

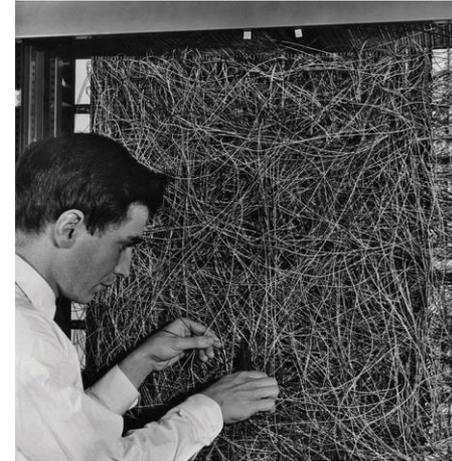
# Introducción a deep learning

# 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**

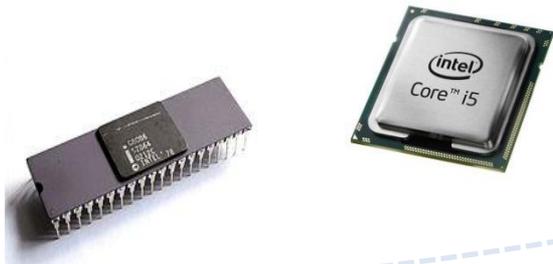
10.2 TFlops (~411 x intel i5)

**Nvidia RTX Titan, 2018**

16 TFlops (~640 x intel i5)

**Nvidia RTX 3090, 2020**

35 TFlops (~1400 x intel i5)



# antecedentes

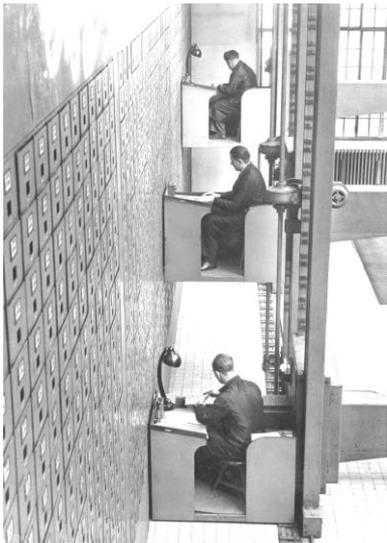
## Velocidad de almacenamiento

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



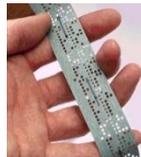
## Almacenamiento

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 papel 1995 /  
Revistas papel*



*Buscadores internet 1998*



TensorFlow

PYTORCH

*Software gratuito  
y  
Toolkits 2010*

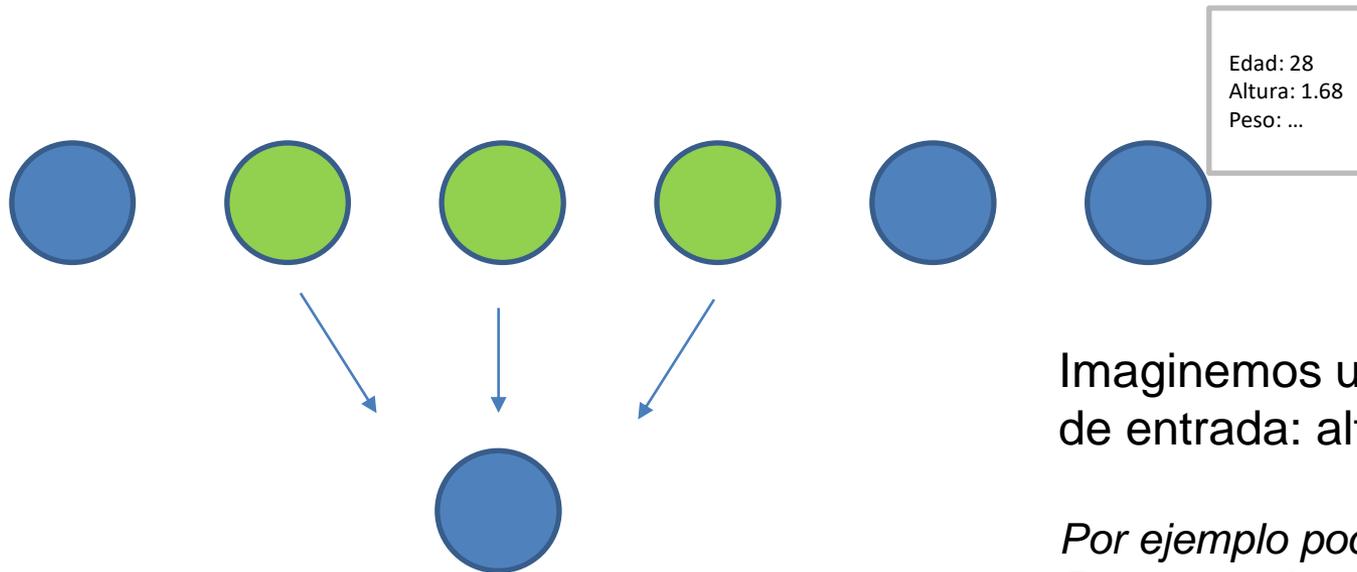


GitHub

*2008 Redes sociales /  
plataformas de desarrollo colaborativo*

# fundamentos

- Procesado en fases, capas

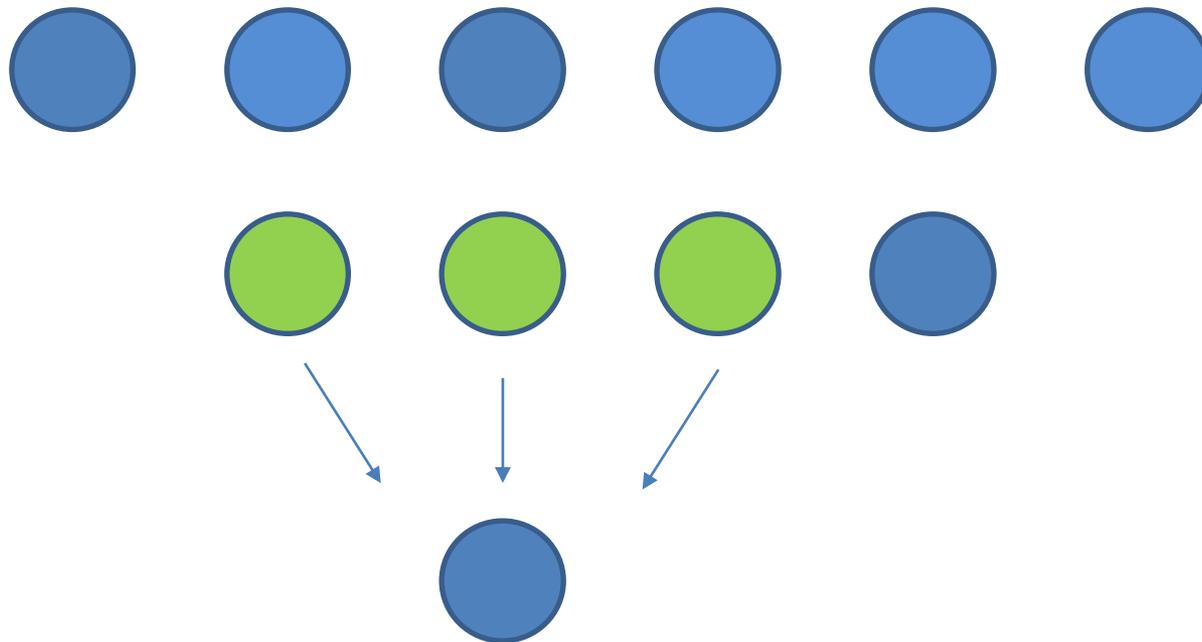


Imaginemos unos datos de entrada: altura, edad,..

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

# fundamentos

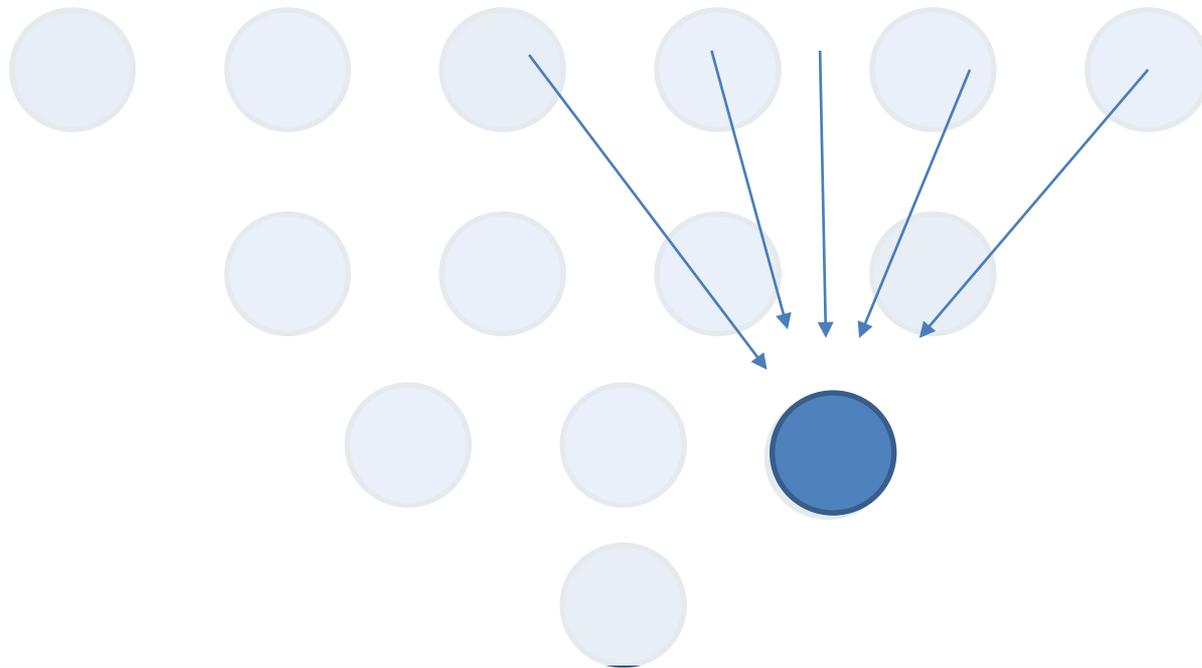
- Procesado en fases, capas



Y repetimos,  
en todas las  
filas...

# fundamentos

- Procesado en fases, capas

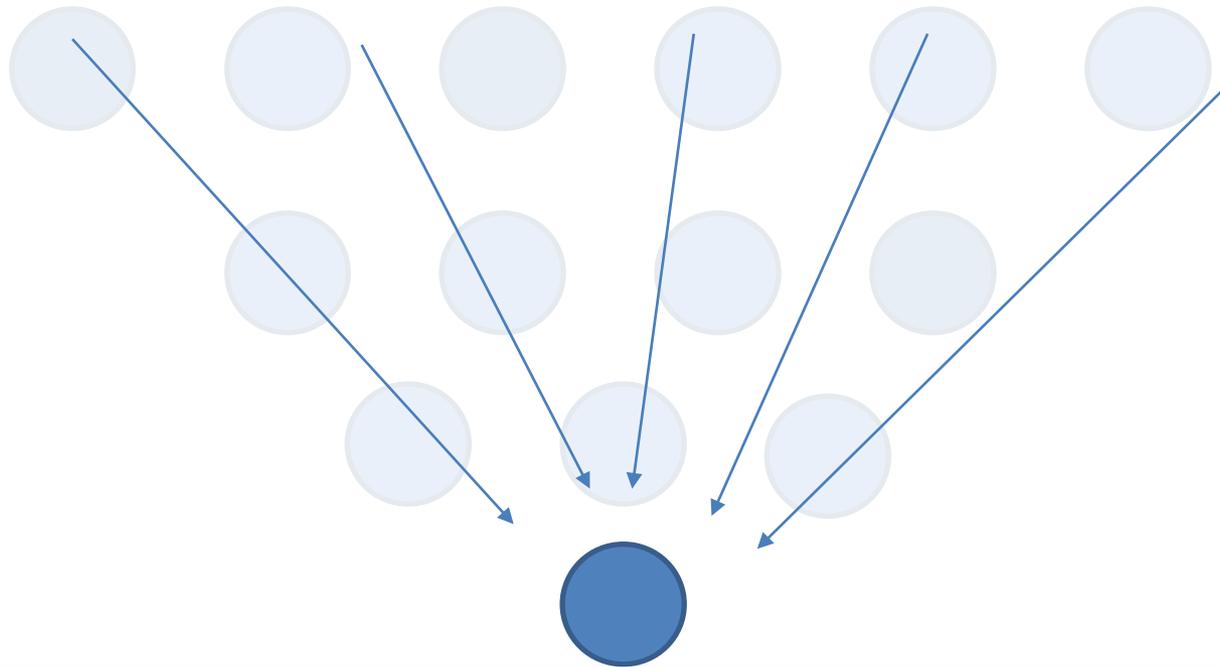


Algunos nodos

reciben  
**parte** de la  
información

# fundamentos

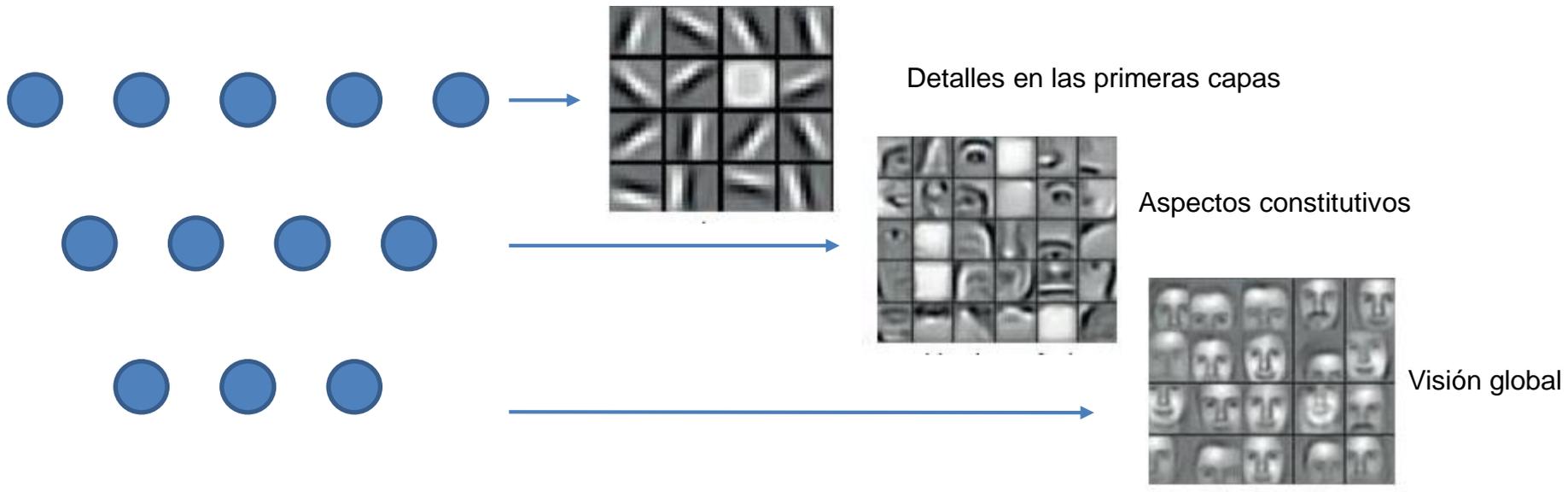
- Procesado en fases, capas



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

# fundamentos

- 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

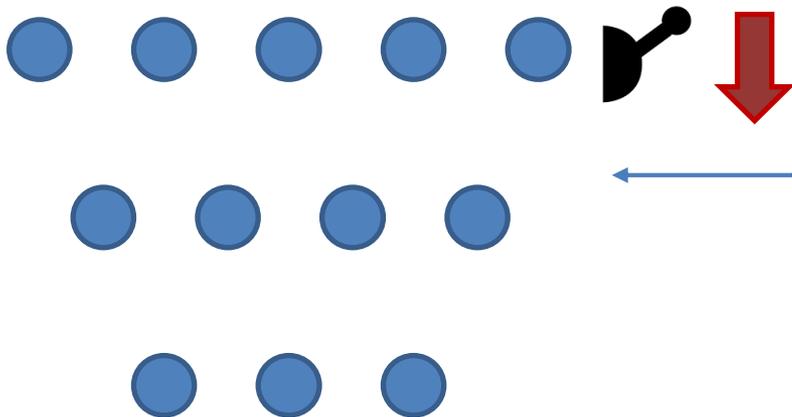


# fundamentos

- **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**



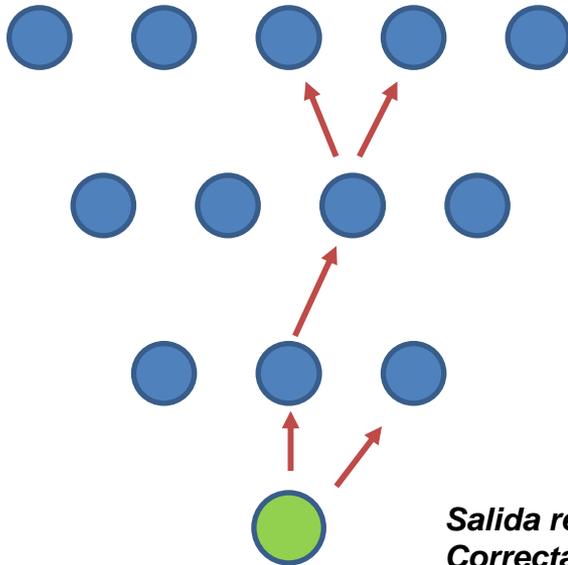
- Depende de lo complicada que sea la tarea pueden ser **millones** de correcciones

- **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**

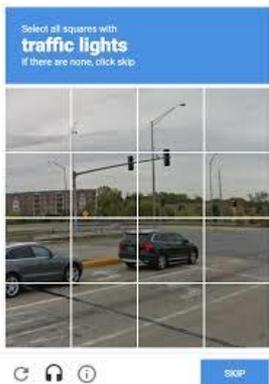
# fundamentos

- **Etiquetado de datos**

- ¿Quién etiqueta?
- Freelance,
  - Mechanical turk amazon
- “Granjas” de datos
- **Todos nosotros**

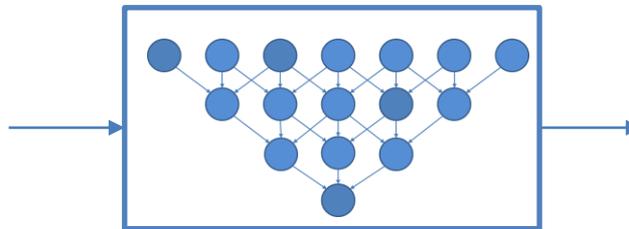
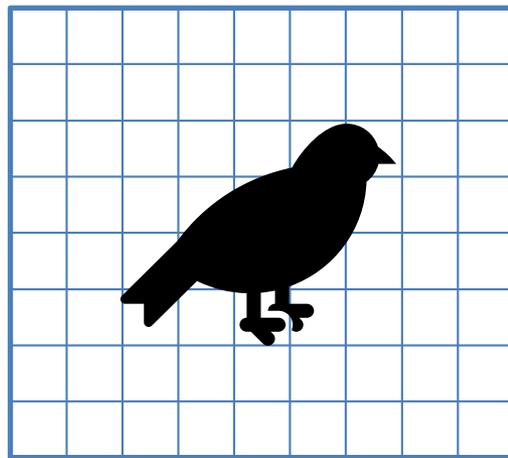


<https://time.com/5518339/china-ai-farm-artificial-intelligence-cybersecurity/>



## Tipos de problemas (1 / 4)

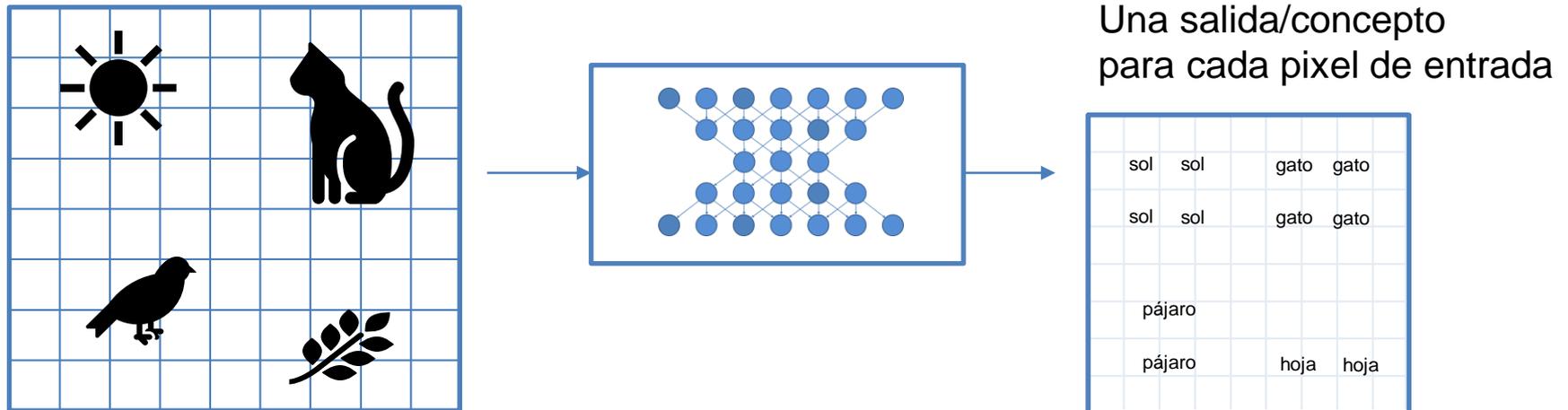
- Aunque cada día hay más variantes las dos formas principales de usar DNNs hoy en día:
  - **Clasificación:**
    - Decir **qué concepto** hay en una imagen/texto/audio



Una sola salida/concepto  
para toda la entrada

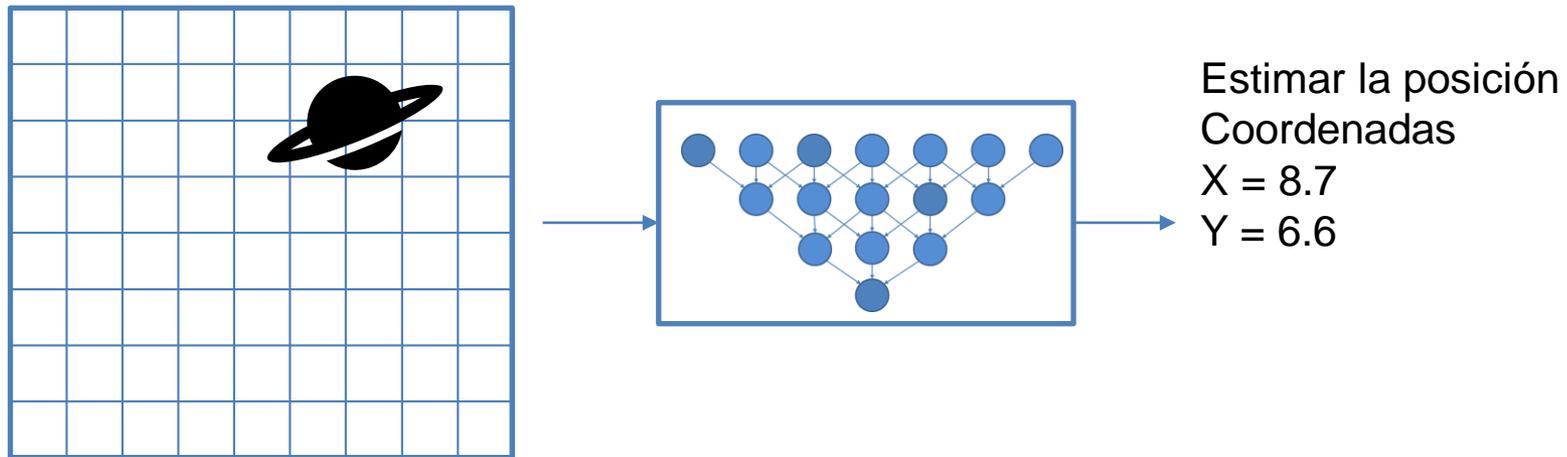
## Tipos de problemas (2 / 4)

- 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



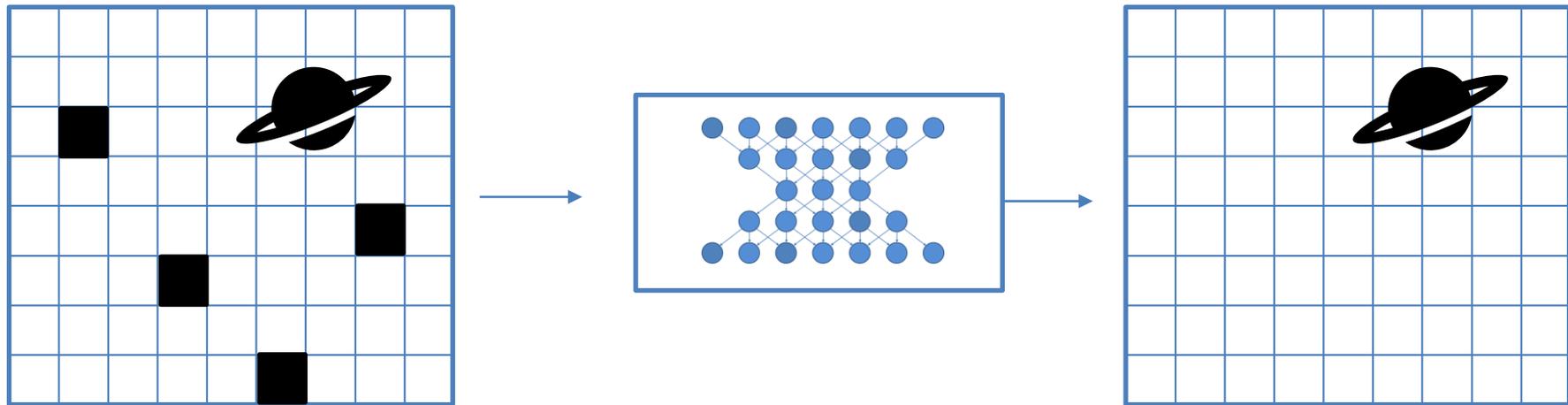
## Tipos de problemas (3 / 4)

- 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**

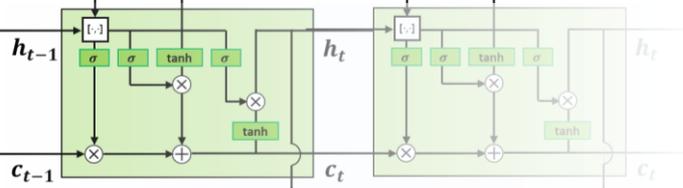


## Tipos de problemas (4/4)

- 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...

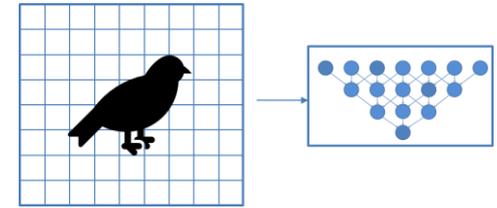


# aplicaciones: análisis



## – Clasificación:

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

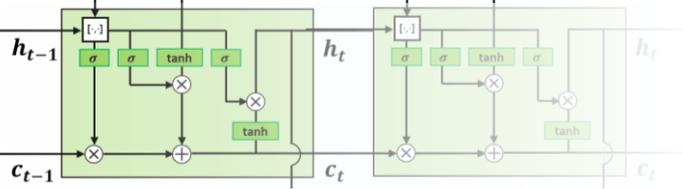


En este ejemplo la red neuronal se prepara para resolver el problema de clasificación:  
**¿Qué hay en esta imagen? -> 1 respuesta**

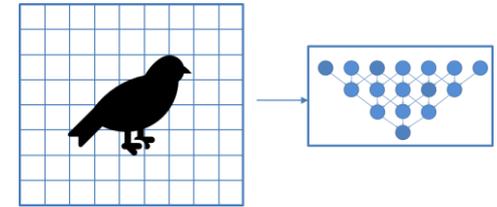
*Entre las posibles respuestas hay 120 razas de perro  
En 2012 el error top5 era del 25%,  
Hinton y Krizhevsky red de 7 capas 15%  
Hoy en día decenas, cientos de capas, alrededor del 2%,*



# aplicaciones: análisis

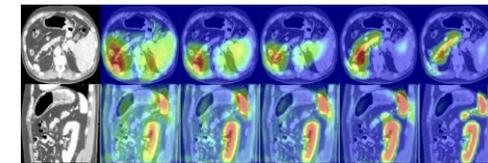


– **Clasificación:** ¿ nos podemos fiar ?



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

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

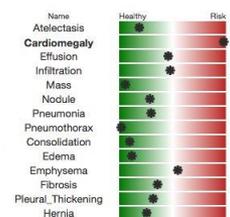


Out Of Distribution reconstruction error  
Heatmap where the image varies from the training distribution.

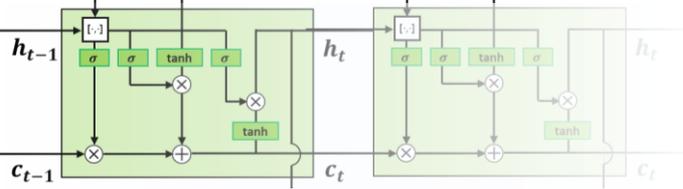


recScore:0.27, ssim:0.39

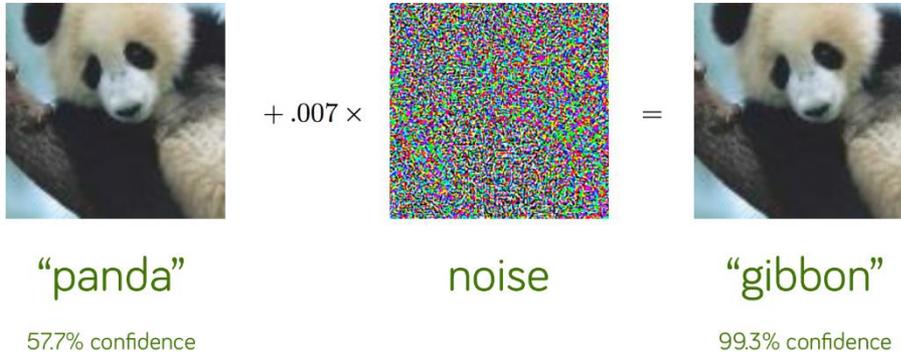
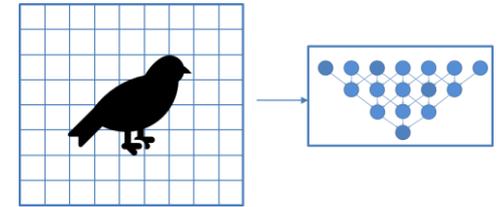
Disease Predictions  
Probability of a disease.



# 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)*

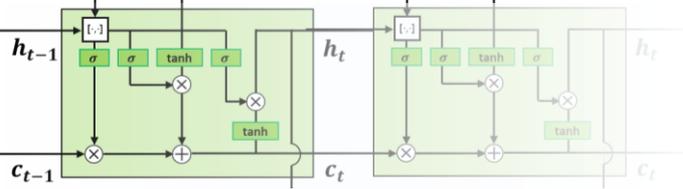


Speed Limit 80  
(88% confidence)

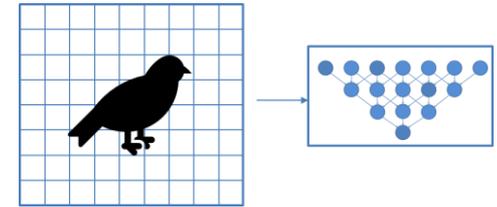
**Robust physical-world attacks on deep learning visual classification.**

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).

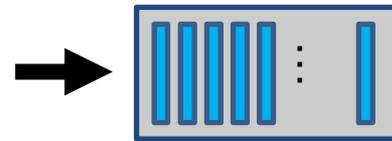
# aplicaciones: análisis



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



[paperswithcode.com](http://paperswithcode.com)

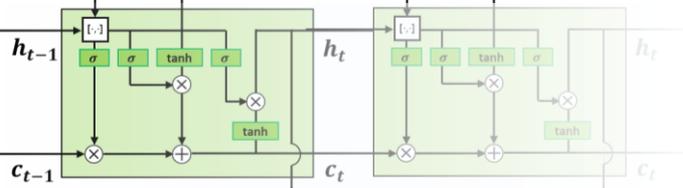


Entrenamiento

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

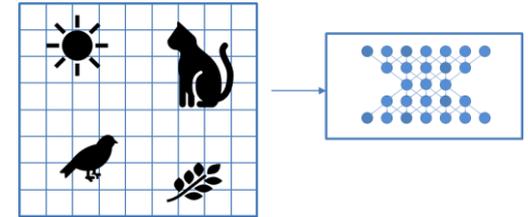
Se conoce como:  
**Aumento de datos**

# aplicaciones: análisis



## – Clasificación múltiple:

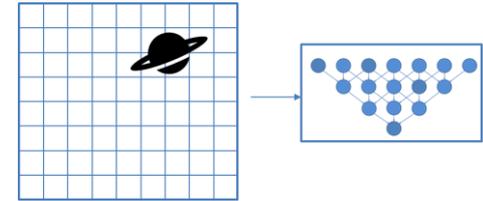
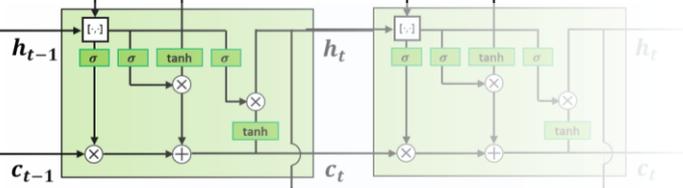
- **varias propiedades/conceptos** de una imagen/texto/audio



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

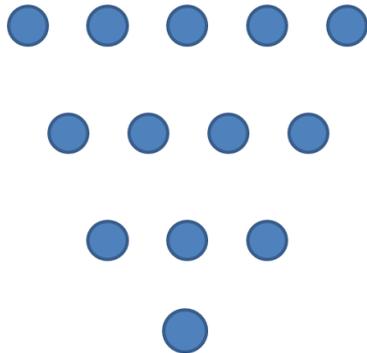
¿Hay un perro?	No
¿Hay un gato?	Sí
¿Hay árboles?	No
¿Hay un pájaro?	Sí
¿Hay cielo?	No
¿Hay hierba?	Sí

# 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:

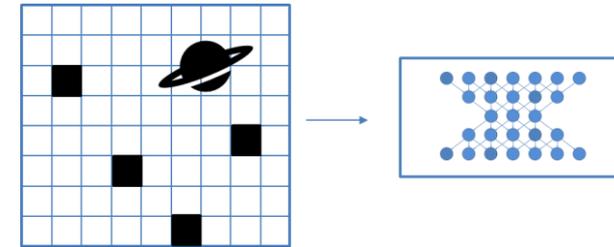
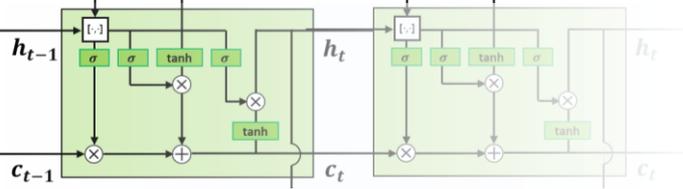
**¿Qué edad tienen estas persona?**

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:

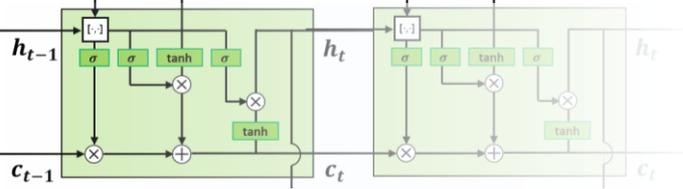
**Convertir una imagen de BN en color**



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

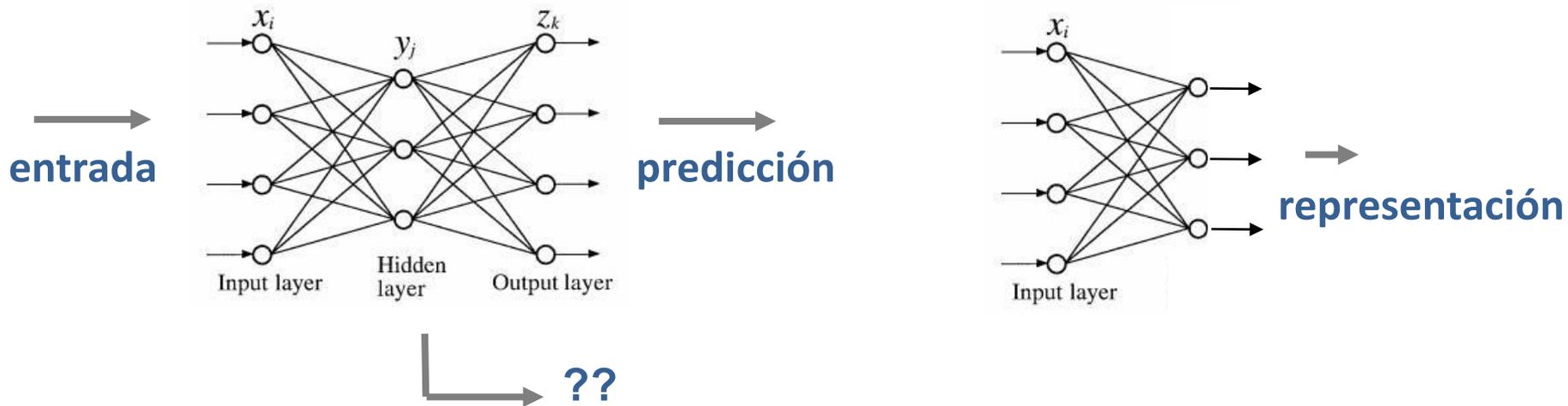
**Mejorar la calidad de la imagen**

# aplicaciones: análisis

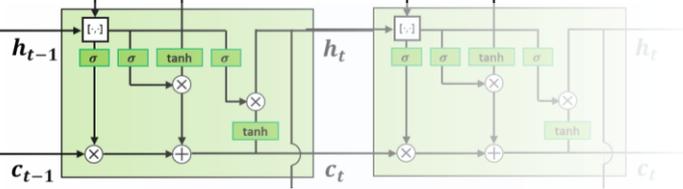


## ■ Representation learning

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

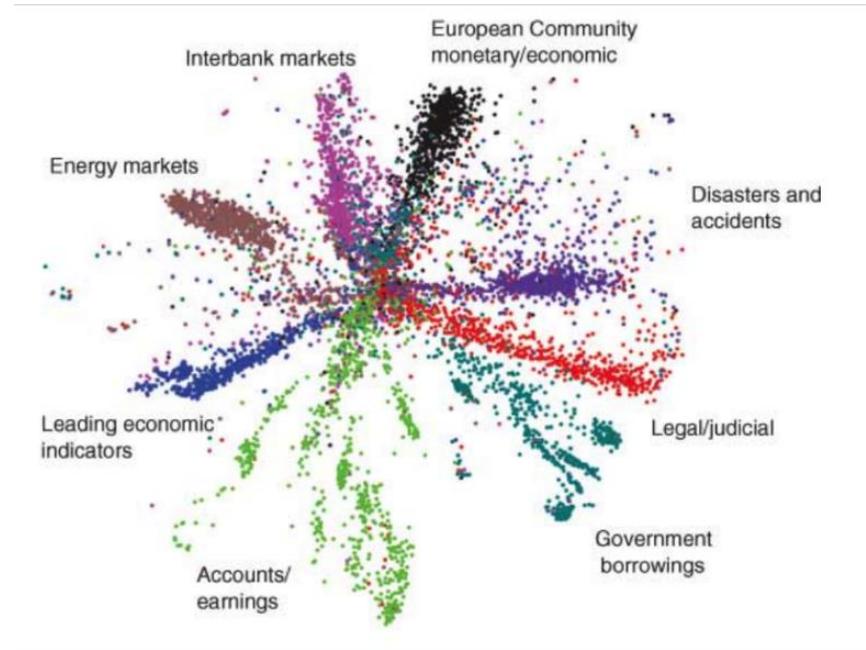


# aplicaciones: análisis

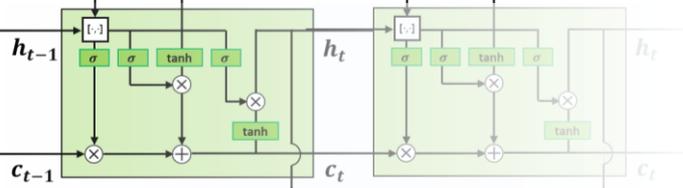


## ■ Representation learning

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



# aplicaciones: análisis

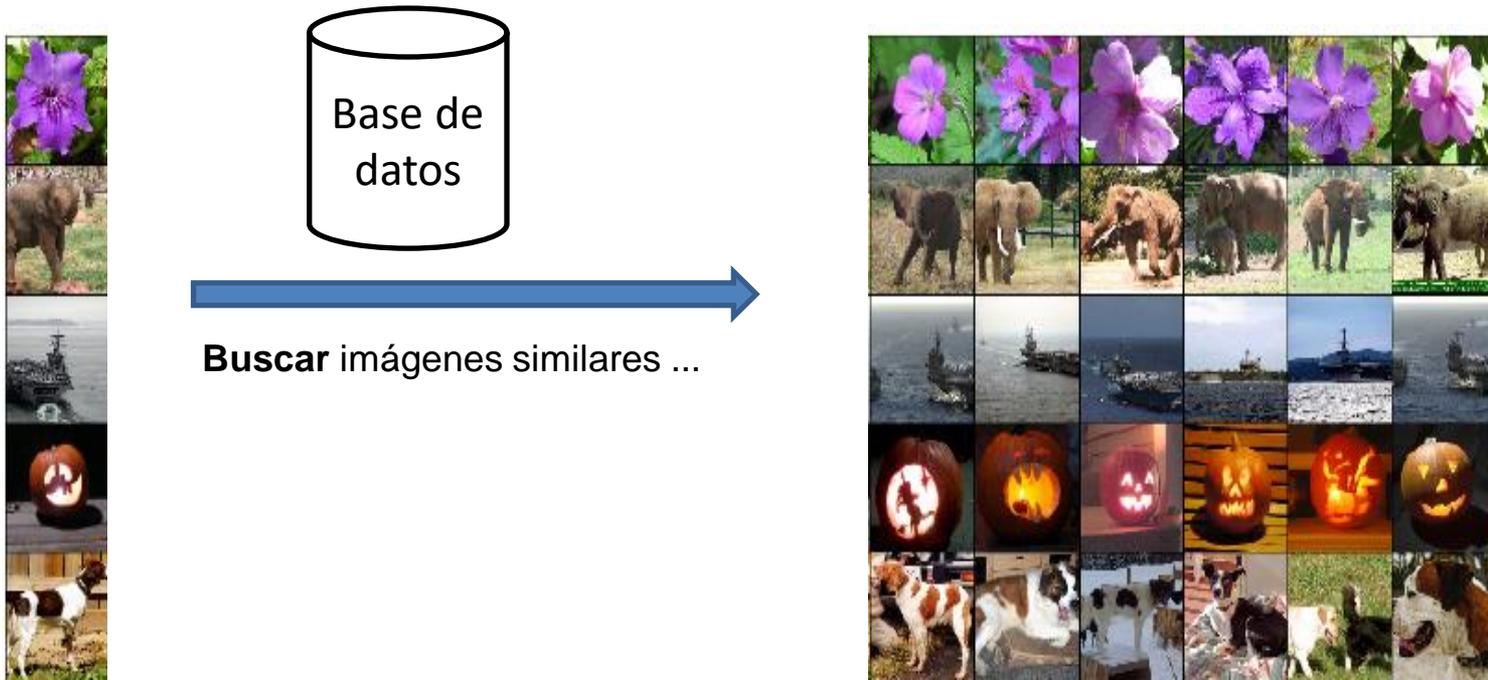


Imagenet classification with deep convolutional neural networks

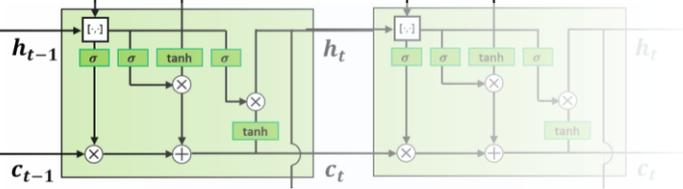
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).. *Advances in neural information processing systems*, 25.

## ■ Representation learning

- Podemos utilizar representaciones internas de la red

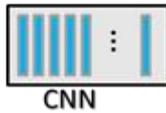


# aplicaciones: análisis

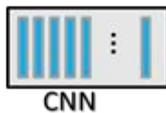


## ■ Representation learning

- Podemos utilizar representaciones internas de la red



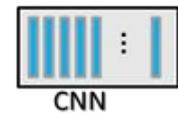
CNN



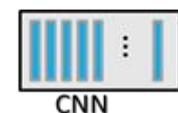
CNN



**Comparar** si dos imágenes corresponden a la misma identidad



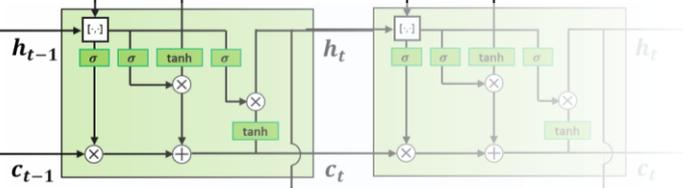
CNN



CNN

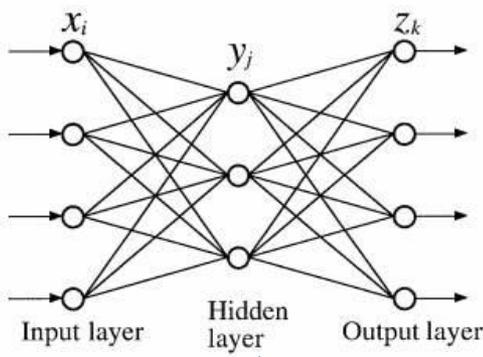


# aplicaciones: análisis



## ■ Generación

- Podemos aprender a manipular las imágenes



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



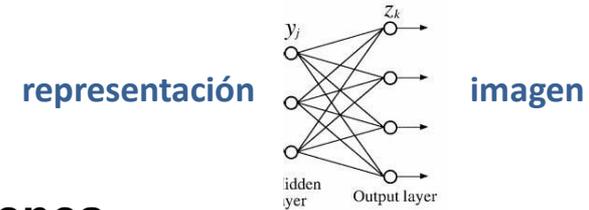
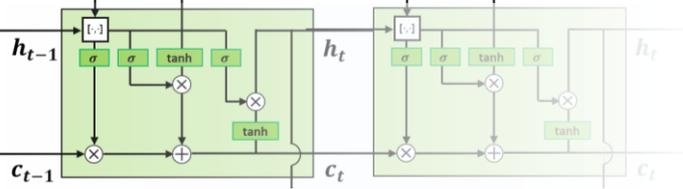
representación

Azimuth									
0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

Stroke thickness									
0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

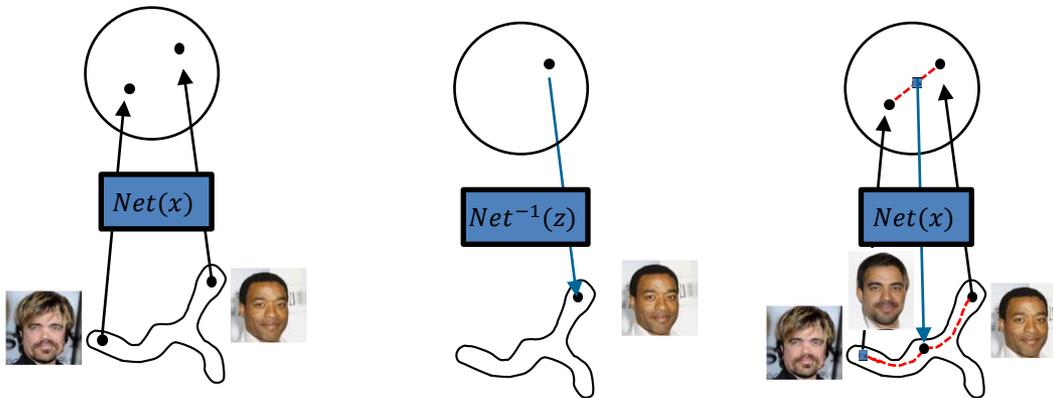
Width									
0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

# aplicaciones: síntesis



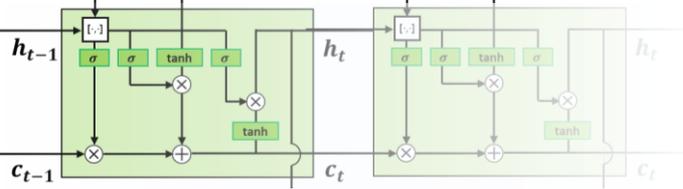
## ■ Generación

- 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

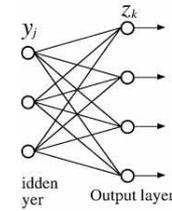
# aplicaciones: síntesis



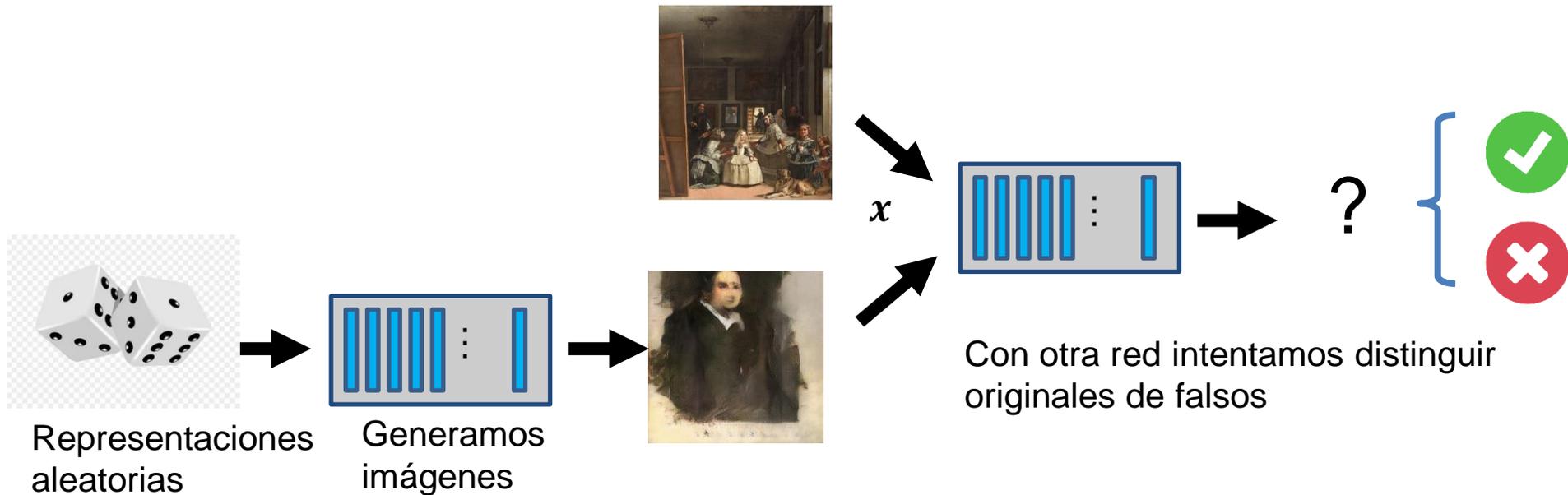
## ■ Generación

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

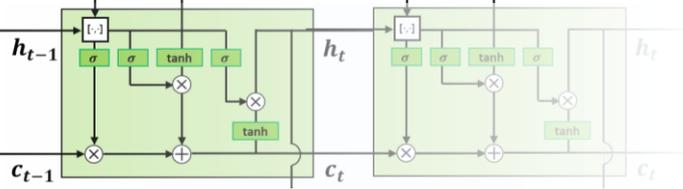
representación



imagen



# 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

# 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](#)

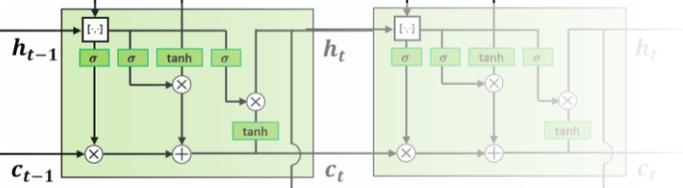


A wooden spanish laptop of 1650  
found the library of El Escorial



Medieval 1230 book page  
illustrating monks playing basketball

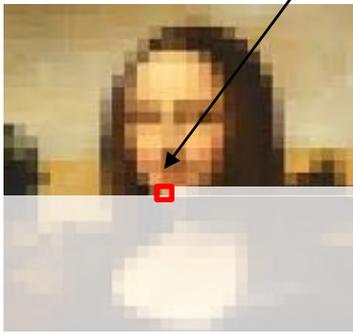
# aplicaciones: síntesis



## ■ 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 ?



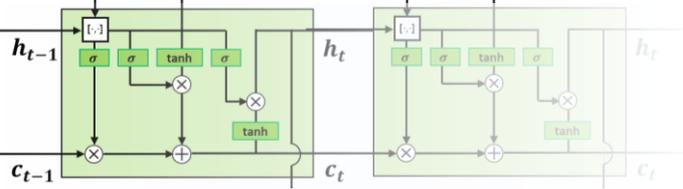
occluded

completions

original

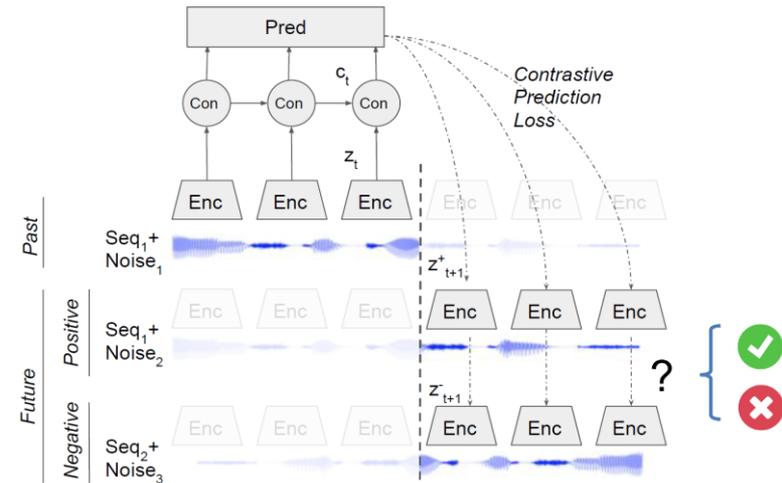
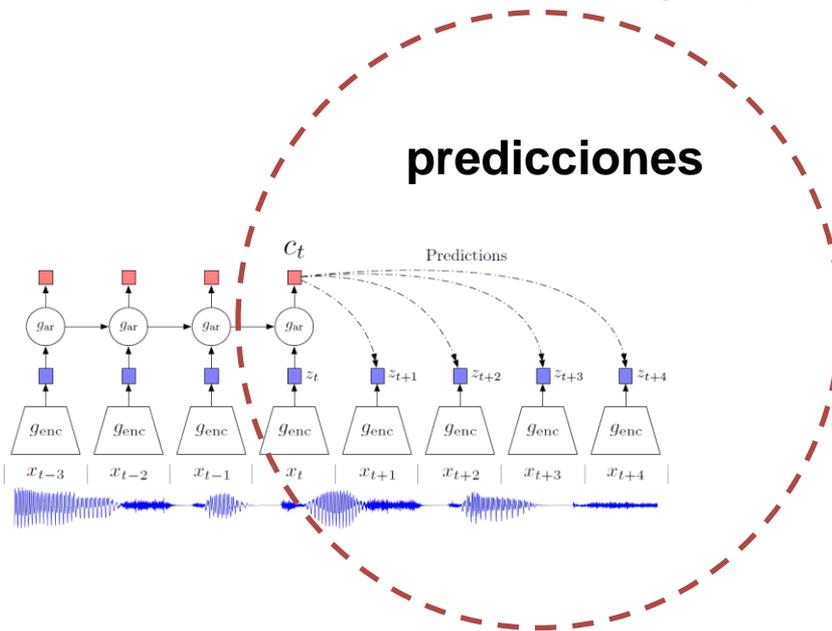


# aplicaciones: análisis



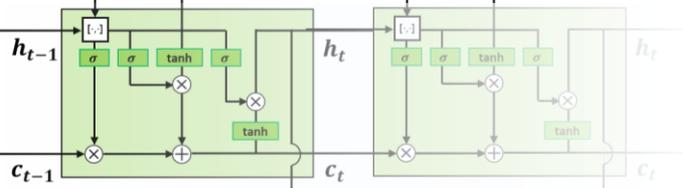
## ■ Aprendiendo sin etiquetas: no supervisado

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



Una estrategia es dar varias opciones como si fuera un examen

# aplicaciones: análisis



## ▪ Aprendiendo sin etiquetas: no supervisado

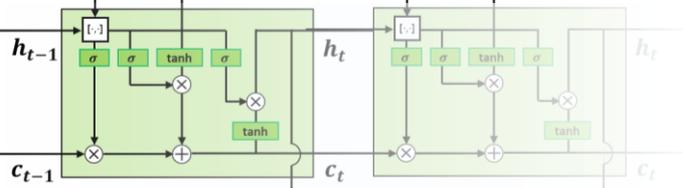
- Motivación: hay muchos datos no etiquetados
- Las representaciones obtenidas se pueden usar en otras tareas



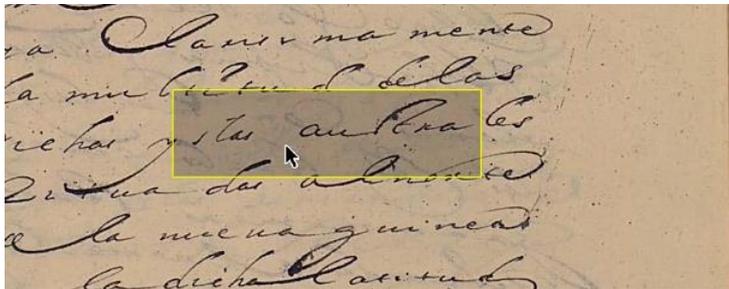
Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. arXiv preprint arXiv:2104.14294..



# 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**