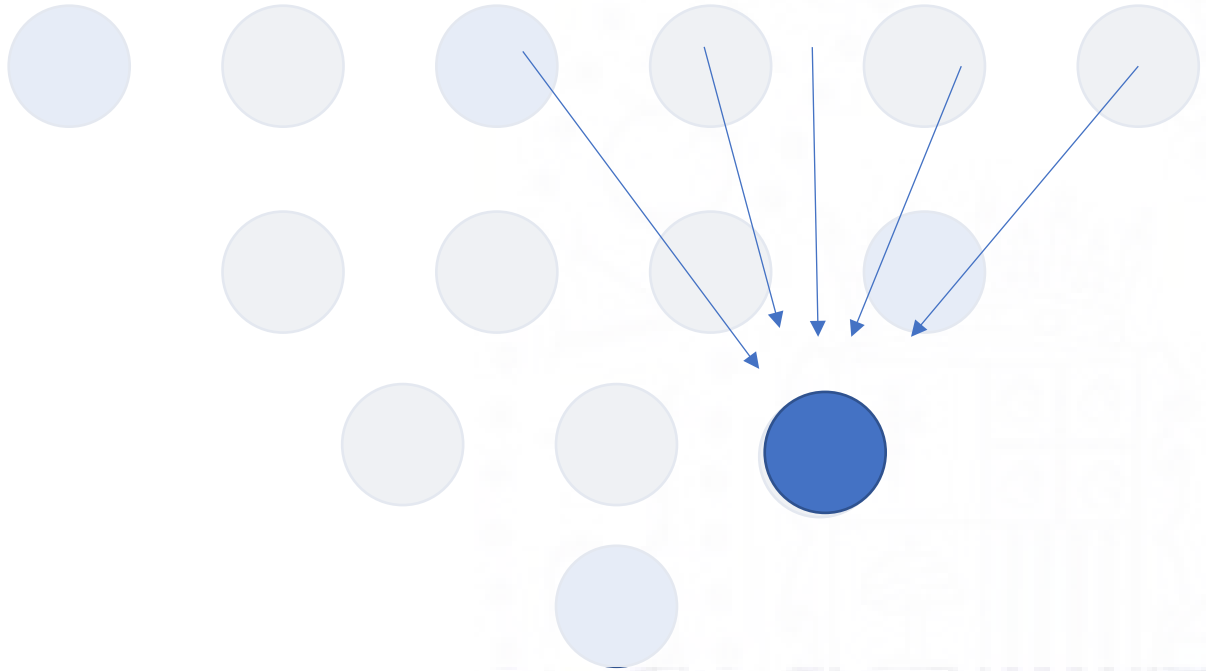
- **Procesado en fases, capas**



Y repetimos,
en todas las filas…

Fundamentos

- **Procesado en fases, capas**

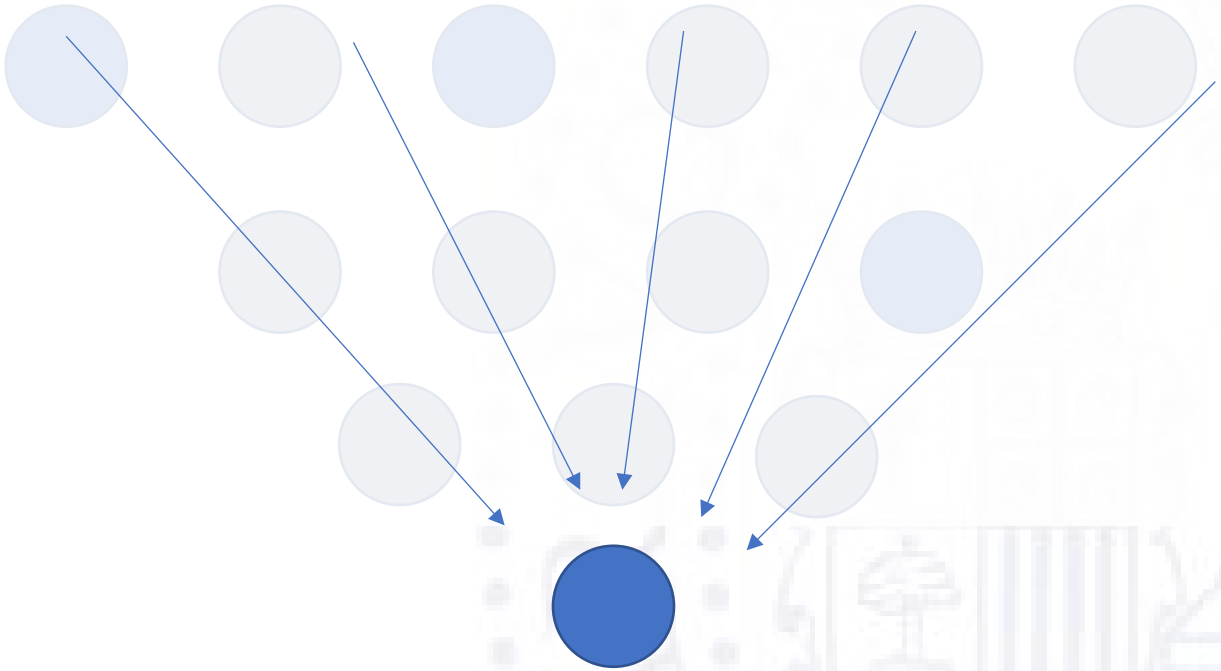


Algunos nodos

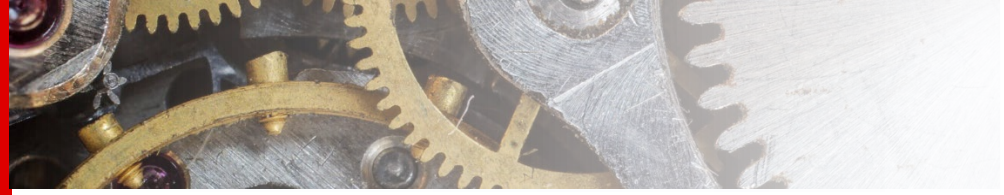
reciben **parte** de la información

Fundamentos

- **Procesado en fases, capas**

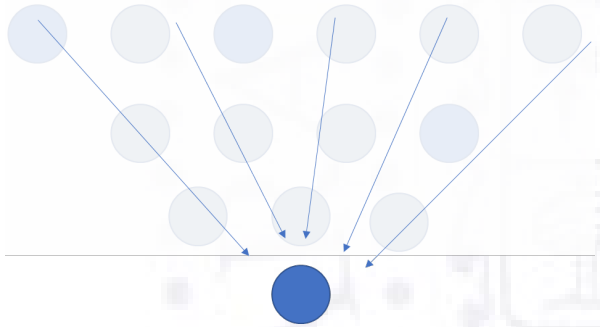


Ahora al último le podríamos preguntar, **¿quién es el más joven ?**
Ha recibido **toda** la información

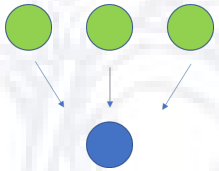


Fundamentos

- **Procesado en fases, capas**

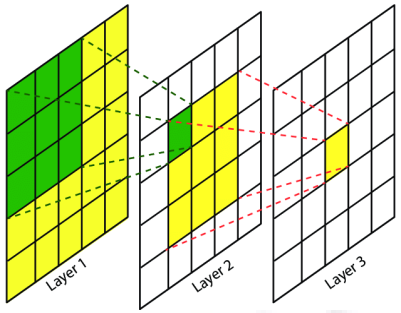


- En una red convolucional cada píxel es procesado por una entrada
- La información se va combinando en capas sucesivas (capas/layers)
- Cada unidad está conecta a varias unidades de las etapas anteriores

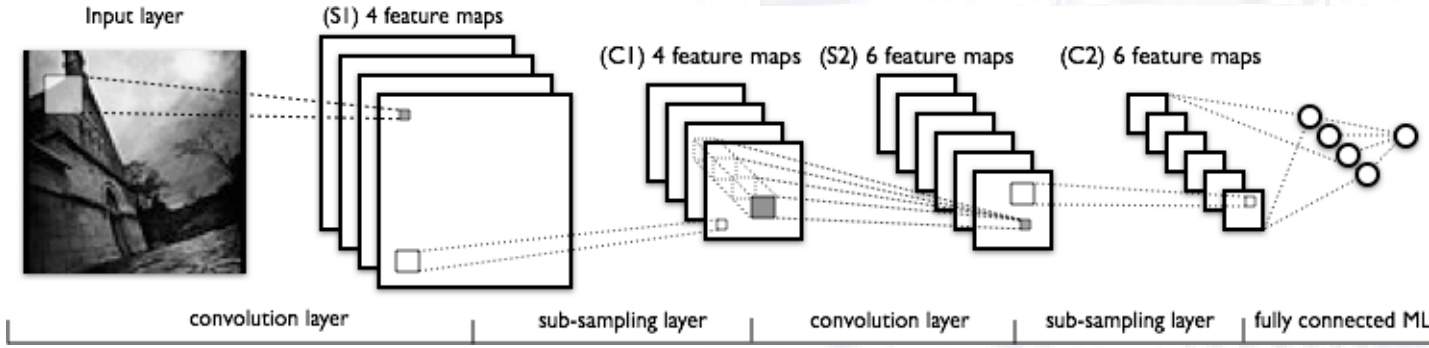


Fundamentos

- **Redes convolucionales**
  - Cada capa suma varios valores de entrada con distinto peso, normalmente 9 entradas: 3x3



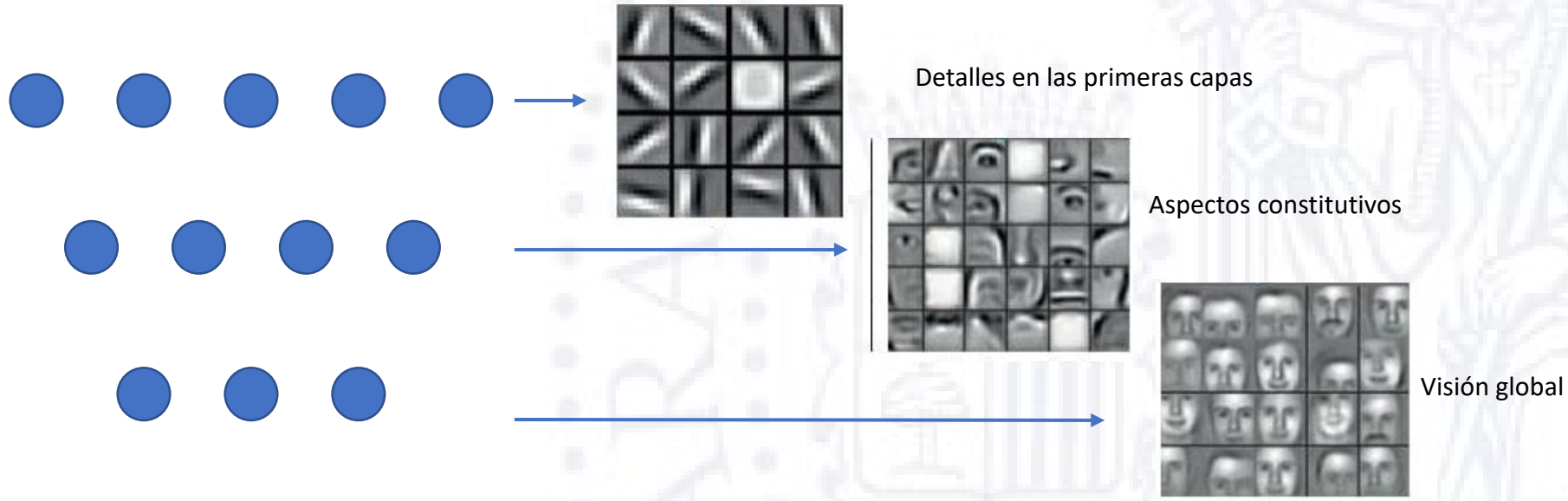
Yann LeCun



LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

Fundamentos

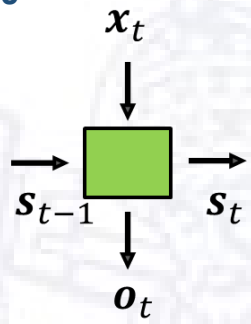
- **Redes convolucionales**
  - Con mayor **profundidad (depth)** se logra mayor abstracción **(Deep learning)**
  - Las primeras redes profundas tenían 7 capas
  - Hoy en día en cuestión de minutos se tiene acceso a redes de más de 100 capas ya entrenadas



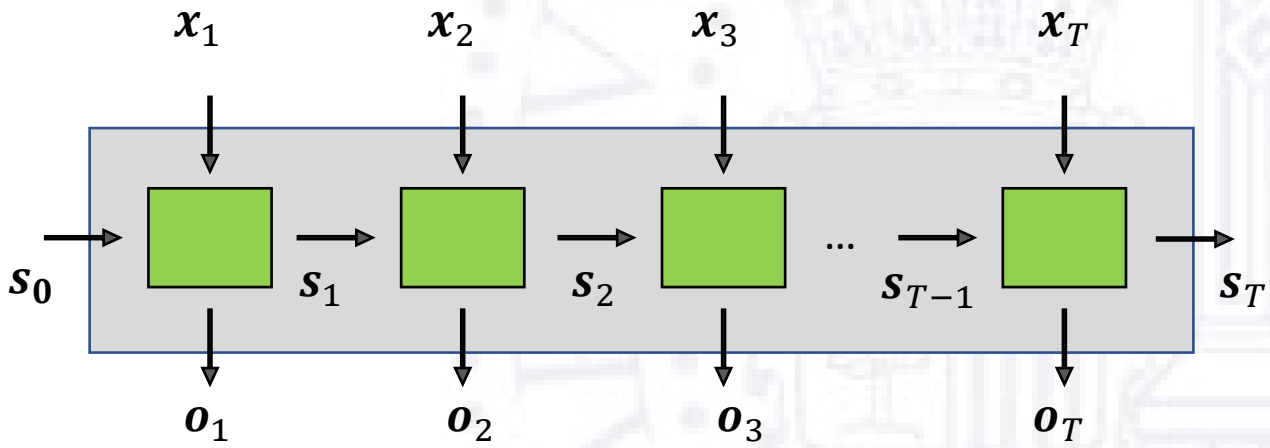
Fundamentos

- **Redes recurrentes (RNN)**
- **Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780**

  - Analizan la entrada en orden (sentido temporal, orden del texto)
  - Cada celda recibe información del pasado y escribe nueva información
  - Memoria finita: problema de cuello de botella

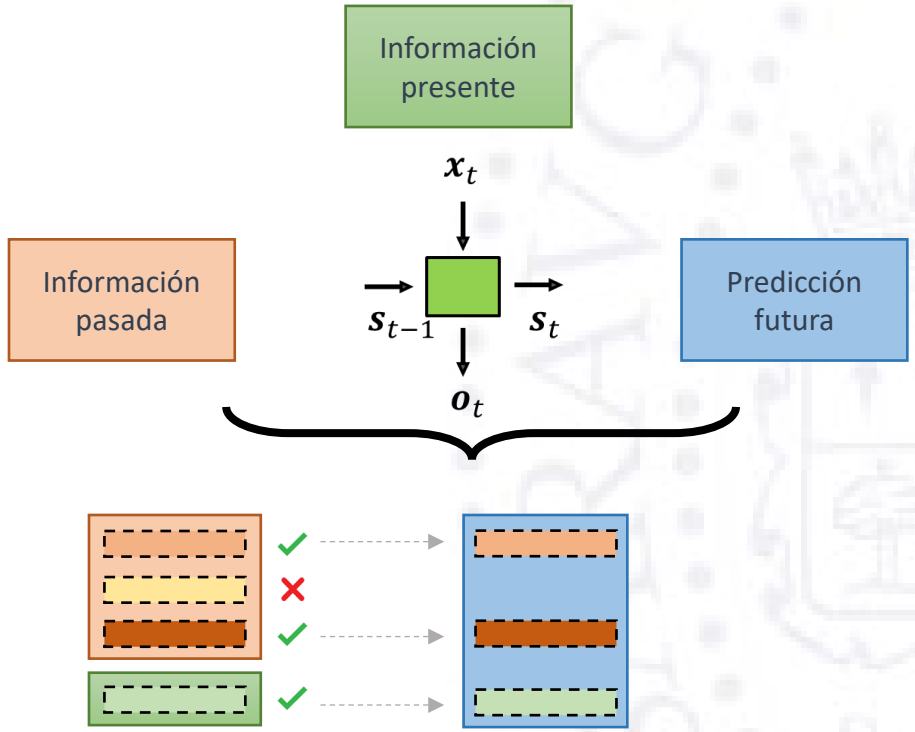


RNN cell



Fundamentos

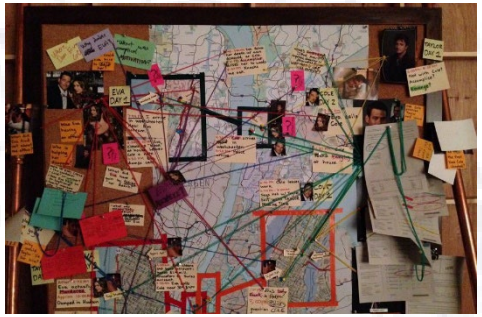
- **Redes recurrentes (RNN)**
- **Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780**
  - Pueden retener o borrar información del pasado
  - Añade la información del presente que sea interesante
  - Con todo ello genera una predicción futura



La celda procesa la información pasada y presente:

- Puede olvidar información pasada
- Puede mantener información pasada
- Puede añadir información presente

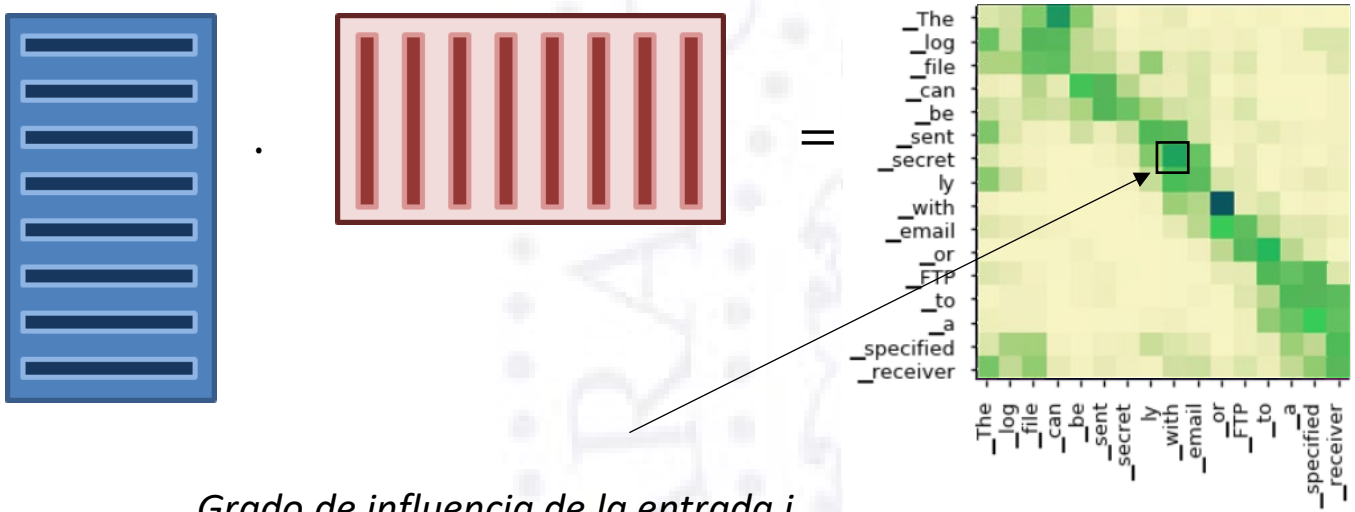
Fundamentos



- **Transformers** (187k citas)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. Kaiser L, Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30, 5998-6008

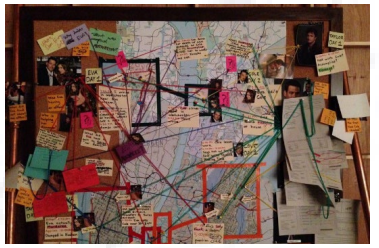
  - Son capaces de analizar las relaciones entre todas las entradas

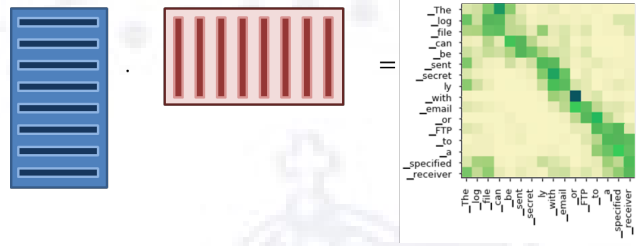


Grado de influencia de la entrada i sobre la j

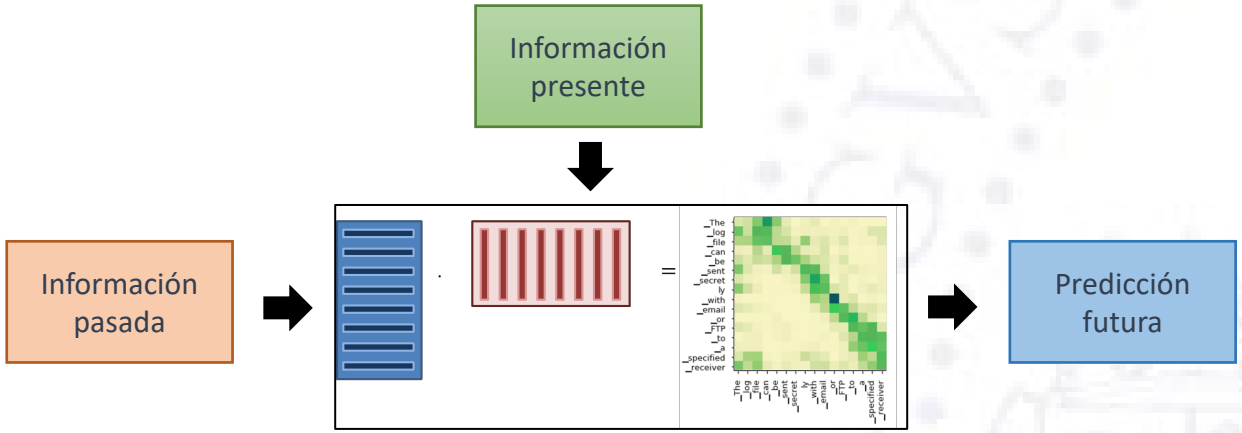
https://nlp.seas.harvard.edu/2018/04/03/attention.html

Fundamentos

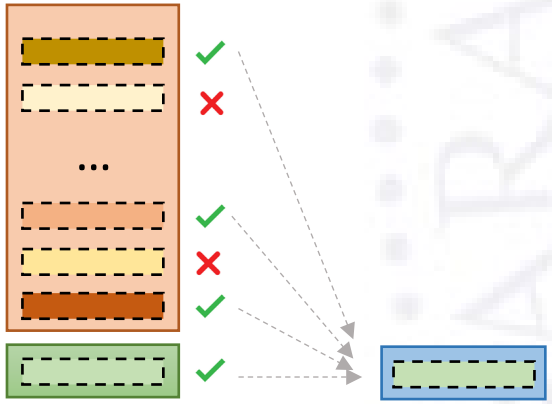




- **Transformers** (187k citas)
  - Relaciones entre todas las entradas?

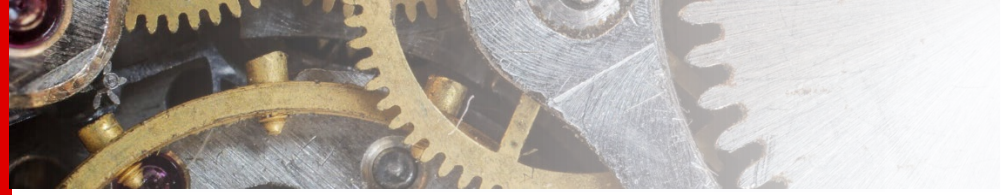


La información pasada puede alcanzar >100k palabras



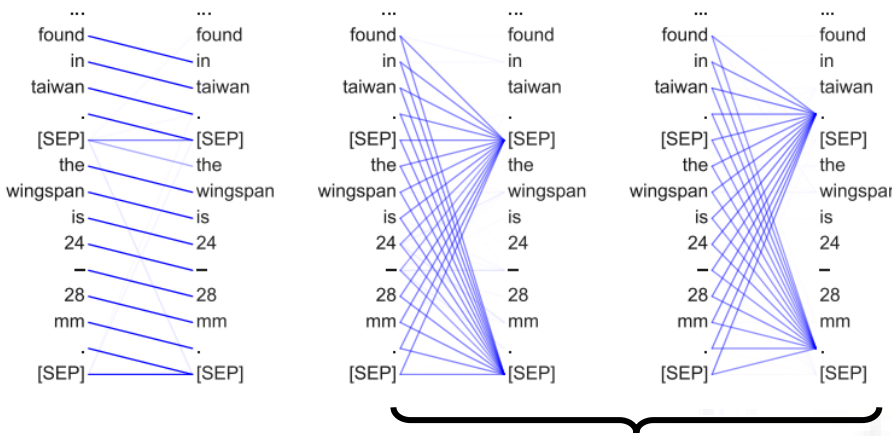
El mecanismo de atención (self attention)

- Analiza **toda la información** del pasado (contexto)
- Los modelos actuals grandes (large) pueden mirar cientos de miles de palabras previas (**contexto**)
- Es **costoso**, necesita grandes recursos e infraestructura en inferencia



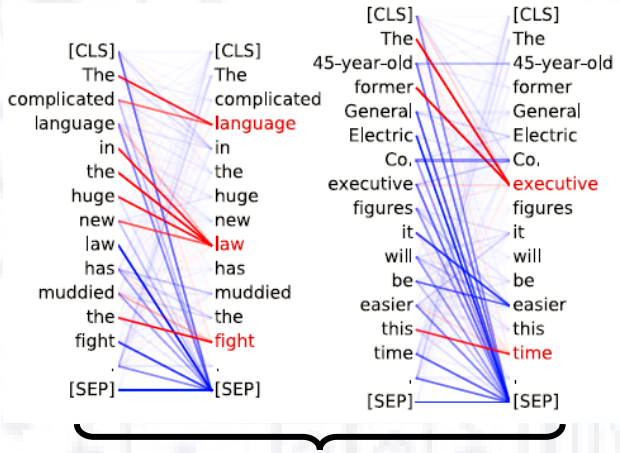
Fundamentos

- **Transformers**

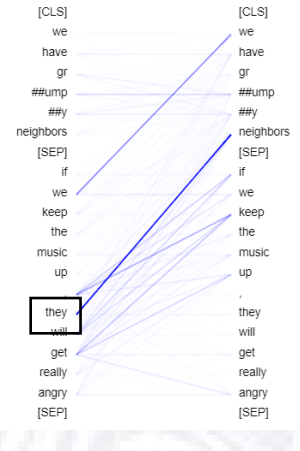


palabra anterior

Final de frase



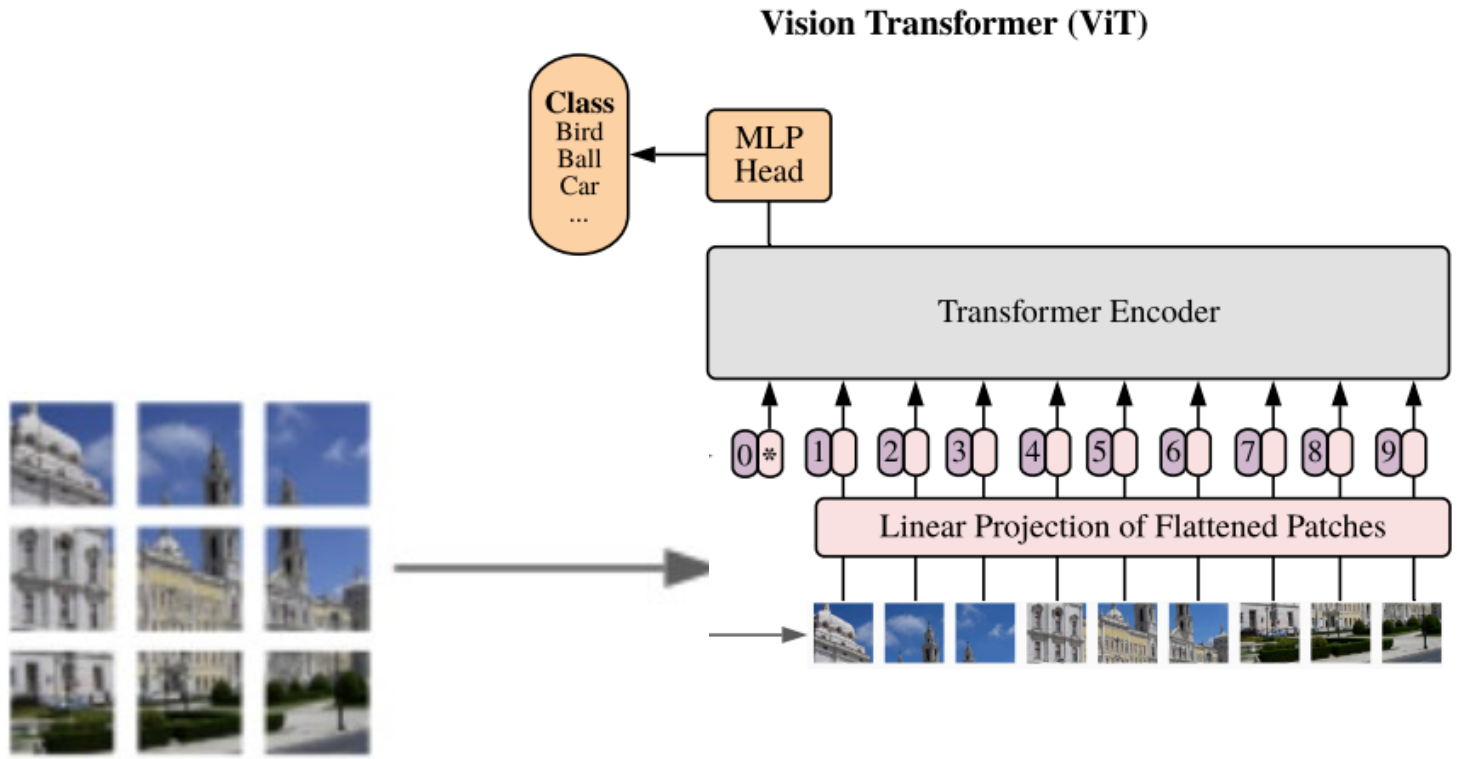
Determinantes y modificadores de un nombre



Desambigüación ellos -> vecinos

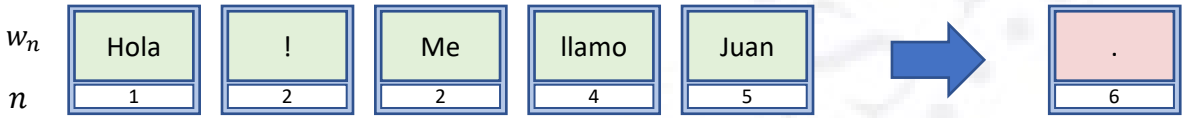
Fundamentos

- **Visual Transformer(Dosovitskiy 2020)**
- **Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Uszkoreit, J. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.**
  - Se descompone la imagen en parches (16x16 patches) de izquierda a derecha y los procesa un transformer

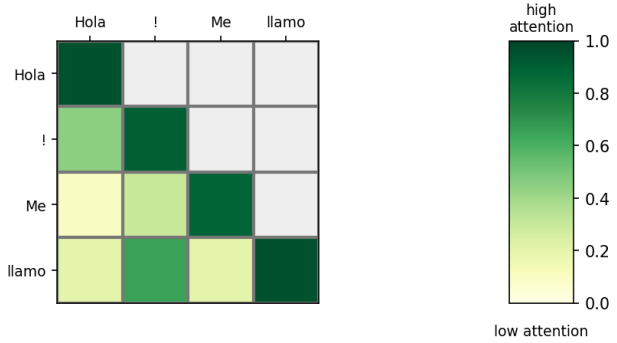


Fundamentos

- **Transformer**



N^2 cálculos



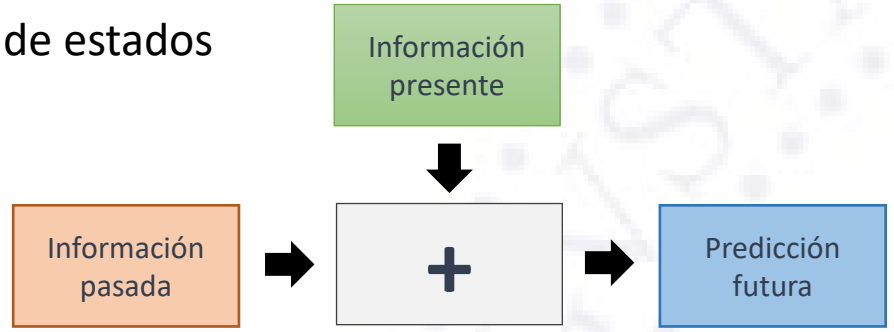
- Generar una secuencia de longitud N → N^2.
  - Esto es un problema si la secuencia es muy larga.
- Entrenamiento: rápido (se puede paralelizar)
- Inferencia: lenta (escala cuadráticamente).

Fundamentos

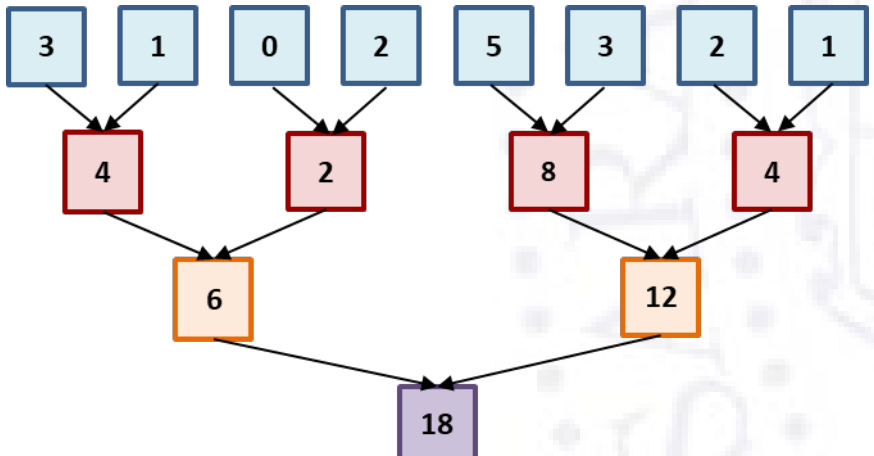
- **Modelo: Mamba**

Gu, A., Goel, K., & Ré, C. (2021). Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396...

  - Espacio de estados

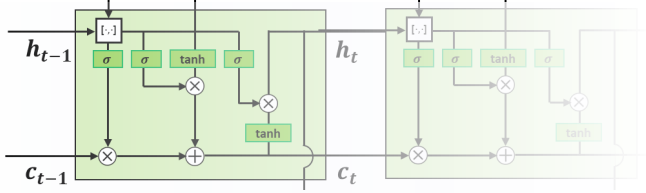


- **Cada respuesta se basa una parte en la anterior (pasado) y otra en lo que lee**
- **Como las redes recurrentes anteriores, entonces son igual de lentos?**



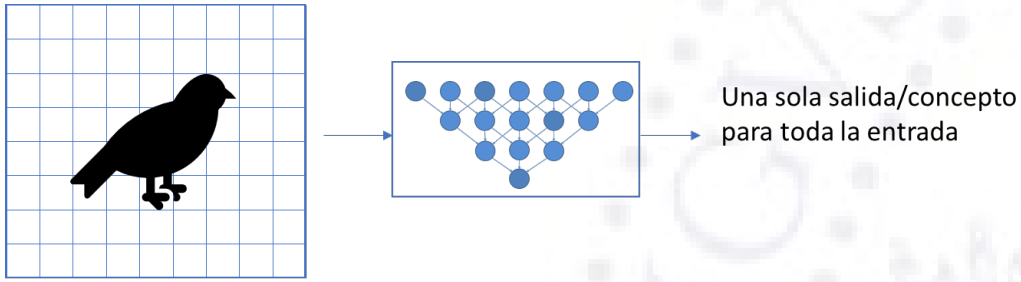
En lugar de sumar la información en orden se pueden hacer las sumas en paralelo:

Algoritmo Associative Scan (Escaneo Asociativo)

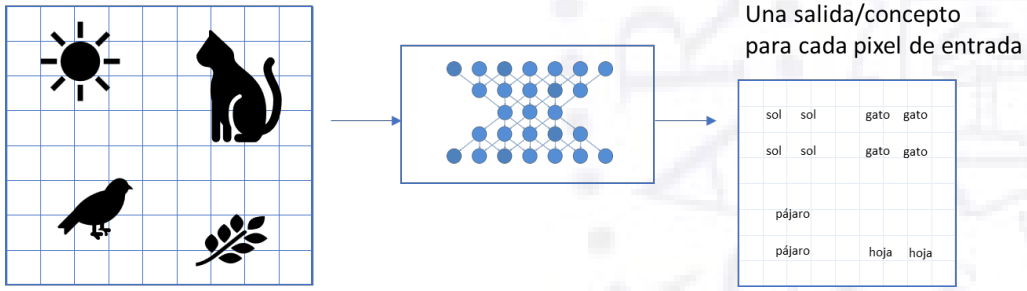


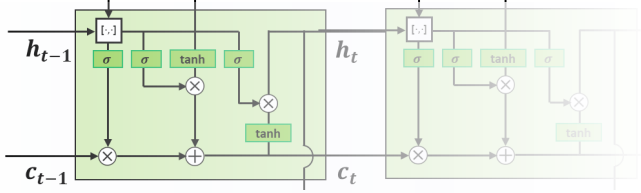
Aplicaciones

- Clasificación:
  - Predicción: qué concepto hay en una imagen/texto/audio



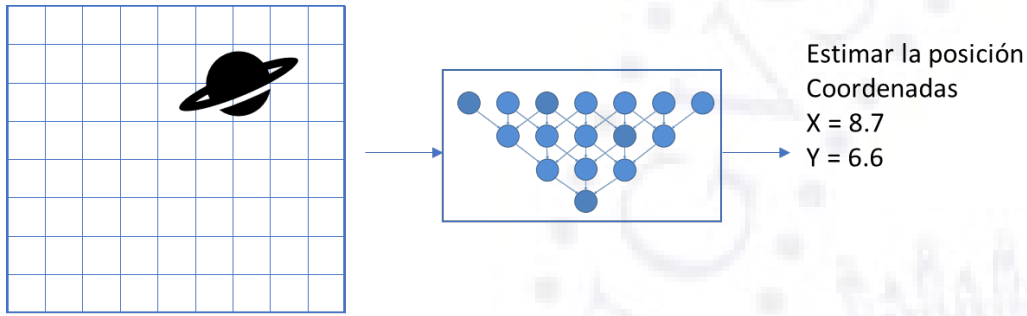
  - Predicción: qué concepto hay en cada zona/pixel: imagen/texto/audio



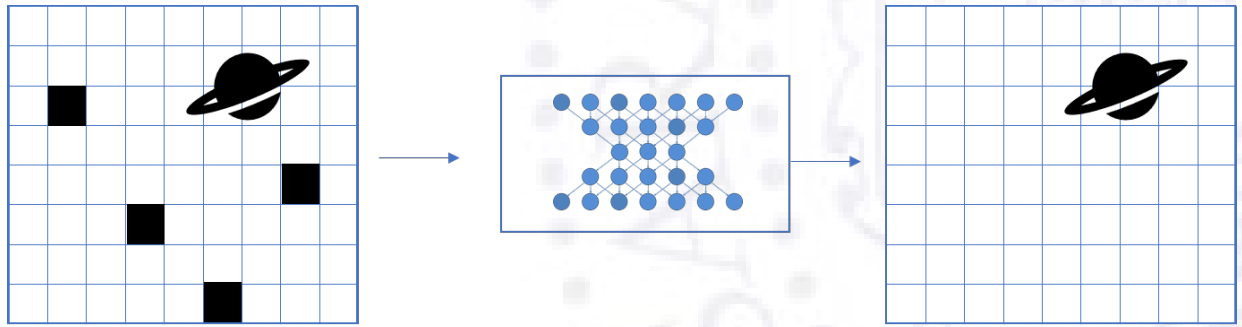


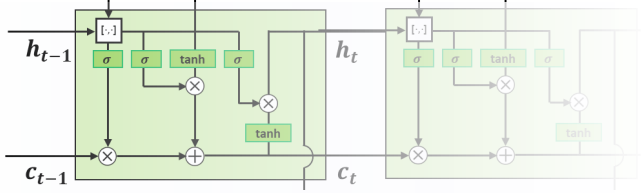
Aplicaciones

- Regresión:
  - Utilizar los datos para obtener algún tipo de predicción numérica

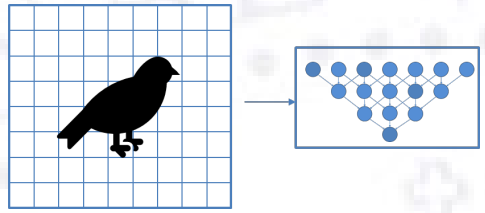


  - Predecimos varios valores numéricos: por cada zona, pixel…

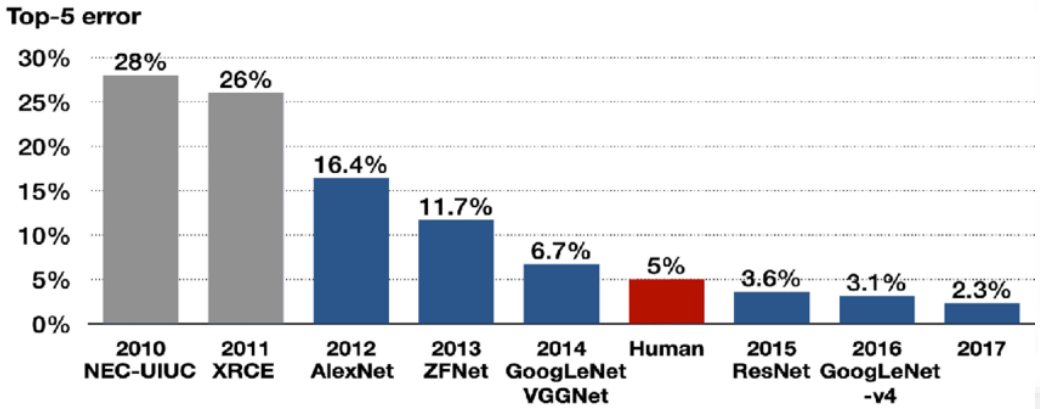


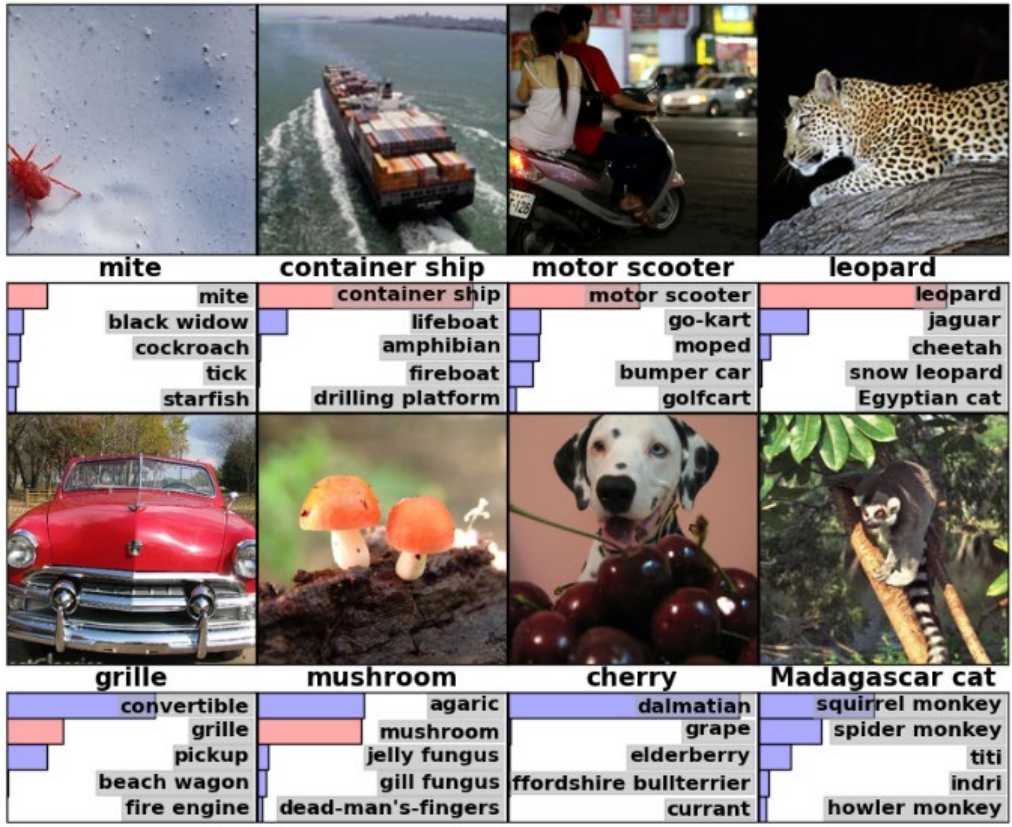


Aplicaciones



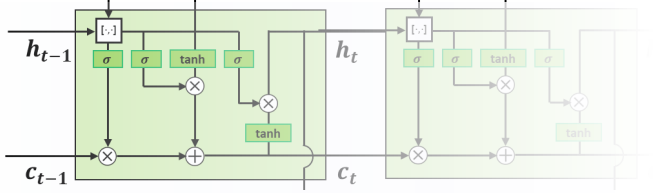
- **Imagenet challenge**
  - problema de clasificación:
    - **¿Qué hay en esta imagen? -> 1 respuesta**
  - 1000 clases, imágenes RGB
    - Hay ambigüedad en las etiquetas
    - por eso a veces se da la tasa top5
    - (error si no está entre las 5 primeras opciones)





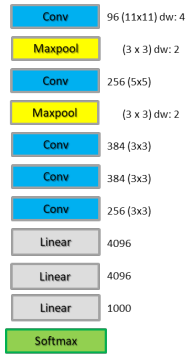
VIVOLAB

¿Qué tiene la IA para mi archivo?



Aplicaciones

- Mejora gradual de prestaciones

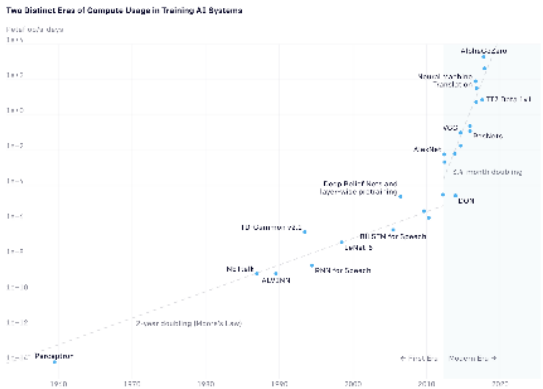


Geoffrey Hinton

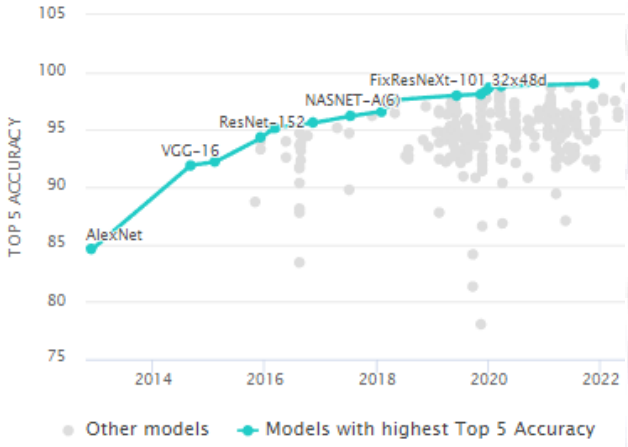
*2012: **Alexnet** 7 capas*
84.6 % aciertos (Top 5)
clasificación de imagen (Imagenet)

*2014: **Inception** 25 capas*
93.3% aciertos

*2015: **Resnet** >100 capas*
96.43% aciertos

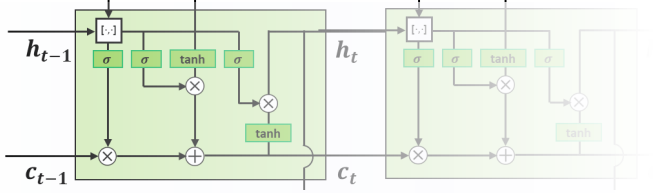


recursos

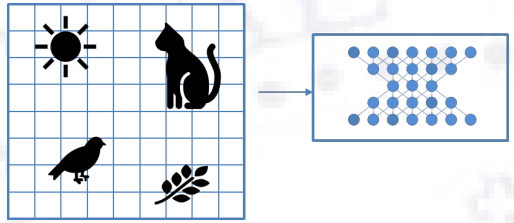


rendimiento

Universidad Zaragoza



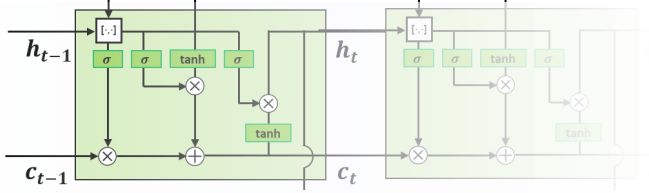
Aplicaciones



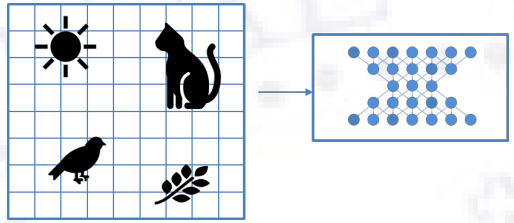
- Clasificación múltiple:
  - varias propiedades/conceptos de una imagen/texto/audio

En este ejemplo la red neuronal se prepara para resolver muchas respuestas sí o no:

¿Hay un perro?	**No**
¿Hay un gato?	Sí
¿Hay árboles?	**No**
¿Hay un pájaro?	Sí
¿Hay cielo?	**No**
¿Hay hierba?	Sí



Aplicaciones

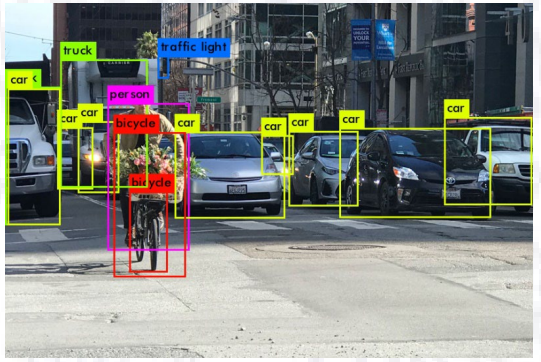


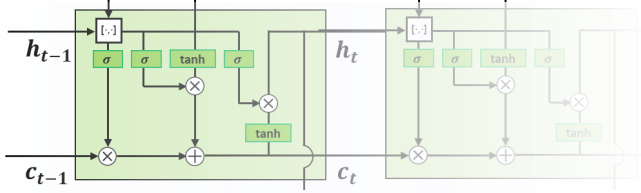
- **Clasificación múltiple**
  - varias propiedades/conceptos de una imagen/texto/audio

En este ejemplo la red neuronal se prepara para resolver muchas respuestas sí o no:

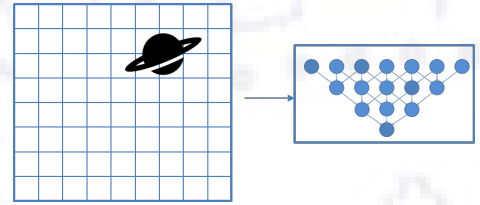
¿Hay un perro?	**No**
¿Hay un gato?	Sí
¿Hay árboles?	**No**
¿Hay un pájaro?	Sí
¿Hay cielo?	**No**
¿Hay hierba?	Sí

- YOLO, You Only Look Once: Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
  - La red procesa la imagen una vez por ejemplo la (YOLO You Only Look Once)
  - Por cada zona final predice los objetos que hay en esa zona
  - Muy rápido, pero puede perder objetos

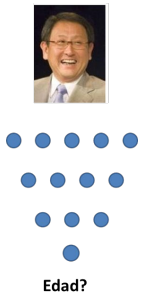




Aplicaciones



- Regresión



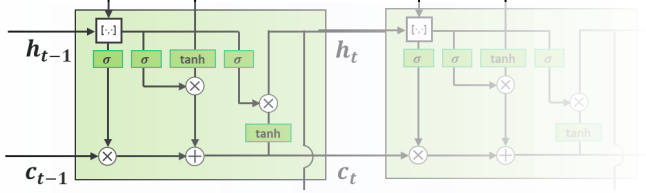
¿Qué edad tiene estas persona?

- La respuesta sería un número con la edad en años
- Entrenaríamos el sistema con muchas imágenes
- El algoritmo de gradiente aplica las correcciones necesarias cuando la red se equivoca

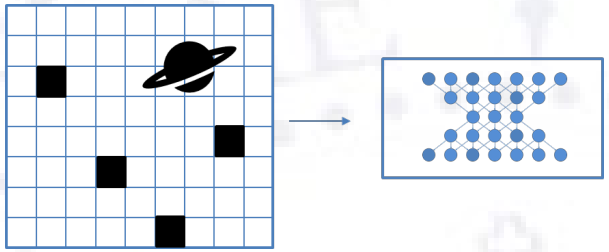
- Ejemplo: estimación de fecha fotografías

GT: 1962	GT: 1943	GT: 1973	GT: 1946
DEXPERT: 1962	DEXPERT: 1944	DEXPERT: 1973	DEXPERT: 1946
ConvNeXt: 1963	ConvNeXt: 1949	ConvNeXt: 1972	ConvNeXt: 1946
ResNet-50: 1965	ResNet-50: 1945	ResNet-50: 1972	ResNet-50: 1944

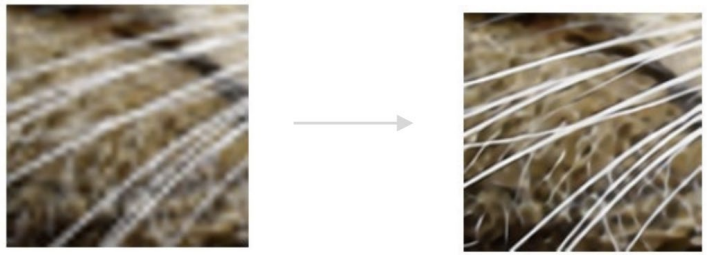
Net, F., Hernández, N., Molina, A., & Gómez, L. (2024, March). A transformer-based object-centric approach for date estimation of historical photographs. In *European Conference on Information Retrieval* (pp. 137-150). Cham: Springer Nature Switzerland.
https://github.com/cesc47/DEXPERT



Aplicaciones

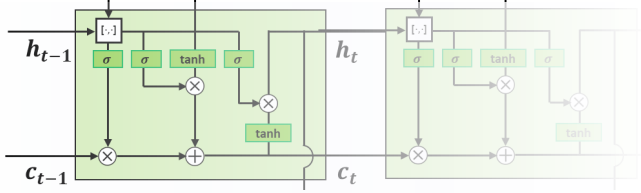


- **Ejemplos regression múltiple**
  - Image super-resolution
    - Predecir una versión de mayor resolución de una imagen
    - Podemos generar las etiquetas automáticamente
  - Image colorization
    - Predecir una versión a color de una imagen en blanco y negro
    - Usando imágenes a color, eliminamos el color para generarlas
    - muestras de entrenamiento

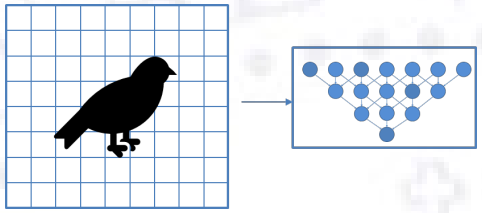


rtve

FIAT/IFTA Archive Achievement Awards 2023 - Shortlisted Nominee Category: Excellence in Media Management Title: **"AI + Colour = New Life to the RTVE Archive"**



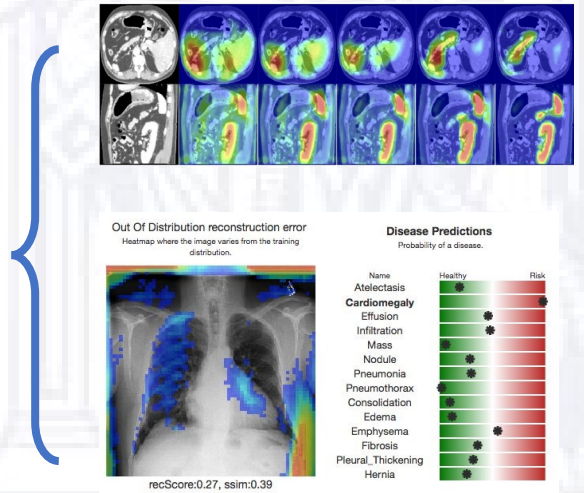
Aplicaciones

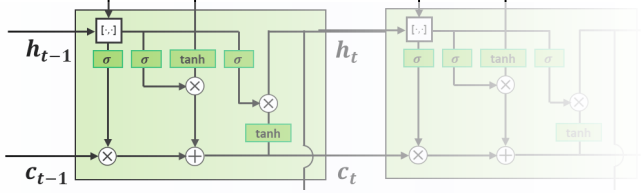


- Clasificación: ¿ nos podemos fiar ?

¿cómo es ese pequeño porcentaje de fallos… ?

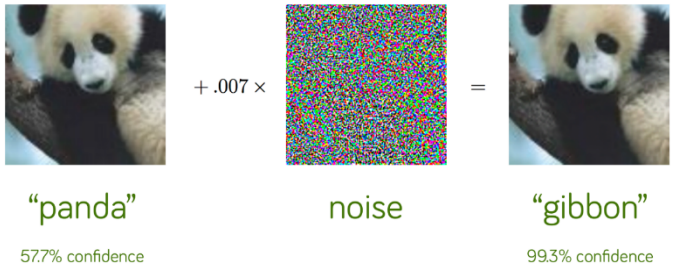
Hay modelos que pueden mostrar **qué zonas** han considerado más



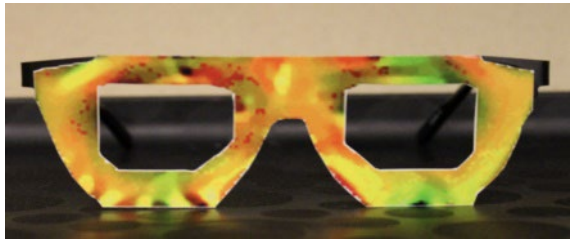


Aplicaciones

- **Adversarial attacks**
  - Un pequeño cambio en la imagen puede hacer que cambie la predicción



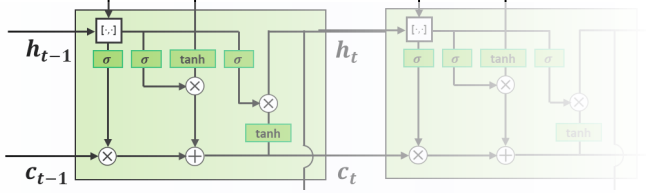
Asversarial attacks
Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.



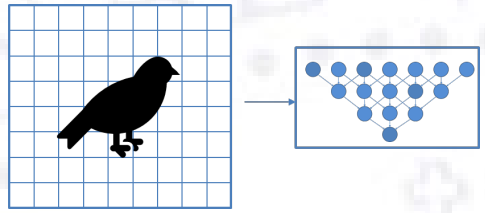
Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition
Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, Michael K. Reiter
ACM Conference on Computer and Communications Security (CCS 2016)

Speed Limit 80
(88% confidence)

Robust physical-world attacks on deep learning visual classification.
Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., ... & Song, D. (2018). In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1625-1634).

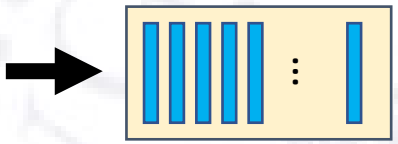


Aplicaciones



- Clasificación: ¿ nos podemos fiar ?

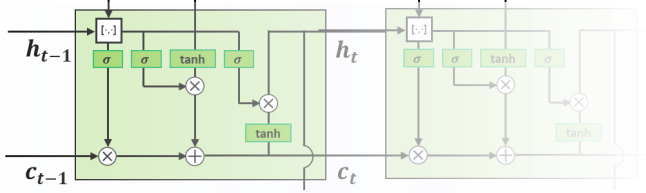
paperswithcode.com



Entrenamiento

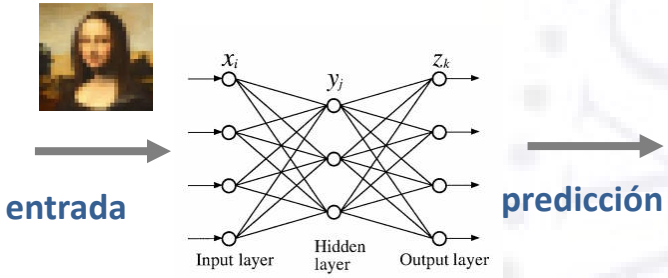
Hoy en día se entrenan facilitando múltiples versiones de las imágenes/sonidos

Se conoce como:
Aumento de datos



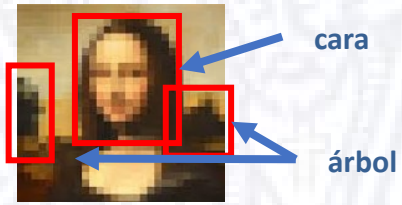
Aplicaciones

- Resumen: aprendizaje supervisado



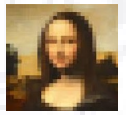
Clasificación (global): Mona Lisa

Clasificación Múltiple:



Regresión (global): Edad ? -> xx años

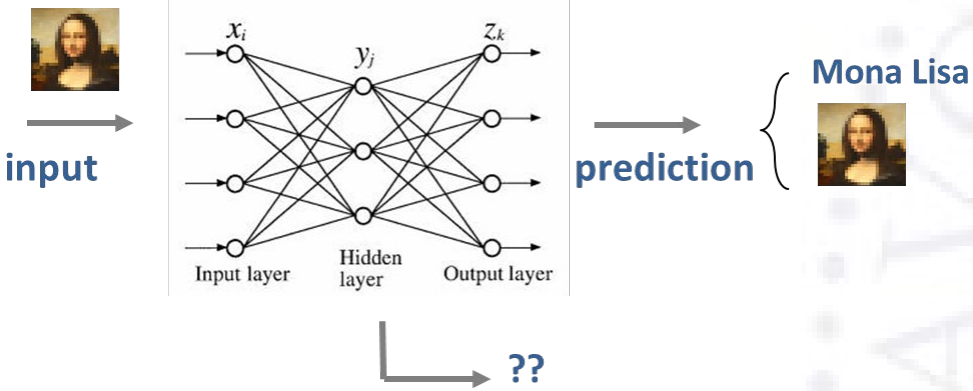
Regresión múltiple:

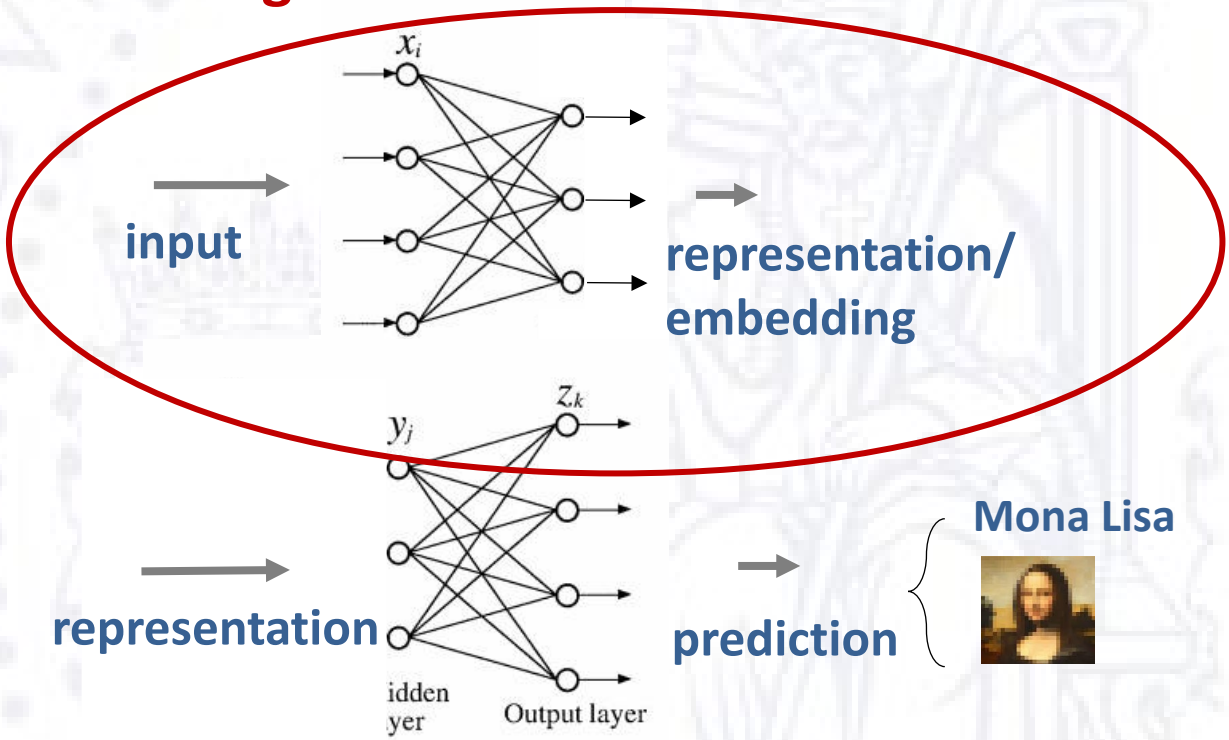


Colorización
super-resolución

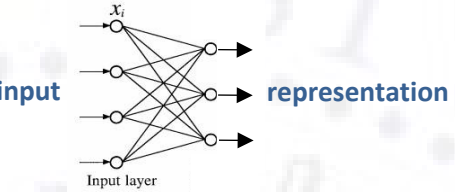
Representaciones

- **Representation learning**
  - Podemos utilizar representaciones internas de la red
  - Vectores de representación o **embeddings**

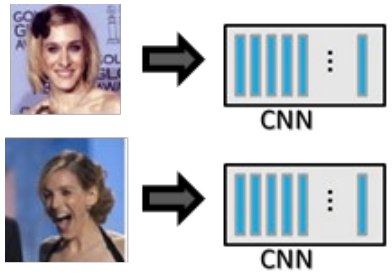




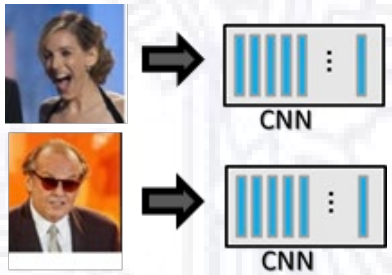
Representaciones



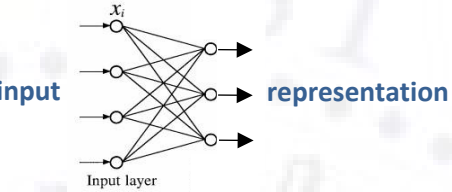
- **Representation learning**
  - Podemos utilizar representaciones internas de la red



Comparar si dos imágenes corresponden a la misma identidad

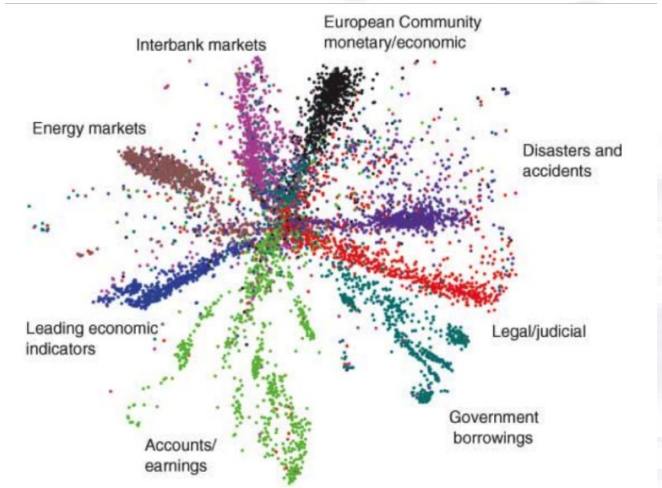


Representaciones

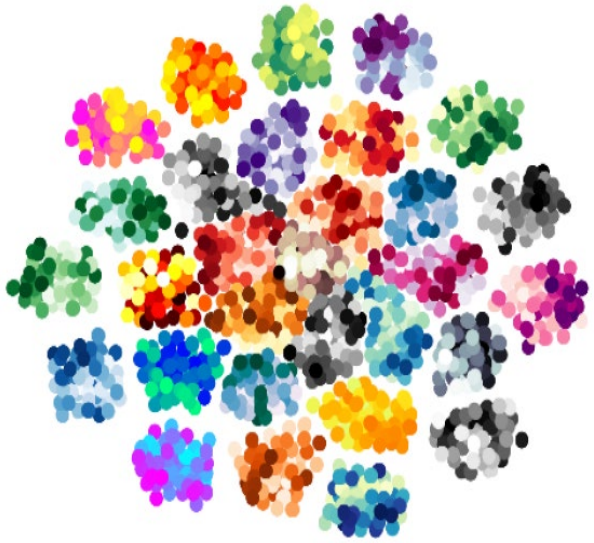


- **Representation learning**
  - **Para representarlos se suelen usar aproximaciones en 2D que mantienen las distancias**
    - imágenes/sonidos/textos similares están más próximos en ese espacio
    - Ejemplos: t-sne, umap, …

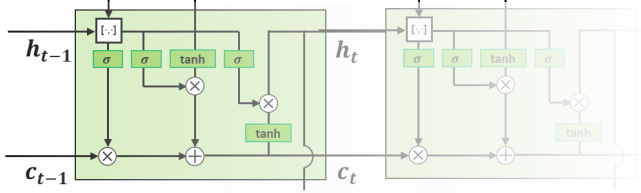
Handwritten digits,
A. Karpathy, Stanford University



Text topic classification
G. Hinton, Toronto University

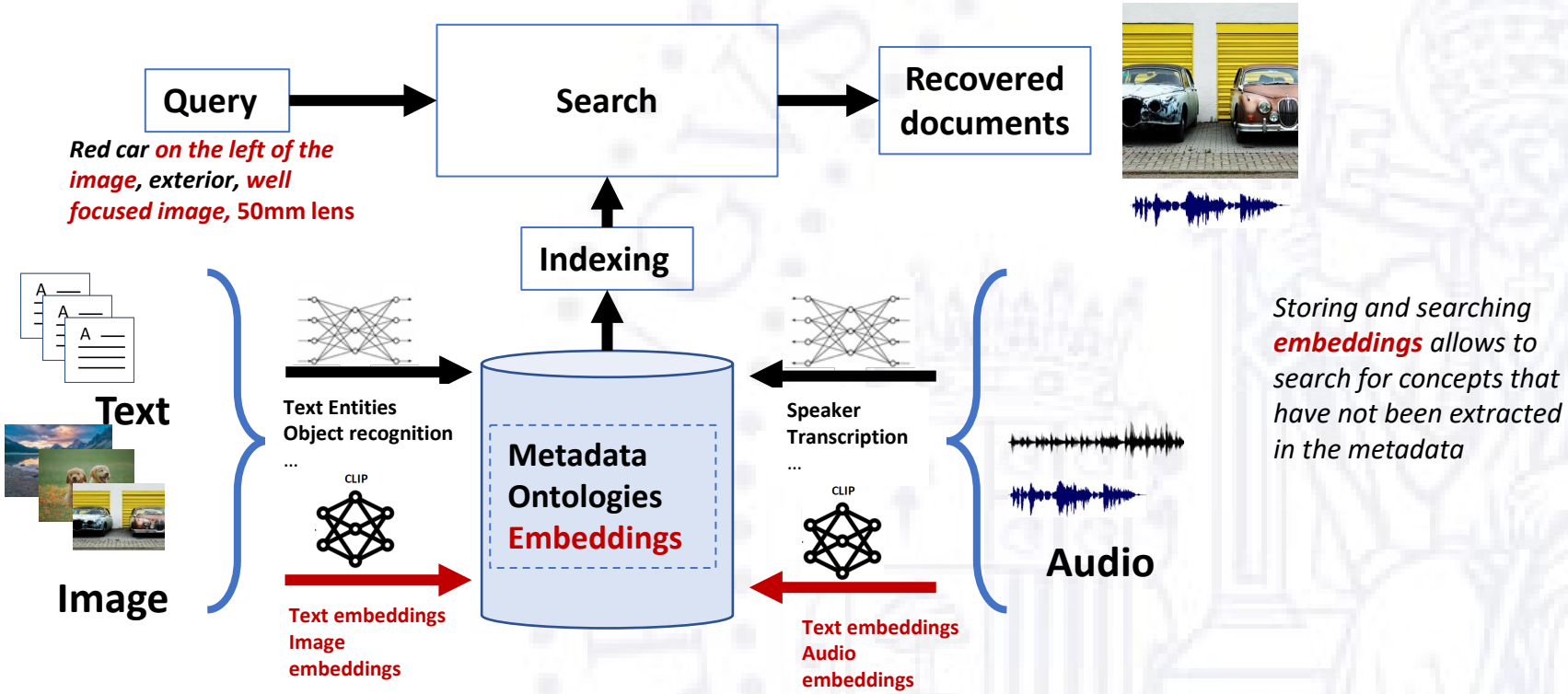


Representación de diferentes locutores en un sistema de identificación de identidad basaso en texto
Mingote, V., Miguel, A., Ortega, A., & Lleida, E. (2019). Supervector extraction for encoding speaker and phrase information with neural networks for text-dependent speaker verification. Applied Sciences, 9(16), 3295.



Aplicaciones

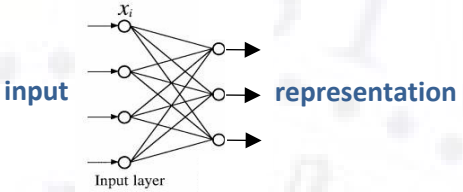
- **Búsqueda y recuperación de información multimodal:**
  - basada en embeddings



*Storing and searching **embeddings** allows to search for concepts that have not been extracted in the metadata*

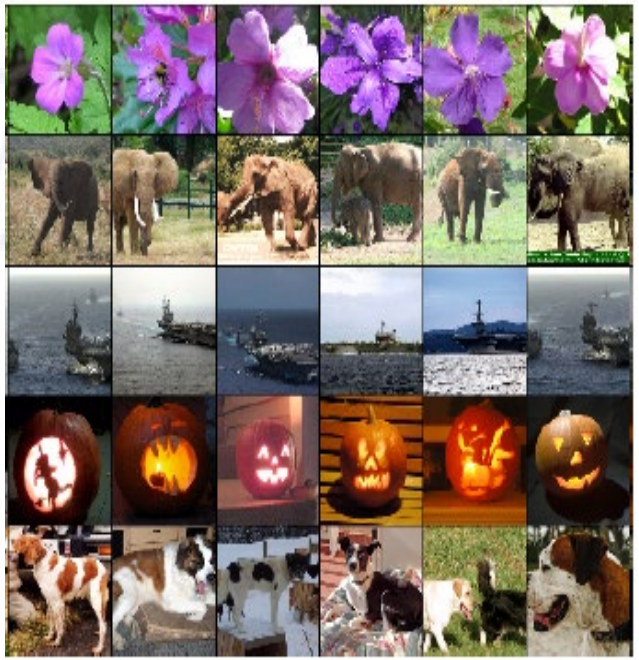
https://colab.research.google.com/drive/1J-pqo2oPDSqqvdBDg5LN82MRRCFMx1gi?usp=sharing
https://colab.research.google.com/drive/1rw1pg_lVbgUwKwtMin17cl4VFUPrfAoI?usp=sharing

Representaciones

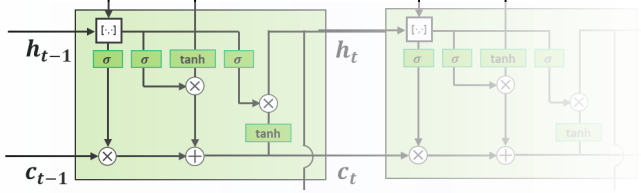


- **Representation learning**
  - Podemos utilizar representaciones internas de la red para buscar imágenes similares

Buscar imágenes similares ...



Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

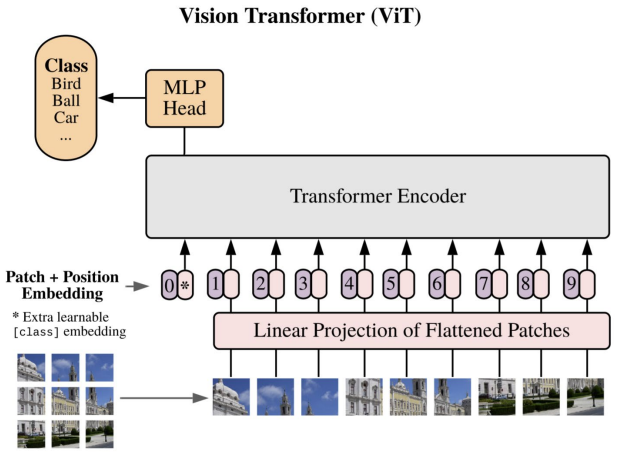


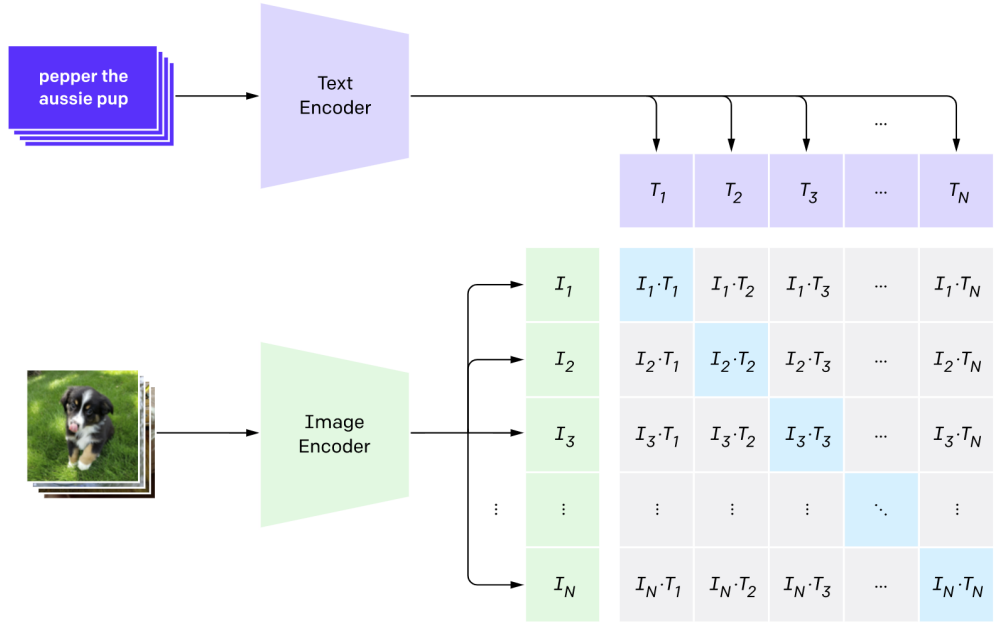
Aplicaciones

- **CLIP (Contrastive Language-Image Pre-Training)**

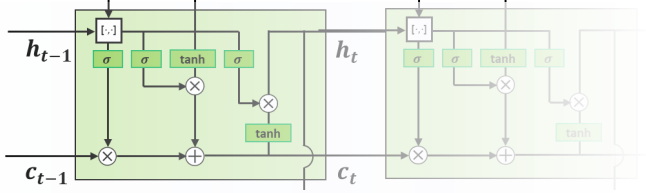
  Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020.

- Permite relacionar la representación de diferentes modalidades por ejemplo texto e imagen
- Comparación entre imágenes y textos para entrenar:





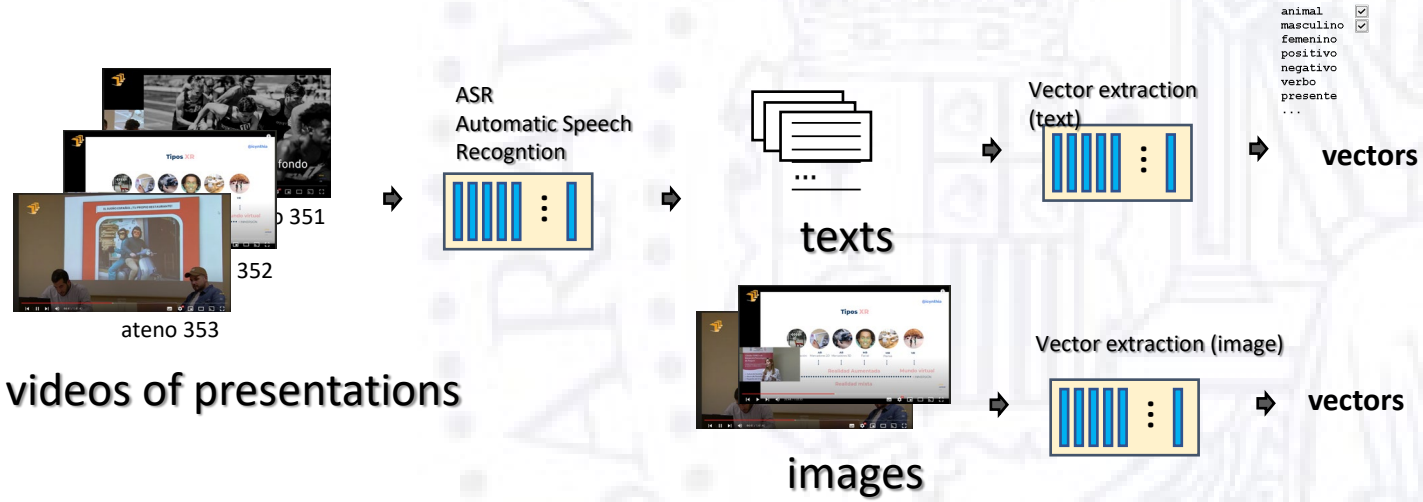
https://openai.com/blog/clip/

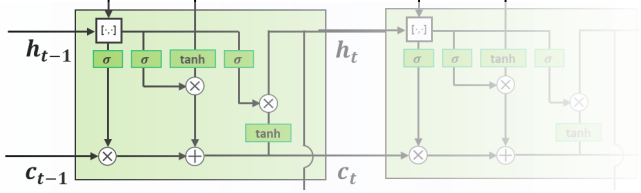


Aplicaciones

Búsqueda: charlas Ateneo EINA

- Transcribimos las reuniones mediante un sistema de última generación (basado en transformadores).
- Obtenemos **representaciones vectoriales** de los textos y de sus resúmenes (transformador).
- Al generar los resúmenes, muchos errores de reconocimiento se corrigen gracias al contexto.





Aplicaciones

Búsqueda: charlas Ateneo EINA

- **Ejemplo de búsqueda de texto a imagen**

'image with a person standing next to a slide with hands raised'

- Ateneo 275, título/autor/año: LA AERONAÚTICA DESDE UNA NUEVA PERSPECTIVA DE GÉNERO.
Alejandro Ibrahim Perera. 01/02/2017. data/es/ateneo/90rMFE6Fi6M/90rMFE6Fi6M.img/r200154.jpg

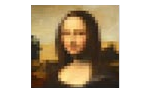
- Ateneo 260, título/autor/año: ATAPUERCA Y EVOLUCIÓN HUMANA: ¿CÓMO SABEMOS QUIÉN ES QUIÉN Y CUÁL ES SU ANTIGÜEDAD?.
Gloria Cuenca Bescós. 18/11/2015.

- Ateneo 306, título/autor/año: EMPRENDIMIENTO EN LA ERA DE LA TRANSFORMACIÓN DIGITAL.
Ana Monreal Vidal. 4/12/2019

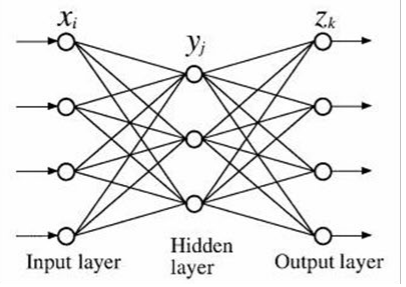
- Ateneo 329, título/autor/año: SERVICIOS E INFRAESTRUCTURA DE AWS Y LA PRÓXIMA REGIÓN EN ARAGÓN.
Javier Ramírez. 15/12/21 .

Generación

- **Generación**
  - Utilizando vectores de representación o **embeddings** como entrada
  - Generar imágenes/sonidos/texto



input

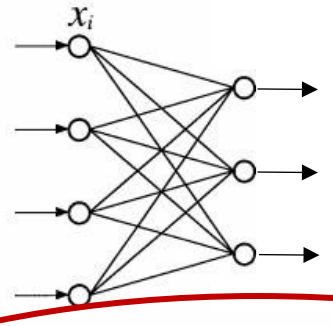


prediction

Mona Lisa

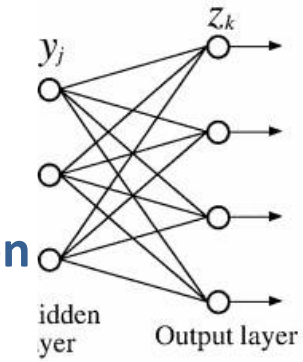
??

input



representation/ embedding

representation

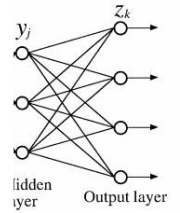


prediction

Mona Lisa

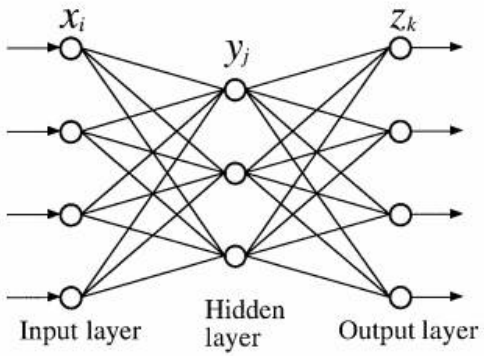
Generación

representación

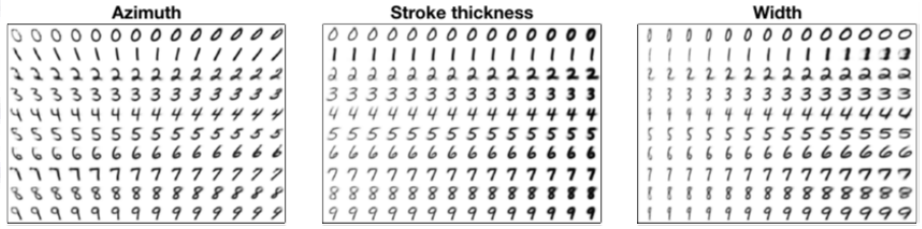


imagen

- **Generación**
  - Podemos aprender a manipular las imágenes
    - ¿Qué pasa si cambio la representación para conseguir otra imagen distinta?



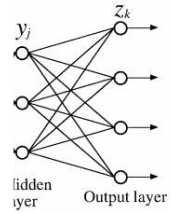
representación



Antoran, J., & Miguel, A. (2019, December). Disentangling and Learning Robust Representations with Natural Clustering. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 694-699). IEEE.

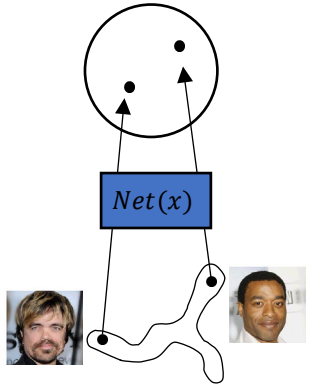
Generación

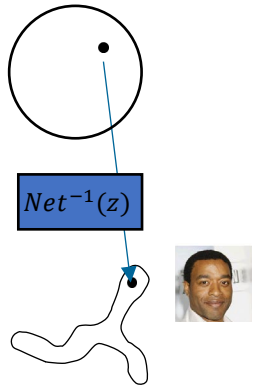
representación

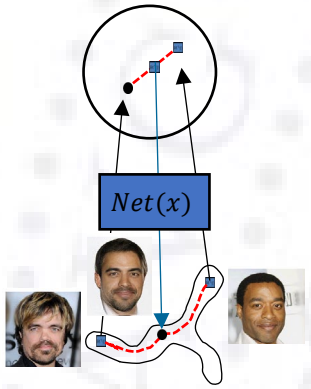


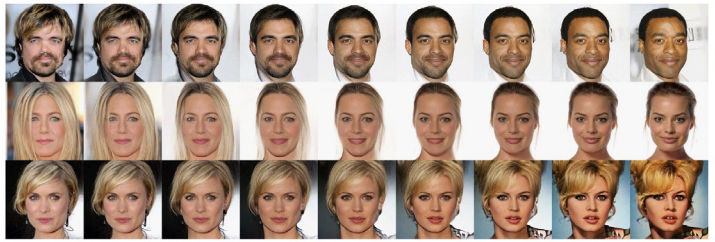
imagen

- **Generación**
  - **Podemos aprender a manipular las imágenes**
    - ¿Qué pasa si cambio la representación para conseguir otra imagen distinta?





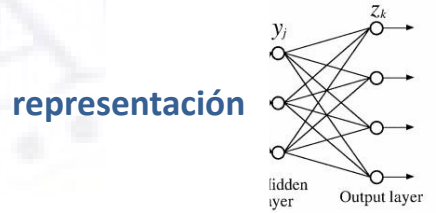




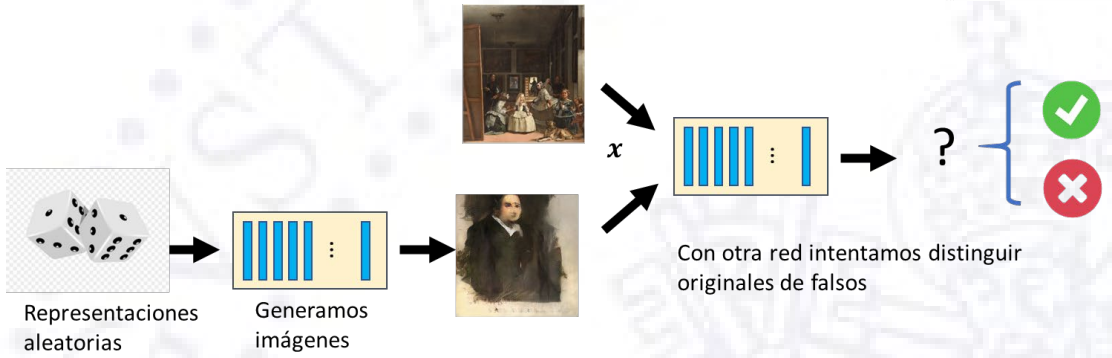
Generación de nuevas imágenes que nunca han existido

Kingma, D. P., & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. In Advances in neural information processing systems (pp. 10215-10224). Kingma, D. P., & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. In Advances in neural information processing systems (pp. 10215-10224).

Generación



- **Generación**
  - Hay modelos en los que directamente se aprende a generar imágenes
  - **Generative adversarial network**

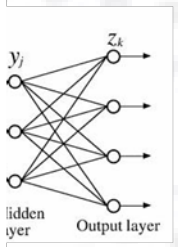


Goodfellow et al., 2014; Radford et al., 2016; Liu & Tuzel, 2016; Karras et al., 2018; Karras et al., 2019; Goodfellow, 2019; Karras et al., 2020, Karras 2021

StyleGAN3 (Karras 2021)

Generación

representación



imagen

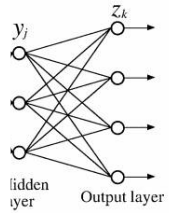
- **Detección de imágenes generadas: Deep fakes**

¿Qué imagen es artificial?
One hour of imaginary celebrities
https://www.youtube.com/watch?v=36lE9tV9vm0

Diversos Deep fakes y “reenactments” virales de los últimos años

Generación

representación



imagen

Generación: Modelos de difusión

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. arXiv preprint arXiv:2006.11239

Podemos añadir ruido hasta que no se reconozca la imagen

Con una red aprendemos a “limpiar” ese ruido

Son capaces de generar imágenes tan realistas como los GANs

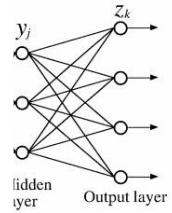
the angel of air. unreal engine @arankomatsuzaki

treehouse in the style of studio Ghibli animation @danielrussruss

A wooden spanish laptop of 1650 found the library of El Escorial

Medieval 1230 book page illustrating monks playing basketball

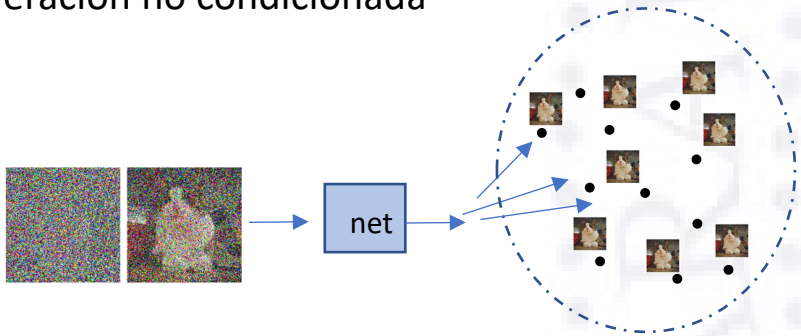
Generación



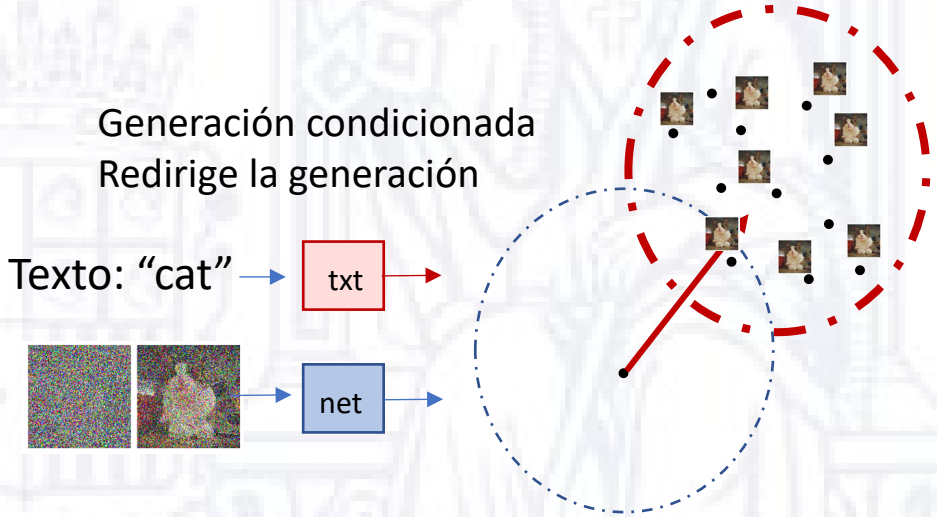
- **Generación guiada por texto**

Cómo incorporamos información del texto o de una clase?

Generación no condicionada

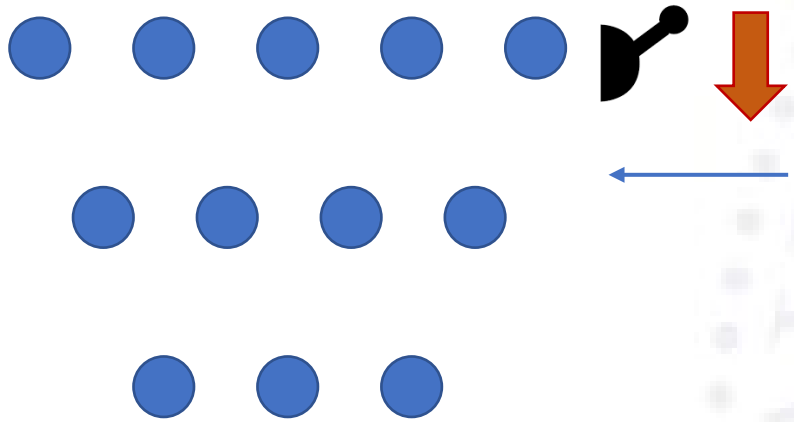


Generación condicionada
Redirige la generación



Aprendizaje

- **Aprendizaje:**

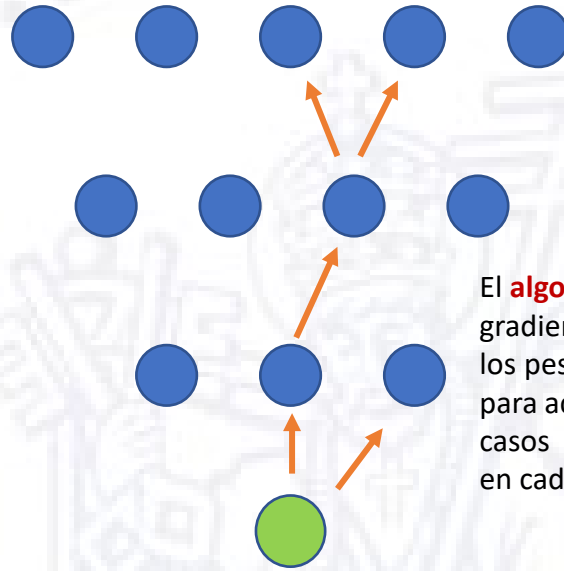


Un modelo de deep learning actual puede tener desde unos pocos millones de parámetros a **miles de millones**!!

Se podría probar prueba y error hasta que se encontrara alguna buena combinación de todas las palancas … pero tardaríamos demasiado

Aprendizaje

- **Aprendizaje Automático/Machine Learing**
  - Proceso de corrección
    - **Milllones de veces** en redes grandes
  - Hay que disponer de datos y respuestas, **coste etiquetado?**
    - Corpus, bases de datos
    - Miles o millones de ejemplos con su etiqueta
  - Problema **sesgos** en los datos
    - Si mostramos más veces un ejemplo aparecerá un sesgo en el sistema



El **algoritmo** de gradiente corrige los pesos ajustables para acertar más casos
en cada iteración

Todos nosotros etiquetamos / Instalaciones especializadas

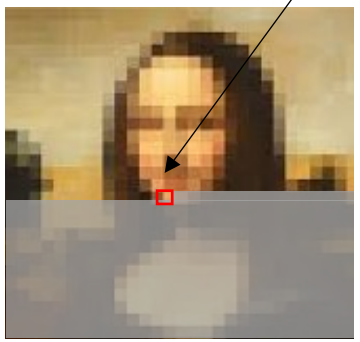
Múltiples denominaciones
- Entrenamiento
- Ajuste
- Estimación
- Aprendizaje

Aprendiendo sin etiquetas

Aprendiendo sin etiquetas: no supervisado

- Podemos conseguir que los sistemas automáticos comprendan los datos **forzando a que hagan predicciones** sobre lo que no han visto

Viendo los pixels anteriores: ¿cómo es el siguiente ?

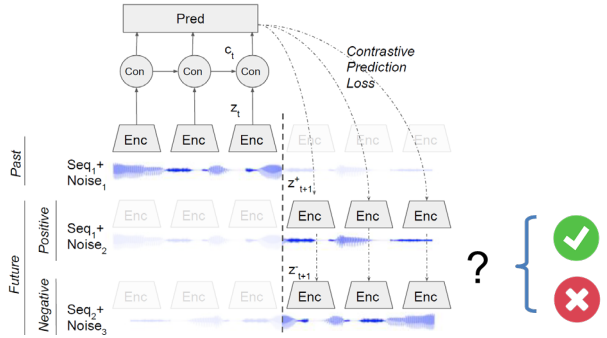


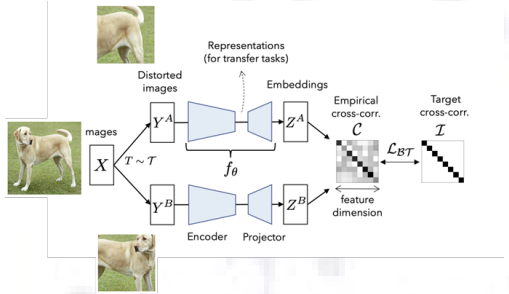
A Oord, N Kalchbrenner, O Vinyals, L Espeholt, A Graves, K Kavukcuoglu
Conditional Image Generation with PixelCNN Decoders 2016

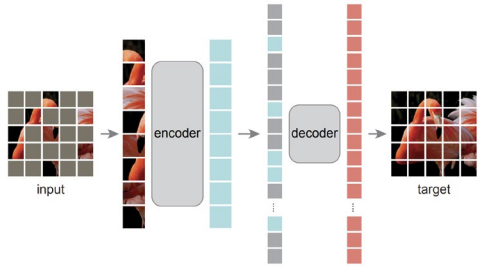
- **Clustering:** Agrupamos los datos por parecido
  - por ejemplo color similar

Aprendiendo sin etiquetas

- **Aprendiendo sin etiquetas: no supervisado**



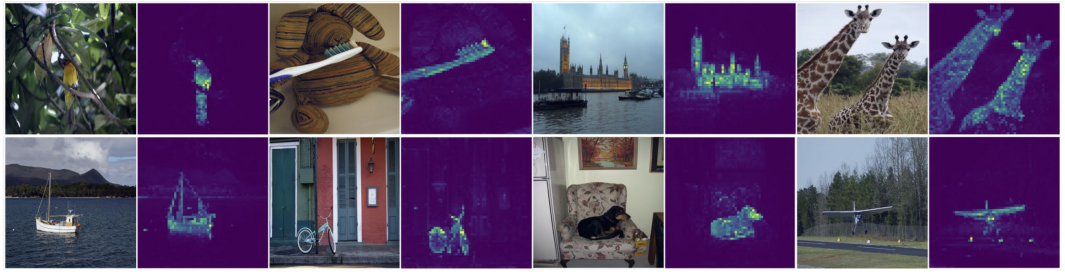




Una estrategia es dar varias opciones
como si fuera un examen

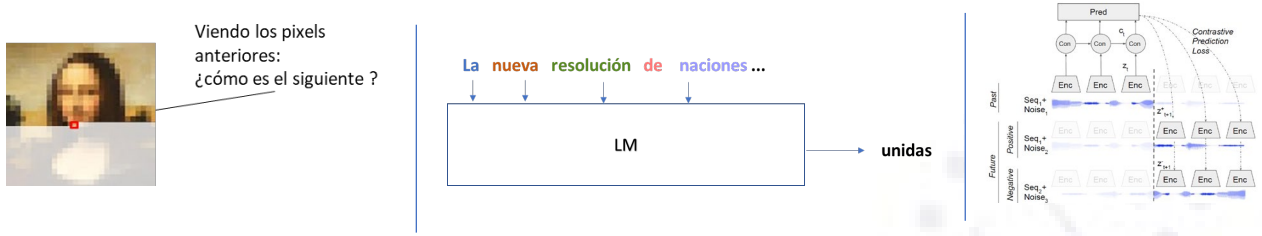
Resolver la pregunta:
¿ son partes de la misma imagen ?

Reconstruir la imagen
a partir de una con oclusiones

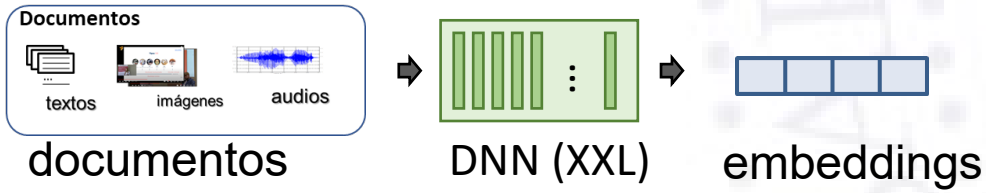


Aprendiendo sin etiquetas

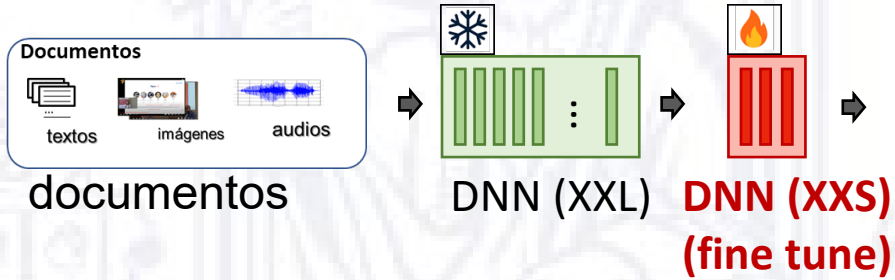
- **Entrenamiento (autosupervisado/no supervisado)**
  - **Usado en texto, imagen, audio …**



  - **Se diseñan redes de gran tamaño**
    - Aprende a representar los datos / embeddings
    - Coste computacional alto
    - Entrenar un LLM puede suponer usar cientos o miles de GPUs durante semanas (costes de millones de $)
  - **Autosupervisado / No supervisado**
    - Lo que lo hace viable es que **no se necesitan etiquetas**
    - Gran cantidad de textos recopilados de distintas fuentes:
      - Wikipedia, Periódicos, Web, Github



- **Reajuste / Fine tuning (supervisado)**
  - **Disponiendo de un conjunto pequeño de etiquetas**
    - Podemos aprovechar las representaciones aprendidas
    - Se reajusta una pequeña parte de la red o una red adicional de pequeño tamaño



Múltiples denominaciones
- Reajuste
- Adaptación
- Ajuste fino
- Fine tuning

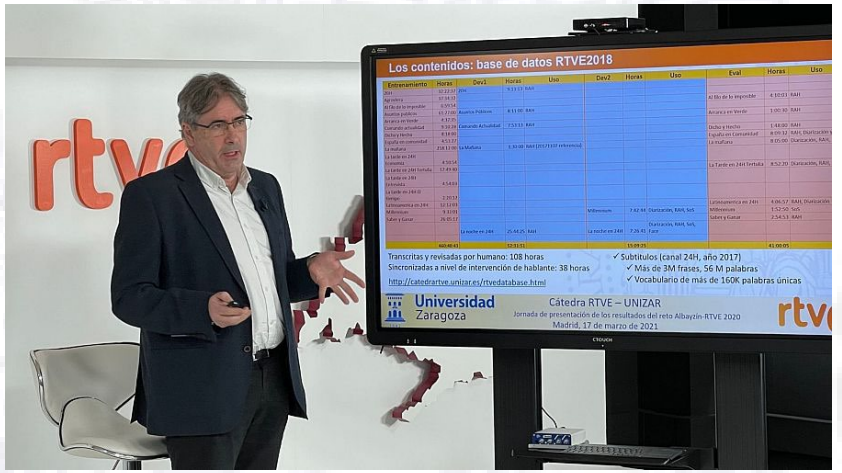
Conclusiones

- **Comunicar, comparar y aprender**
  - Congresos y conferencias cientifico-técnicas como Iberspeech / Interspeech / ICCASP
    - ICCASP 2025
      - 3,200 papers 2,661 participantes
    - NEURIPS 2025
      - 5,290 papers 26,382 participantes (>21000 enviados)

Conclusiones

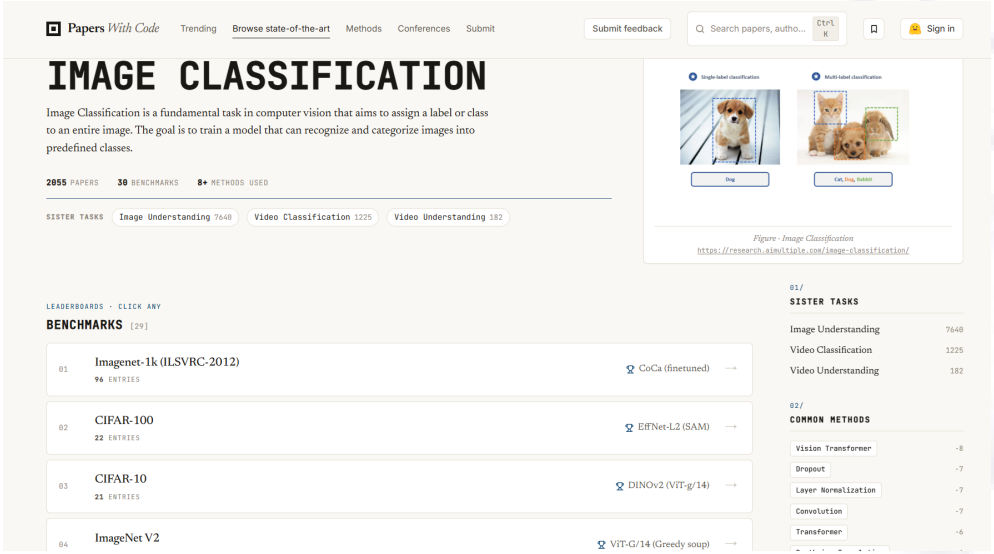
- **Comunicar, comparar y aprender**
  - Congresos y conferencias cientifico-técnicas como Iberspeech / Interspeech / ICCASP
  - Los **retos Albayzín-RTVE** 2018, 2020, 2022, 2024 organizados por la Cátedra RTVE de la Universidad de Zaragoza

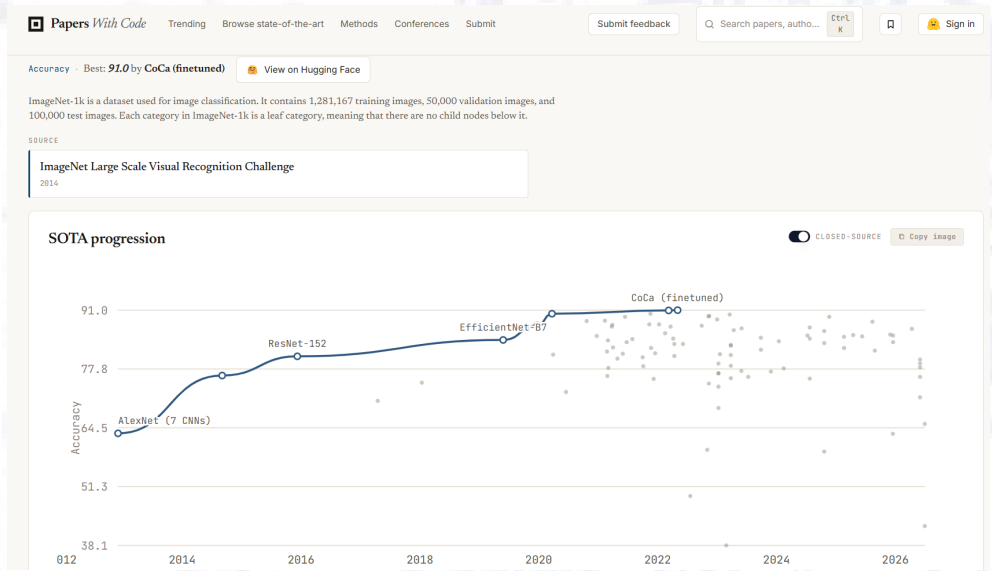
- Diagnóstico Tecnológico Actual
  - Han permitido evaluar el **nivel de madurez** real de la Inteligencia Artificial en español.
  - Medición de la precisión de los sistemas e identificación de las limitaciones actuales en **entornos audiovisuales reales**.
- Fomento de la Innovación Abierta
  - Han **conectado** la investigación universitaria con las necesidades de la industria.
  - Han acelerado la **transferencia tecnológica** mediante la competición colaborativa.



Conclusiones

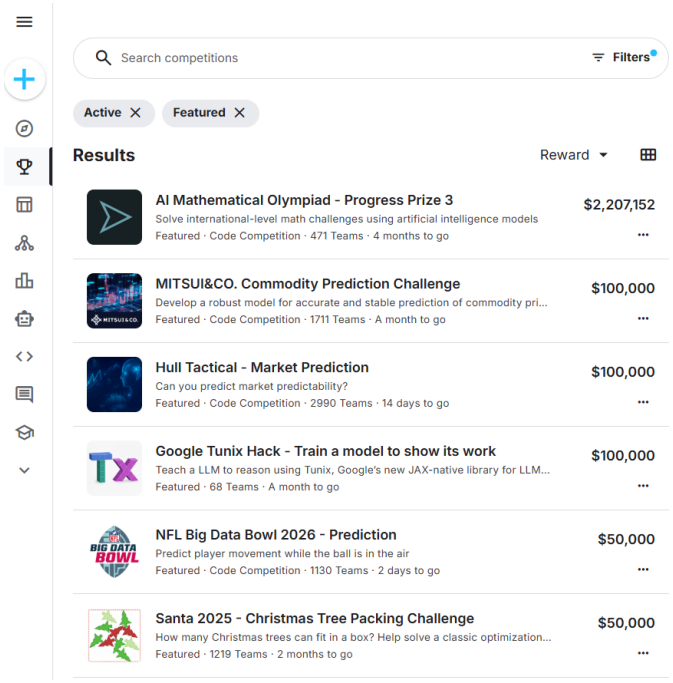
- **Comunicar, comparar y aprender**
  - Congresos y conferencias cientifico-técnicas como Iberspeech / Interspeech / ICCASP
  - Los **retos Albayzín-RTVE** 2018, 2020, 2022, 2024 organizados por la Cátedra RTVE de la Universidad de Zaragoza
  - https://paperswithcode.co/benchmark/imagenet-1k?task=image-classification
    - Repositorio de papers y tareas para comparar resultados
    - También en https://opencodepapers-b7572d.gitlab.io/

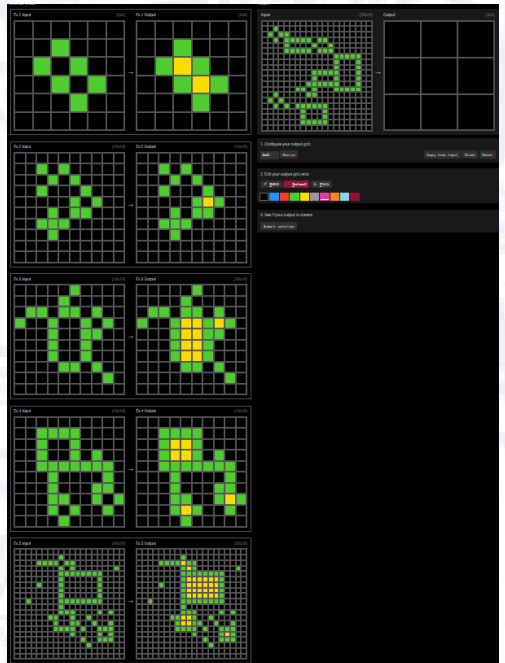




Conclusiones

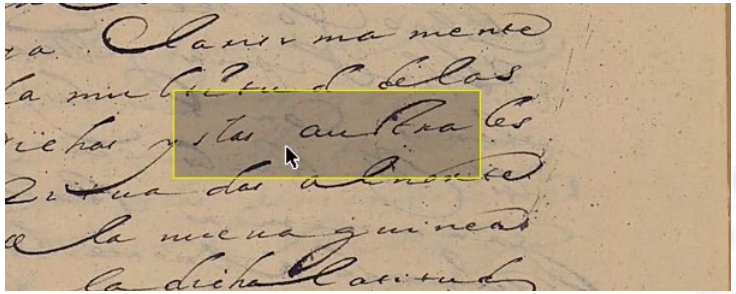
- **Comunicar, comparar y aprender**
  - Congresos y conferencias cientifico-técnicas como Iberspeech / Interspeech / ICCASP
  - Los **retos Albayzín-RTVE** 2018, 2020, 2022, 2024 organizados por la Cátedra RTVE de la Universidad de Zaragoza
  - https://www.kaggle.com/competitions
    - Organización de retos con o sin premio económico
    - Se ha convertido en un referente en el intercambio de ideas y soluciones





Conclusiones

- **La inteligencia artificial halla rastros del descubrimiento español de Australia**



Un grupo de investigación en la UPV lleva años desarrollando sistemas de reconocimiento de texto manuscrito antiguo.

¿Se puede usar ya la tecnología?

- En un texto concreto un experto es más fiable
- La tecnología actual puede permitir **buscar**
- "escalar" un sistema básico permite hacer frente a documentos que no podrían ser tratados.
- **El Archivo General de Indias**, tiene 80 millones de páginas que no se han procesado en su totalidad.
- **Objetivo:** **asistir al profesional**