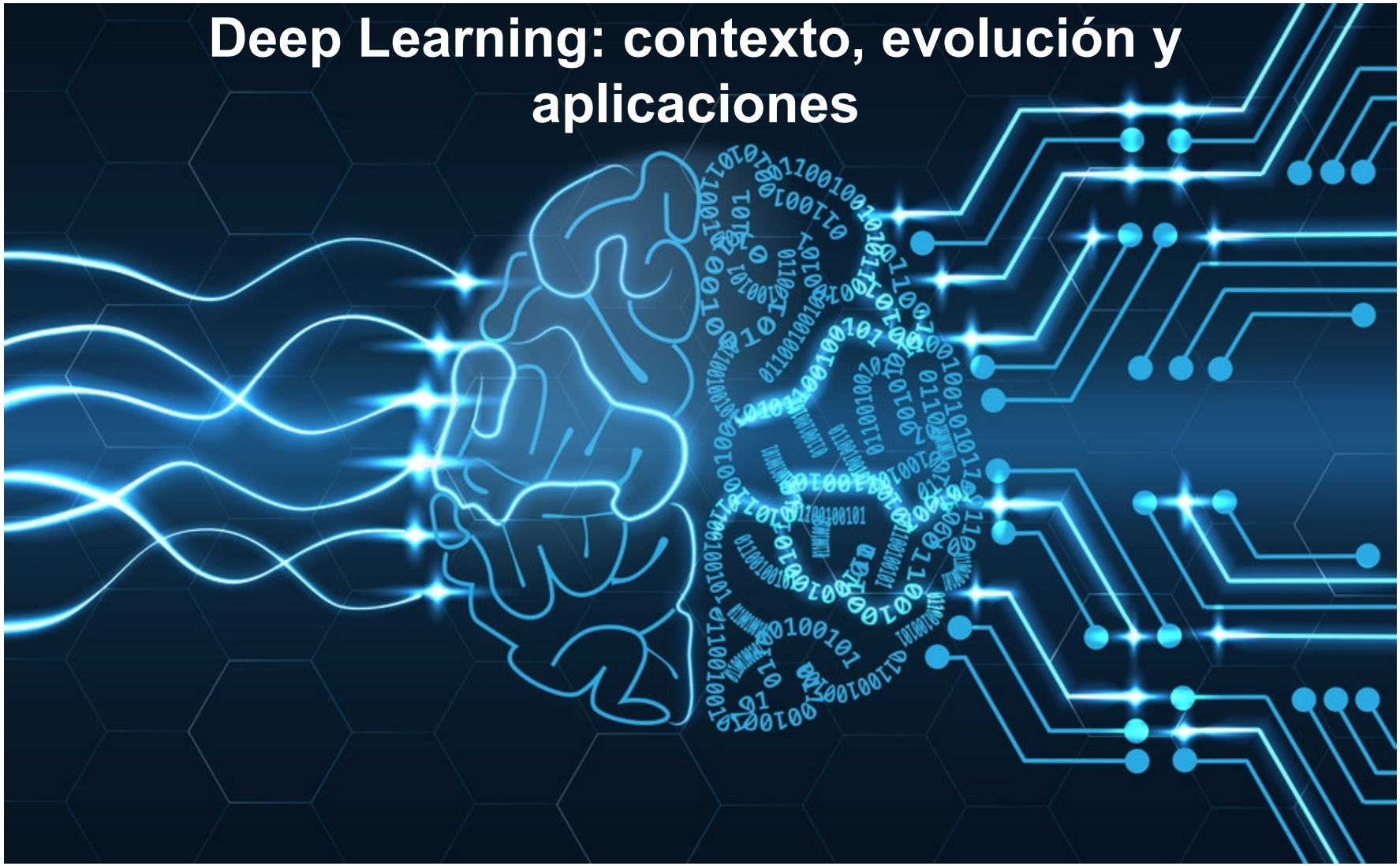
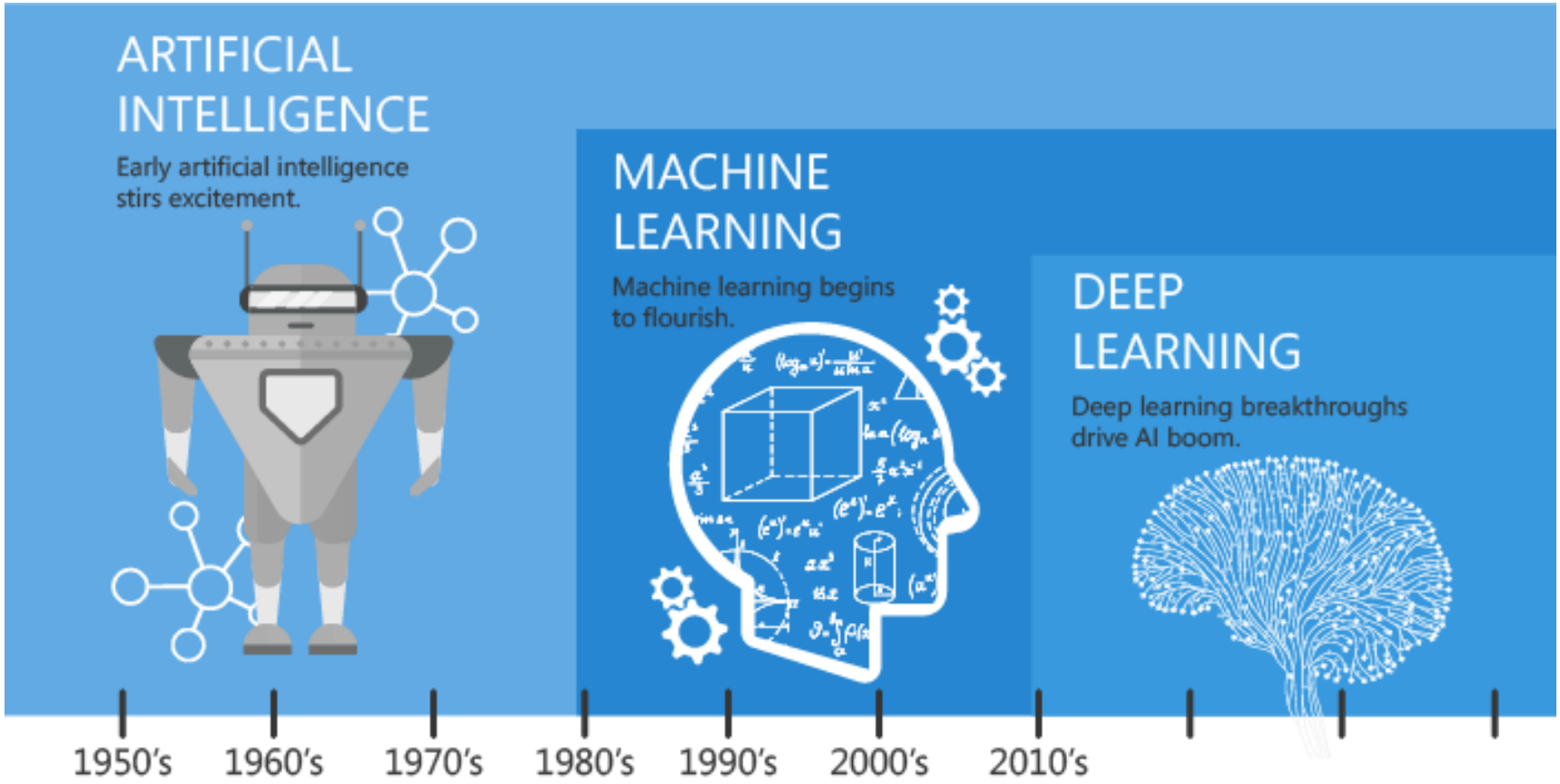
INVETT Group at University of Alcalá



David Fernández Llorca (david.fernandezl@uah.es)

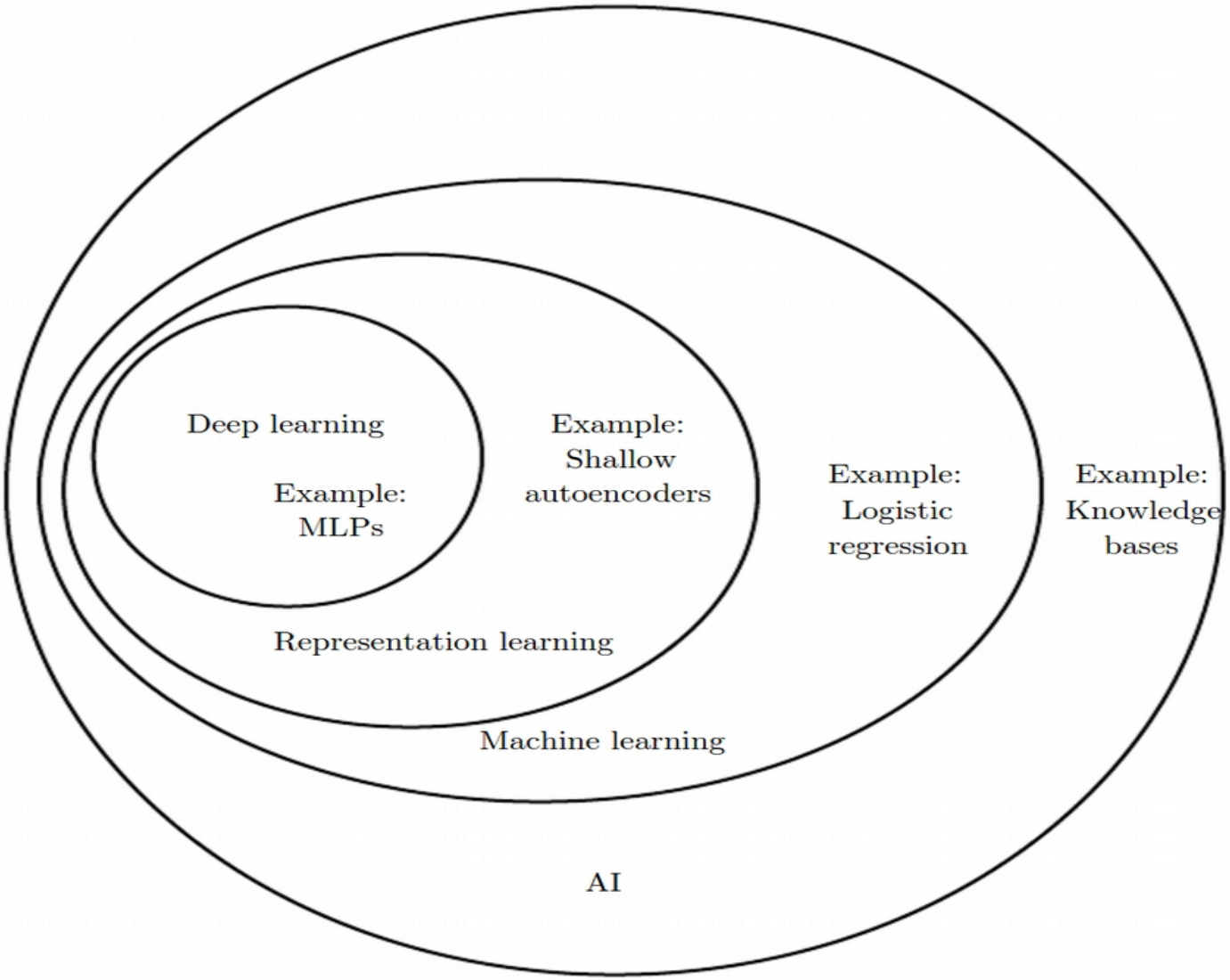
INVETT Group at University of Alcalá



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

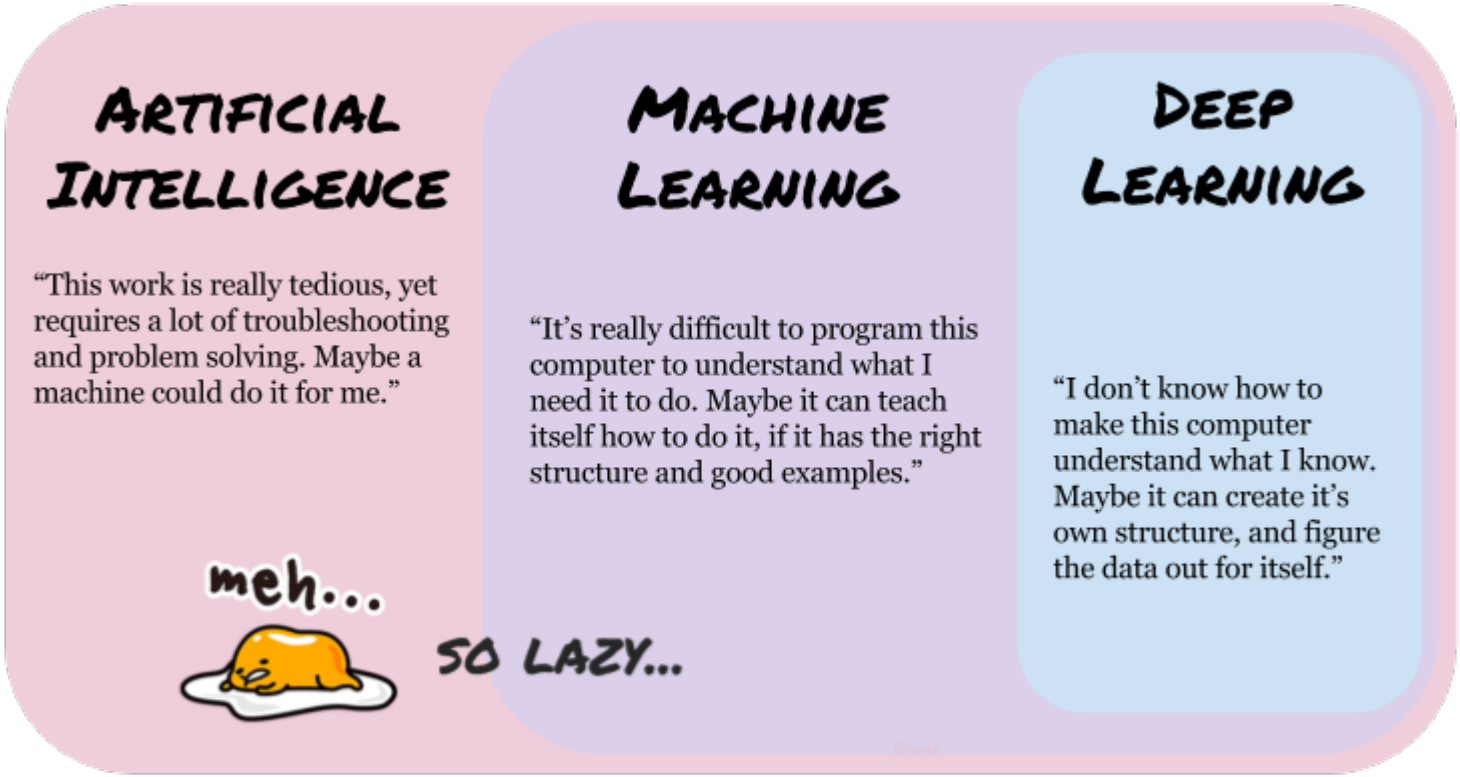
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá



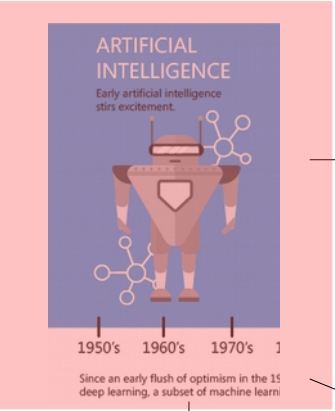
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

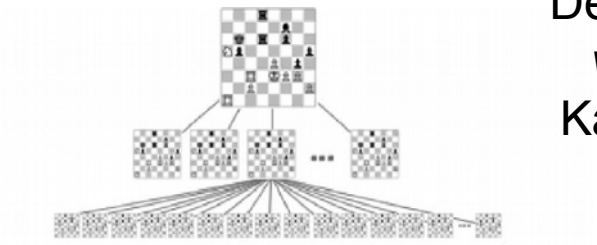


David Fernández Llorca (david.fernandezl@uah.es)

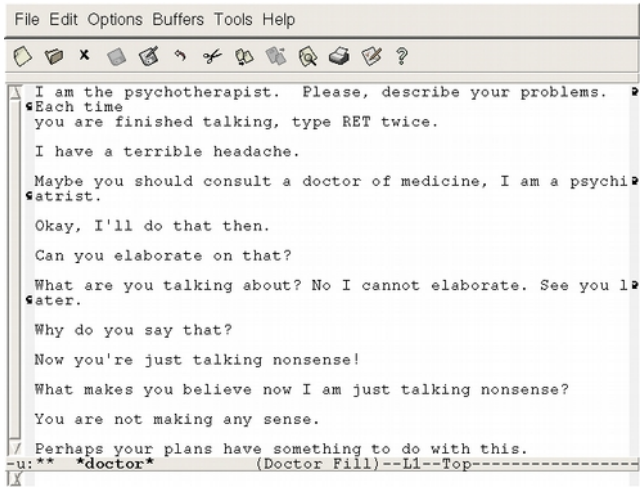
INVETT Group at University of Alcalá



Robots industriales (autómatas)



Deep Blue
versus
Kasparov
(1997)

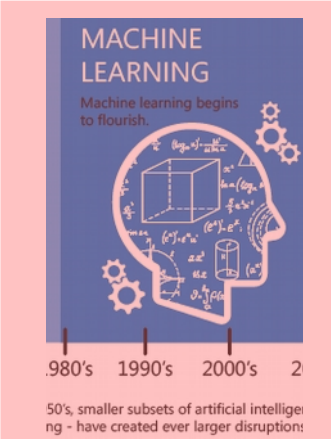


ELIZA: 1er Bot Conversacional
(1966)

David Fernández Llorca (david.fernandezl@uah.es)

Machine learning

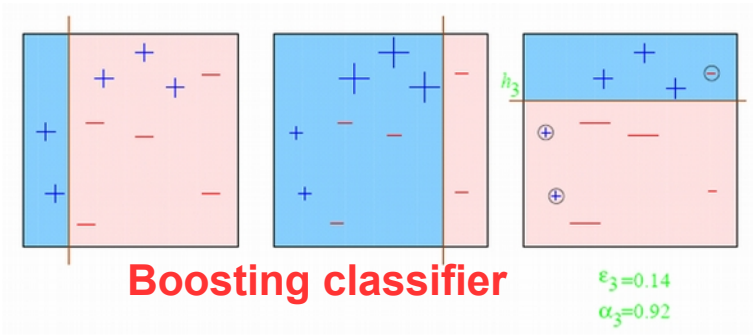
- Habilidad de las máquinas de adquirir conocimiento.
- ¿Conocimiento?
  - **Modelo** que hay que definir (cada modelo tiene un **número de parámetros**)
- ¿De quién o qué aprendo?
  - **Datos** en crudo (raw data): ***bases de datos***
  - **Representación de los datos** (***features***)
  - **Objetivos** (targets) o **salida deseada**.
- ¿Cómo aprendo?
  - Algoritmos matemáticos de **optimización**.



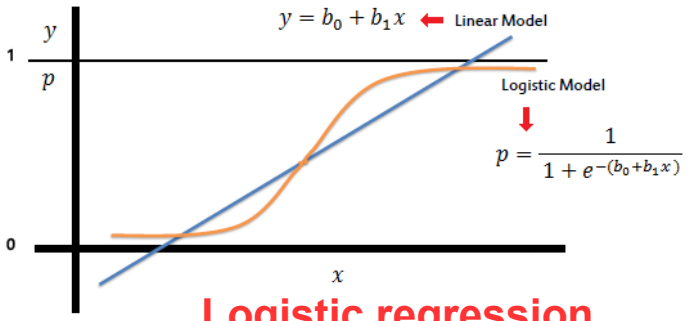
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

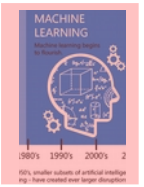
Modelos (metodologías)



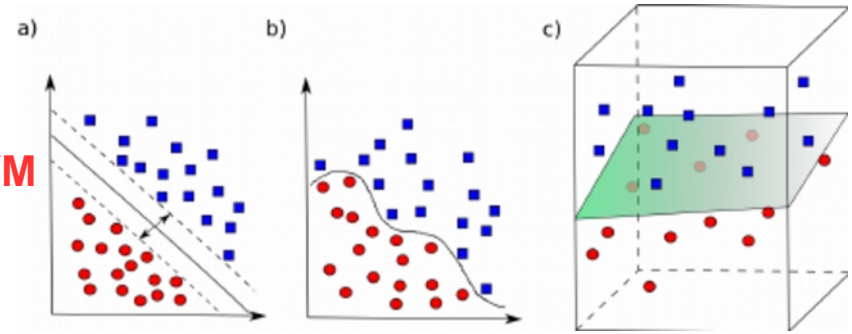
Boosting classifier

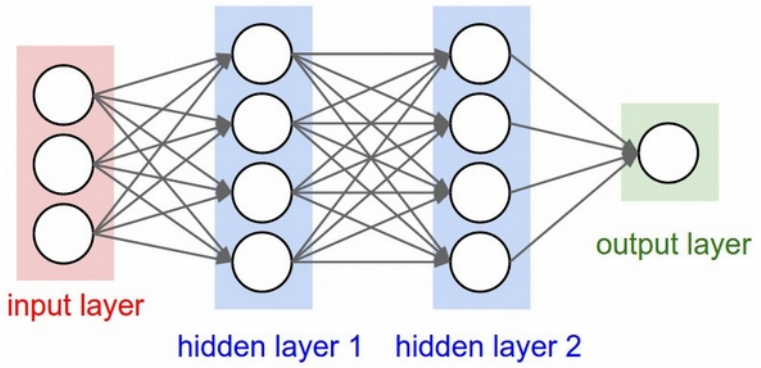


Logistic regression

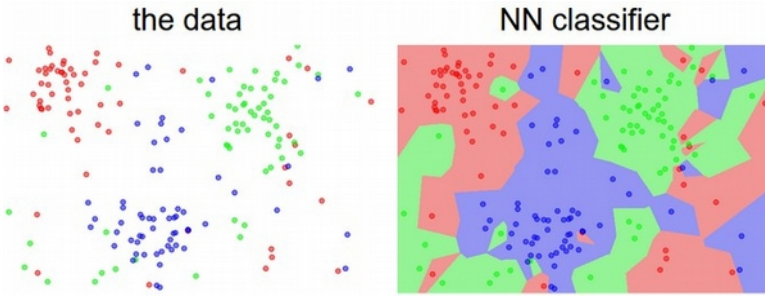


SVM



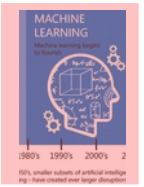


Neural networks



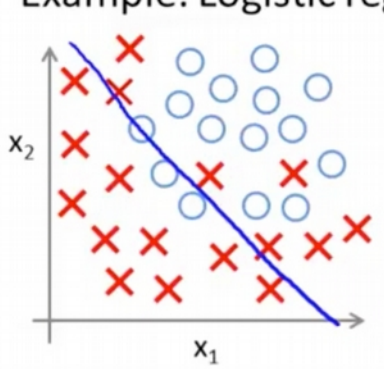
Nearest Neighbor

David Fernández Llorca (david.fernandezl@uah.es)



Parámetros

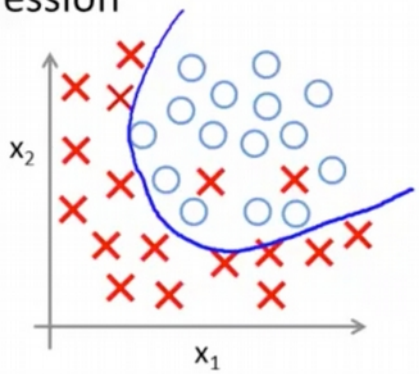
Example: Logistic regression



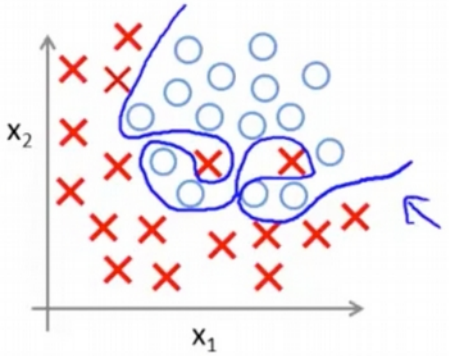
$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

(g = sigmoid function)

"Underfit"



$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$



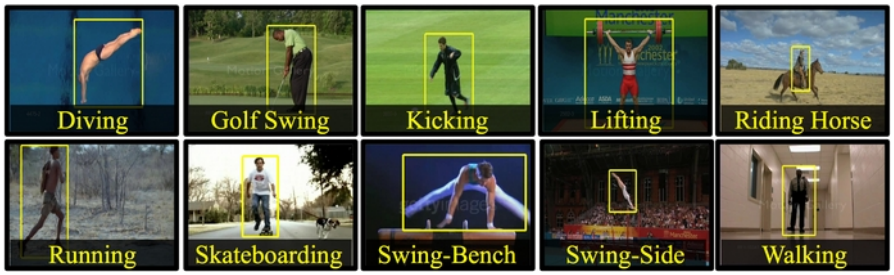
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$

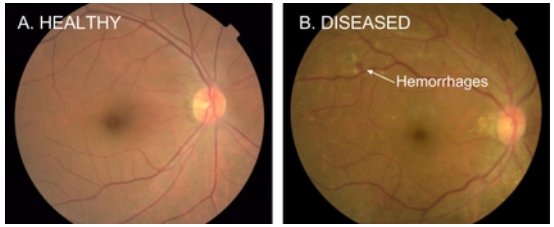
"Overfit"

David Fernández Llorca (david.fernandezl@uah.es)

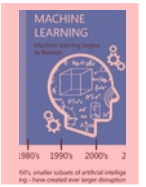
INVETT Group at University of Alcalá

Bases de datos



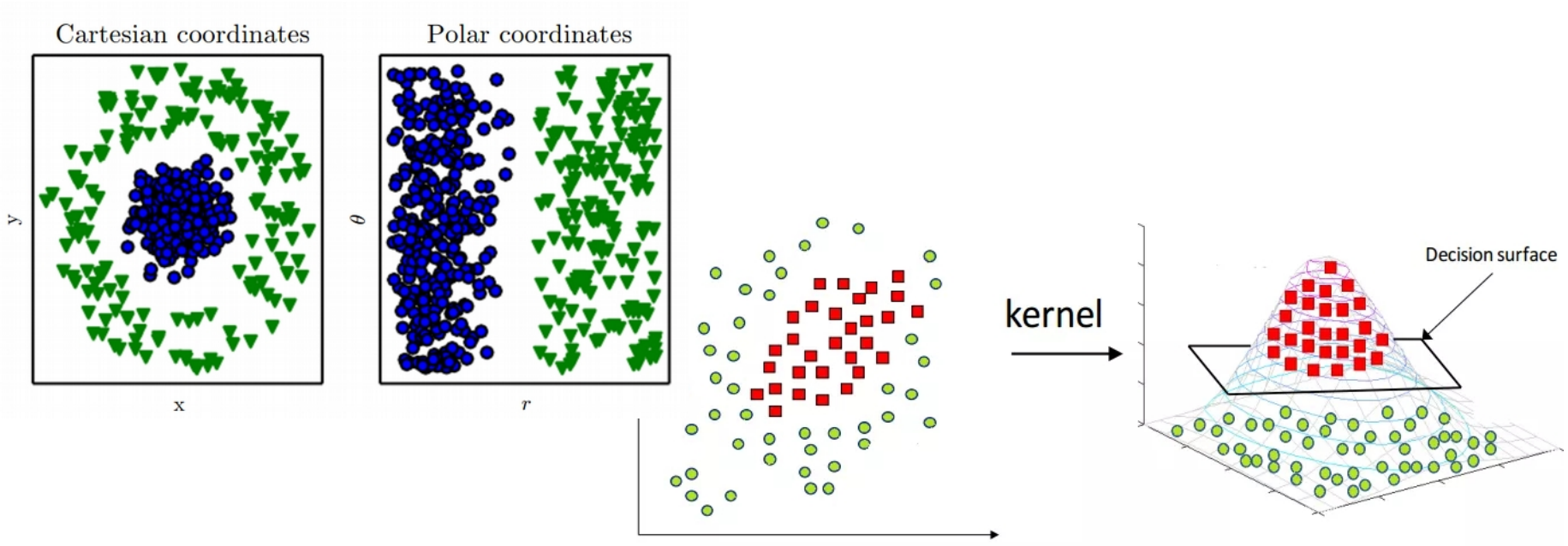


David Fernández Llorca (david.fernandezl@uah.es)



Representación de los datos: *key issue*

- La eficiencia en la representación de los datos es fundamental para el aprendizaje máquina
  - 1250 + 750 = 2000 o MCCL + DCCL = MM



David Fernández Llorca (david.fernandezl@uah.es)

Representación de los datos: *key issue*

- No aprendemos datos en crudo → extraemos *features*

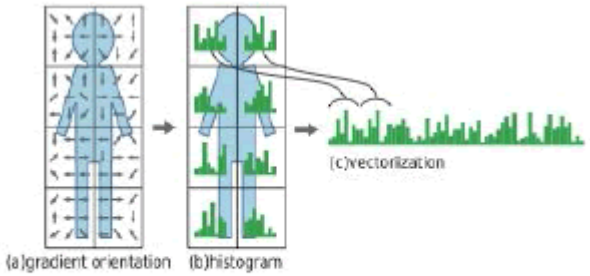
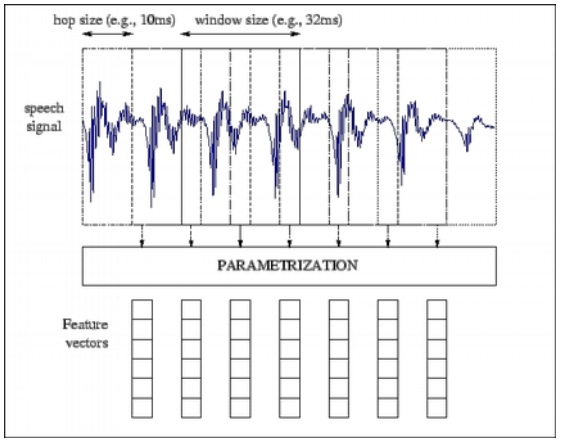
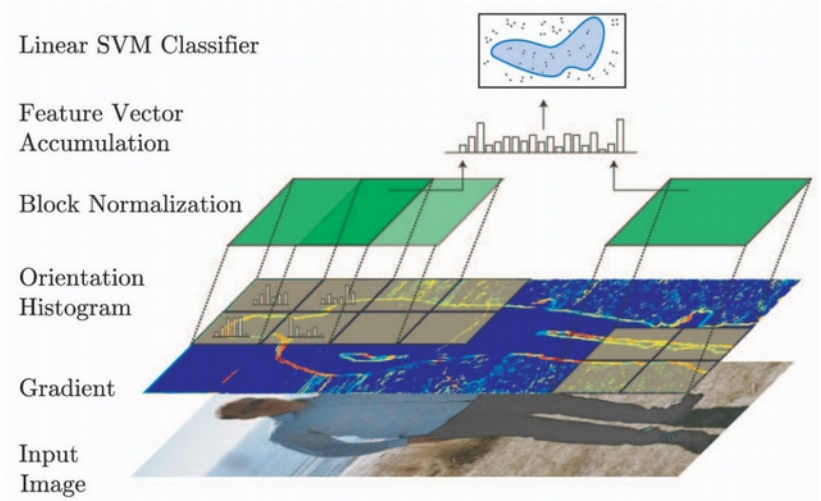


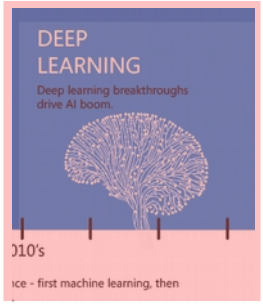
Figure 3.5 (a) (b) Overview of HOG calculation





David Fernández Llorca (david.fernandezl@uah.es)

Deep learning

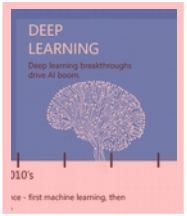


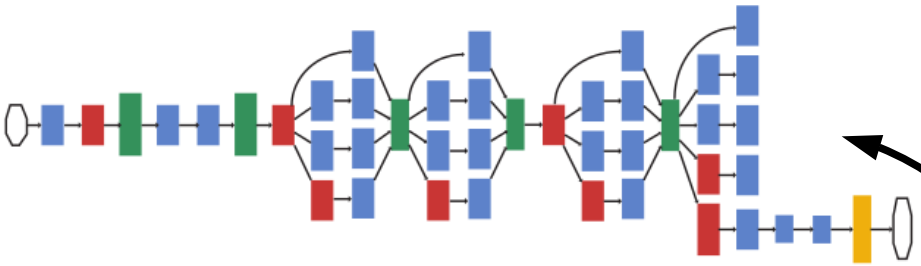
- **1º** Nueva forma de afrontar el problema de la representación
  - Se **aprenden** características relevantes (**features**)
  - Se **aprenden representaciones complejas** a partir de **representaciones** más **simples**.
- **2º** Si agrupamos en un grafo la relación de representaciones en capas conectadas, el grafo es muy profundo (***deep***) con muchas capas: **muchos** (pero muchos) **parámetros**.
- **3º** Para poder aprender los valores de los muchos (pero muchos) parámetros del modelo se necesitan muchos (muchos, muchos) datos: ***big data***.
- **4º** Para poder entrenar estos mega modelos con cantidades ingentes de datos se necesita un **HW** muy (pero muy) **potente**

David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

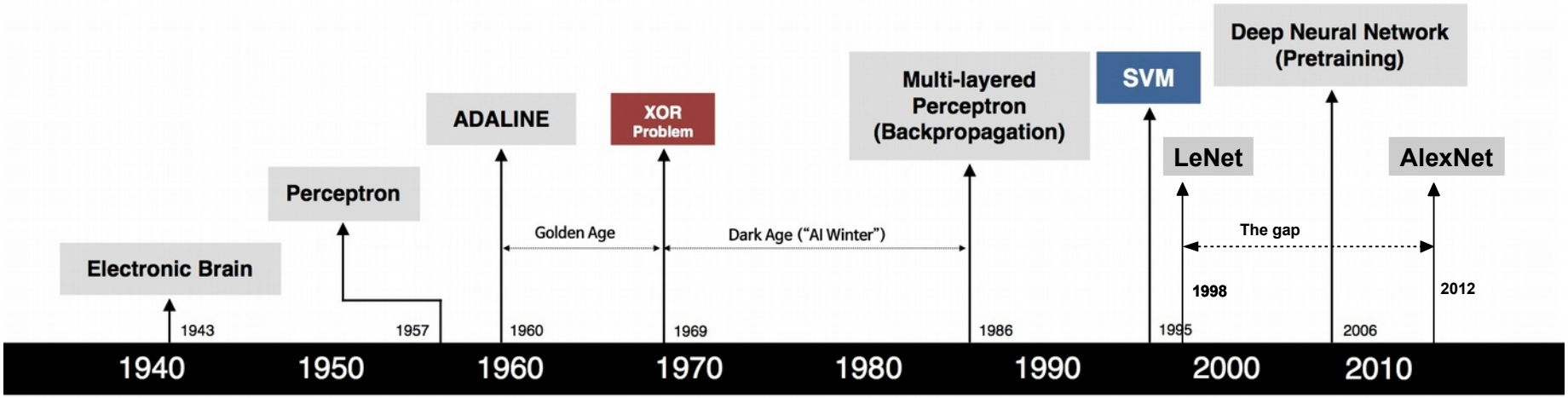
El círculo virtuoso del Deep Learning





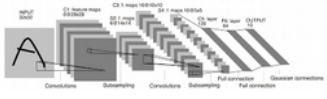
David Fernández Llorca (david.fernandezl@uah.es)

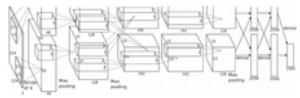
INVETT Group at University of Alcalá




Yann LeCUn


Alex Krizhevsky



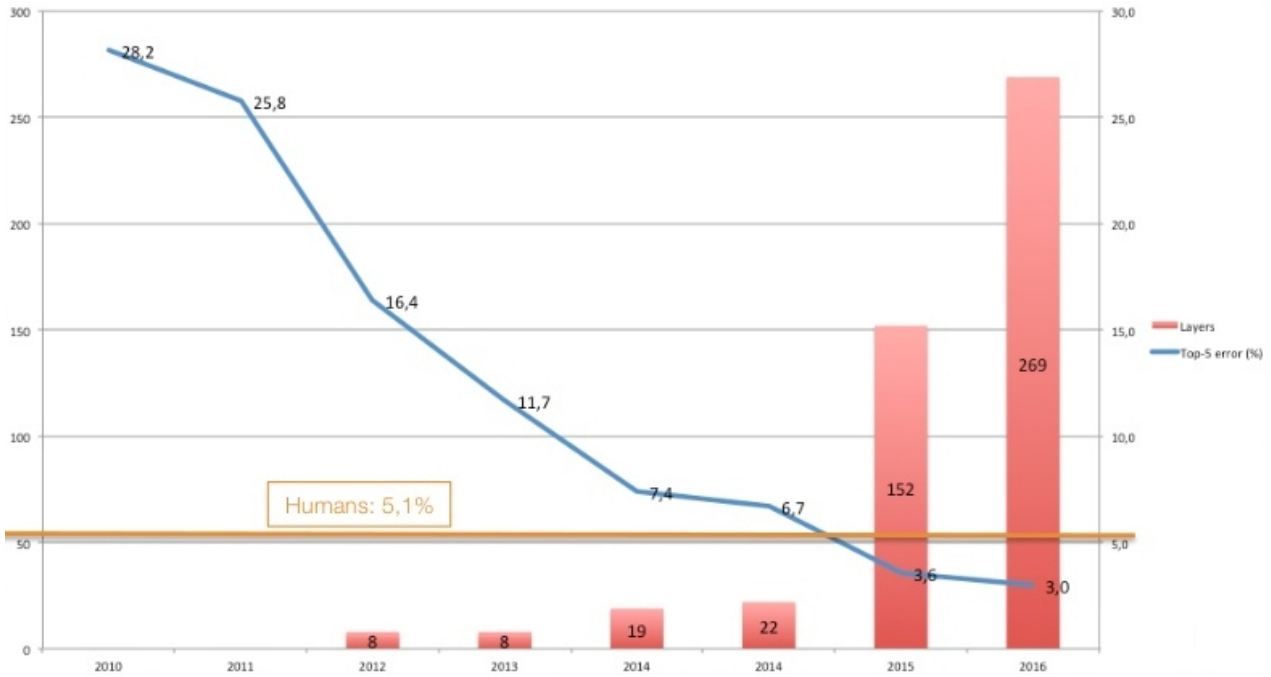


David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

ILSVRC

- 1000 categorías
- 1.2M-14M imágenes de entrenamiento
- 150K imágenes de test



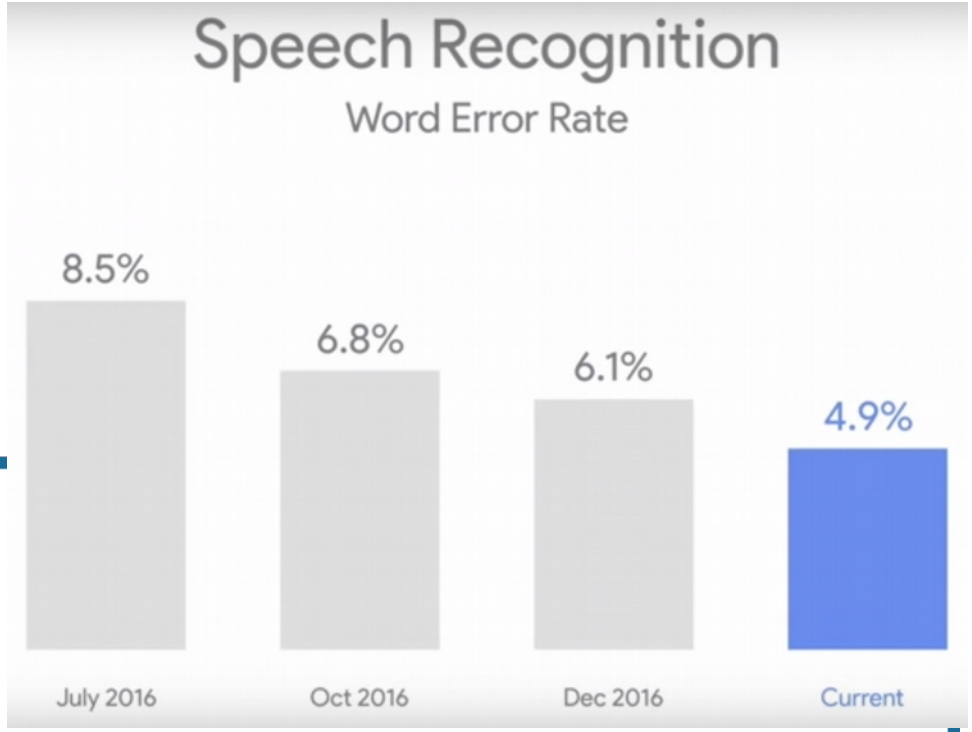
- **2017: top-5 error 2,25%**
- **Desde 2018 este challenge ya no se hace (se da por resuelto)**

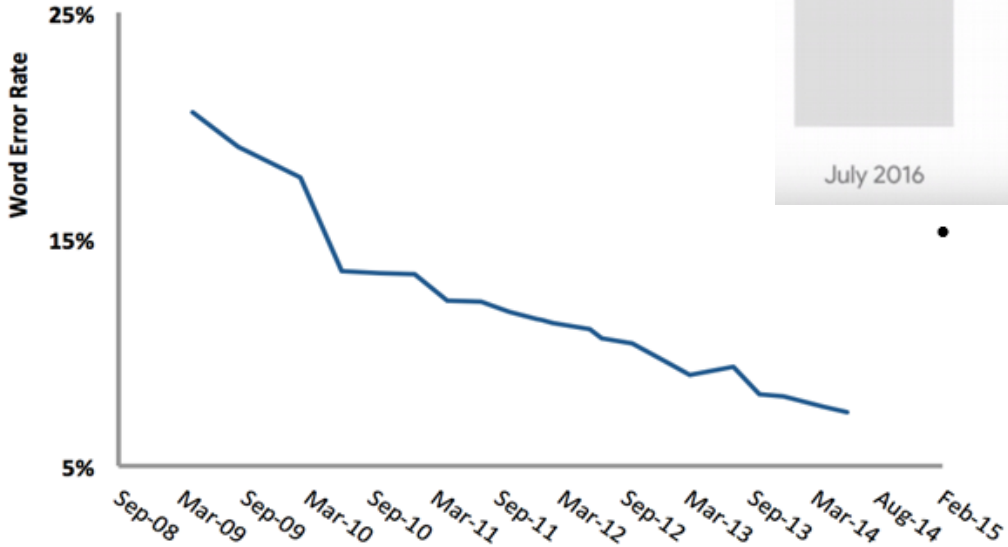
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

Speech recognition

- Word Error Rate (%)
- Google Assistant < 5%
- Microsoft ~ 5%
- Human error 2-4%

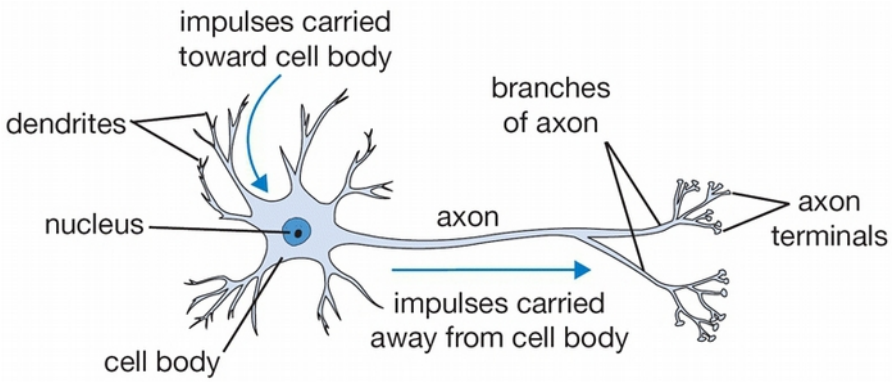


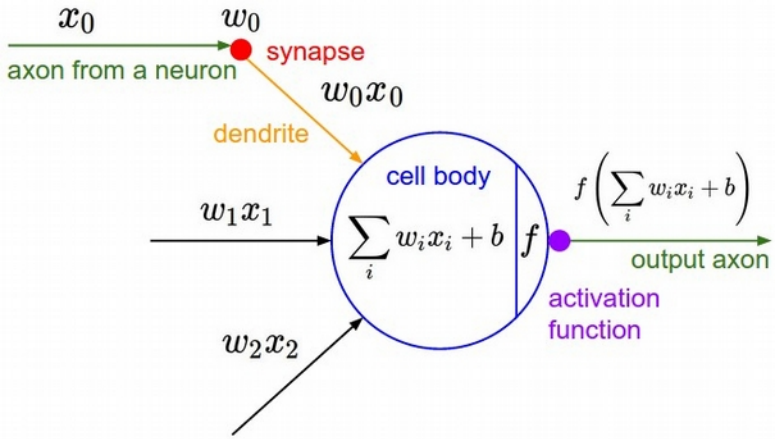


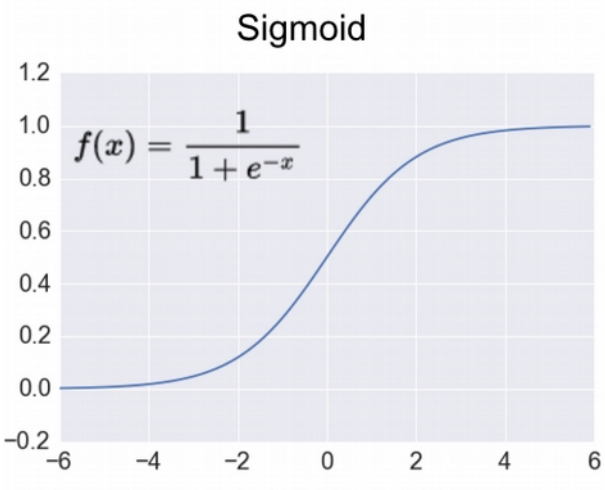
- Deep learning (DNNs) is driving recent performance improvements in ASR and meaning extraction

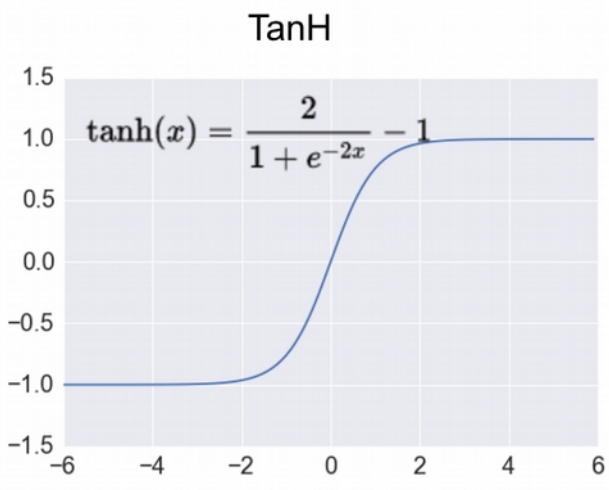
David Fernández Llorca (david.fernandezl@uah.es)

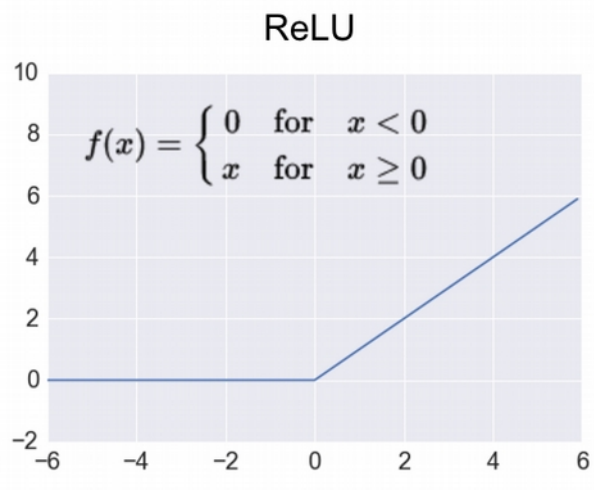
INVETT Group at University of Alcalá



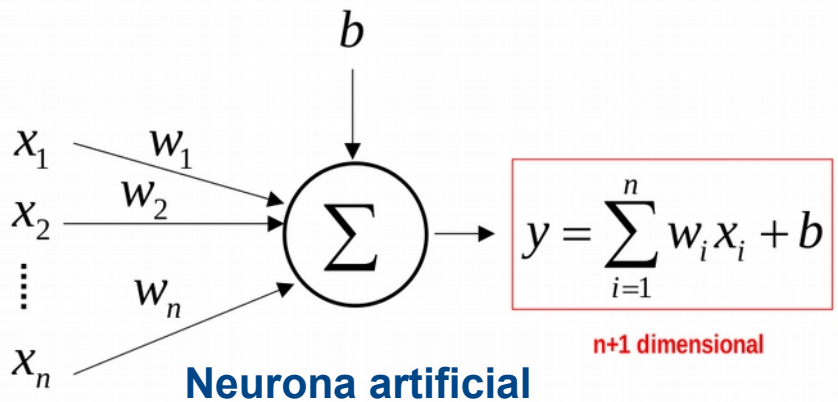






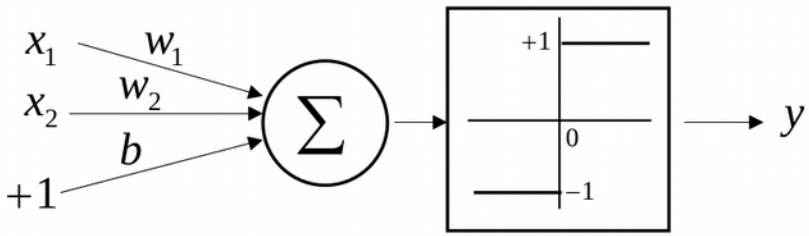


David Fernández Llorca (david.fernandezl@uah.es)



Neurona artificial (ADALINE)

PERCEPTRON

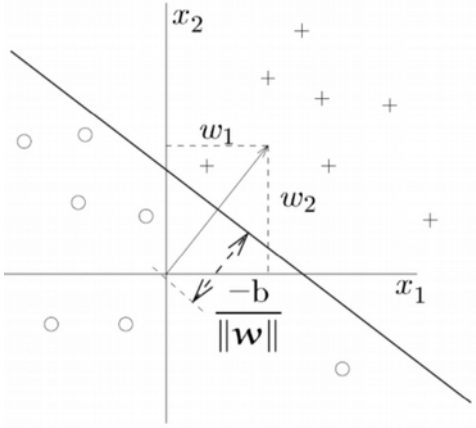


$$y = f\left(\sum_{i=1}^{2} w_i x_i + b\right)$$

$$f(s) = \begin{cases} 1 & if \quad s > 0 \\ -1 & otherwise \end{cases}$$

Ejemplo: 2 dimensiones

$$w_1 x_1 + w_2 x_2 + b = 0 \longrightarrow x_2 = -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}$$



Weights determine the slope of the line

Bias determines the offset (how far the line is from the origin)

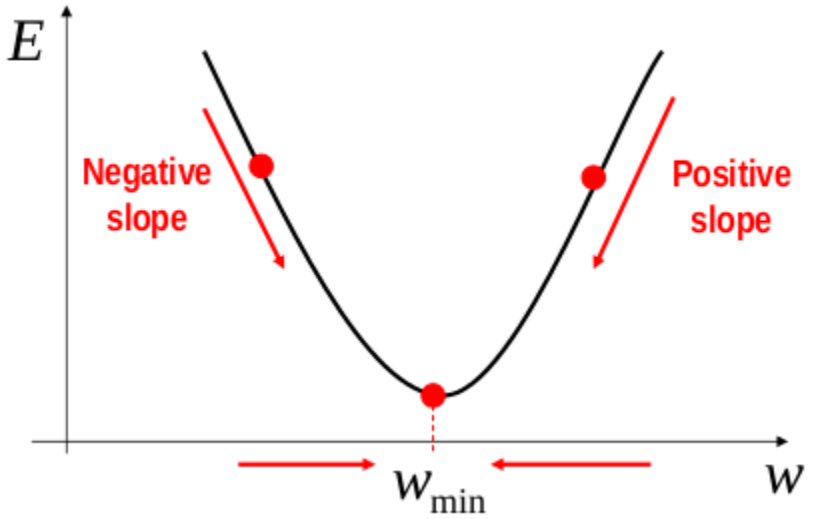
David Fernández Llorca (david.fernandezl@uah.es)

Entrenamiento

- Estimar los parámetros (pesos y offset) a partir de N datos (entradas **x** y salida deseada **d**)
- **Descenso del gradiente**:
  - Función de coste a minimizar: **Error**

$$E = \sum_{p=1}^{N} E^p = \frac{1}{2} \sum_{p=1}^{N} \left(d^p - \sum_{i=1}^{n} w_i x_i^p \right)^2$$

  - Los **pesos** se actualizan en un proceso **iterativo** de forma **proporcional** a **(-) la derivada del Error respecto de los pesos**



$$w^{new} = w^{old} - \gamma \frac{\partial E}{\partial w}$$

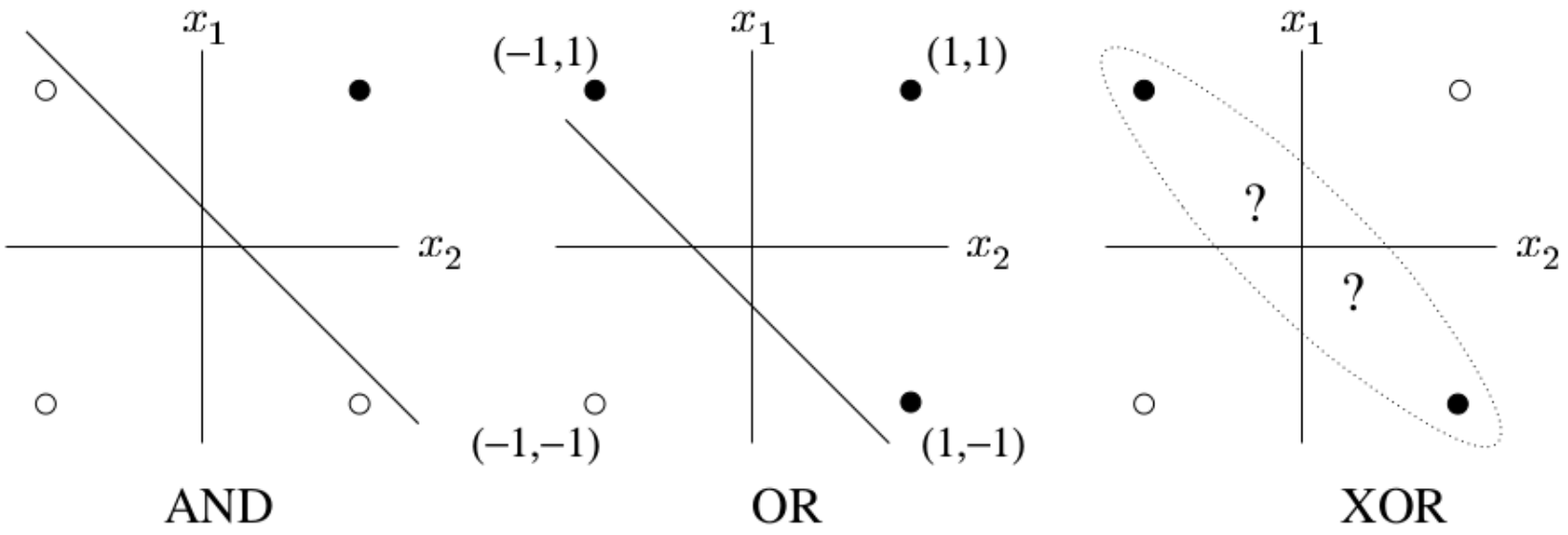
$$\Delta w = -\gamma \frac{\partial E}{\partial w}$$

$$w^{new} = w^{old} + \Delta w$$

David Fernández Llorca (david.fernandezl@uah.es)

El problema de la XOR

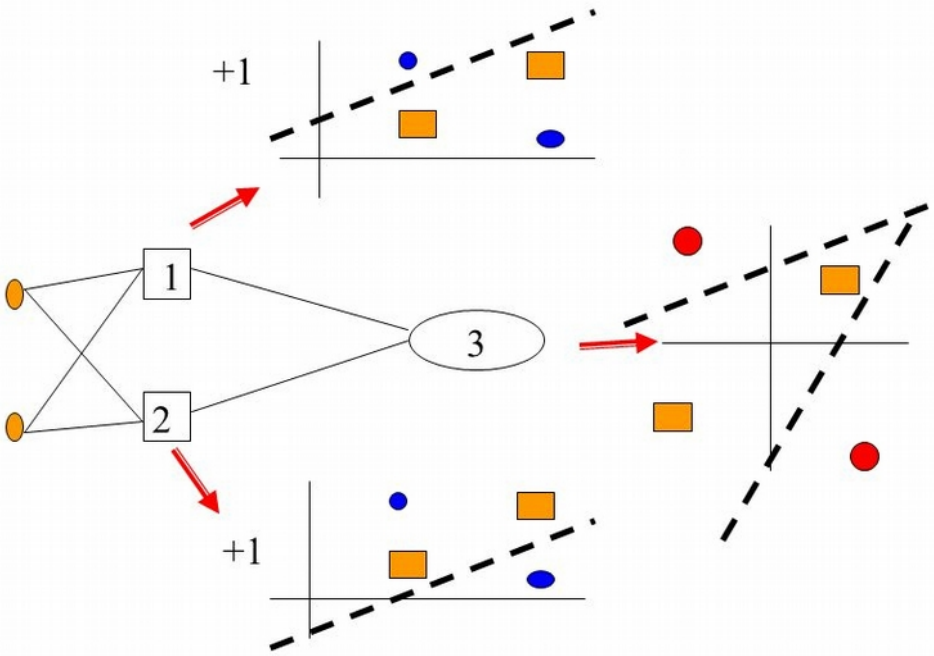
- Partimos de una red con **una sola capa** de neuronas: solo se pueden resolver problemas separables linealmente.
- Hasta este momento (1969) no se sabe como entrenar redes **con más capas**. Hasta 1986...

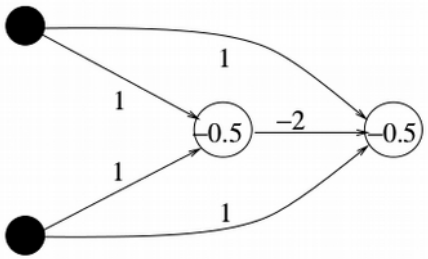


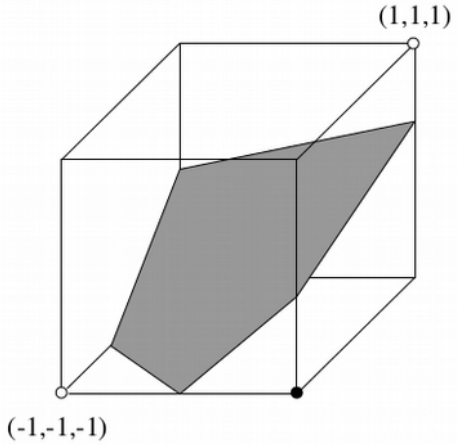
David Fernández Llorca (david.fernandezl@uah.es)

El problema de la XOR

- **Backpropagation** (1986): método para propagar el error desde las capas de salida hasta la capa de entrada.
  - Permite crear modelos con capas intermedias.
  - Añadir capas implica añadir dimensionalidad al problema y capacidad de discriminación del modelo discriminante.



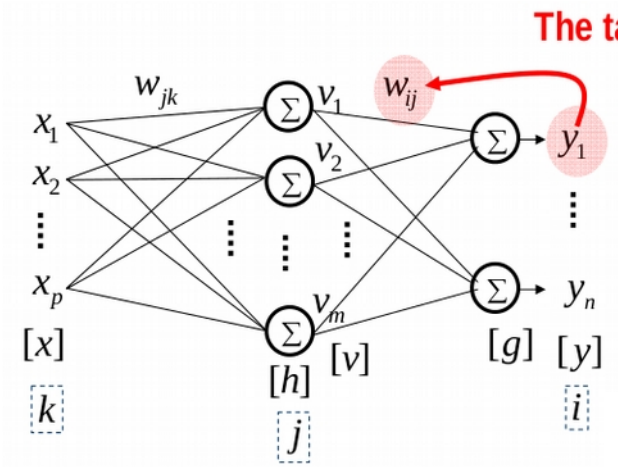




David Fernández Llorca (david.fernandezl@uah.es)

Backpropagation

– Aplicación recursiva de la regla de la cadena de la derivada



The target error is "back-propagated" to update the weights

$$\delta_i = (d_i(t) - y_i(t))$$

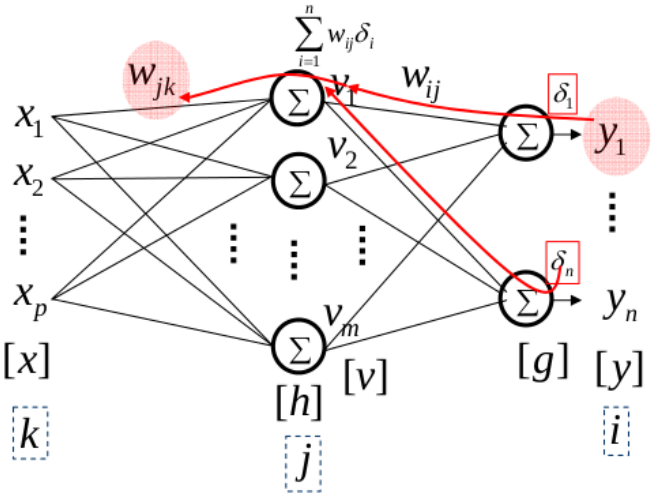
$$w_{ij}(t+1) = w_{ij}(t) + \gamma \delta v_j(t)$$

$$f(x,y,z) = (x+y)z$$

$$f = qz \qquad q = x + y$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial x}$$

¡¡ Muy fácil de implementar y de paralelizar (GPUs) !!



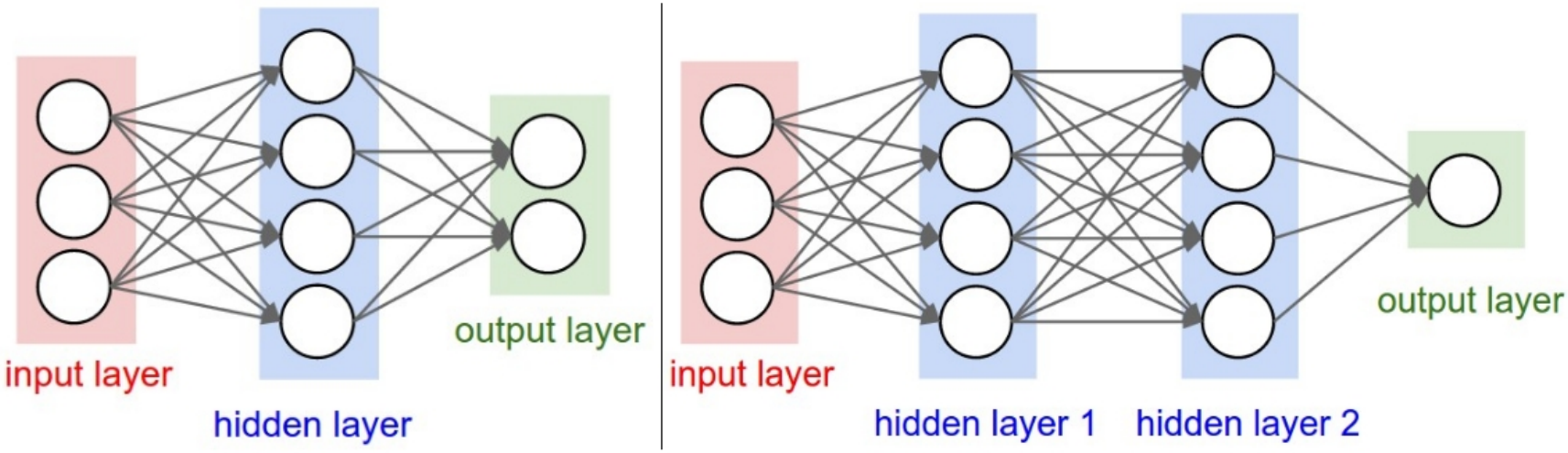
The target error is "back-propagated" multiplied by the weights between hidden and output layer

$$\Delta w_{jk} = \gamma \delta_j x_k$$

$$\delta_j = \sum_{i=1}^{n} w_{ij}\delta_i$$

David Fernández Llorca (david.fernandezl@uah.es)

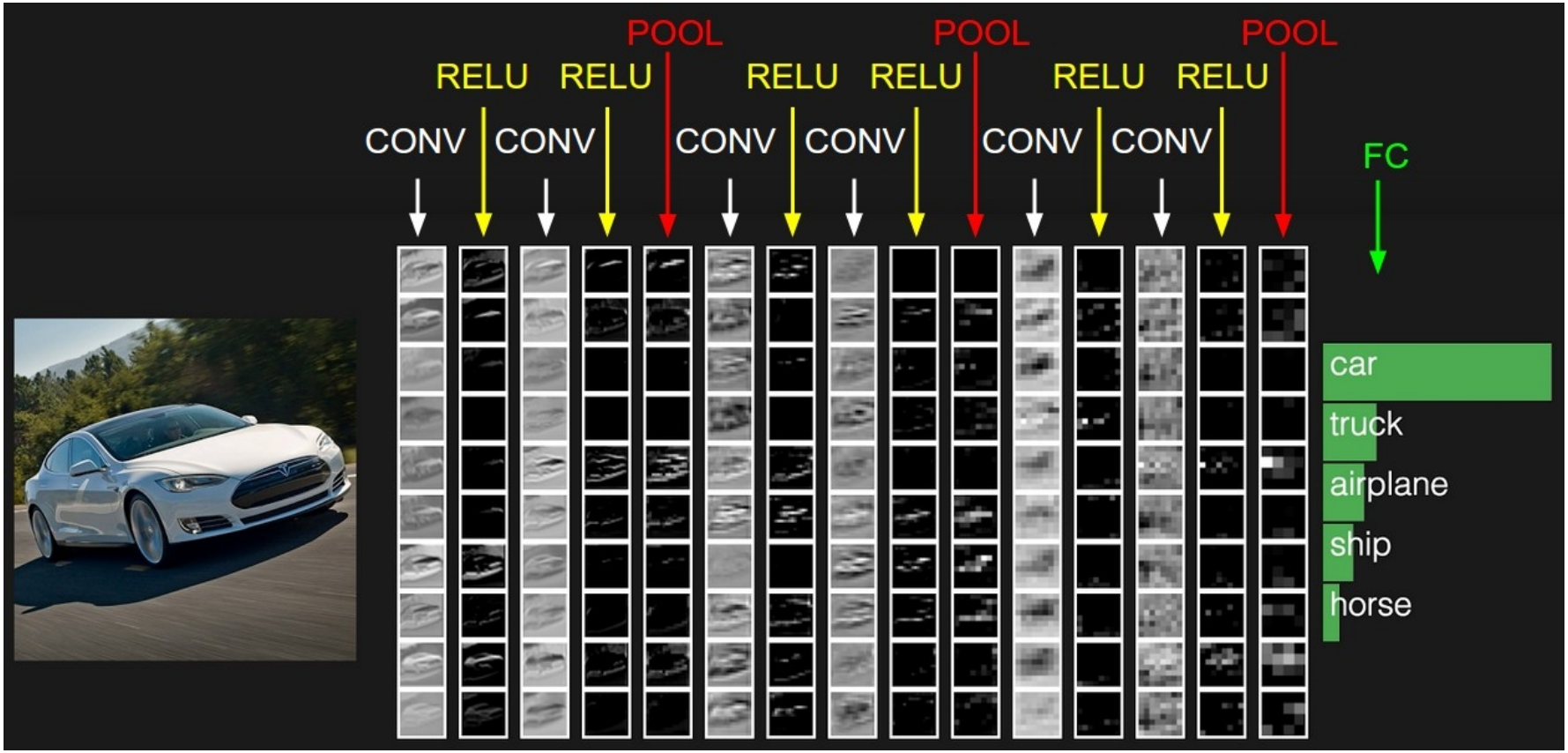
Redes Neuronales (Fully Connected FC)



Left: A 2-layer Neural Network (one hidden layer of 4 neurons (or units) and one output layer with 2 neurons), and three inputs.
Right: A 3-layer neural network with three inputs, two hidden layers of 4 neurons each and one output layer. Notice that in both cases there are connections (synapses) between neurons across layers, but not within a layer.

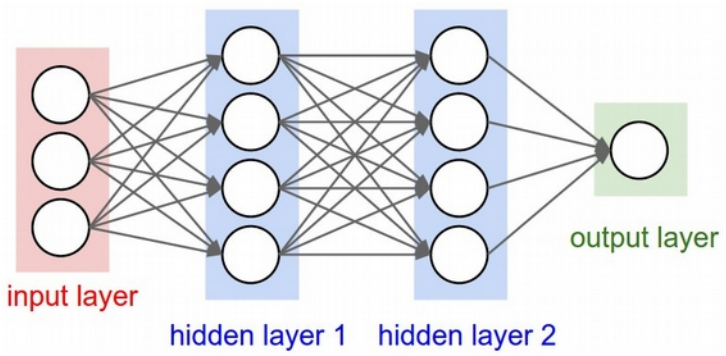
David Fernández Llorca (david.fernandezl@uah.es)

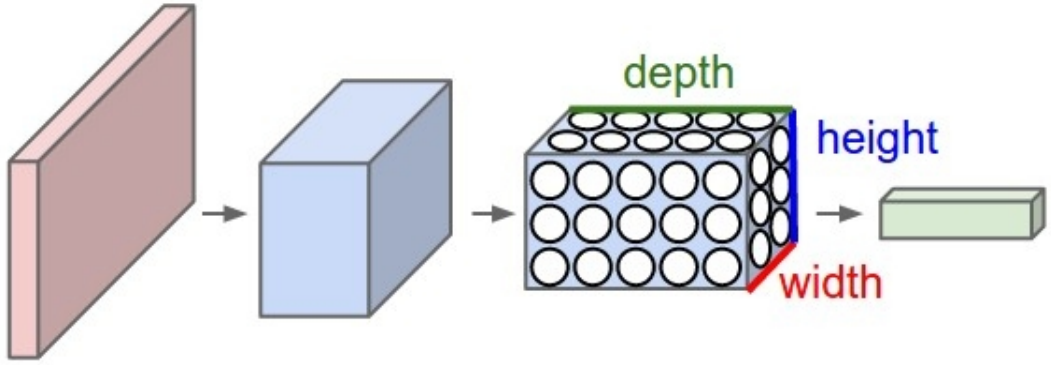
Redes Neuronales CONVOLUCIONALES

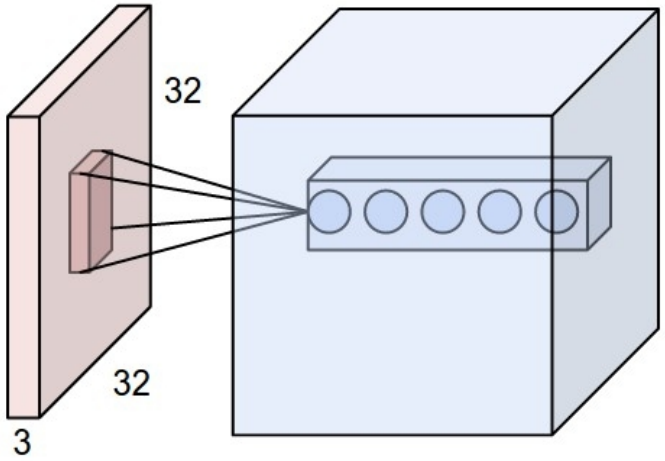


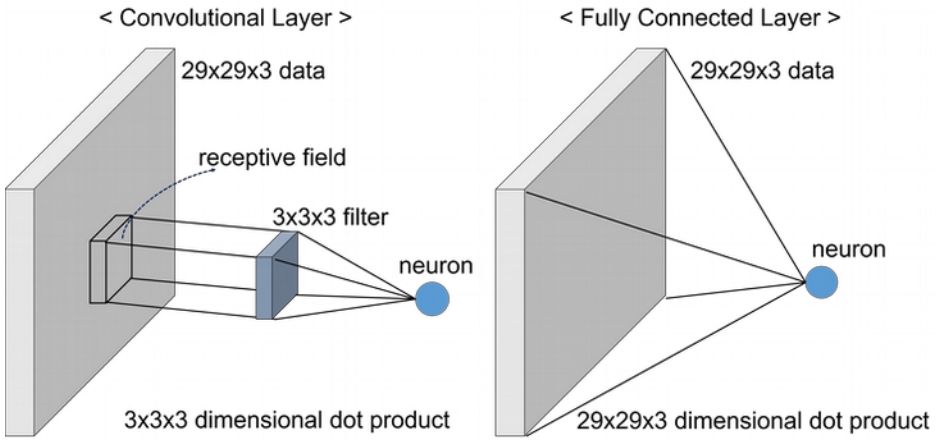
David Fernández Llorca (david.fernandezl@uah.es)

Capas CONVOLUCIONALES



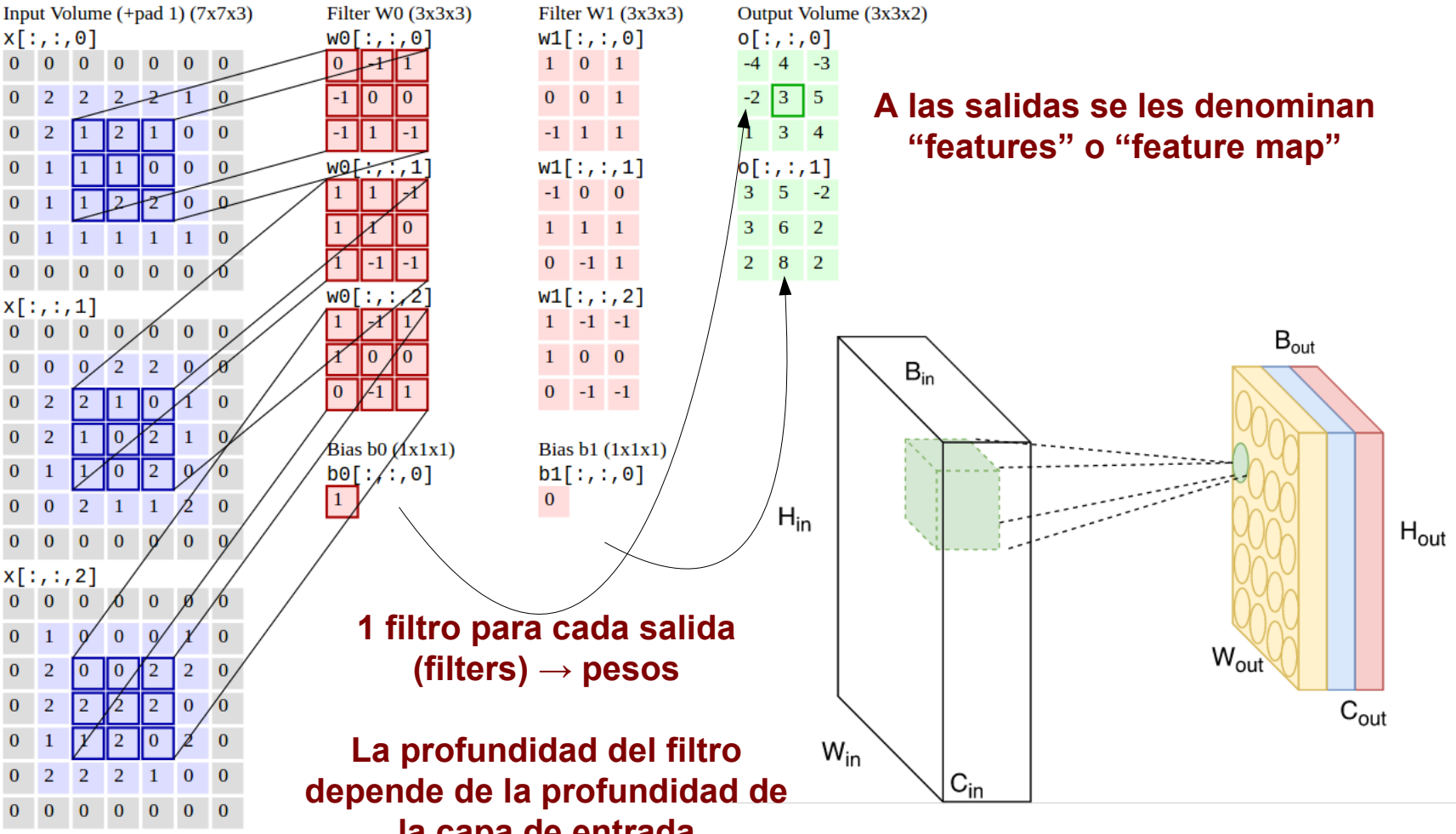






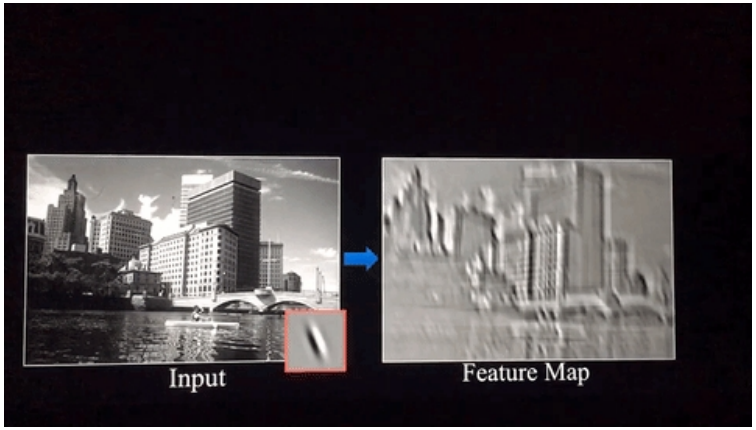
David Fernández Llorca (david.fernandezl@uah.es)

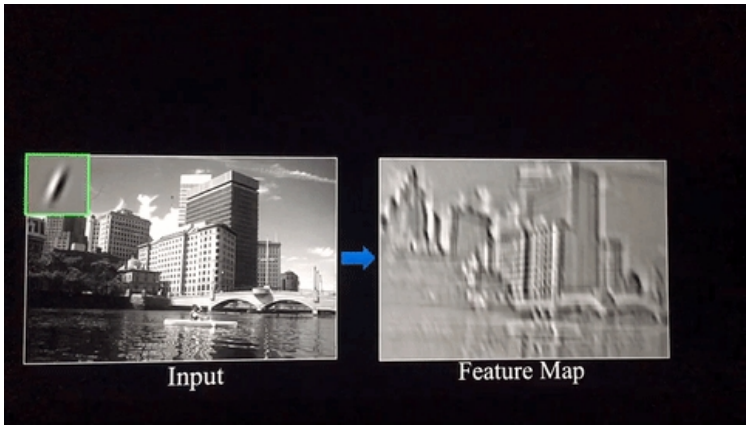
Capas CONVOLUCIONALES

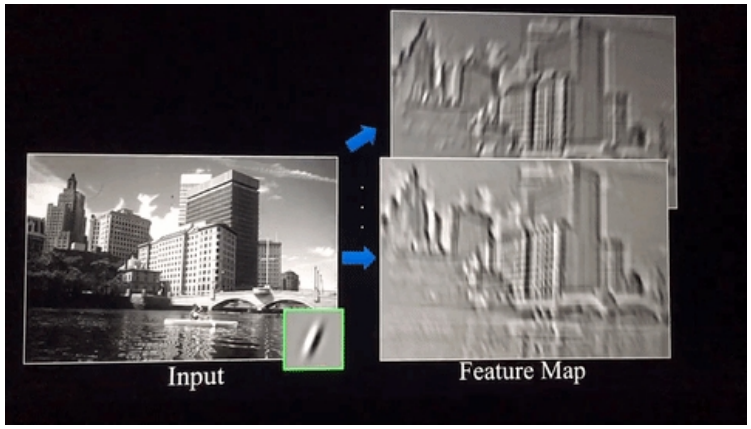


David Fernández Llorca (david.fernandezl@uah.es)

Capas CONVOLUCIONALES



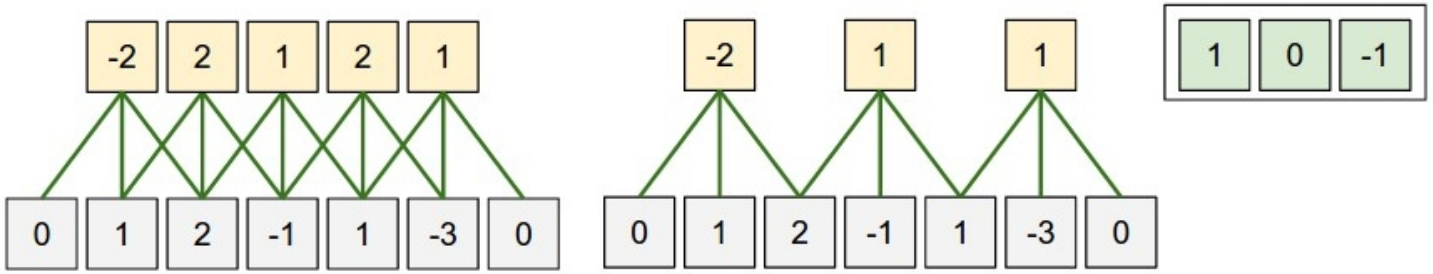




David Fernández Llorca (david.fernandezl@uah.es)

Capas CONVOLUCIONALES: parámetros principales

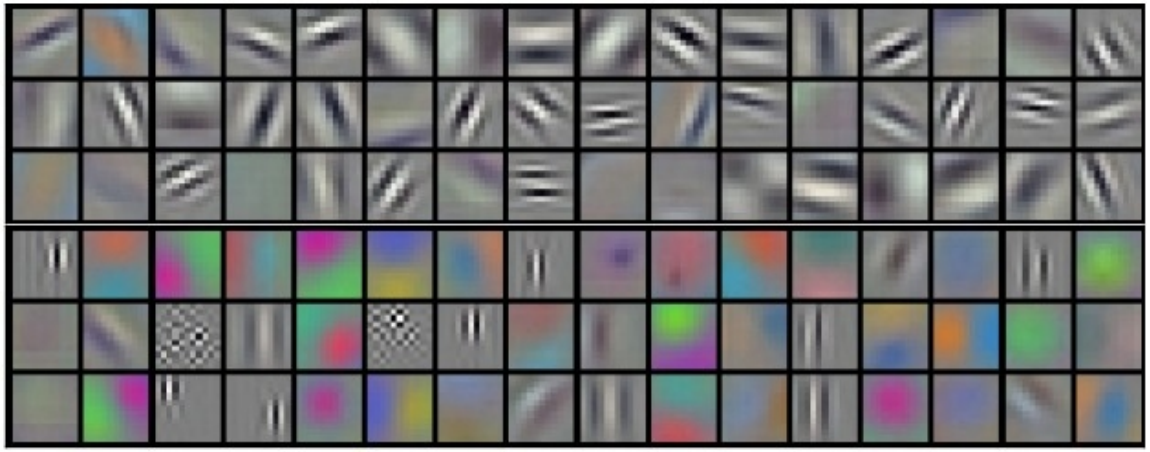
- Tamaño del volumen de entrada (anterior): ***W x W x Cin***
- Tamaño del filtro (receptive field): FxF
  - La profundidad es la de la capa de entrada (anterior): Cin
  - Total: ***F x F x Cin*** **(pesos de la red!!)**
- Stride o salto de pixel del filtro sobre el volumen de entrada: ***S***
- Padding: puede ser necesario rellenar con ceros el borde del volumen de entrada: ***P***
- Tamaño del volumen de salida:
  - El número de filtros ***Cout*** es un parámetro de diseño
  - El ancho y alto vendrá dado por: ***(W-F+2P)/S + 1*** **(debe dar un entero!!)**

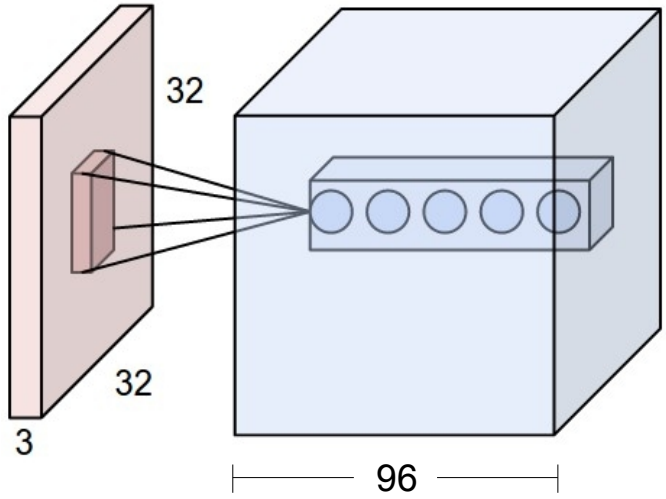


David Fernández Llorca (david.fernandezl@uah.es)

Capas CONVOLUCIONALES

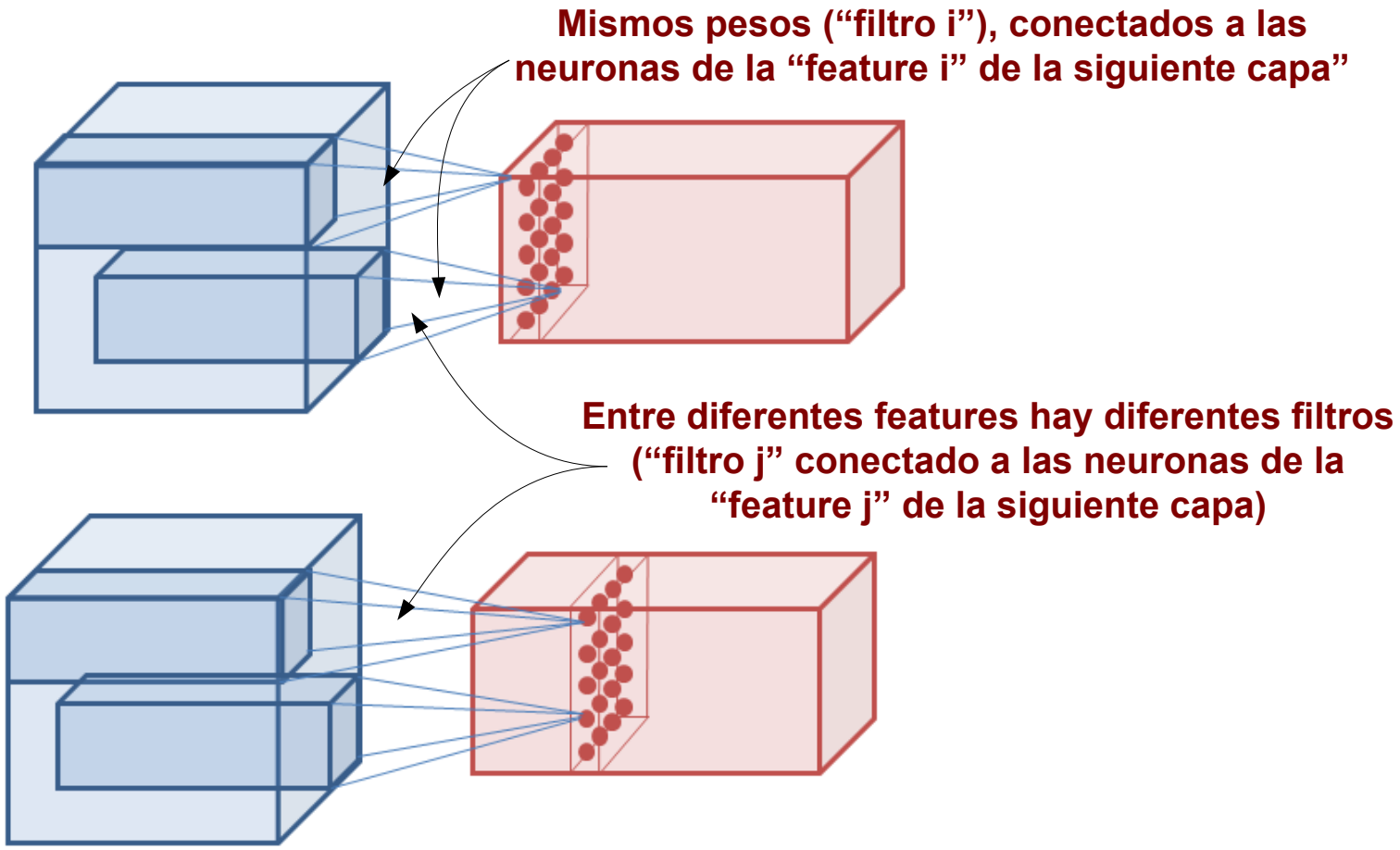
- **Ejemplo de filtros (pesos) aprendidos**
  - 96 filtros de 11x11x3 de la primera capa de una red
  - Los filtros se comparten para todas las neuronas de cada "feature map" del bloque de salida.





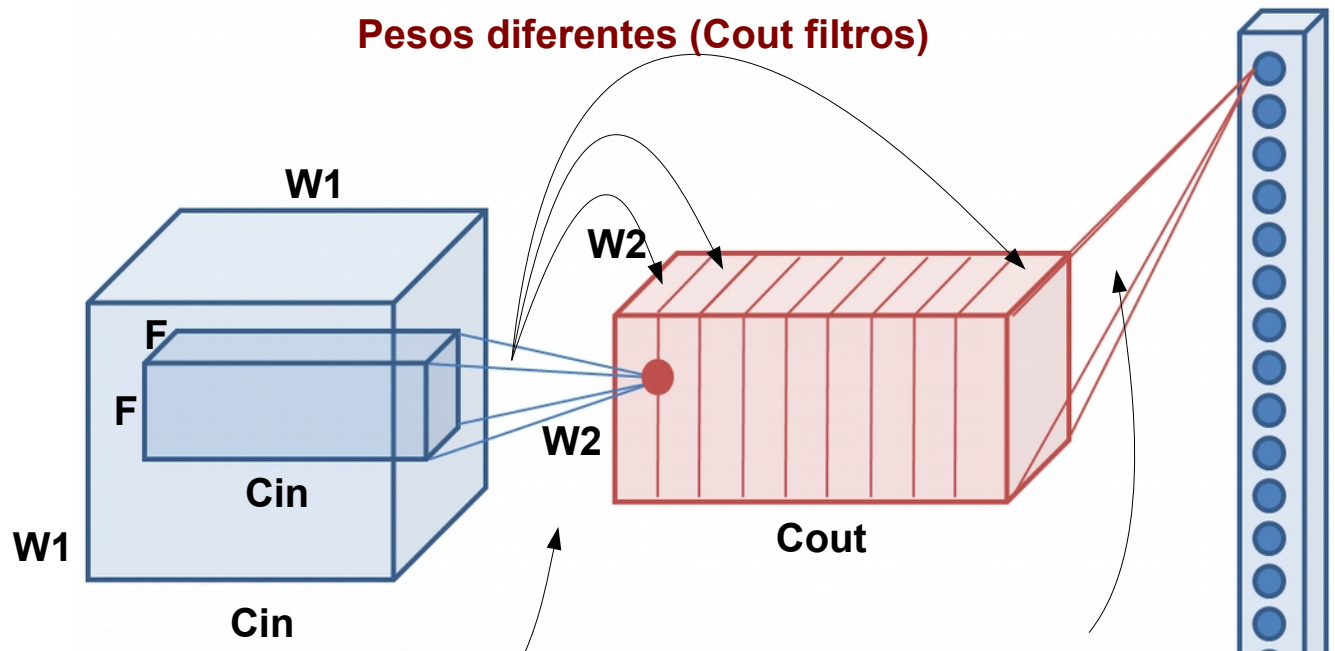
David Fernández Llorca (david.fernandezl@uah.es)

Capas CONVOLUCIONALES



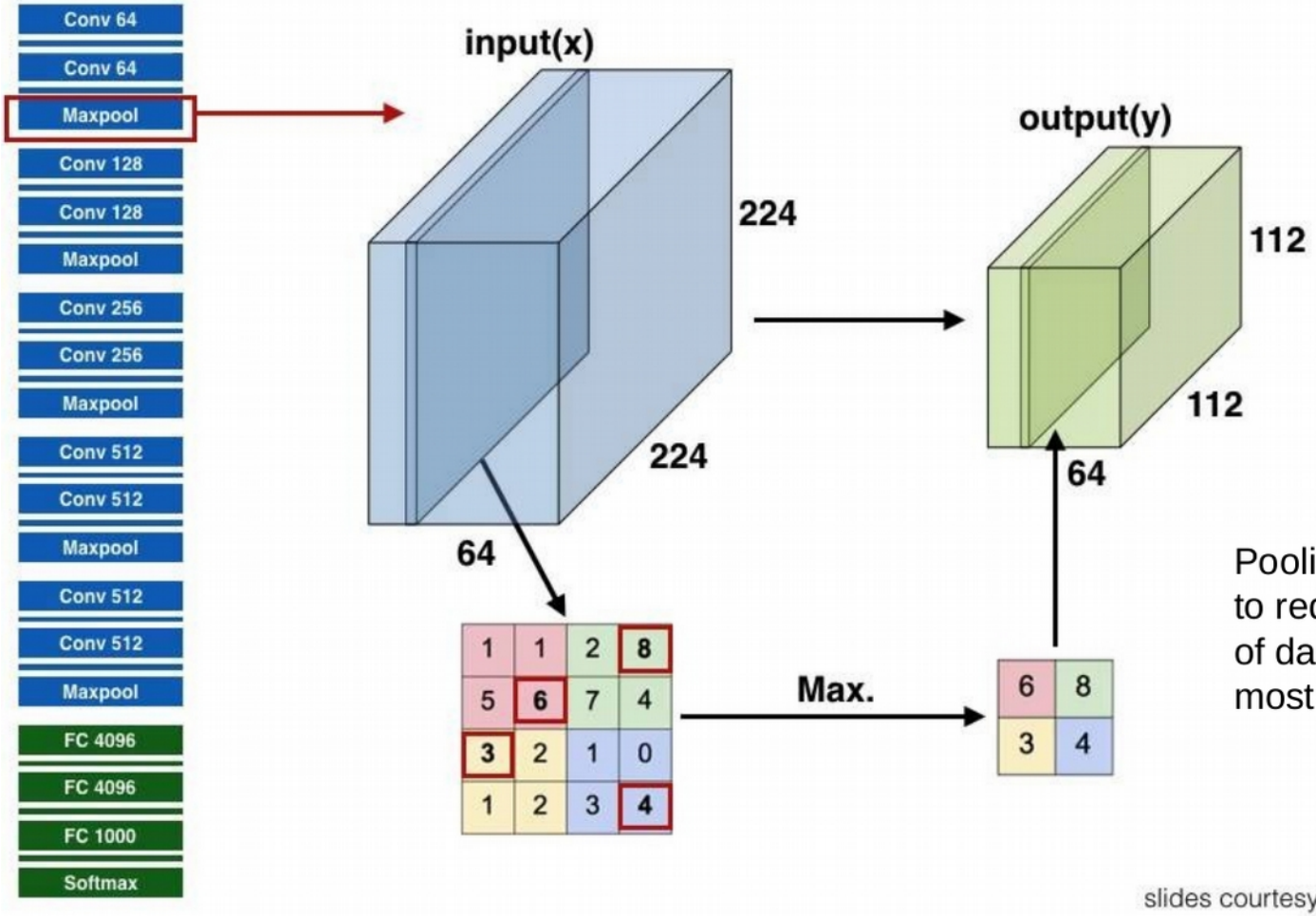
David Fernández Llorca (david.fernandezl@uah.es)

Capas CONVOLUCIONALES



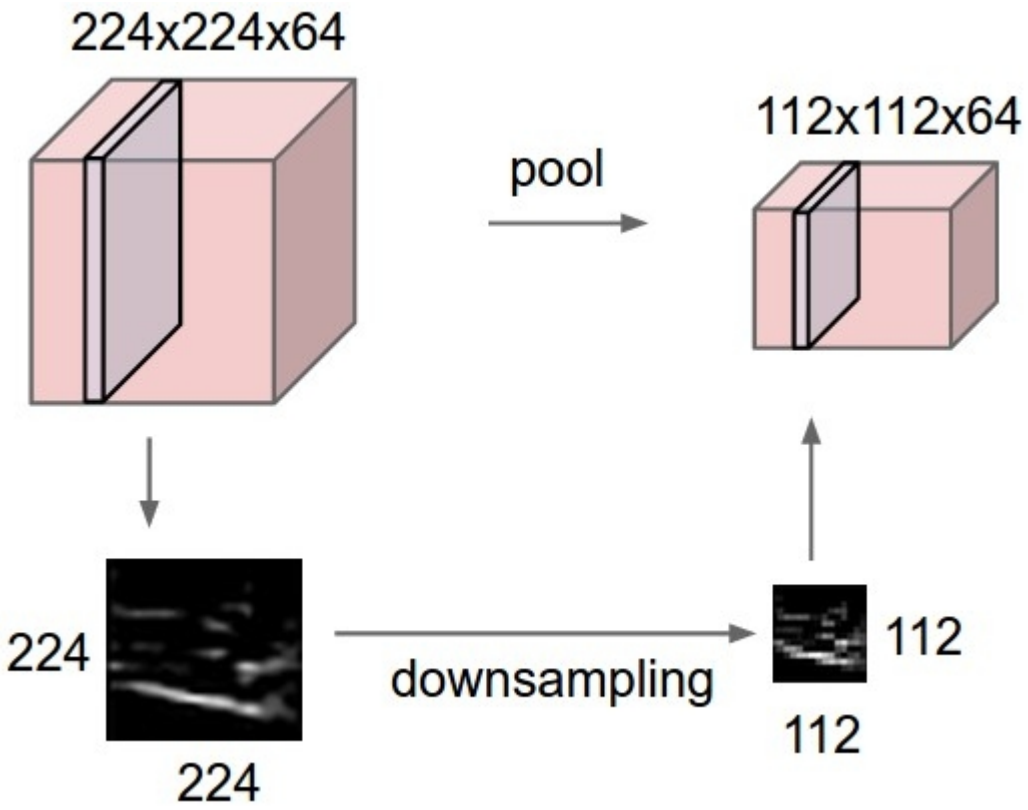
David Fernández Llorca (david.fernandezl@uah.es)

Capas POOLING



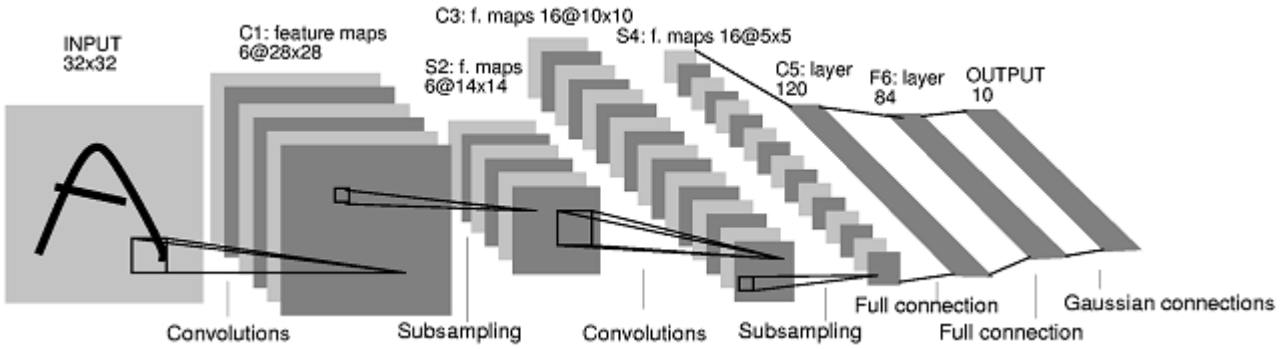
David Fernández Llorca (david.fernandezl@uah.es)

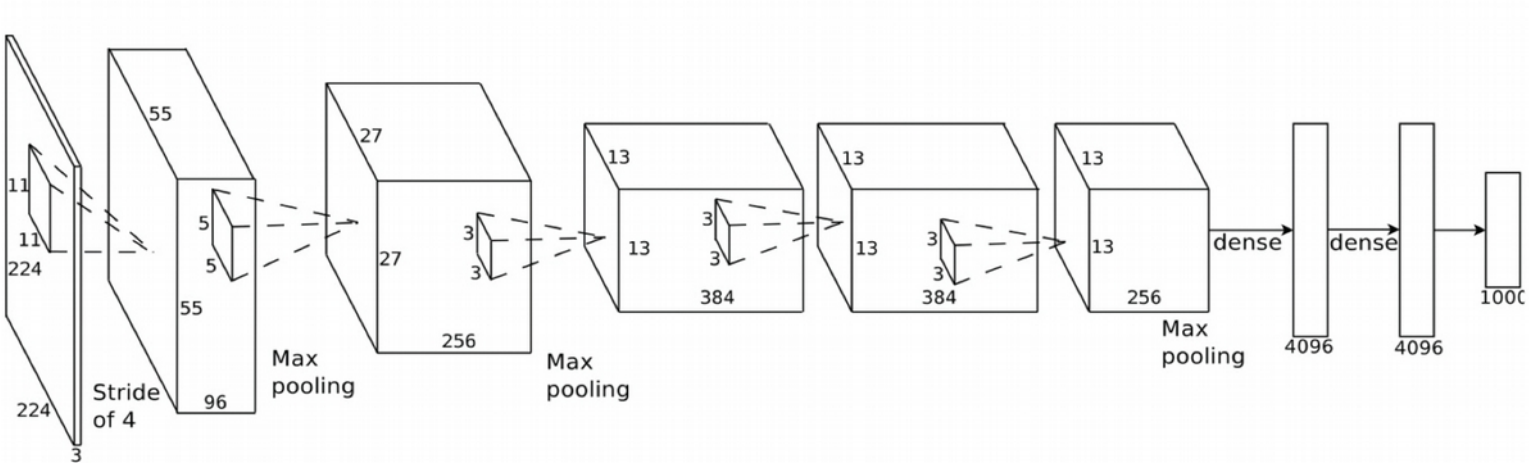
Capas POOLING



David Fernández Llorca (david.fernandezl@uah.es)

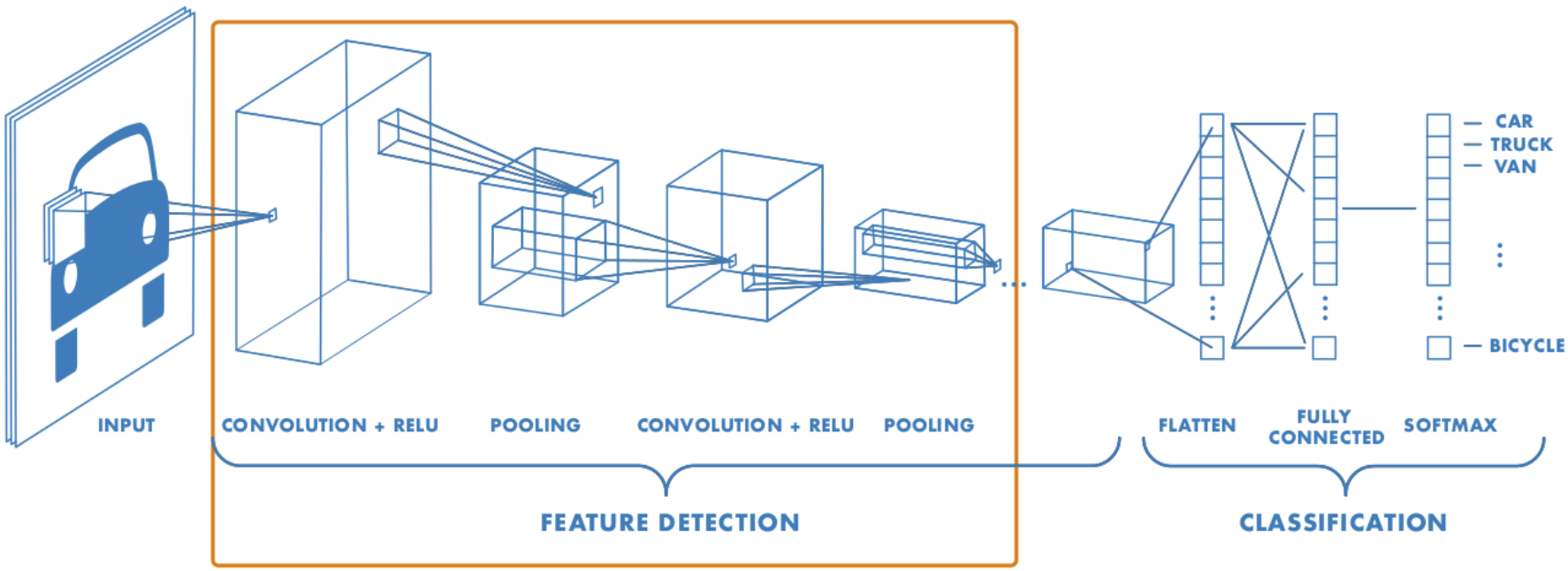
INVETT Group at University of Alcalá





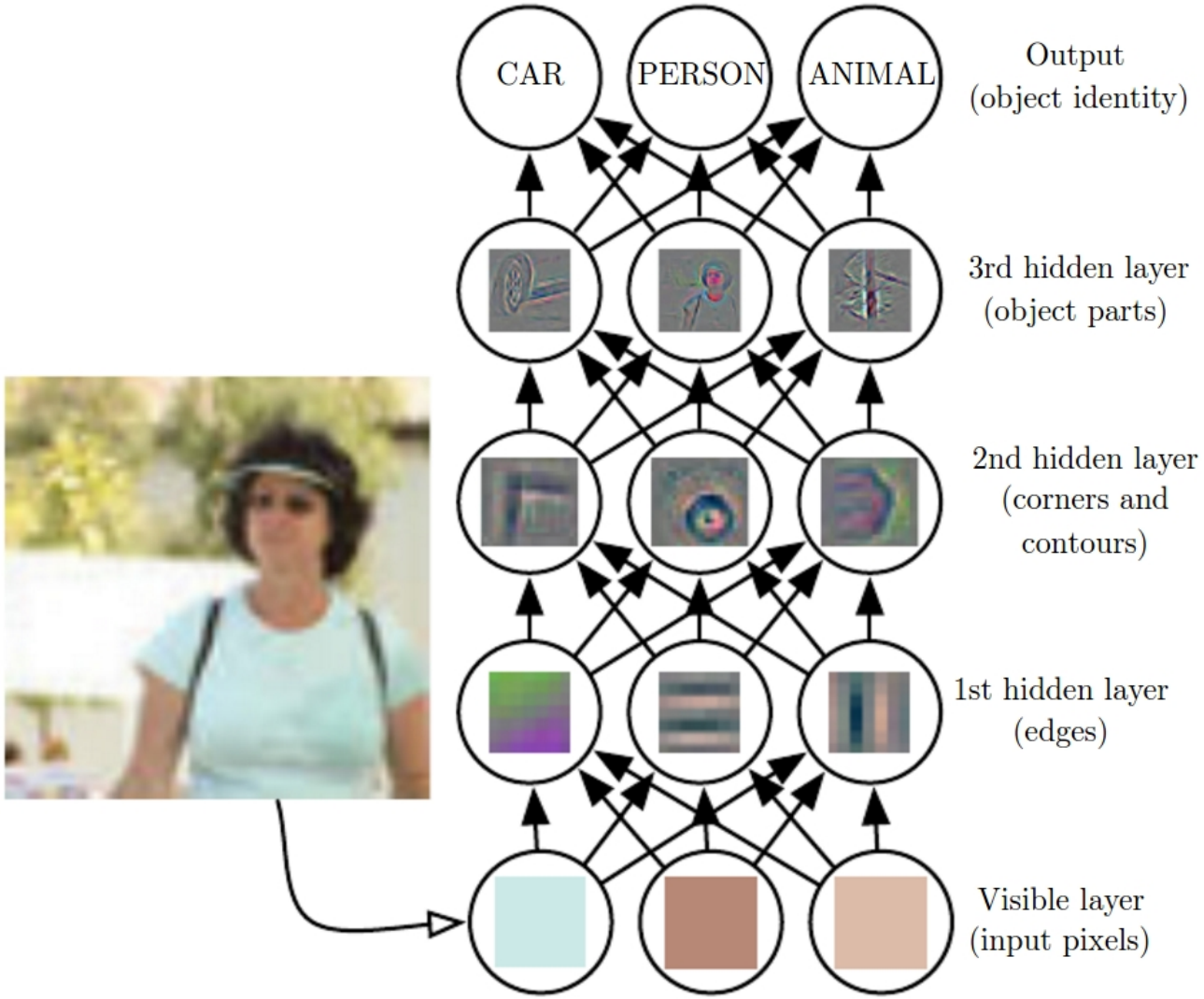
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá



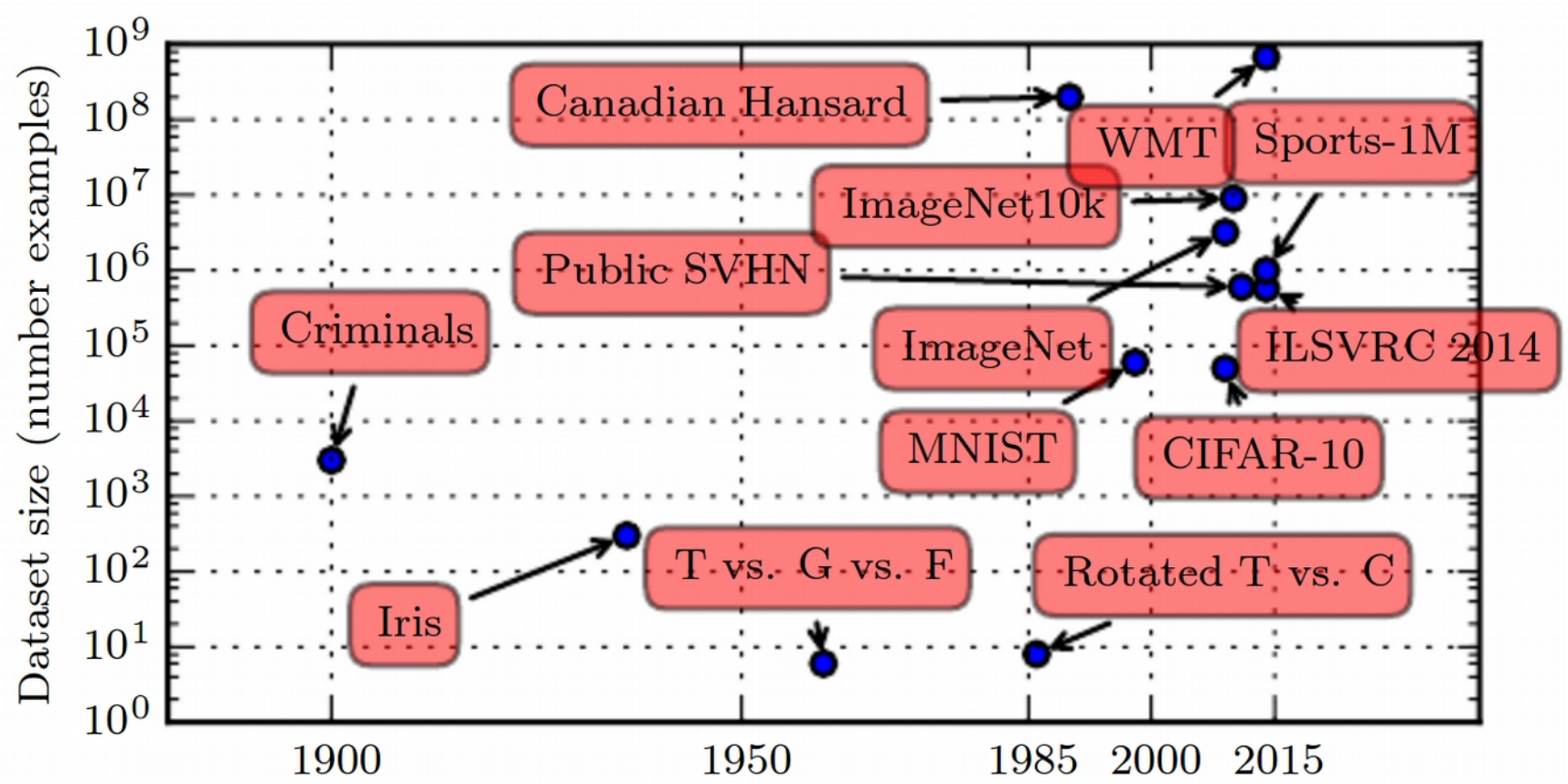
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá



David Fernández Llorca (david.fernandezl@uah.es)

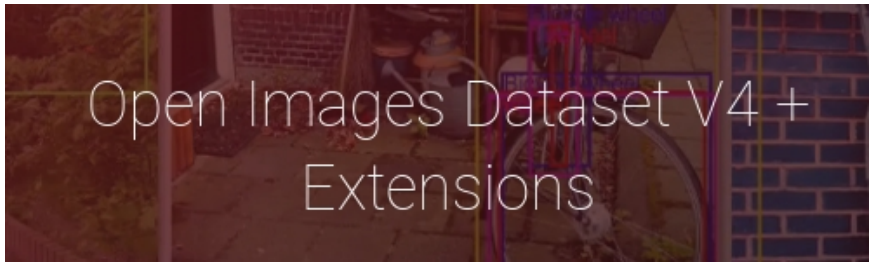
INVETT Group at University of Alcalá



David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

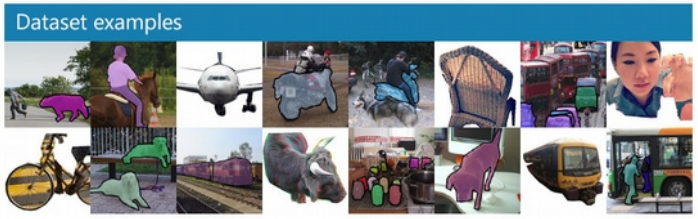
14 millones de imágenes
1000 categorías (clasificación)



30 millones de imágenes
15 millones de regiones
600 categorías (clasificación + localización)

Google

coco
Common Objects in Context



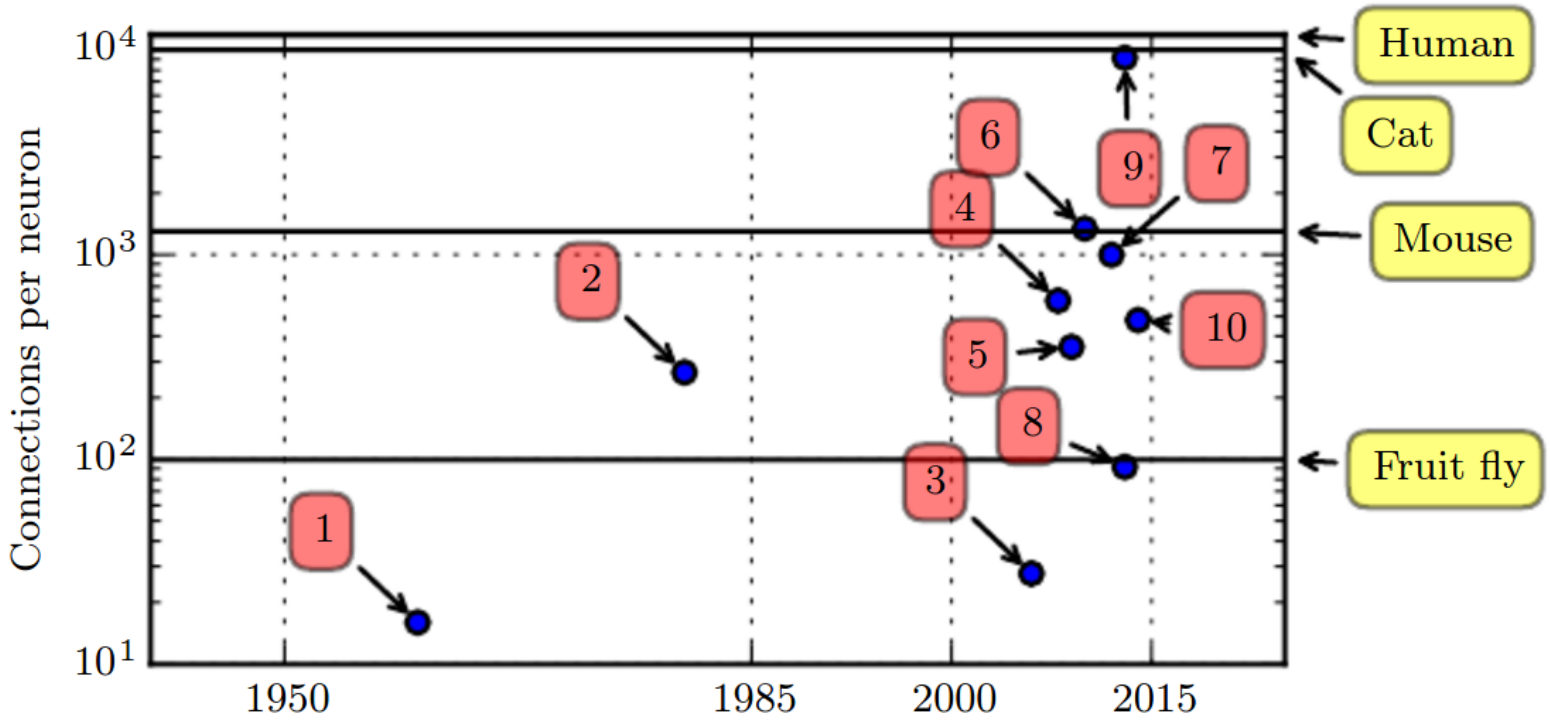
330K imágenes; 1.5M objetos;
250K personas con keypoints;
Segmentación de objetos+pose

Microsoft

facebook

David Fernández Llorca (david.fernandezl@uah.es)

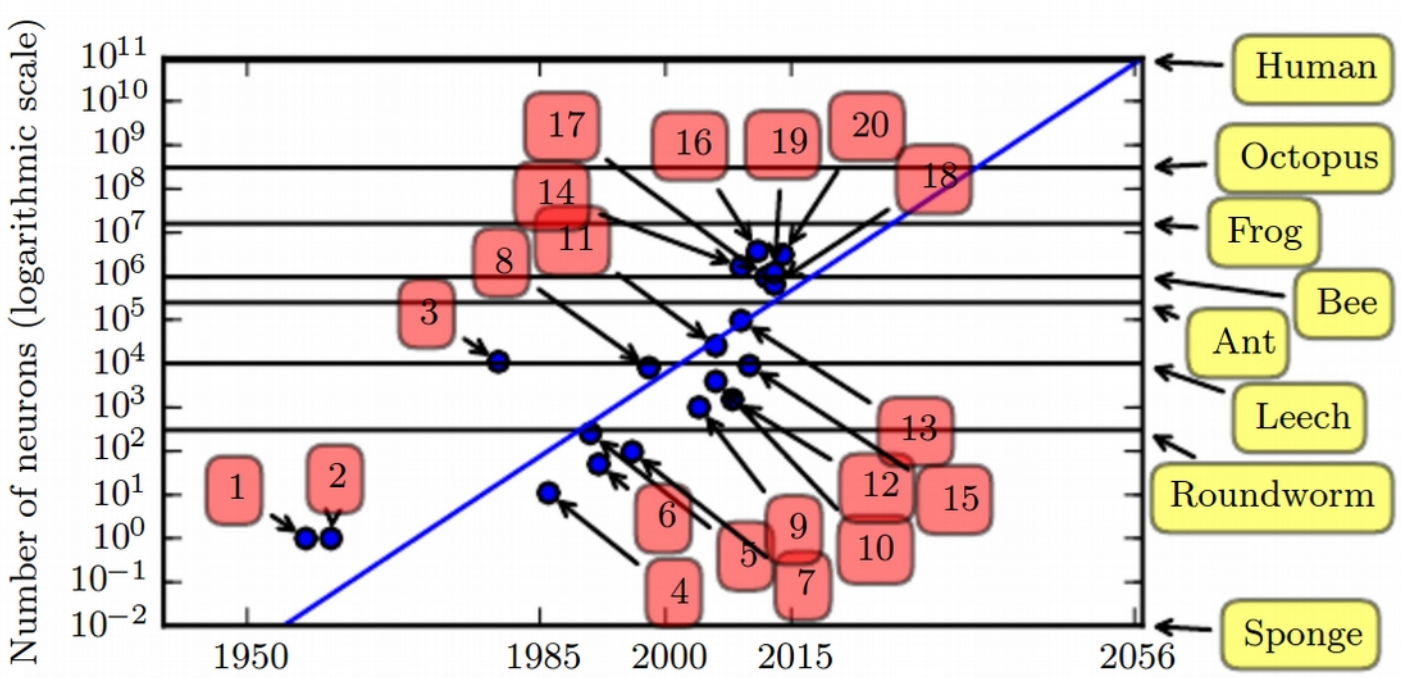
INVETT Group at University of Alcalá



1. Adaptive linear element (Widrow and Hoff, 1960)
2. Neocognitron (Fukushima, 1980)
3. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
4. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
5. Unsupervised convolutional network (Jarrett *et al.*, 2009)
6. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
7. Distributed autoencoder (Le *et al.*, 2012)
8. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
9. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
10. GoogLeNet (Szegedy *et al.*, 2014a)

David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

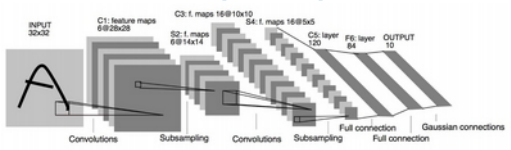


1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart *et al.*, 1986b)
5. Recurrent neural network for speech recognition (Robinson and Fallside, 1991
6. Multilayer perceptron for speech recognition (Bengio *et al.*, 1991)
7. Mean field sigmoid belief network (Saul *et al.*, 1996)
8. LeNet-5 (LeCun *et al.*, 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton *et al.*, 2006)
11. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
12. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
13. GPU-accelerated deep belief network (Raina *et al.*, 2009)
14. Unsupervised convolutional network (Jarrett *et al.*, 2009)
15. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le *et al.*, 2012)
18. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
19. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
20. GoogLeNet (Szegedy *et al.*, 2014a)

David Fernández Llorca (david.fernandezl@uah.es)

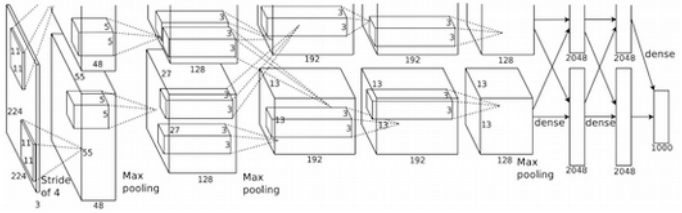
INVETT Group at University of Alcalá

LeNet (1998)

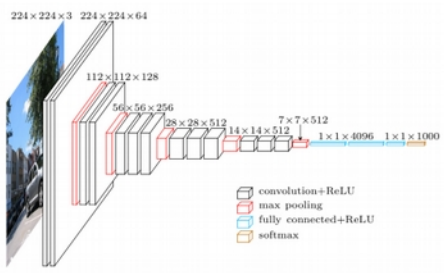


The gap

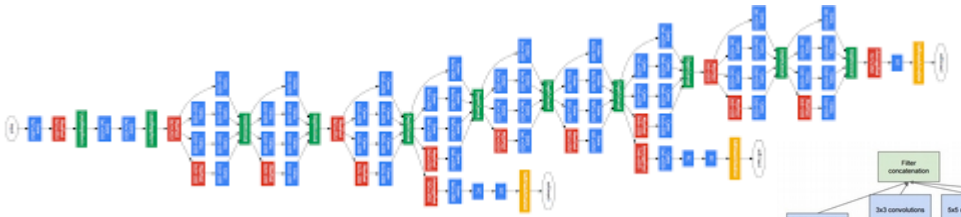
AlexNet (2010)

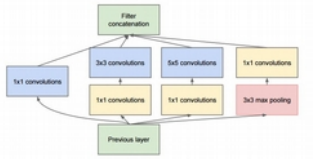


VGG (2013)

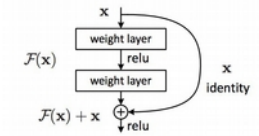


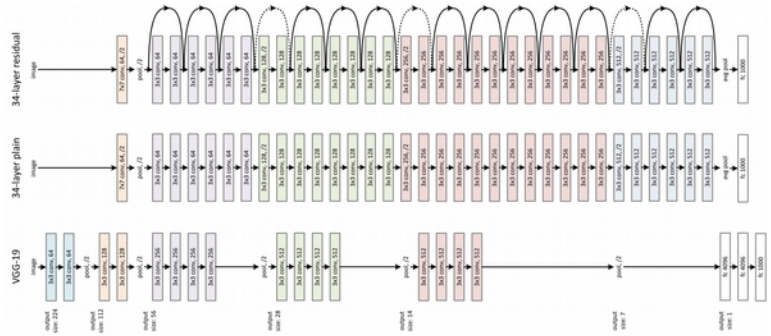
GoogleLeNet - Inception (V1 2014 – V4 2016)





ResNet (2015)



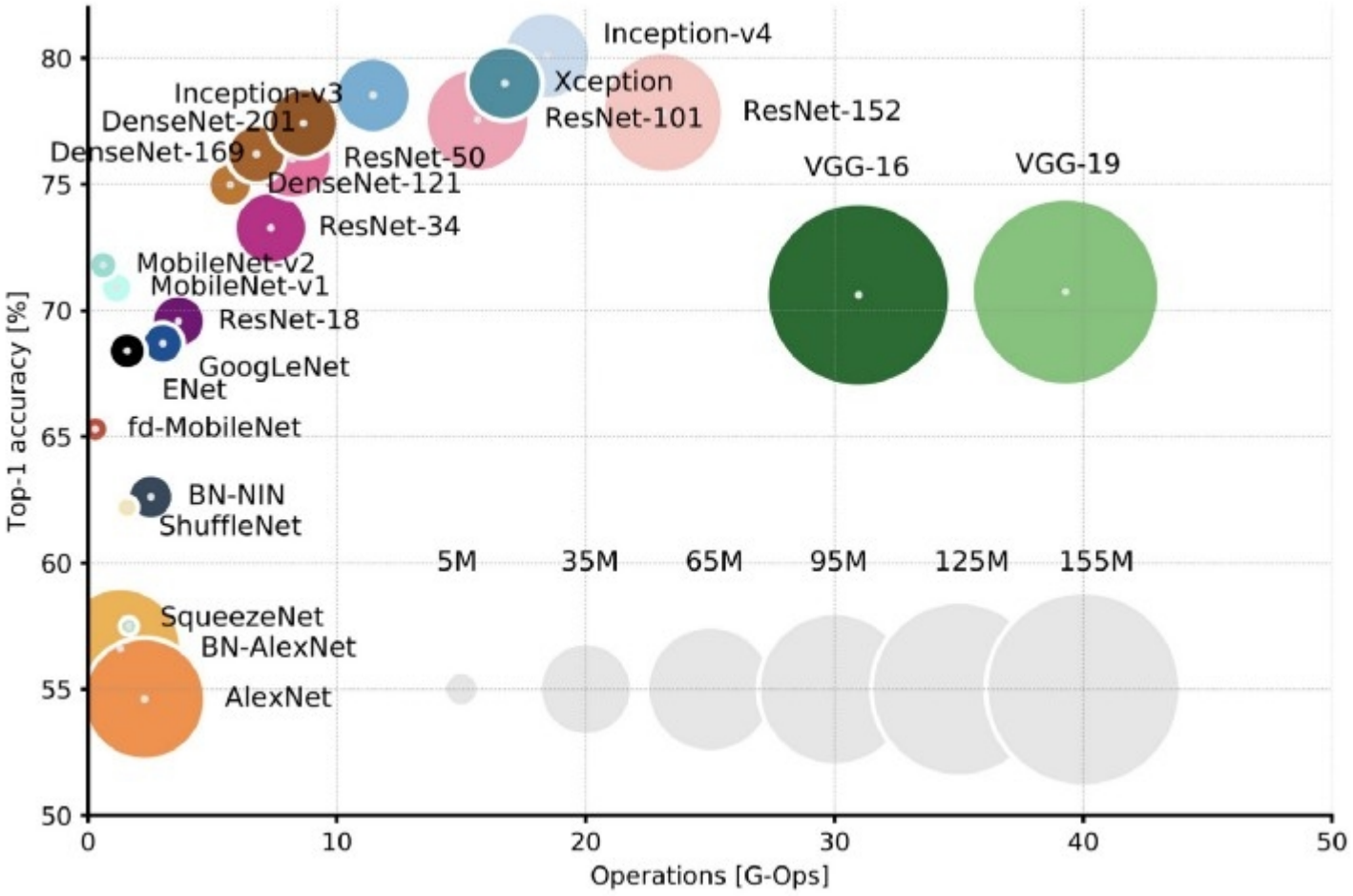


SqueezeNet (2016); ENet (2016); Xception (2017); MobileNet (2017); DenseNet (2017); ShuffleNet (2017)

Reducción del número de parámetros, número de operaciones, mejoras en la arquitectura, etc., manteniendo la precisión

David Fernández Llorca (david.fernandezl@uah.es)

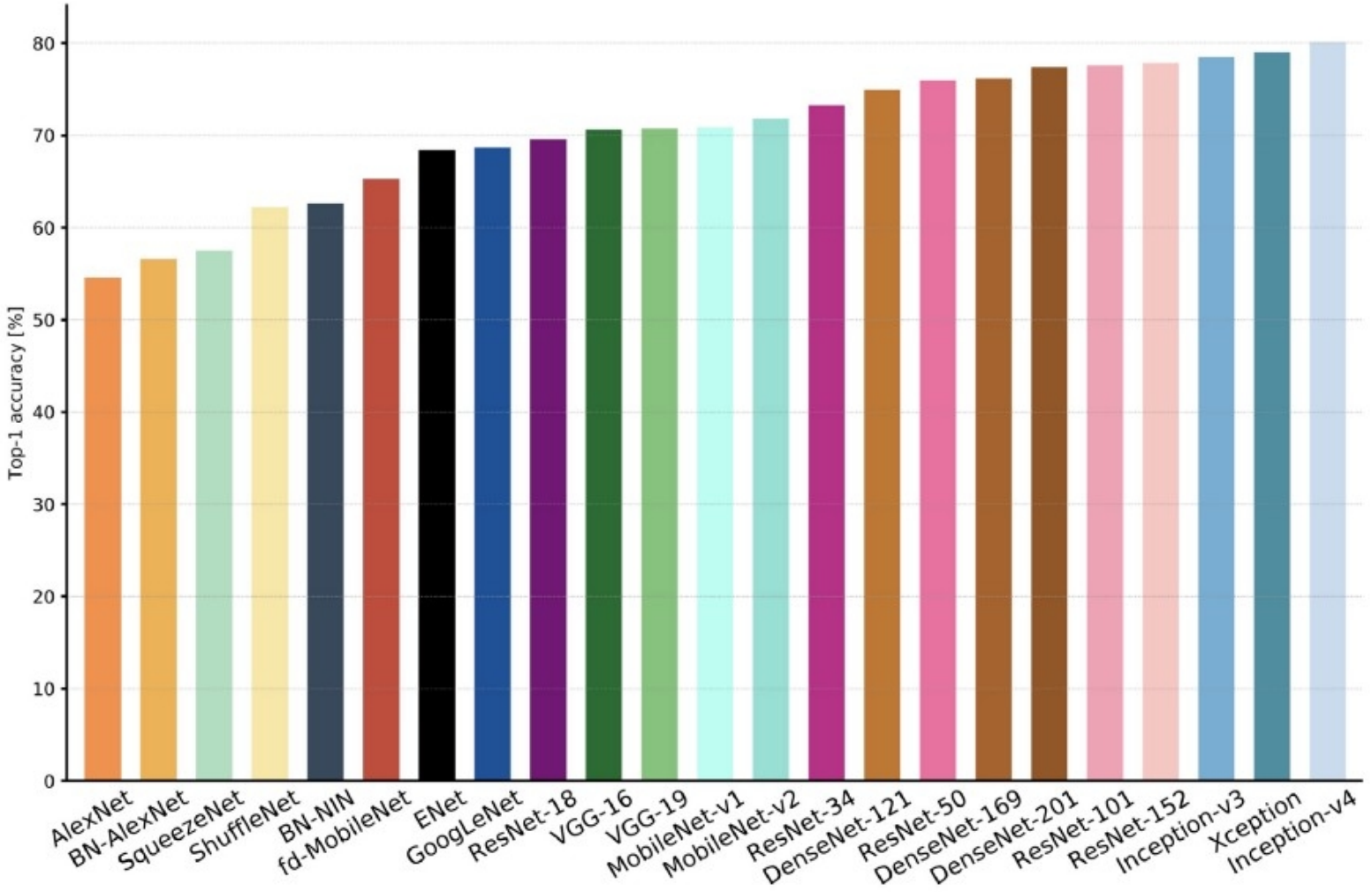
INVETT Group at University of Alcalá



https://medium.com/@culurciello/analysis-of-deep-neural-networks-dcf398e71aae

David Fernández Llorca (david.fernandezl@uah.es)

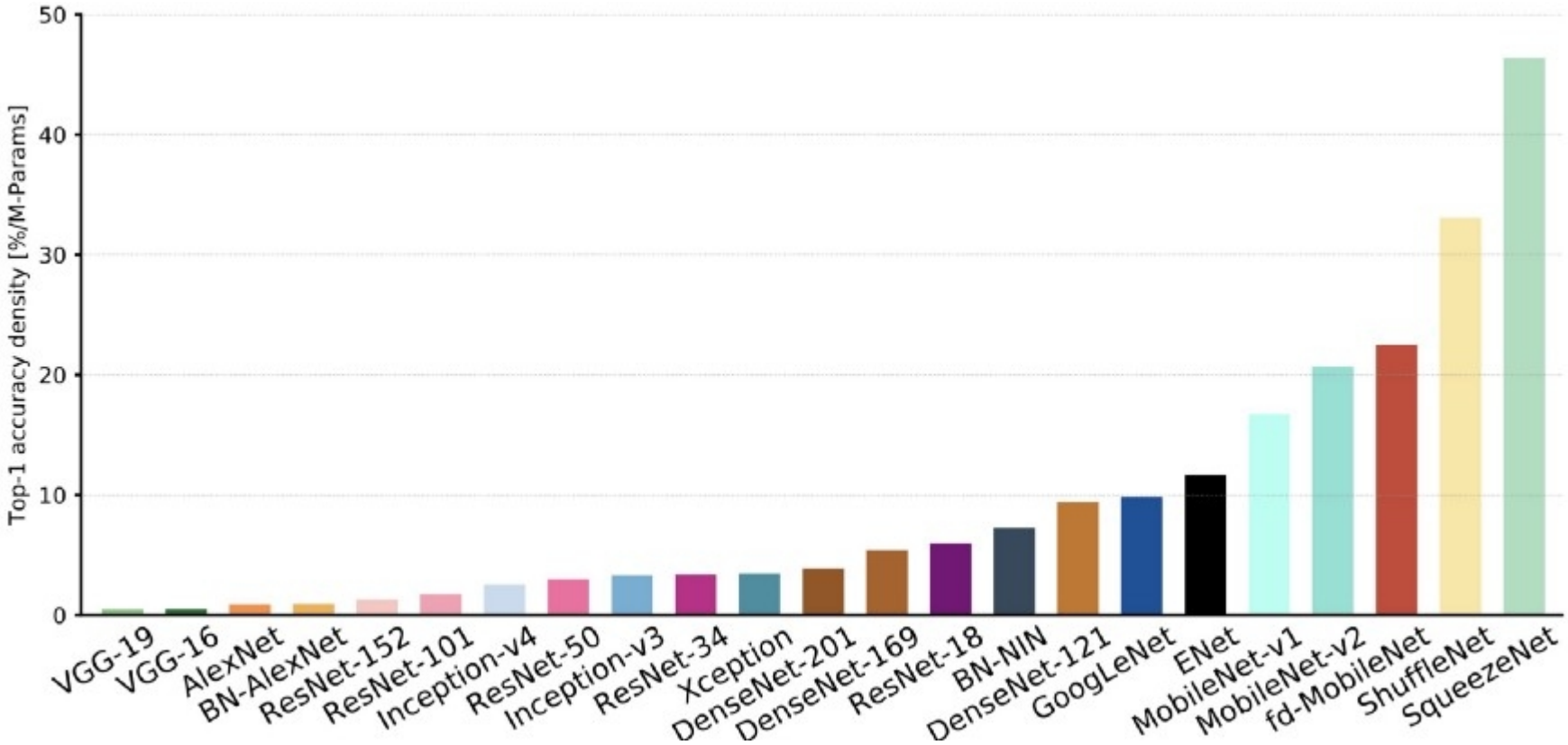
INVETT Group at University of Alcalá



https://medium.com/@culurciello/analysis-of-deep-neural-networks-dcf398e71aae

David Fernández Llorca (david.fernandezl@uah.es)

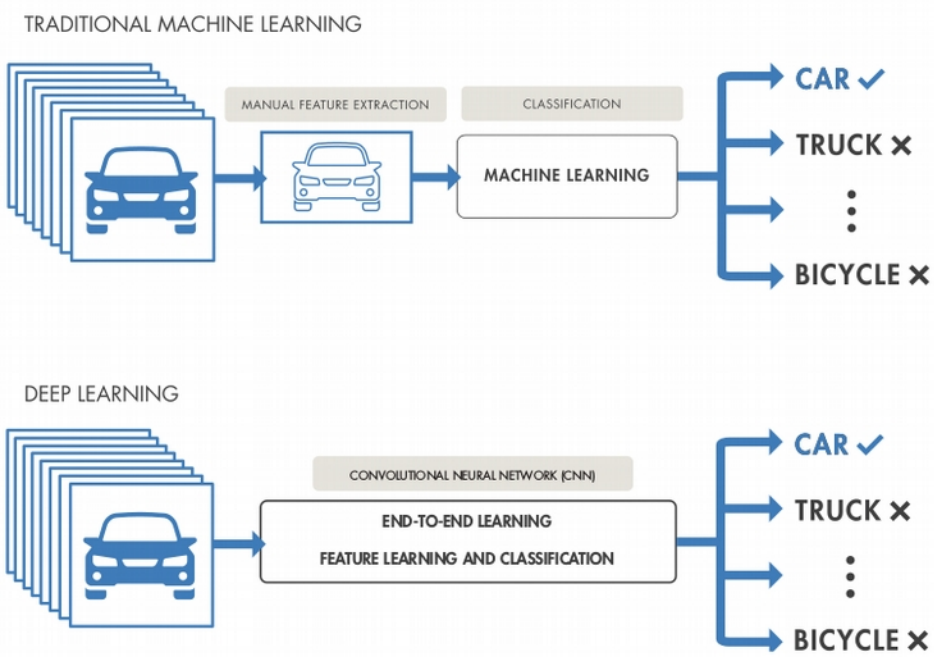
INVETT Group at University of Alcalá



https://medium.com/@culurciello/analysis-of-deep-neural-networks-dcf398e71aae

David Fernández Llorca (david.fernandezl@uah.es)

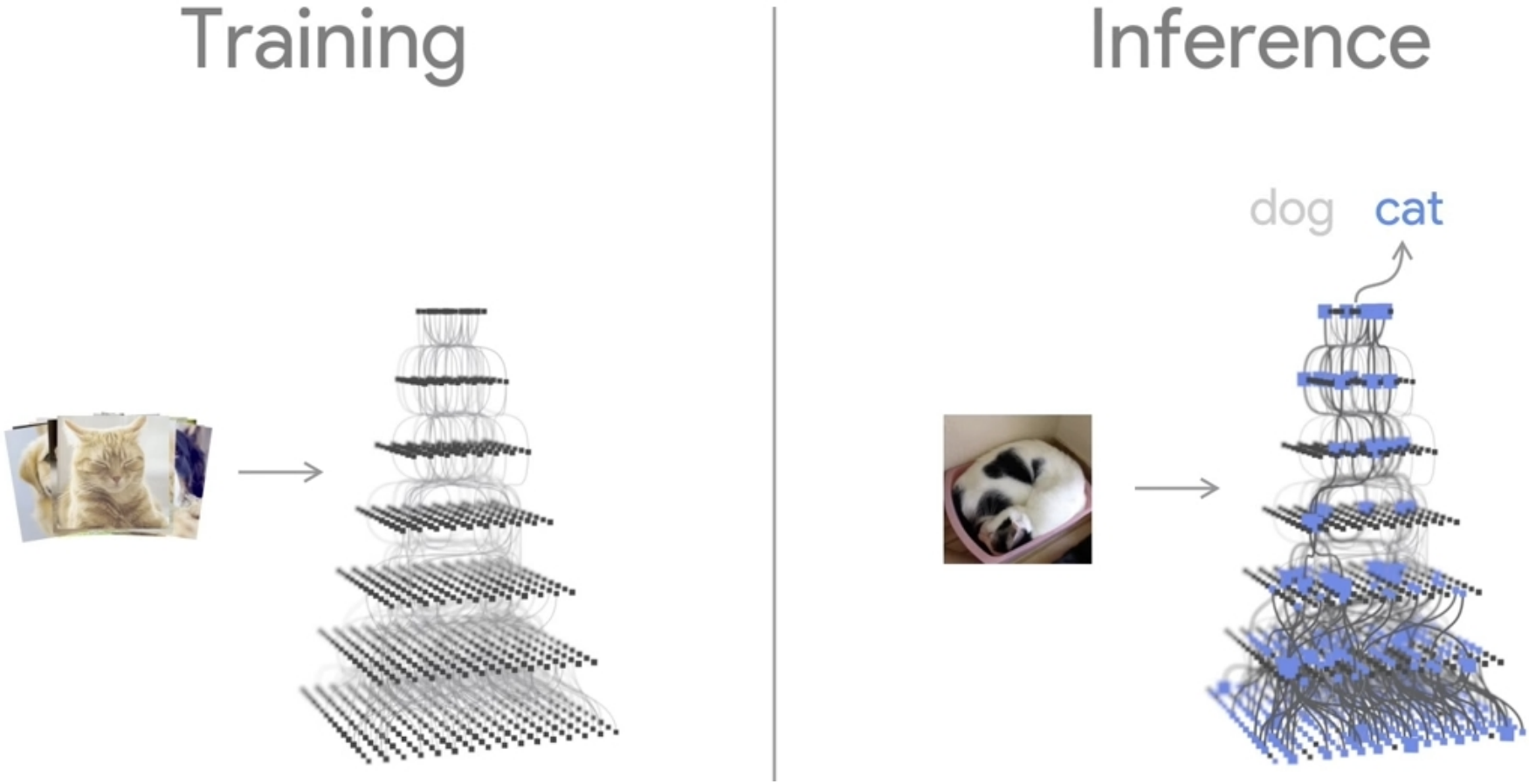
INVETT Group at University of Alcalá



Machine Learning	Deep Learning
+ Good results with small data sets	— Requires very large data sets
+ Quick to train a model	— Computationally intensive
— Need to try different features and classifiers to achieve best results	+ Learns features and classifiers automatically
— Accuracy plateaus	+ Accuracy is unlimited

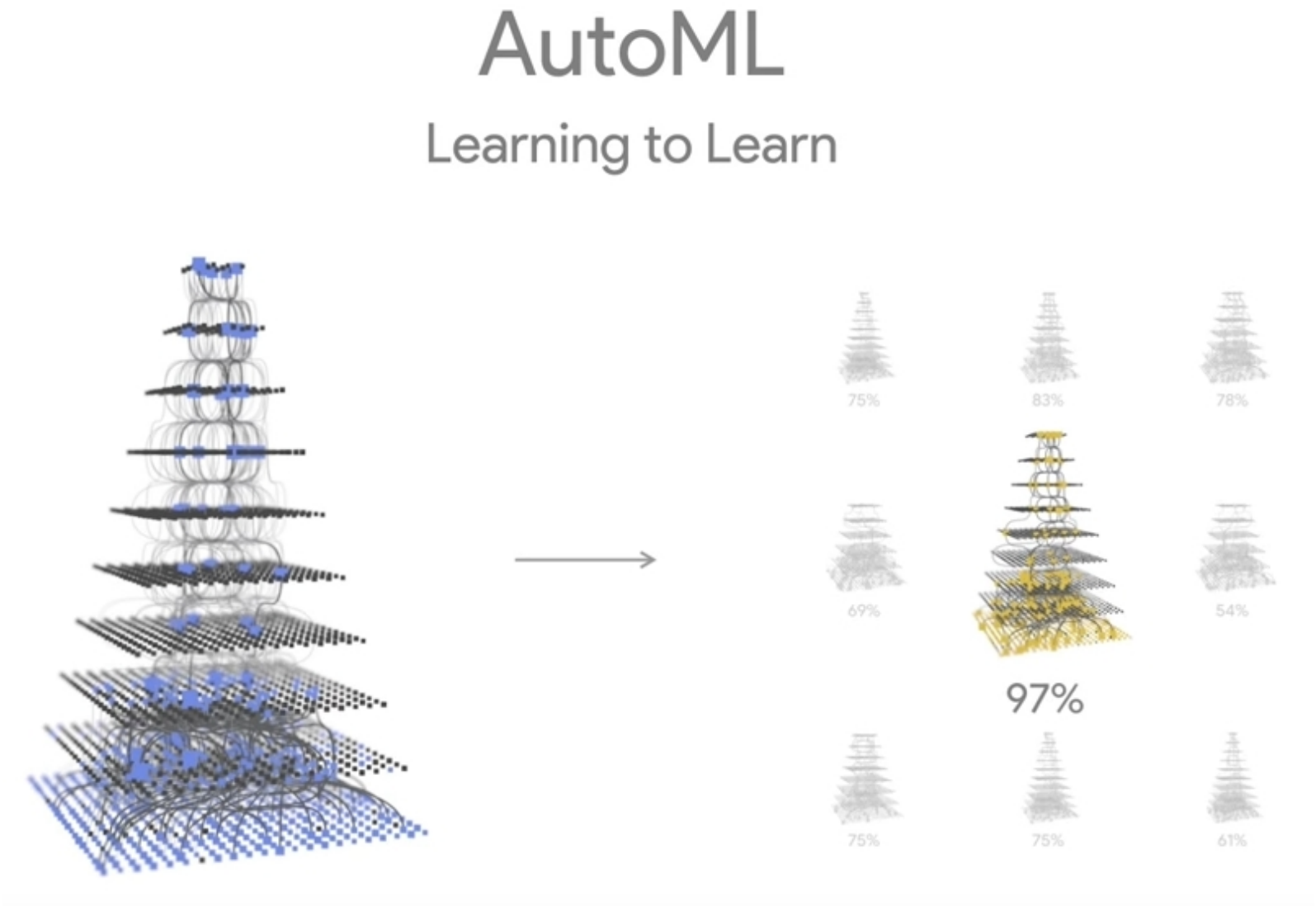
David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá



David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá



David Fernández Llorca (david.fernandezl@uah.es)

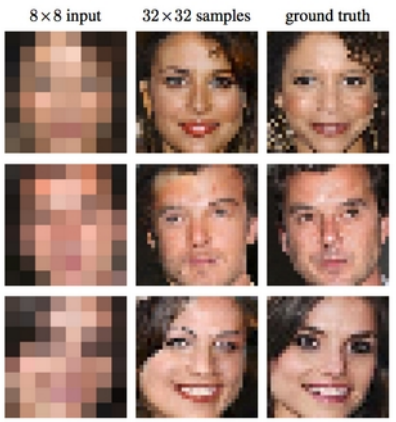
INVETT Group at University of Alcalá

Aplicaciones (I)

- Síntesis de vídeo y audio

- Coloreado de imágenes

- Restauración de píxeles (CSI)

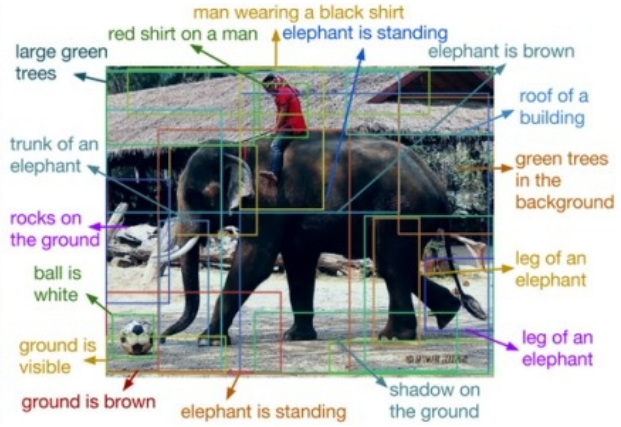


- Estimación pose de personas

David Fernández Llorca (david.fernandezl@uah.es)

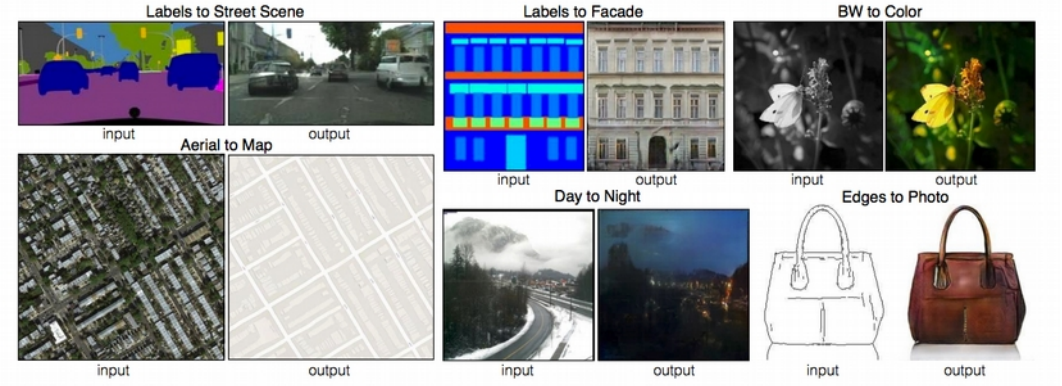
Aplicaciones (II)

- Descripción de fotos



- Traducción idiomática

- Síntesis de imágenes

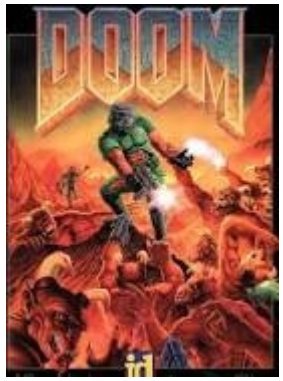


David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

Aplicaciones (III)

– Videojuegos

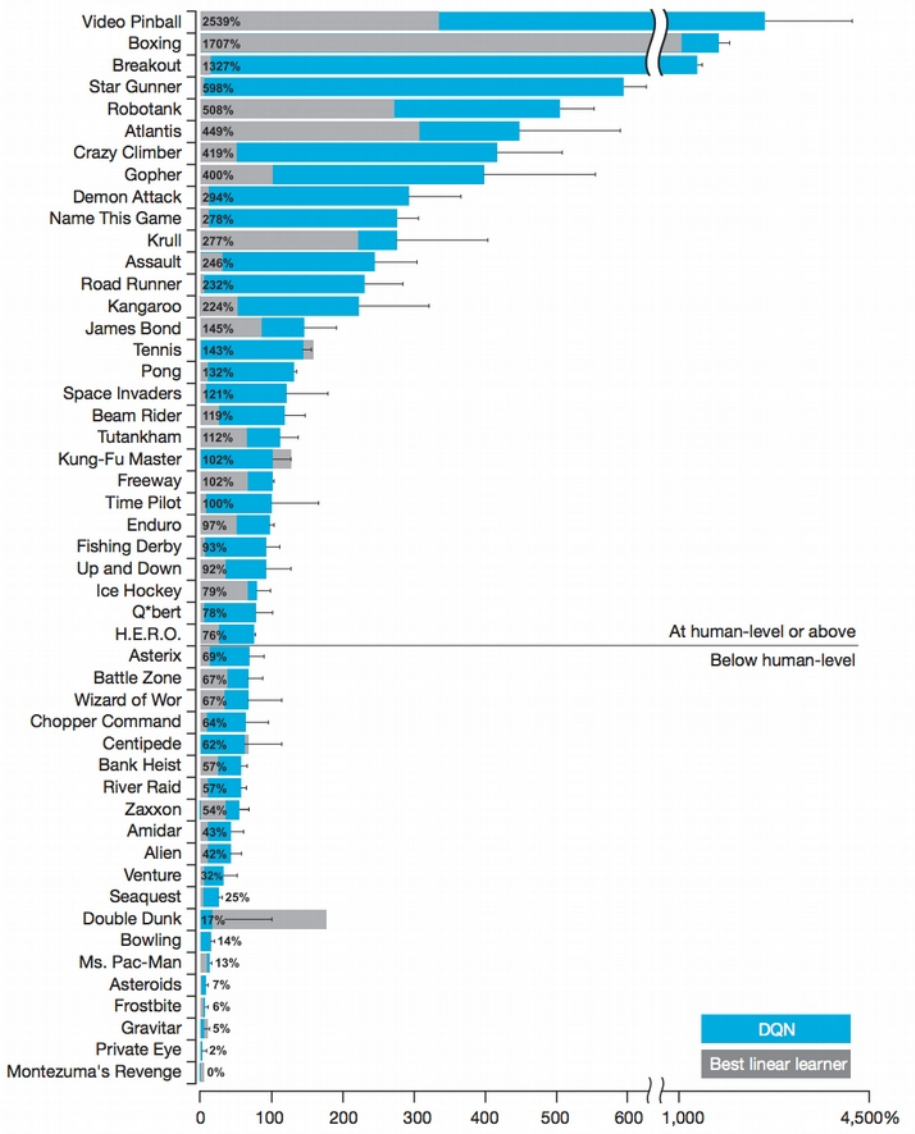


DOOM

	Single Player		Multiplayer	
Evaluation Metric	Human	Agent	Human	Agent
Number of objects	5.2	9.2	6.1	10.5
Number of kills	12.6	27.6	5.5	8.0
Number of deaths	8.3	5.0	11.2	6.0
Number of suicides	3.6	2.0	3.2	0.5
K/D Ratio	1.52	5.12	0.49	1.33

Atari Breakout

Google's Deep Mind



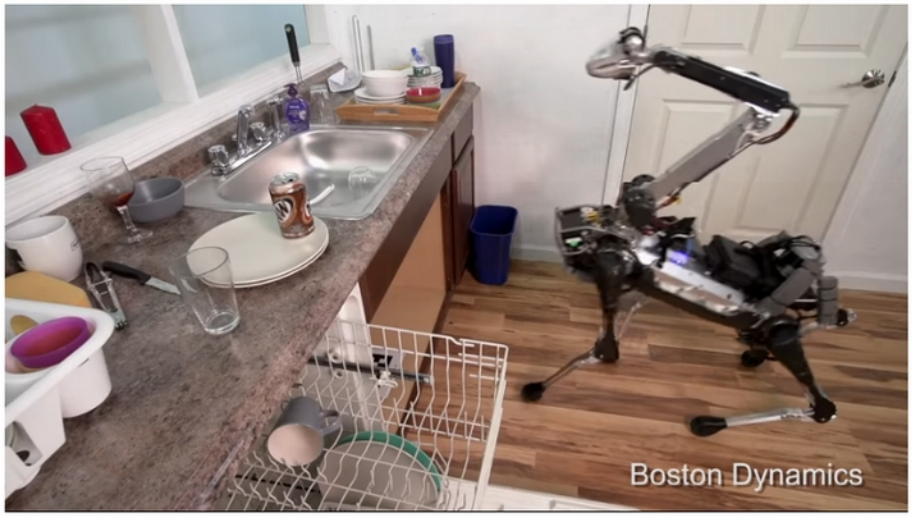
David Fernández Llorca (david.fernandezl@uah.es)

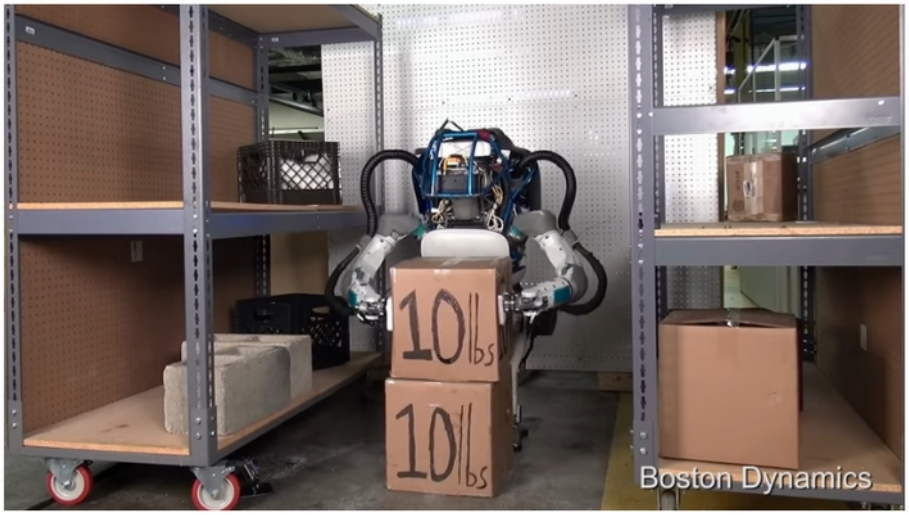
INVETT Group at University of Alcalá

Aplicaciones (IV)

- Vehículos autónomos

- Robótica

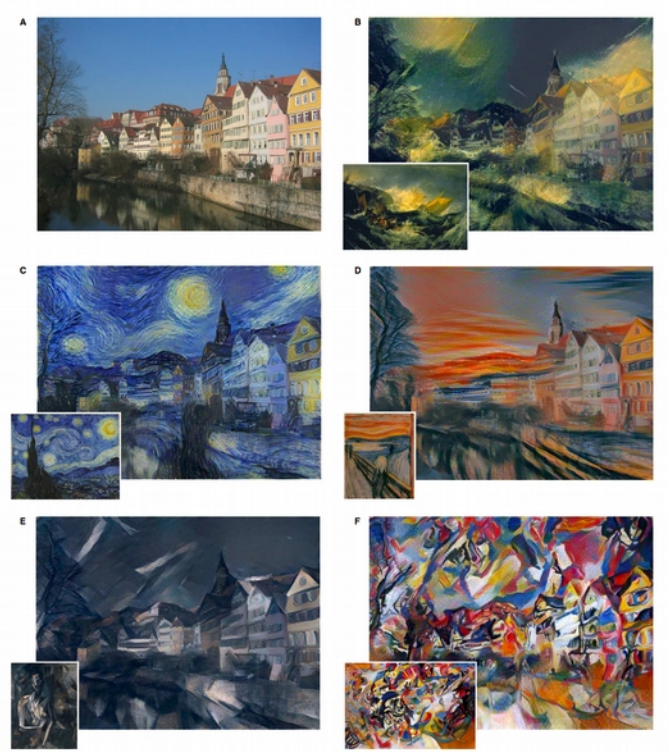


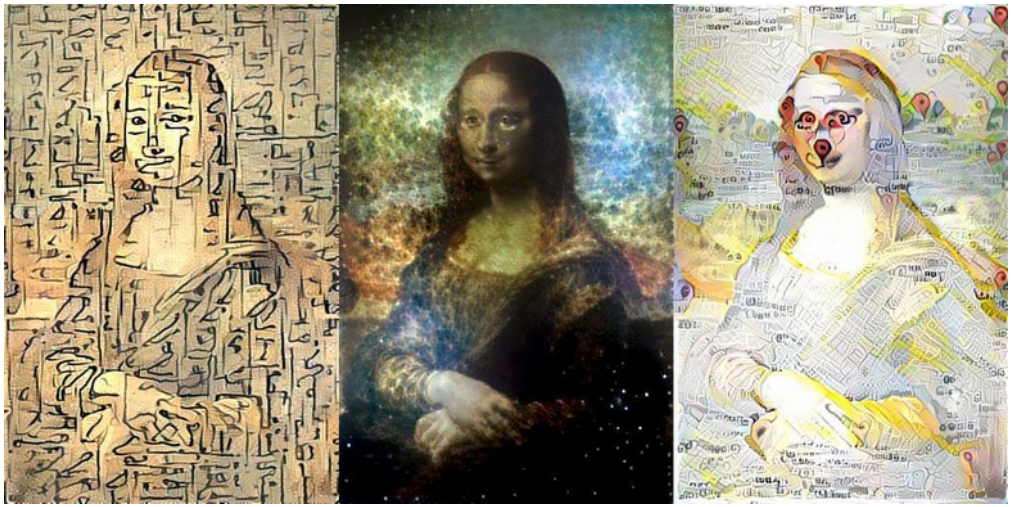


David Fernández Llorca (david.fernandezl@uah.es)

Aplicaciones (V)

- Generación de voz, recomposición de sonidos, composición de música, etc.
- Arte





David Fernández Llorca (david.fernandezl@uah.es)

INVETT Group at University of Alcalá

David Fernández Llorca (david.fernandezl@uah.es)