

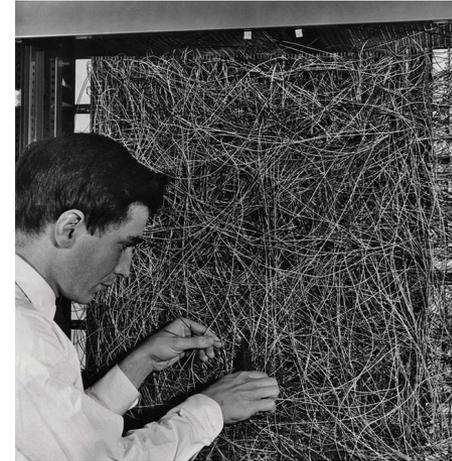
Introducción a deep learning

antecedentes

Máquinas electrónicas

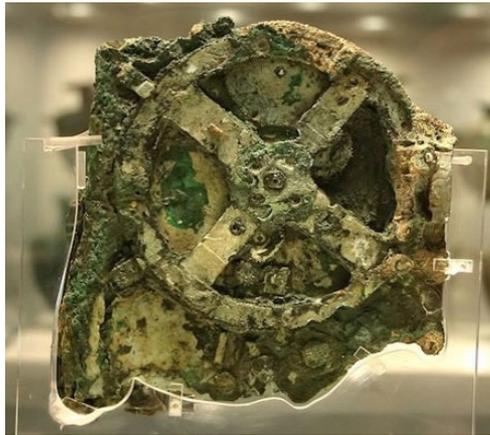
Mark I, 1944

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)



antecedentes

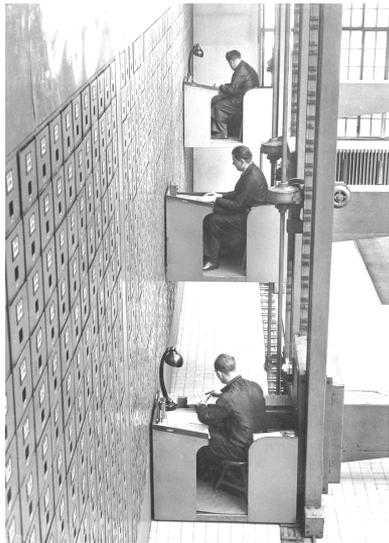
Velocidad de almacenamiento

Disco duro 2000 18GB (48MB/s)
HD estado sólido 2021 1TB (7000 MB/s)



Almacenamiento

Sistema mecánico 1937 (República Checa)



Capacidad de almacenamiento

Cinta perforada 1970 <1 KB
Disco 3 1/2 1987 1.4 MB
DVD 1995 4.7 GB



antecedentes



*Proceedings paper 1995 /
Revistas paper*



Buscadores internet 1998



TensorFlow

PYTORCH

*Software gratuito
y
Toolkits 2010*

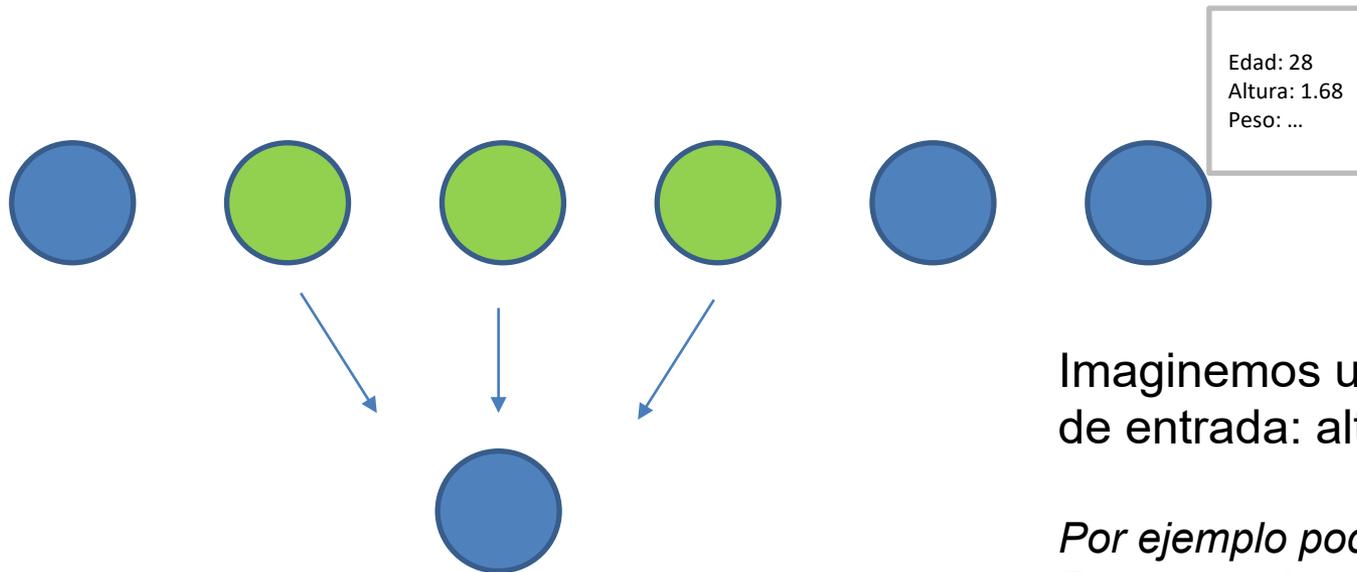


GitHub

*2008 Redes sociales /
plataformas de desarrollo colaborativo*

fundamentos

- Procesado en fases, capas

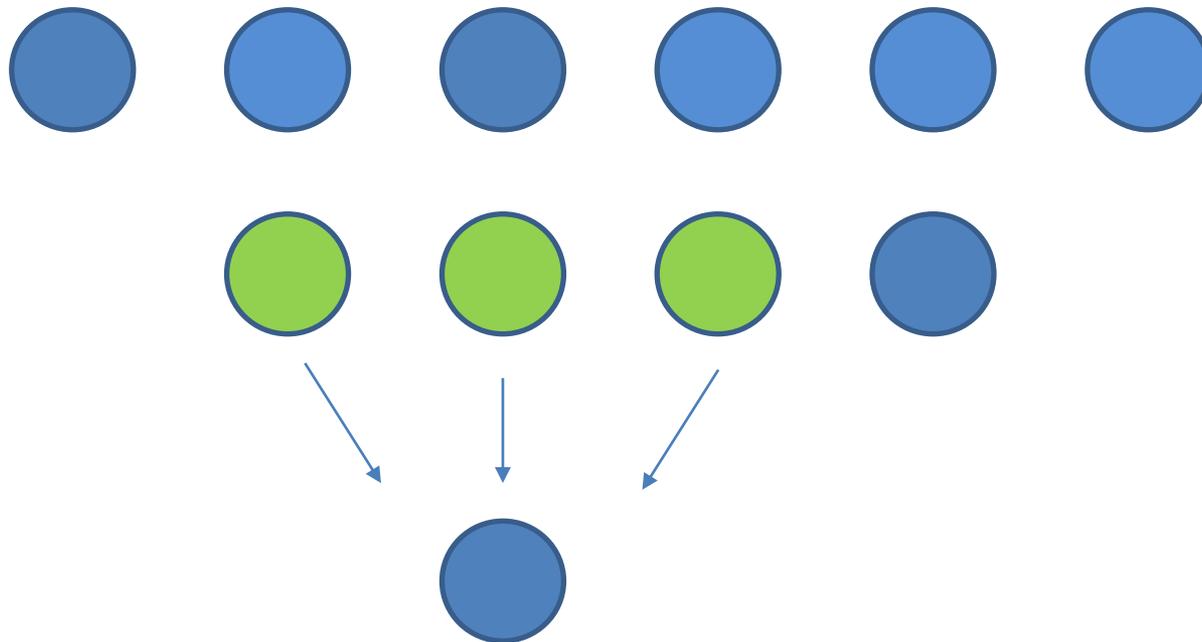


Imaginemos unos datos de entrada: altura, edad,..

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

fundamentos

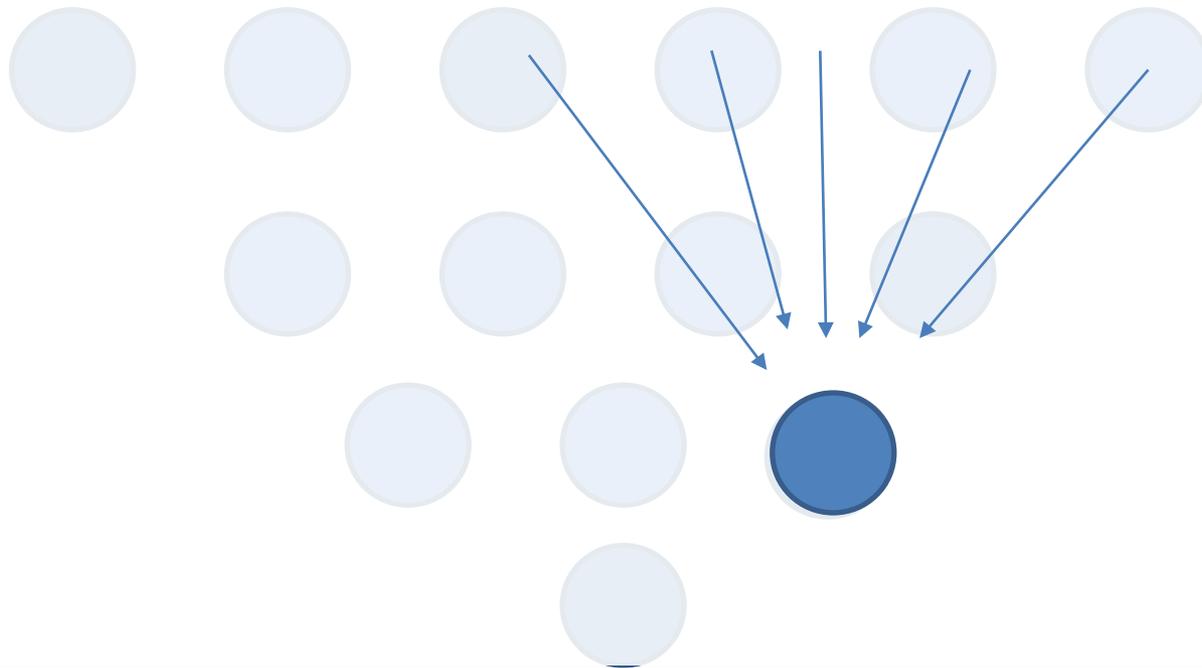
- Procesado en fases, capas



Y repetimos,
en todas las
filas...

fundamentos

- Procesado en fases, capas

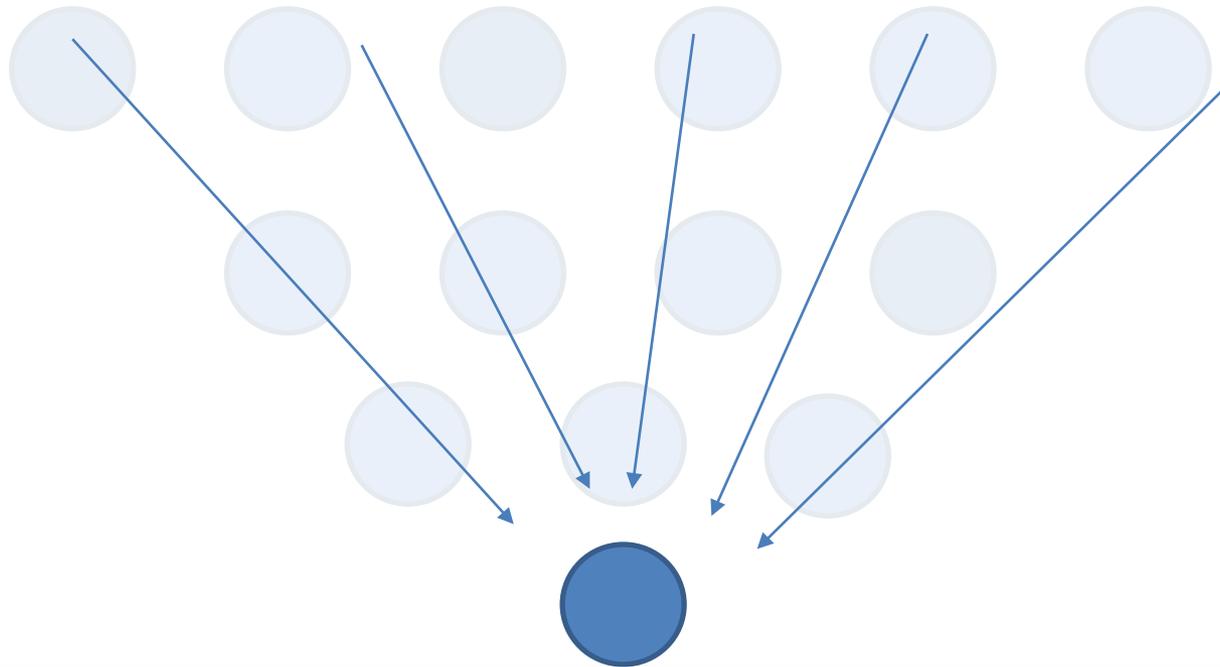


Algunos nodos

reciben
parte de la
información

fundamentos

- Procesado en fases, capas

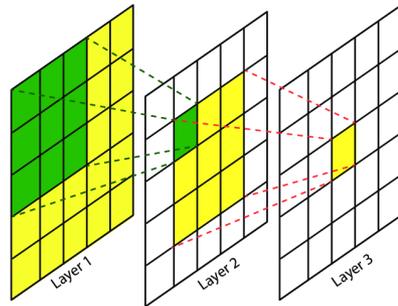


Ahora al último le podríamos preguntar, **¿quién es el más joven ?** Ha recibido **toda** la información

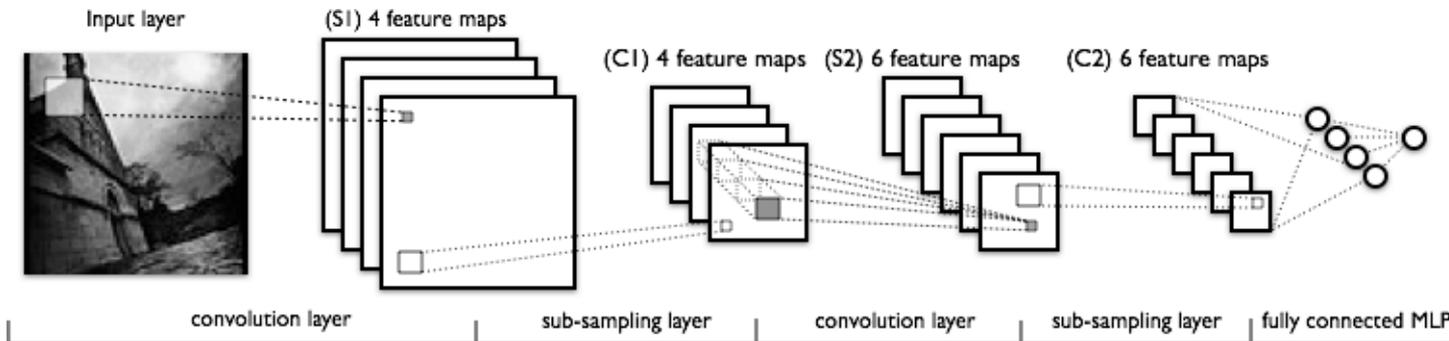
fundamentos

- Redes convolucionales

- Cada capa suma varios valores de entrada con distinto peso, normalmente 9 entradas: 3×3



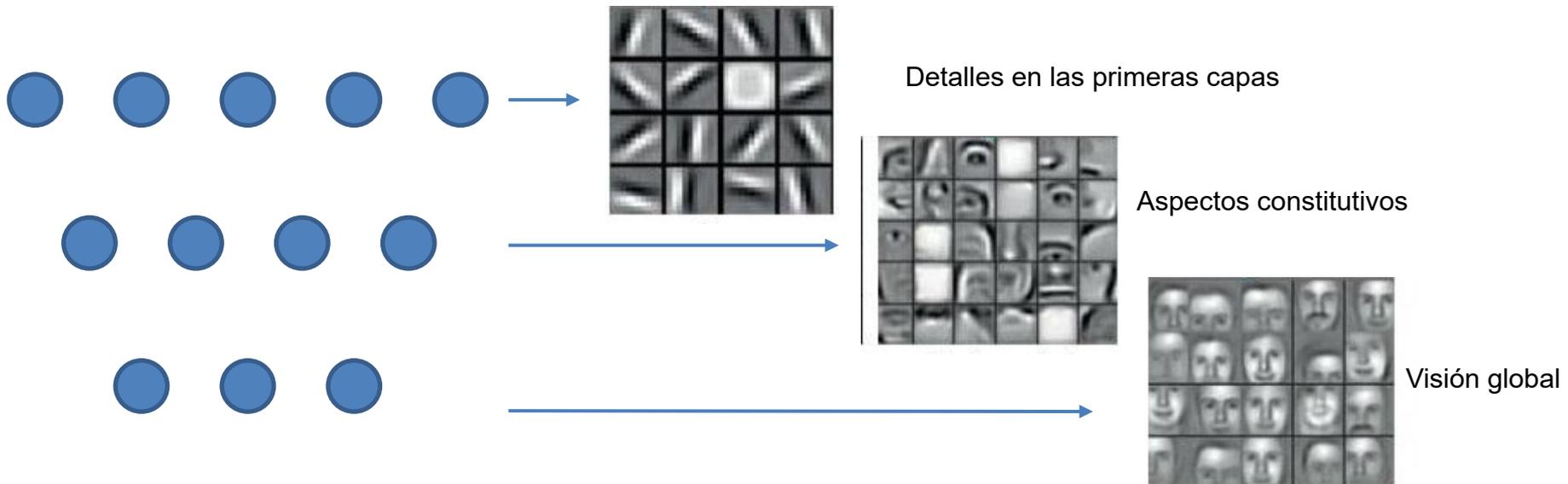
Yann LeCun



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

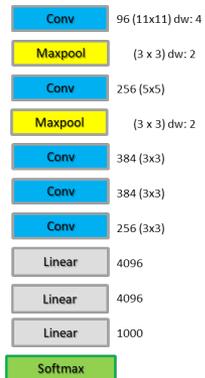
- Redes convolucionales
- Con mayor **profundidad (*depth*)** se logra mayor **abstracción**
 - Las primeras redes profundas tenían 7 capas
 - Hoy en día en cuestión de minutos se tiene acceso a redes de más de 100 capas ya entrenadas



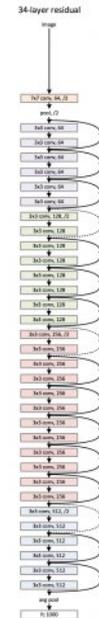
fundamentos

- **Redes convolucionales**

- *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*



Geoffrey Hinton



2012: 7 capas
84.6 % aciertos

2014: Inception 25 capas
93.3% aciertos

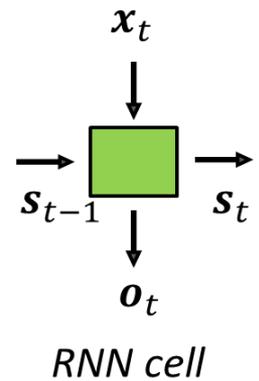
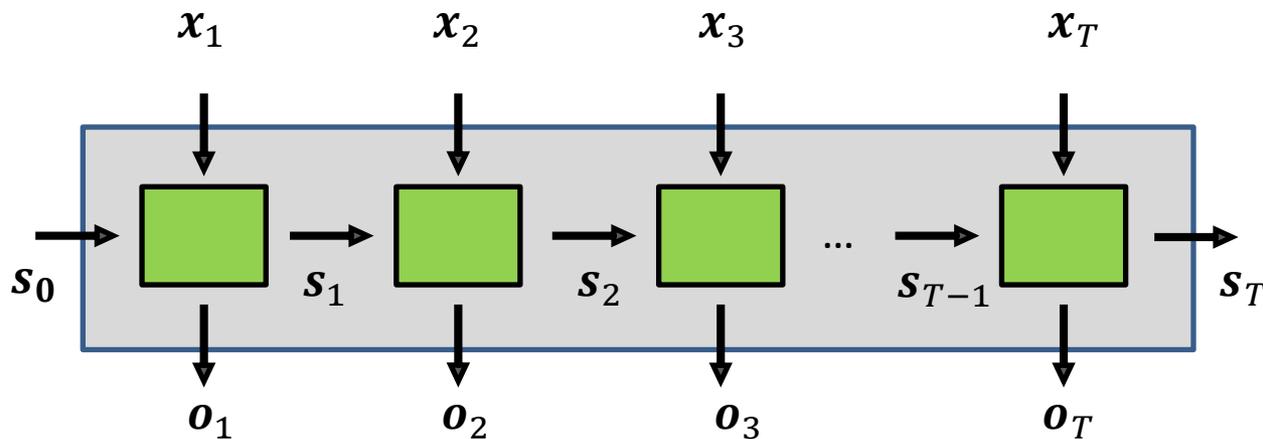
2015: Resnet >100 capas
96.43% aciertos

fundamentos

- **Redes recurrentes**

- 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 tiene una memoria finita para recibir información de los instantes previos y escribir nueva información para el futuro*



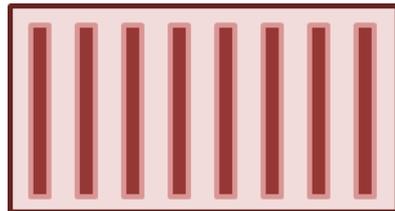
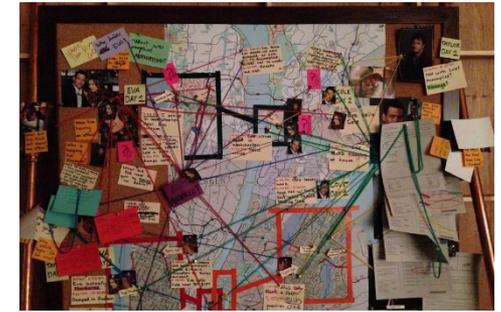
RNN

fundamentos

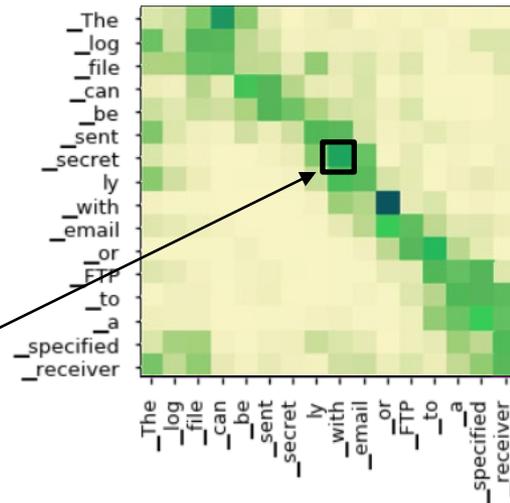
- Transformers

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 la relación de todas las entradas



=



Grado de influencia de la entrada i sobre la j

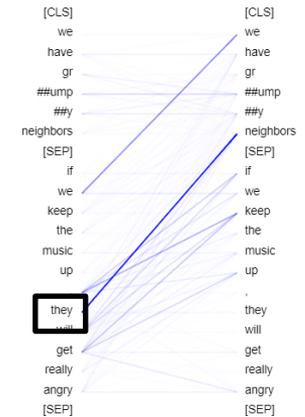
<https://nlp.seas.harvard.edu/2018/04/03/attention.html>

RNN

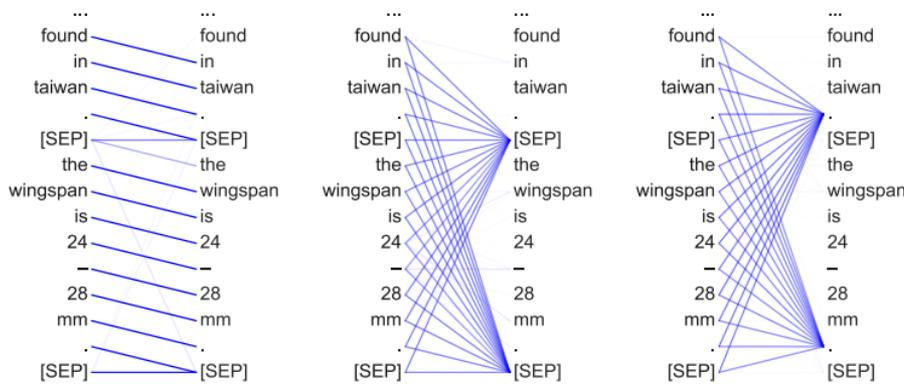
fundamentos

- Transformers

- Son capaces de analizar la relación de todas las entradas*

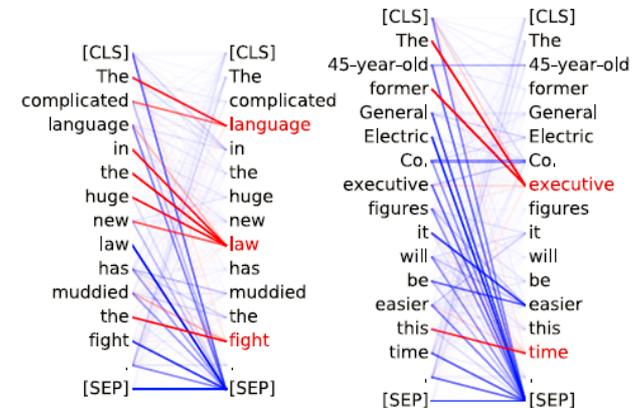


Desambiguación ellos -> vecinos



palabra anterior

Final de frase



Determinantes y modificadores de un nombre

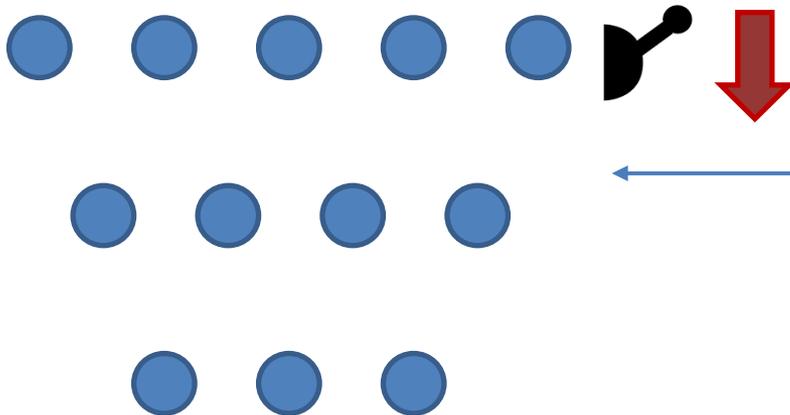
RNN

fundamentos

- **Para aprender a realizar la tarea**

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

fundamentos



• Repetir el proceso de corrección

– miles de veces



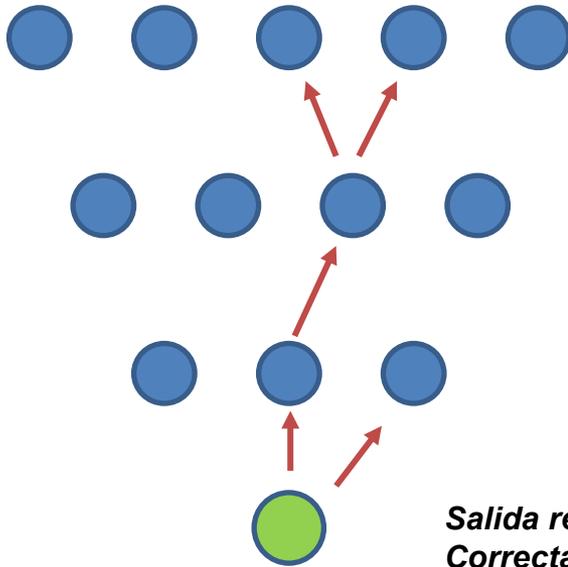
- Depende de lo complicada que sea la tarea pueden ser **millones** de correcciones

– Hay que disponer de datos y respuestas, coste

- Corpus, bases de datos
- Miles o millones de ejemplos con su etiqueta

– Problema sesgos en los datos

- Si mostramos más veces un ejemplo y la respuesta que otros ejemplos aparecerá un sesgo en el sistema



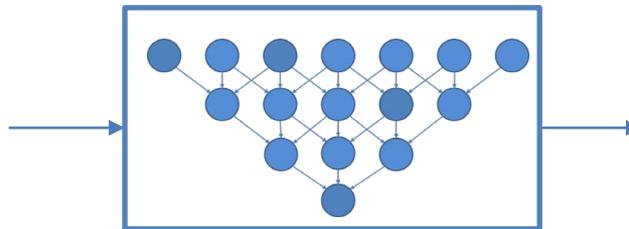
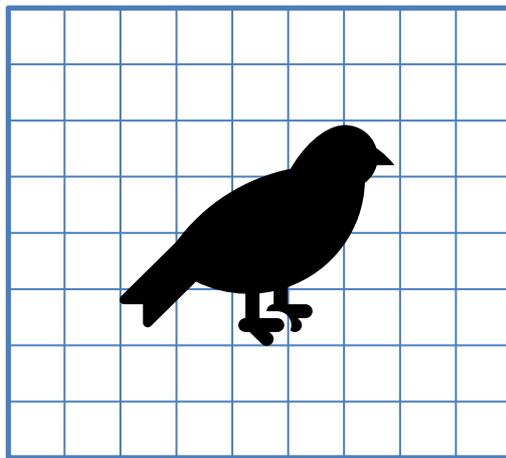
Salida red: oso 99.9%
Correcta: **gato**



Todos nosotros etiquetamos / Instalaciones especializadas

Tipos de problemas

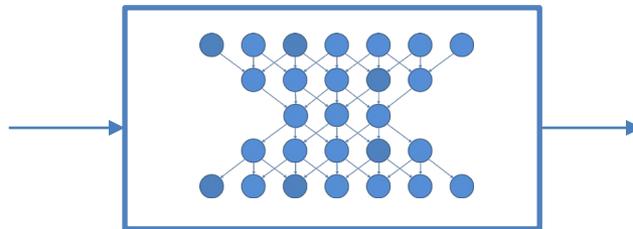
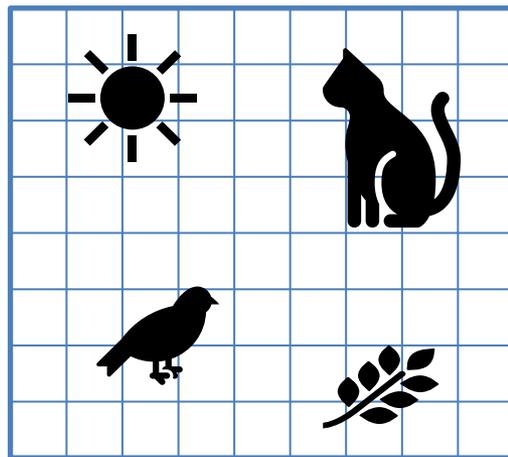
- Aunque cada día hay más variantes las dos formas principales de usar DNNs hoy en día:
 - **Clasificación:**
 - Decir **qué concepto** hay en una imagen/texto/audio



Una sola salida/concepto para toda la entrada

Tipos de problemas (2 / 4)

- Aunque cada día hay más variantes las dos formas principales de usar DNNs hoy en día:
 - **Clasificación múltiple:**
 - Decir **qué concepto** hay en cada zona/pixel: imagen/texto/audio
 - Decir **varias propiedades/conceptos** de una imagen/texto/audio

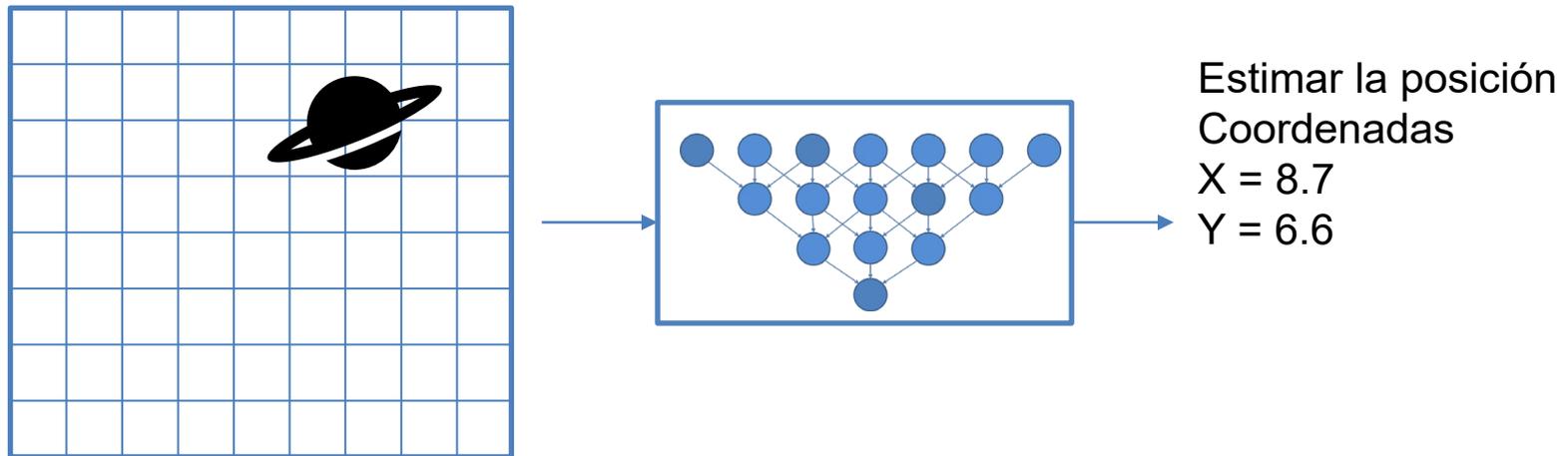


Una salida/concepto para cada pixel de entrada

sol	sol	gato	gato
sol	sol	gato	gato
pájaro			
pájaro	hoja	hoja	

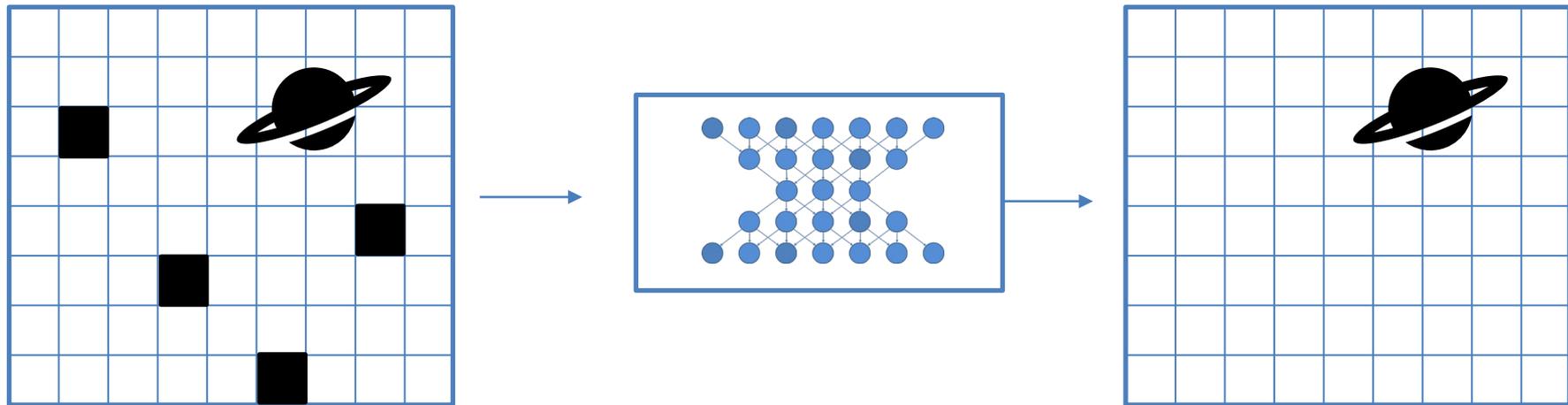
Tipos de problemas (3 / 4)

- Aunque cada día hay más variantes las dos formas principales de usar DNNs hoy en día:
 - **Regresión:**
 - Utilizar los datos para obtener algún tipo de **predicción numérica**

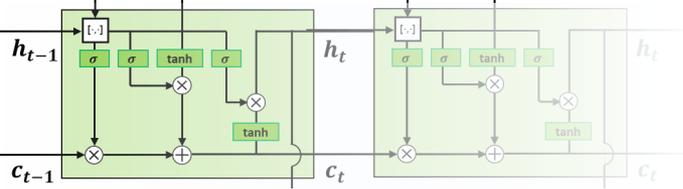


Tipos de problemas (4/4)

- Aunque cada día hay más variantes las dos formas principales de usar DNNs hoy en día:
 - Regresión múltiple:
 - Predecimos varios valores numéricos: por cada zona, pixel...

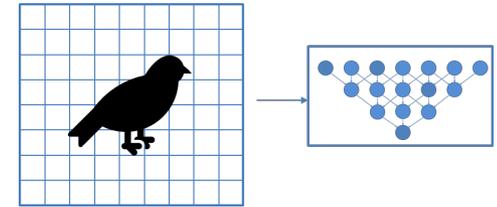


aplicaciones: análisis



– Clasificación:

- Decir **qué concepto** hay en una imagen/texto/audio

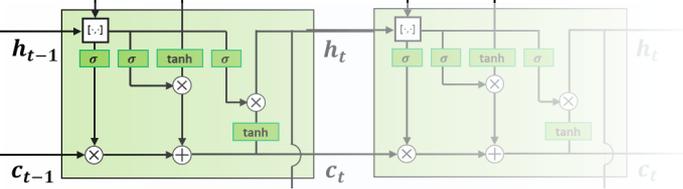


En este ejemplo la red neuronal se prepara para resolver el problema de clasificación:
¿Qué hay en esta imagen? -> 1 respuesta

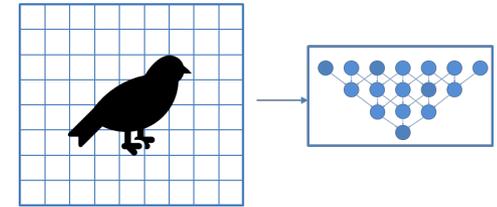
*Entre las posibles respuestas hay 120 razas de perro
En 2012 el error top5 era del 25%,
Hinton y Krizhevsky red de 7 capas 15%
Hoy en día decenas, cientos de capas, alrededor del 2%,*



aplicaciones: análisis

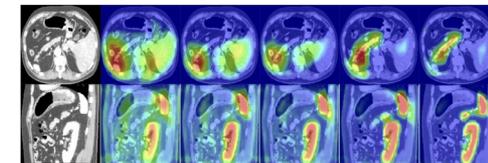


– **Clasificación:** ¿ nos podemos fiar ?

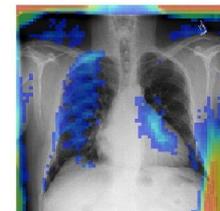


¿ cómo es ese pequeño porcentaje de fallos... ?

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

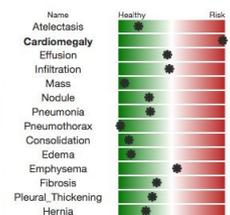


Out Of Distribution reconstruction error
Heatmap where the image varies from the training distribution.

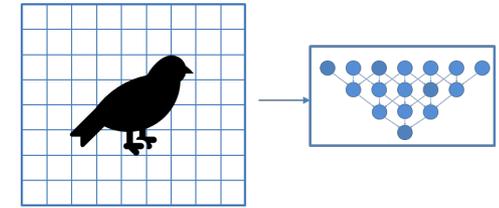
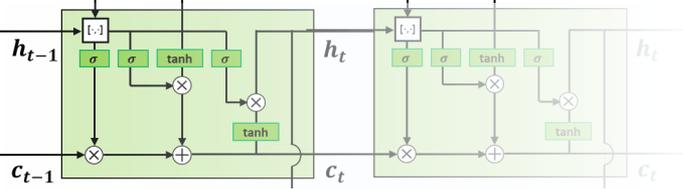


recScore:0.27, ssim:0.39

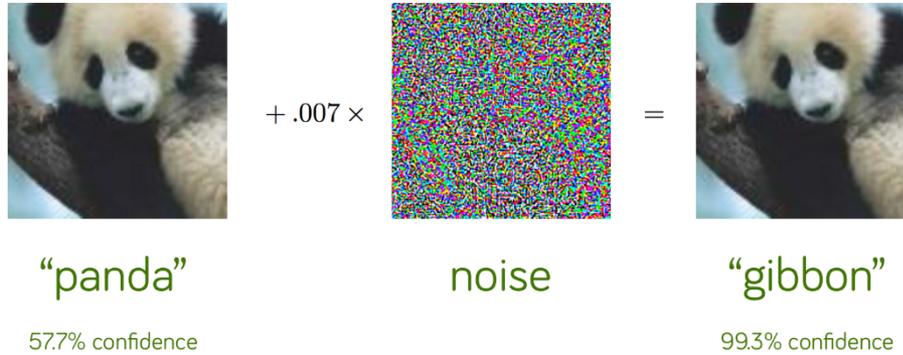
Disease Predictions
Probability of a disease.



aplicaciones: análisis



– **Clasificación:** ¿ nos podemos fiar ?



Ataques adversarios



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)

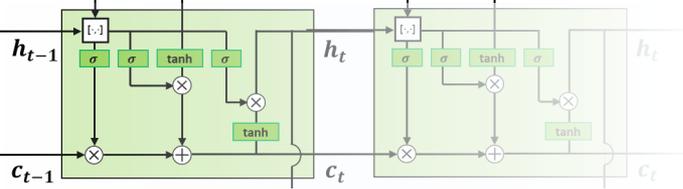


Speed Limit 80
(88% confidence)

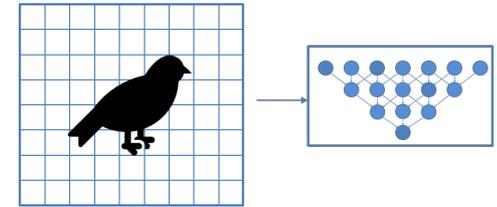
Robust physical-world attacks on deep learning visual classification.

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).

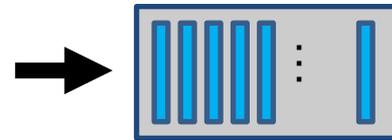
aplicaciones: análisis



– **Clasificación:** ¿ nos podemos fiar ?



paperswithcode.com

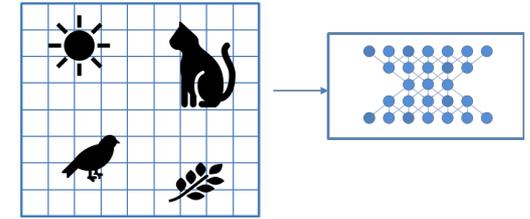
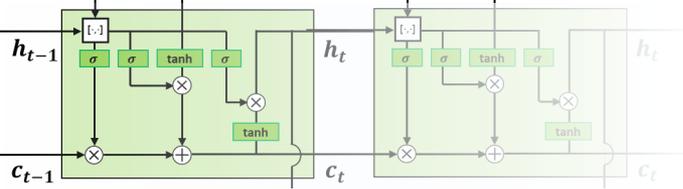


Entrenamiento

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

Se conoce como:
Aumento de datos

aplicaciones: análisis



– Clasificación múltiple:

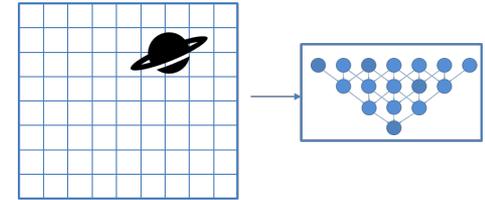
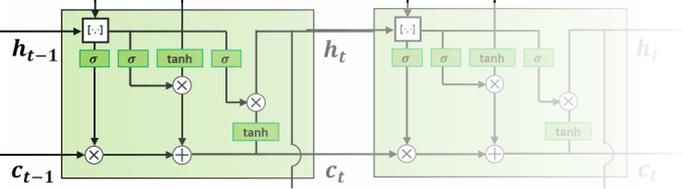
- **varias propiedades/conceptos** de una imagen/texto/audio



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

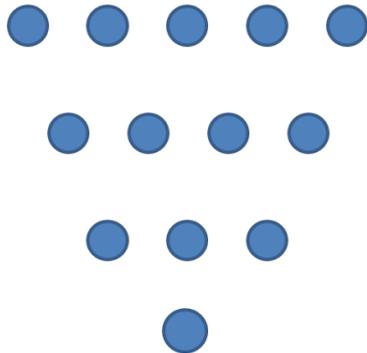
¿Hay un perro?	No
¿Hay un gato?	Sí
¿Hay árboles?	No
¿Hay un pájaro?	Sí
¿Hay cielo?	No
¿Hay hierba?	Sí

aplicaciones: análisis



– Regresión:

- Utilizar los datos para obtener algún tipo de **predicción** numérica



Edad?

En este ejemplo la red neuronal se prepara para resolver el problema:

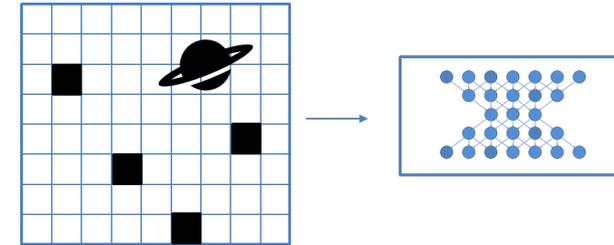
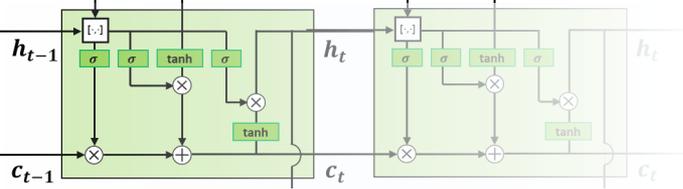
¿Qué edad tienen estas persona?

La respuesta sería un número con la edad en años

Entrenaríamos el sistema con muchas imágenes

Aplicaríamos las correcciones necesarias cuando la red se equivoca

aplicaciones: síntesis



– Regresión múltiple:

- **Transformar los datos** con alguna finalidad, que se parezcan a algo, que mejoren de calidad...



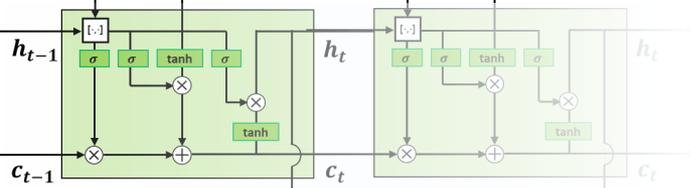
En este ejemplo la red neuronal se prepara para resolver el problema:

Convertir una imagen de BN en color



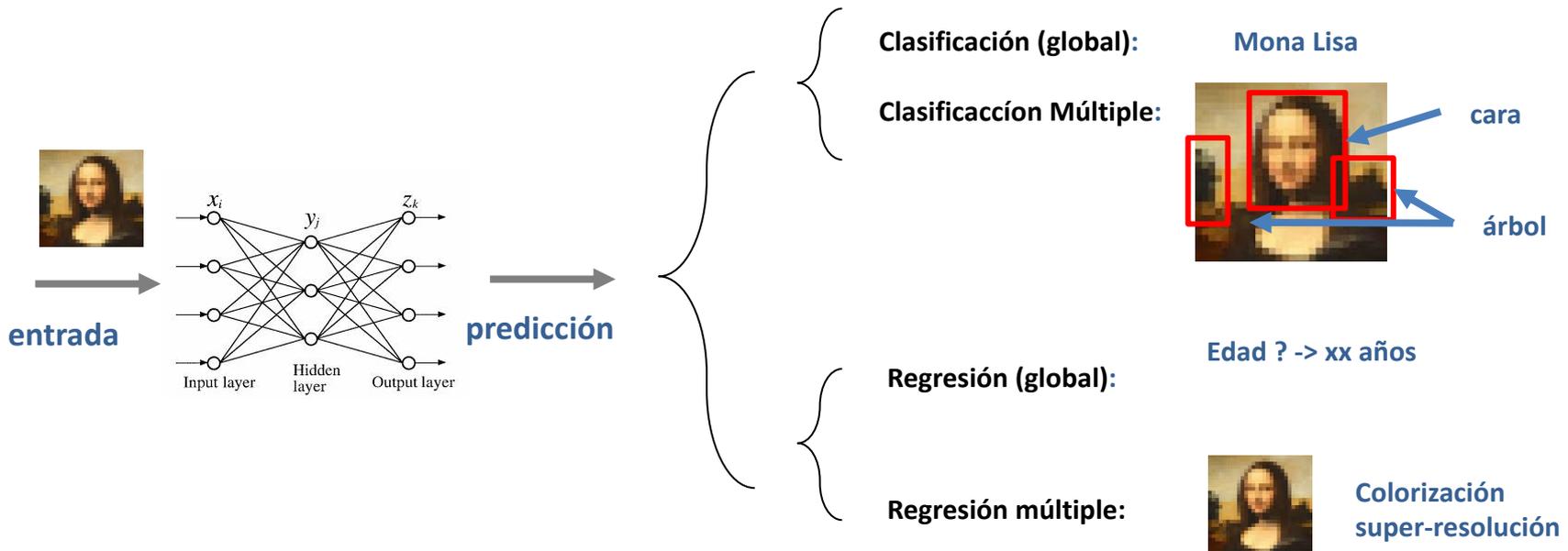
En este ejemplo la red neuronal se prepara para resolver el problema:

Mejorar la calidad de la imagen

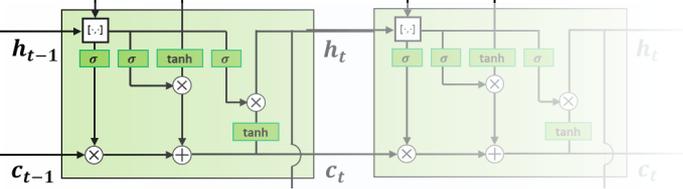


acciones

Resumen: Aprendizaje supervisado



aplicaciones: análisis

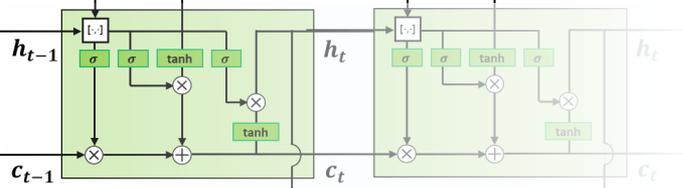


Representation learning

- Podemos utilizar representaciones internas de la red
 - Objetivo comparar imágenes/sonidos/textos



aplicaciones: análisis

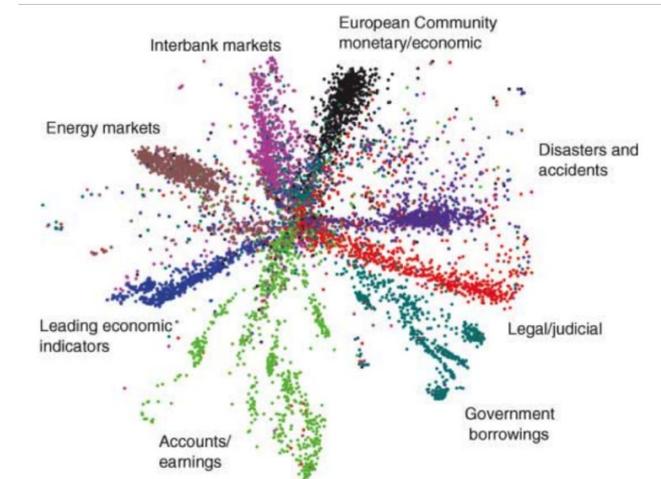


■ Representation learning

- Podemos utilizar representaciones internas de la red
 - imágenes/sonidos/textos similares están más próximos en ese espacio

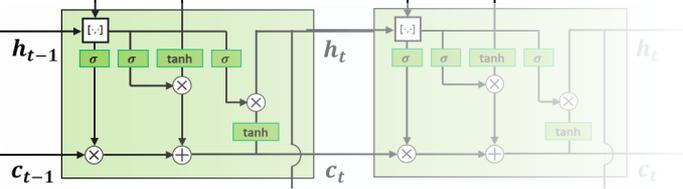


*Handwritten digits,
A. Karpathy, Stanford University*



*Text topic classification
G. Hinton, Toronto University*

aplicaciones: análisis



Imagenet classification with deep convolutional neural networks

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).. *Advances in neural information processing systems*, 25.

■ Representation learning

- Podemos utilizar representaciones internas de la red

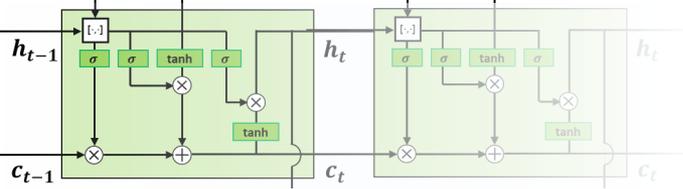


Buscar imágenes similares ...



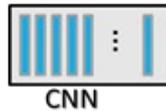
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

aplicaciones: análisis

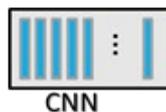


■ Representation learning

- Podemos utilizar representaciones internas de la red

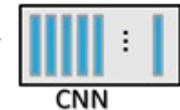


CNN

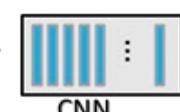


CNN

Comparar si dos imágenes corresponden a la misma identidad



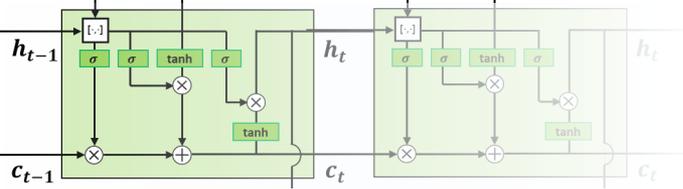
CNN



CNN

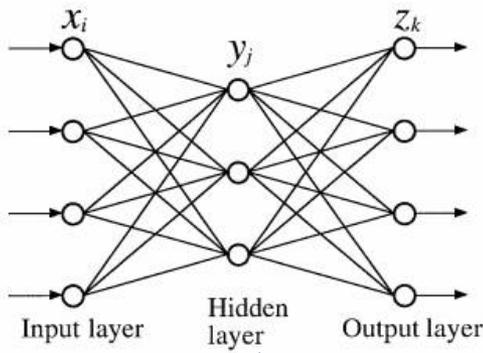


aplicaciones: análisis



■ Generación

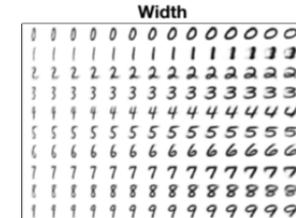
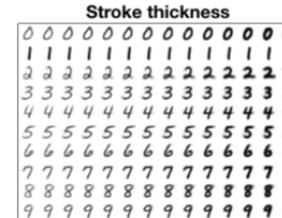
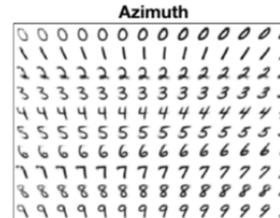
- Podemos aprender a manipular las imágenes



- ¿Qué pasa si cambio la representación para conseguir otra imagen distinta?

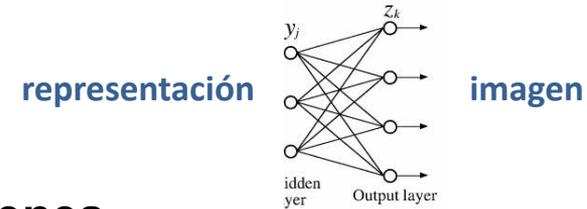
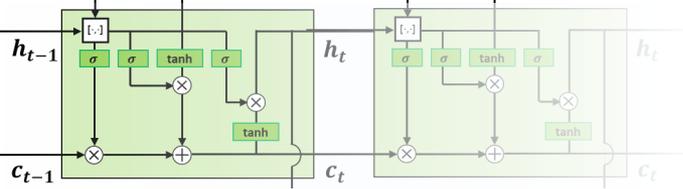


representación



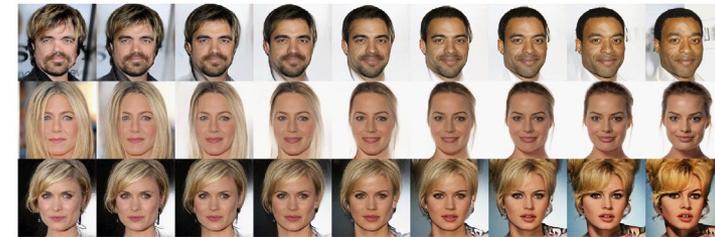
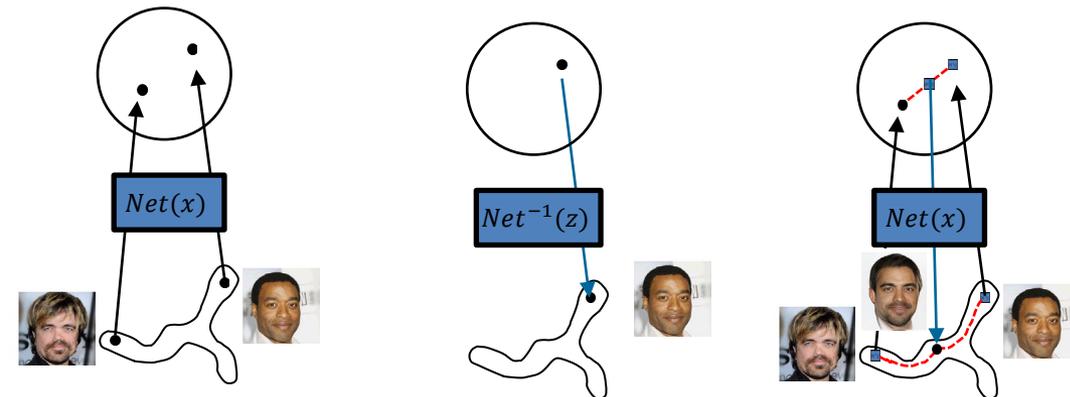
Antoran, J., & Miguel, A. (2019, December). Disentangling and Learning Robust Representations with Natural Clustering. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 694-699). IEEE.

aplicaciones: síntesis



■ Generación

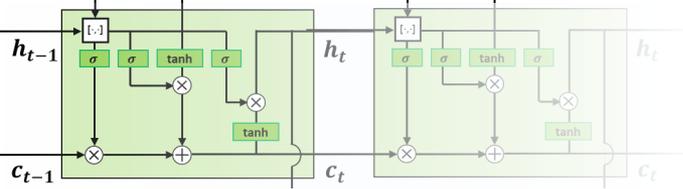
- Podemos aprender a manipular las imágenes
 - ¿Qué pasa si cambio la representación para conseguir otra imagen distinta?



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).

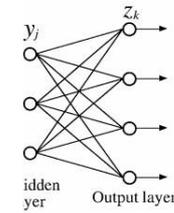
aplicaciones: síntesis



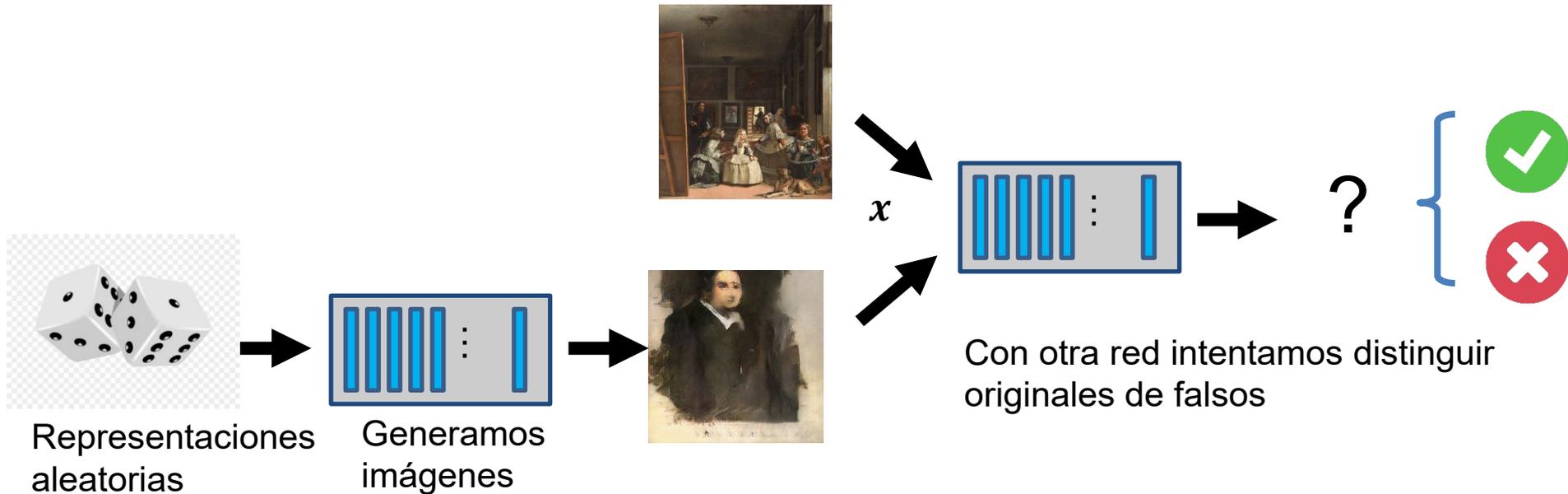
■ Generación

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

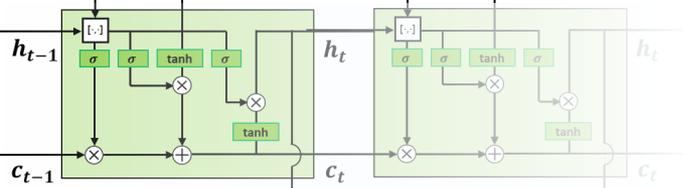
representación



imagen



aplicaciones: síntesis



■ Generación

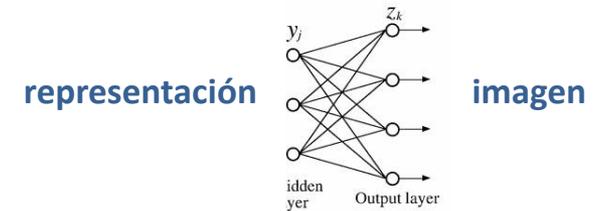
- Hay modelos en los que directamente se aprende a generar imágenes
- **Generative adversarial network**



¿Qué imagen es artificial?

One hour of imaginary celebrities

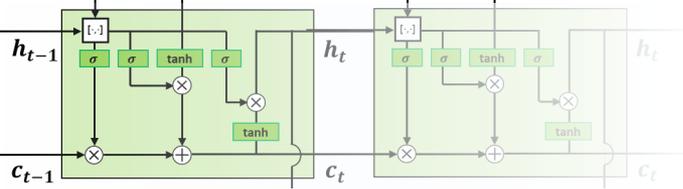
<https://www.youtube.com/watch?v=36lE9tV9vm0>



Dos redes compiten:

- **Red 1 genera imágenes realistas de manera que la Red 2 falle**
- **Red 2 intenta distinguir las imágenes reales de las falsas**
- Es una combinación de generación y clasificación
- Genera datos de forma que sean indistinguibles de los originales
- Generador ilimitado de datos: imágenes, audio, texto...

aplicaciones: síntesis



■ Generación

- Este proceso se ha sofisticado mucho en menos de 10 años

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



2014



2015



2016



2017



2018



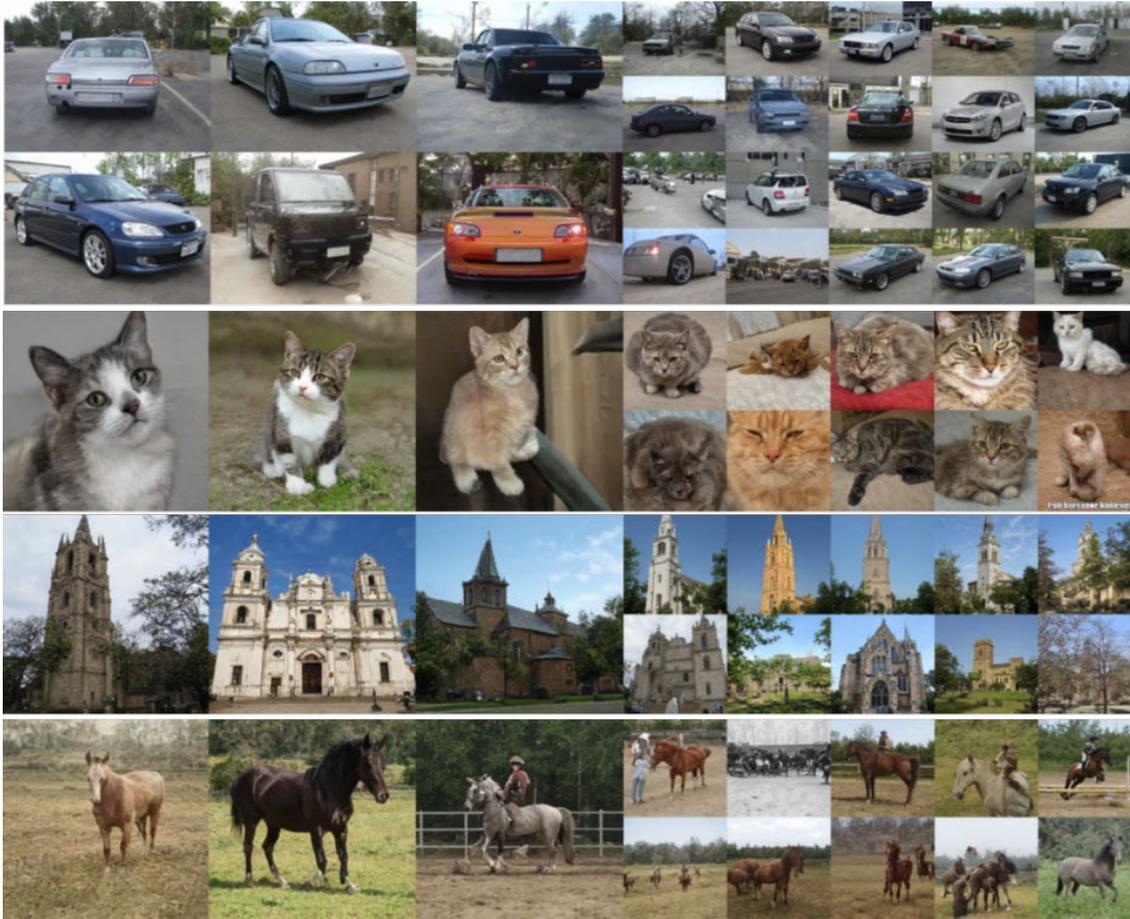
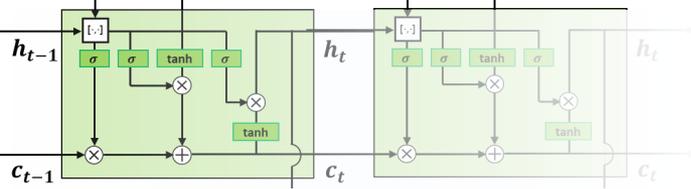
2020



2021

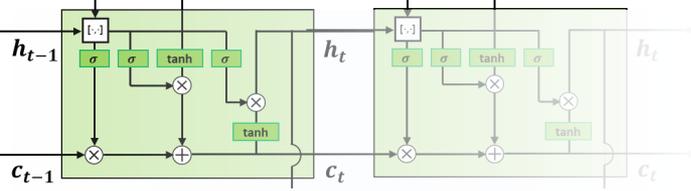
StyleGAN3 (Karras 2021)

aplicaciones: síntesis



StyleGAN2 (Karras 2020)

aplicaciones: síntesis



Monet ↔ Photos

Zebras ↔ Horses

Summer ↔ Winter



Monet → photo



zebra → horse



summer → winter

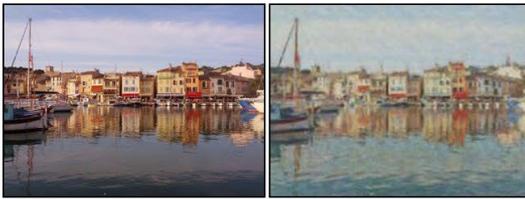


photo → Monet



horse → zebra



winter → summer



Photograph



Monet



Van Gogh



Cezanne

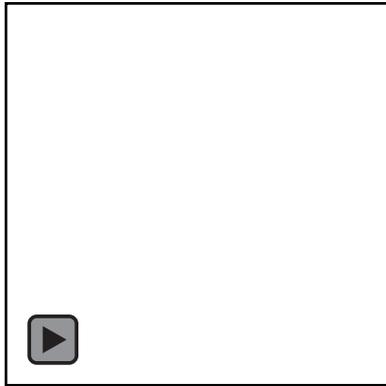


Ukiyo-e

<https://junyanz.github.io/CycleGAN/>

aplicaciones: síntesis

- Generación de vídeos realistas: Deep fakes

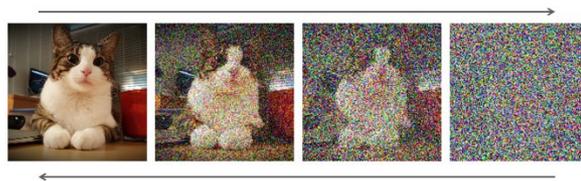


aplicaciones: síntesis

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

■ Generación:

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](#)

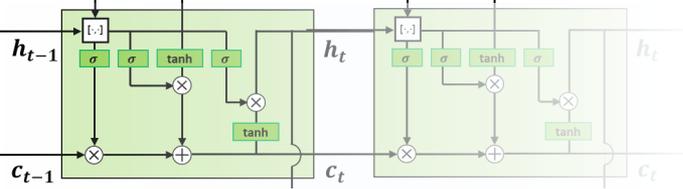


A wooden spanish laptop of 1650
found the library of El Escorial



Medieval 1230 book page illustrating
monks playing basketball

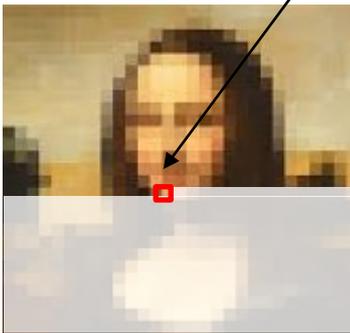
aplicaciones: síntesis



▪ 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

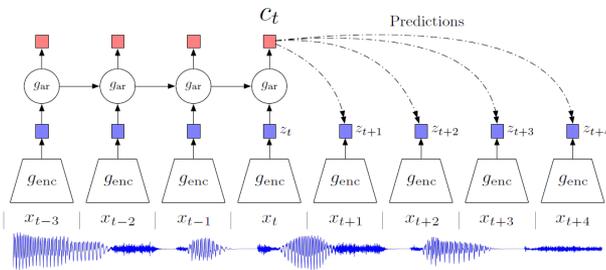
Viendo los pixels anteriores: ¿cómo es el siguiente ?



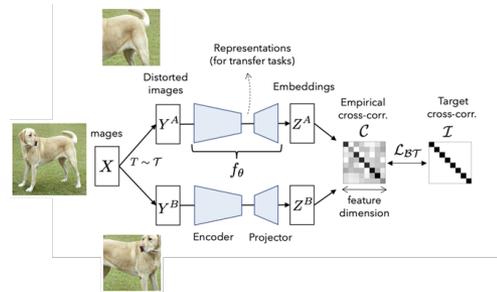
*A Oord, N Kalchbrenner, O Vinyals, L Espeholt, A Graves, K Kavukcuoglu
Conditional Image Generation with PixelCNN Decoders 2016*

aplicaciones: análisis

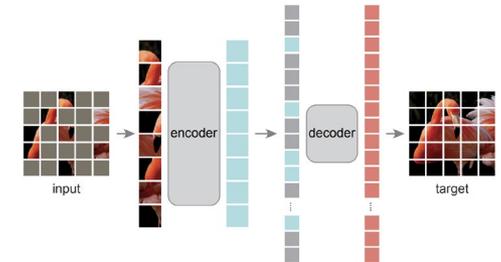
■ Aprendiendo sin etiquetas: no supervisado



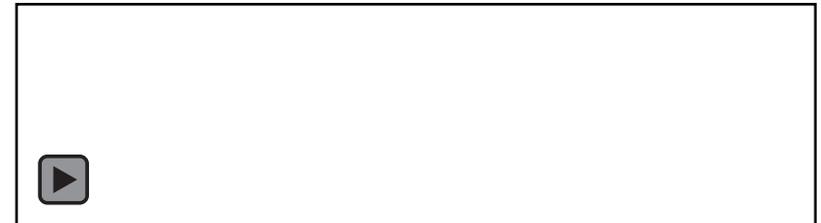
Una estrategia es dar varias opciones como si fuera un examen



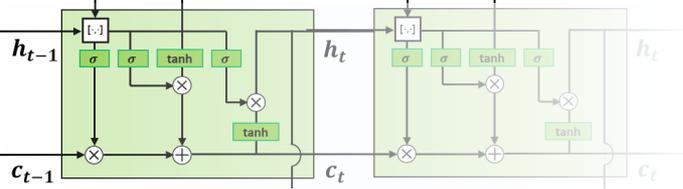
Resolver la pregunta:
¿ son partes de la misma imagen ?



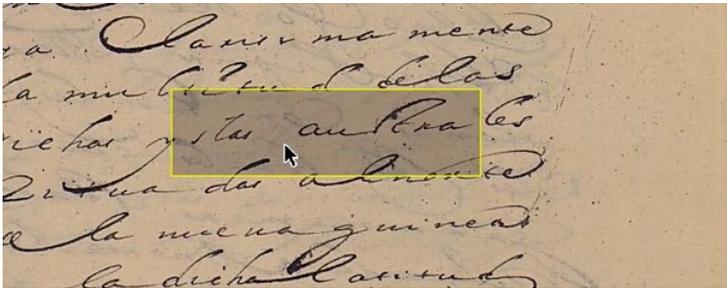
Reconstruir la imagen
a partir de una con oclusiones



aplicaciones



- 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.

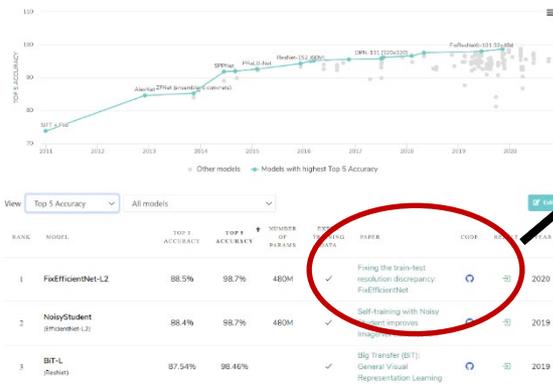
¿Se puede usar ya la tecnología?

- 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**

divulgación

- **Comunicar, comparar y aprender**
 - Congresos y conferencias científico-técnicas como Iberspeech / Interspeech / ICCASP
 - Los retos **Albayzín-RTVE 2018 y 2020** organizados por la **Cátedra RTVE de la Universidad de Zaragoza**
 - Páginas web como papers with code

Image Classification on ImageNet



Code

- google-research/holystudent (official) ★ 414
- tensorflow/tpu ★ 3,964
- Stanley-Zheng/gntonhacks ★ 3
- adventure2169/Summarization_self-training_with_noisy_student_improves_image_classification ★ 2
- thomasyp/PaperTranslation ★ 0

See all 6 implementations

Tasks

- DATA AUGMENTATION
- IMAGE CLASSIFICATION

Results from the Paper

Ranked #3 on Image Classification on ImageNet (using extra training data)

TASK	DATASET	MODEL	METRIC NAME	METRIC VALUE	GLOBAL RANK	USES EXTRA TRAINING DATA	RESULT	BENCHMARK
Image Classification	ImageNet	NoisyStudent (EfficientNet-B0)	Top 1 Accuracy	78.8%	# 85	✓	🏆	Compare
			Top 5 Accuracy	94.5%	# 54	✓	🏆	Compare
			Number of params	5.3M	# 69	✓	🏆	Compare
Image	ImageNet	NoisyStudent	Top 1 Accuracy	88.4%	# 3	✓	🏆	Compare