

Inteligencia Artificial y Visión por Computadora

Pablo Musé

Departamento de Procesamiento de Señales
Instituto de Ingeniería Eléctrica, Facultad de Ingeniería

14 de mayo de 2018



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Aprendizaje Automático y Aprendizaje Estadístico

Definición (Machine learning [Wikipedia])

El aprendizaje automático ("machine learning") es una rama de las ciencias de la computación y de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender.

Definición (Machine learning [Toni Mitchel. "Machine Learning"])

A un programa de computadoras se le asigna aprender de la experiencia E respecto a alguna clase de tarea T y la medida de desempeño P , si su desempeño en la tarea T , medido como P , mejora con la experiencia E .

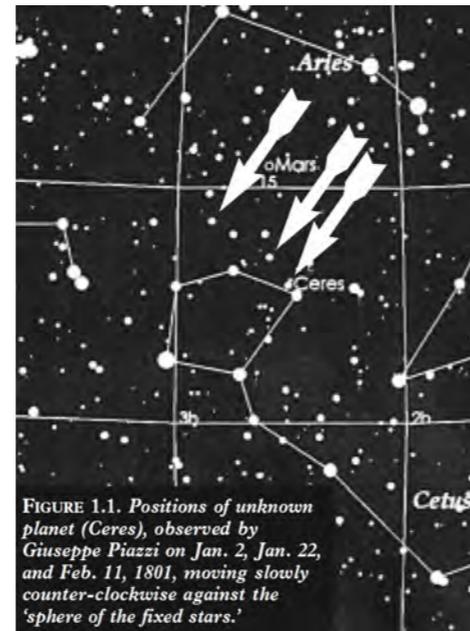
Definición (Aprendizaje estadístico [Wikipedia])

La teoría del aprendizaje estadístico es un marco para el aprendizaje automático basado en la estadística y el análisis funcional. Trata el problema de encontrar una función predictiva basada en datos.

Aprendizaje Automático: ¿revolución o evolución?

1801: Piazzi descubre el planeta enano Ceres. En 42 días mide 19 posiciones, luego se pierde por el resplandor del Sol.

Gauss predice su posición estimando su órbita a partir de esos puntos, ideando el método de mínimos cuadrados.

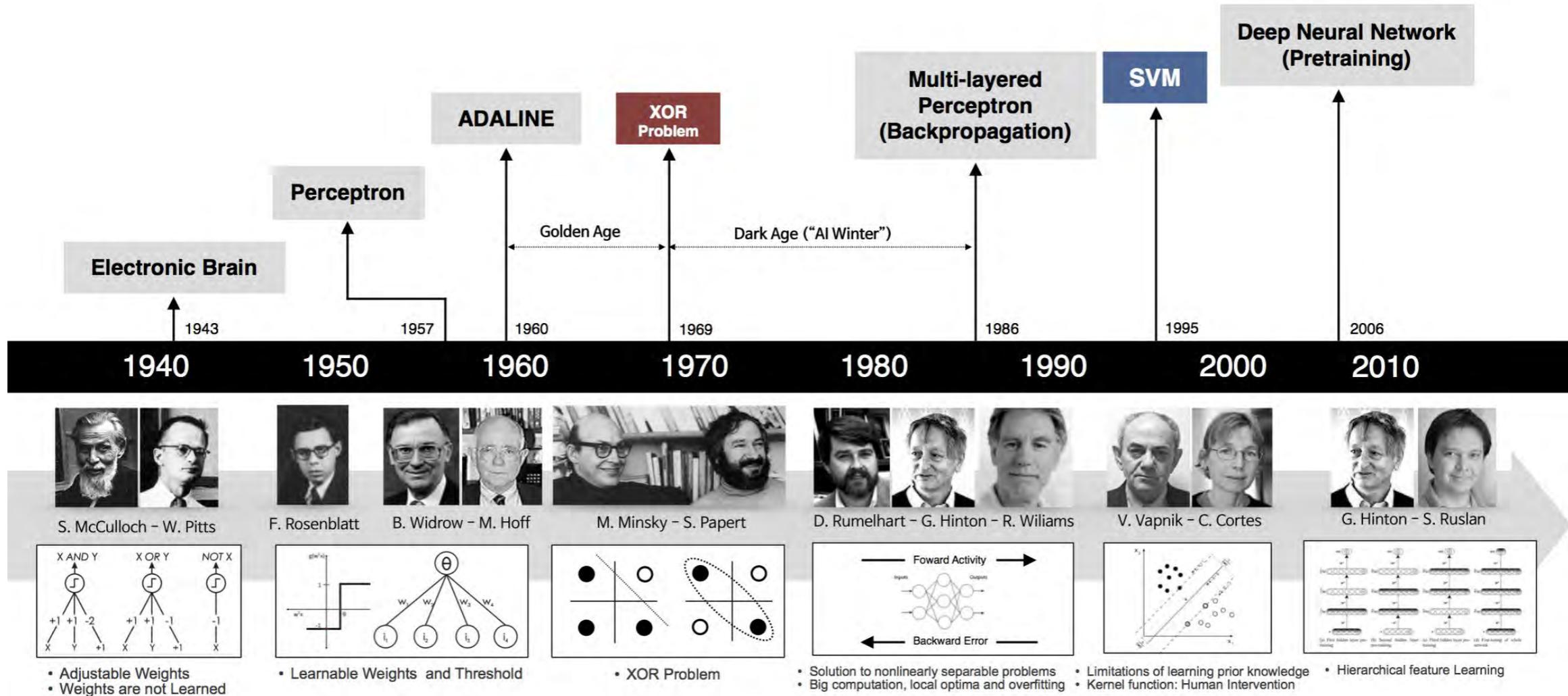


Siglo XIX: comienza a sistematizarse la probabilidad y la estadística (Laplace, Bernoulli, ...).

Aprendizaje Automático: ¿revolución o evolución?

- Los problemas de la ciencia y la industria han impulsado el desarrollo de la estadística desde siempre
- Advenimiento de la computación y la era de la información: explosión en tamaño y complejidad de los problemas estadísticos
- Desafíos tecnológicos: en almacenamiento, organización y búsqueda de datos
- Desafíos científicos: aprender de los datos. Extraer patrones, comportamientos, understand “lo que los datos tienen para decir.”
- Necesidad de un enfoque multidisciplinario: matemática y estadística, computación, ingeniería.

Evolución del Aprendizaje Automático



Evolución del Aprendizaje Automático

Orígenes y evolución del machine learning:

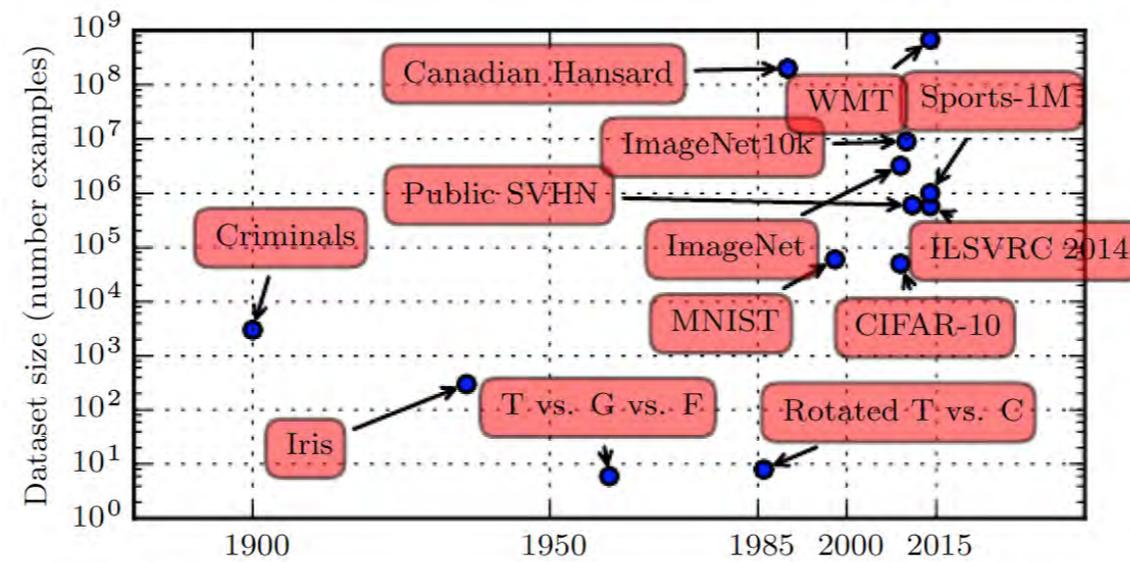
- A pesar de toda la parafernalia: las ideas fundamentales no son nuevas (las primeras publicaciones tienen 60 años).
- Última década: avances revolucionarios posibilitados por:
 - Masividad de datos
 - Potencia de cálculo
 - Avances algorítmicos con base en las redes neuronales simples de los 40' y 50'
- Las técnicas de *deep learning* han sido las responsables de los mayores avances en inteligencia artificial. Superan al hombre en algunas tareas.

La era de la explosión de los datos



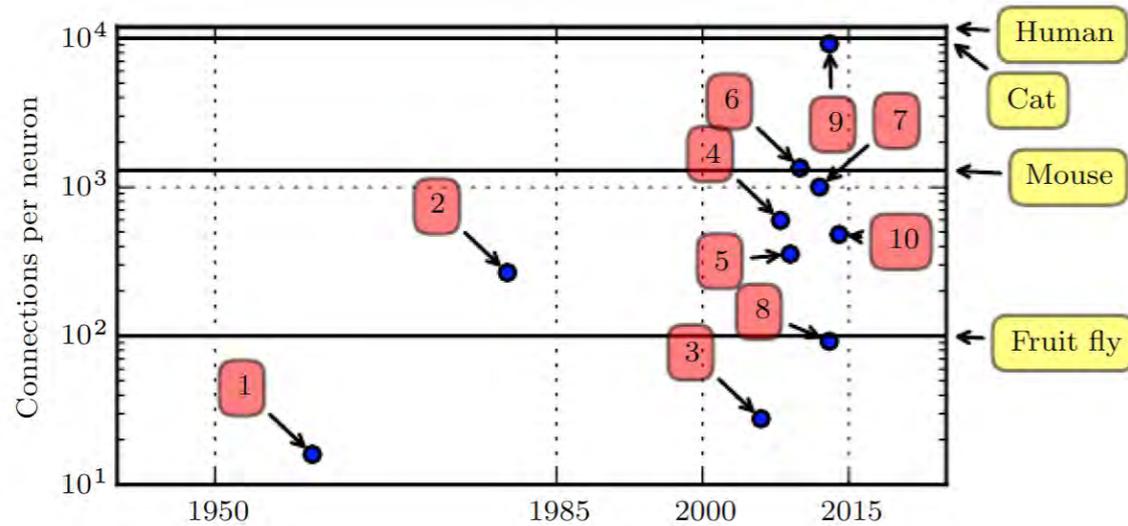
El 90 % de los datos del mundo han sido creados en los últimos 5 años.
Se estima que en mundo se generan 2.5 billones de Gigabytes por día.

Bases de datos (etiquetadas)

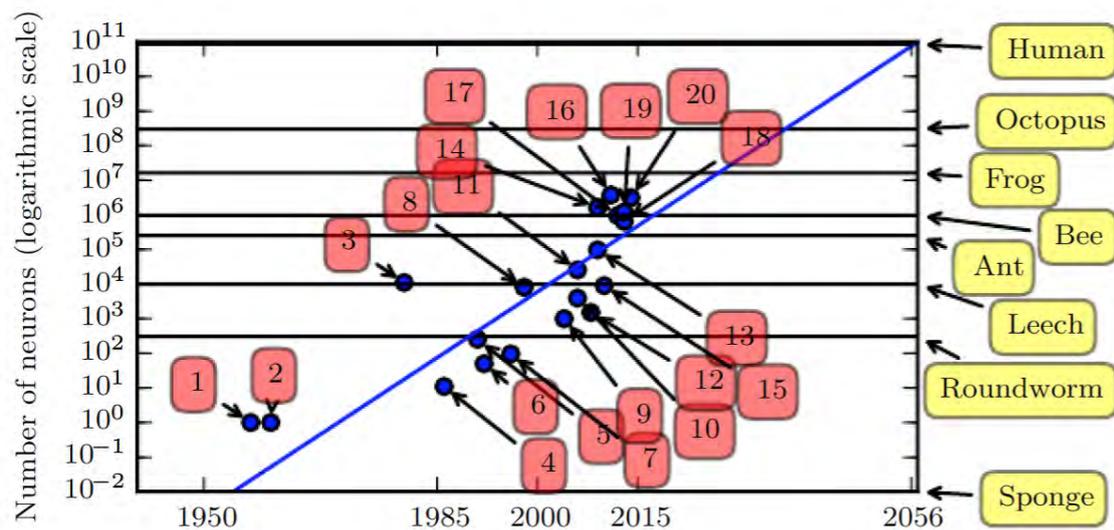


Goodfellow, Bengio & Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>

Complejidad de los sistemas



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](#))



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](#))

Aprendizaje automático en el día a día

- Sistemas de recomendación: nos sugieren productos según nuestros gustos (Amazon, Netflix, Spotify, ...)
- Reconocimiento de voz y respuesta: asistentes personales virtuales (Apple Siri, Google Home, Amazon Alexa, ...)
- Traducción automática
- Reconocimiento y detección de patrones: actividades financieras sospechosas, fraudes en tarjetas de crédito, ...
- Computer vision: etiquetado de fotos, etiquetado de caras, reconocimiento de escritura manual, reconocimiento de objetos.

Limitaciones

A pesar de los impresionantes avances de los últimos años:

- La mayoría de los enfoques en ML se basan en la accesibilidad de gigantescas bases de datos etiquetados: esfuerzo (in)humano
- Existen varias restricciones que conocemos de las leyes naturales (e.g. la física) y matemáticas cuya inclusión en ML no es evidente. Esto permitiría ser más eficiente en el uso de datos.

Inteligencia Artificial vs. Inteligencia "Natural"

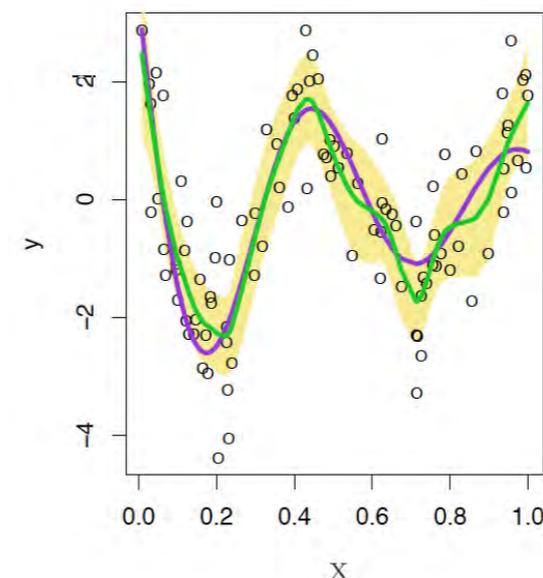
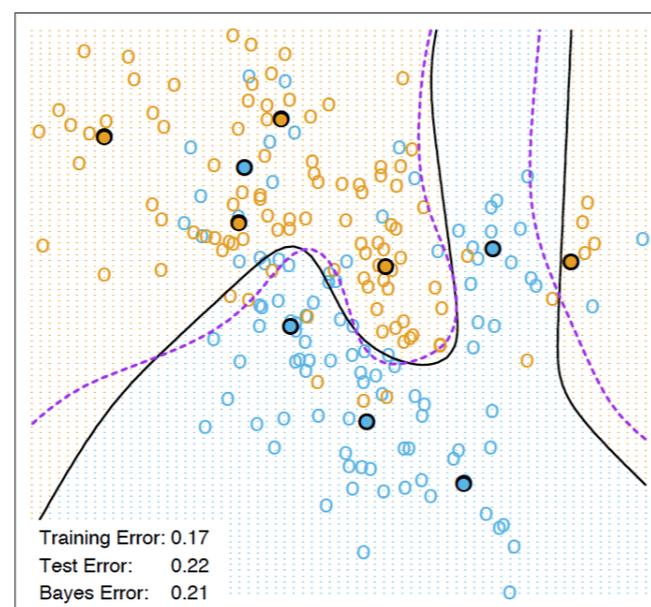
- El aprendizaje automático: *narrow intelligence*. Competencias específicas no se acercan al amplio abanico de capacidades de los seres vivos (*broad intelligence*).
- Comprensión contextual un problema, sentido común, emociones, intuición, instinto: no sabemos aún desarrollar sistemas que los incorporen, y no es claro que podamos.
- Cuando la experiencia falla, el ser humano toma acciones o decisiones que pueden no ser óptimas pero evitan daños mayores. Las máquinas no (y es grave).
- Los humanos somos buenos en transferir ideas de un problema a otro distinto.

Si todavía no entendemos estos "misterios", ¿cómo vamos a poder transferirlos a las máquinas?

Enfoques del Aprendizaje Automático

Aprendizaje supervisado

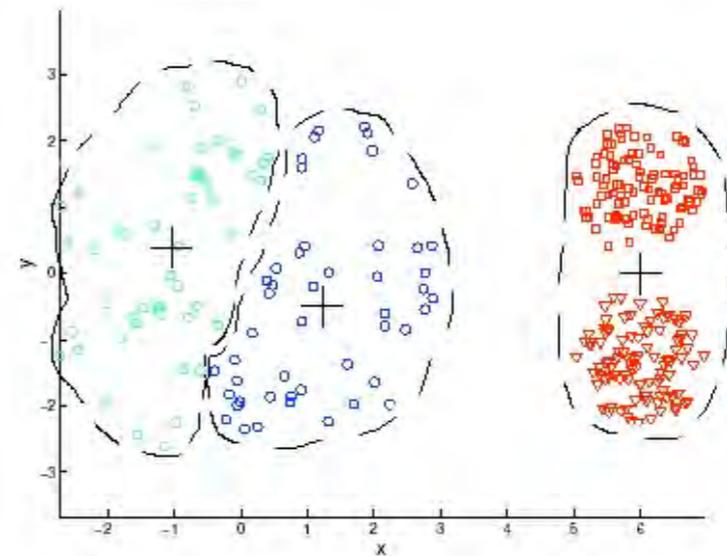
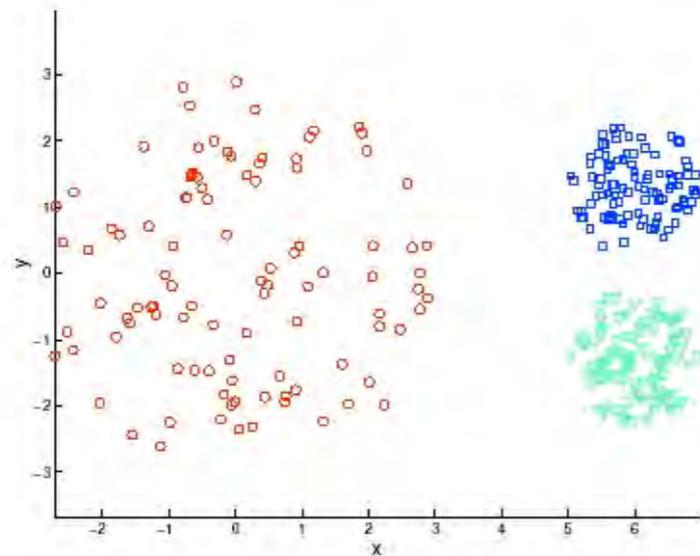
- Se dispone de **datos de entrenamiento etiquetados** (pares entrada-salida conocidos)
- Dada una nueva entrada, **predecir la salida**
- **Clasificación** (datos categóricos): reconocimiento de caras
- **Regresión** (datos continuos): predecir una cosecha dadas las lluvias, la temperatura, etc.



Enfoques del Aprendizaje Automático

Aprendizaje no supervisado

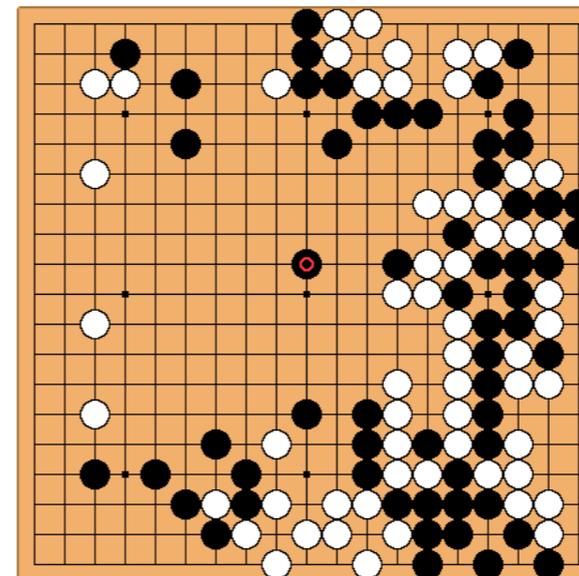
- Se dispone de **datos sin etiquetas**
- Encontrar la estructura que mejor representa los datos
- Descubrir asociaciones o patrones en los datos



Enfoques del Aprendizaje Automático

Aprendizaje por refuerzos (*reinforcement learning*)

- Aprender a elegir acciones con el fin de maximizar alguna noción de recompensa acumulada
- El algoritmo interactúa con un entorno dinámico
- No se cuenta con ejemplos de decisiones óptimas, el algoritmo las descubre por ensayo y error

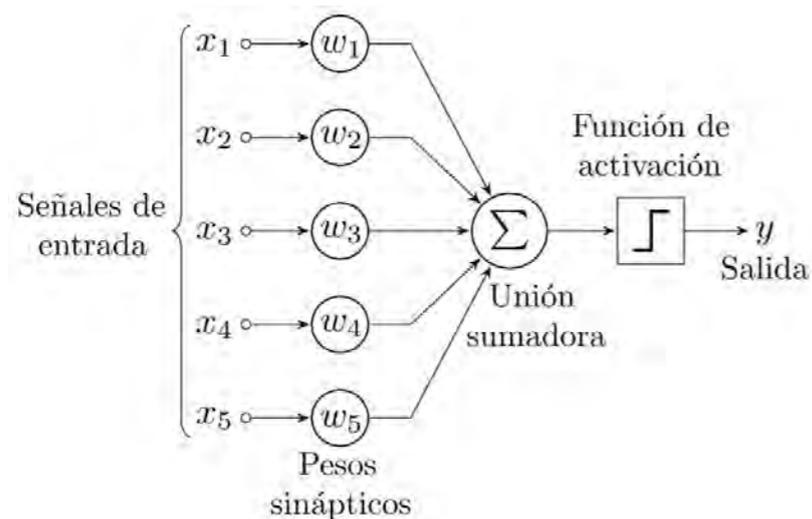


Aprendizaje Profundo (*deep learning*)

Del Perceptrón de Rosenblatt ...

Perceptrón:

- Creado por Frank Rosenblatt en 1957 en la Universidad de Cornell
- De las primeras máquinas capaces de “aprender”
- Una única capa
- Capacidad muy limitada (sólo capaz de clasificar datos linealmente separables).

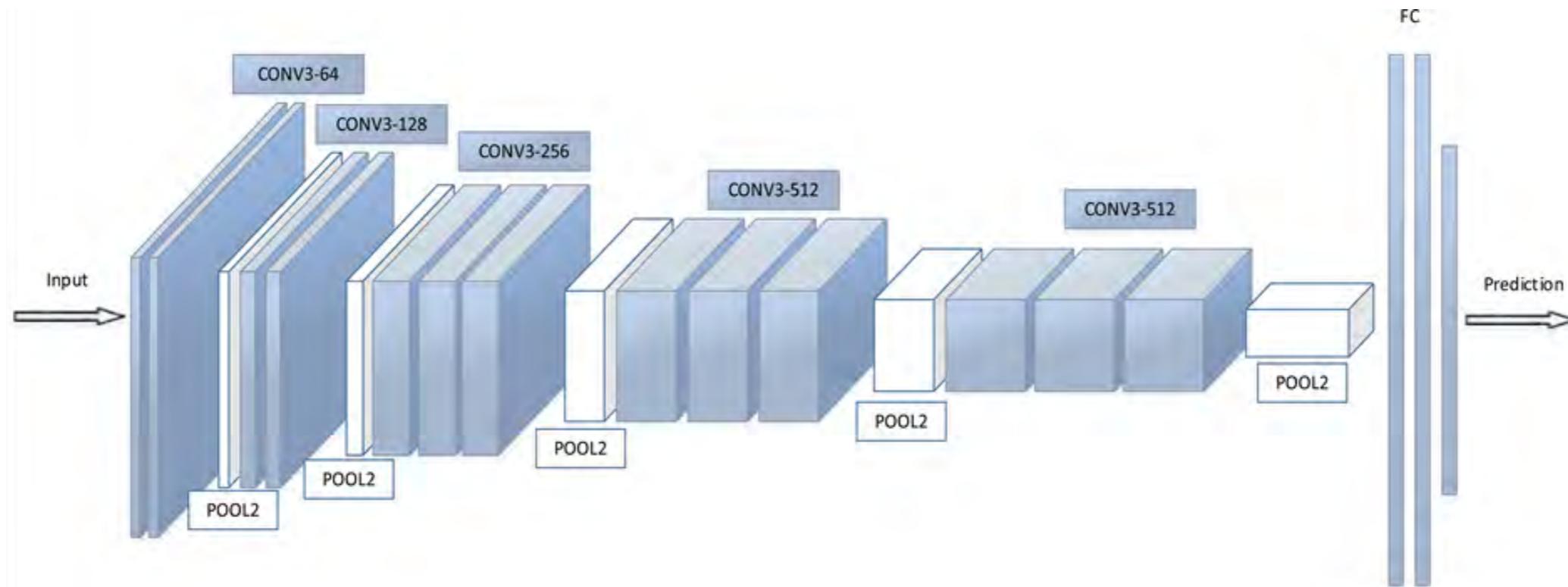


Aprendizaje Profundo (*deep learning*)

...A las Redes Neuronales Profundas

No muy distinto al Perceptrón pero:

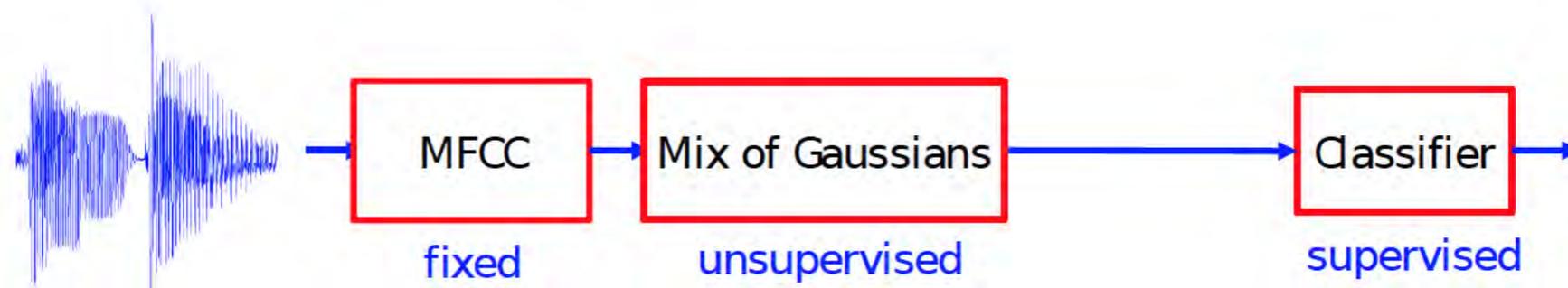
- **Múltiples capas**
- Capas que efectúan operaciones lineales y **no-lineales**



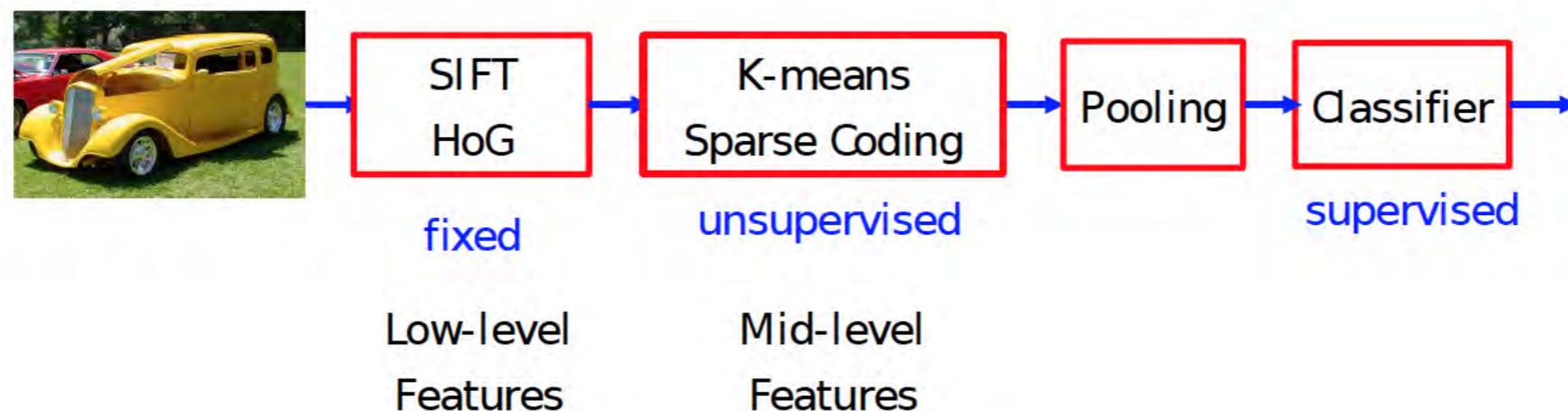
Aprendizaje Automático: Paradigma *Model Driven*

Extracción de Características - Selección de Características - Clasificación

Reconocimiento de audio, principios de los 90 hasta el 2012:



Reconocimiento de imágenes, principios del 2000 hasta el 2012:

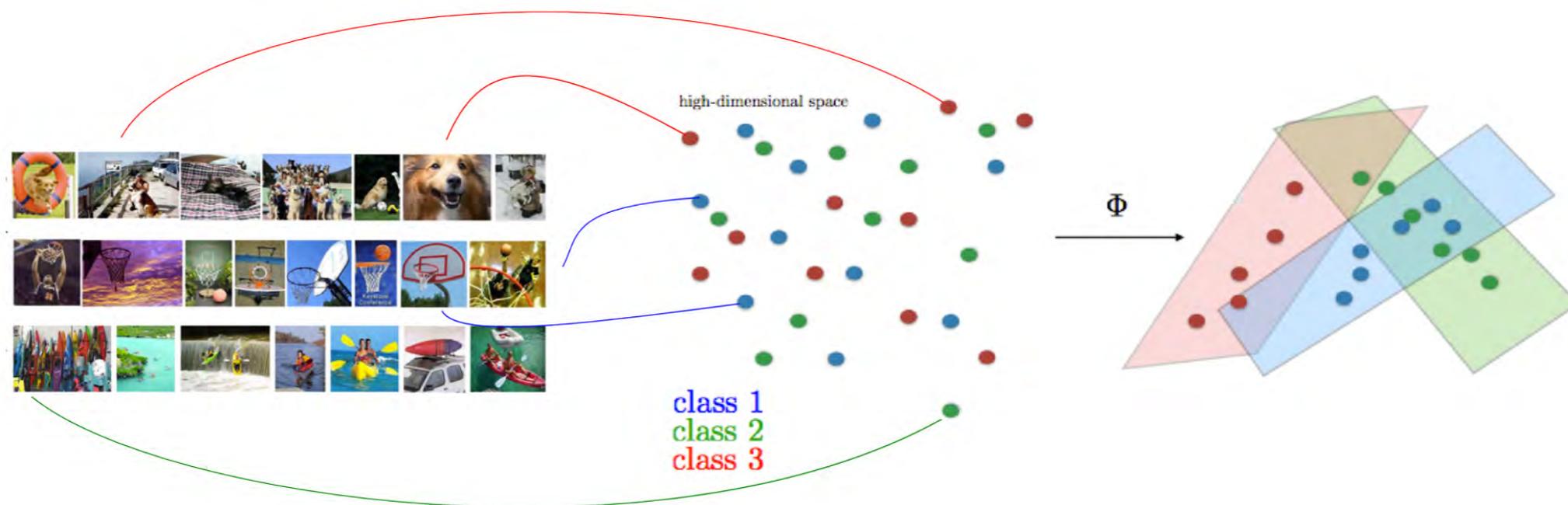


Aprendizaje Automático: Paradigma *Data Driven*

Deep Learning: Extracción y Clasificación implícitas

Deep Neural Networks:

Clase de **representaciones paramétricas no lineales** capaces de **codificar** las características del problema y de ser **optimizadas** de forma eficiente usando métodos de optimización estocástica.

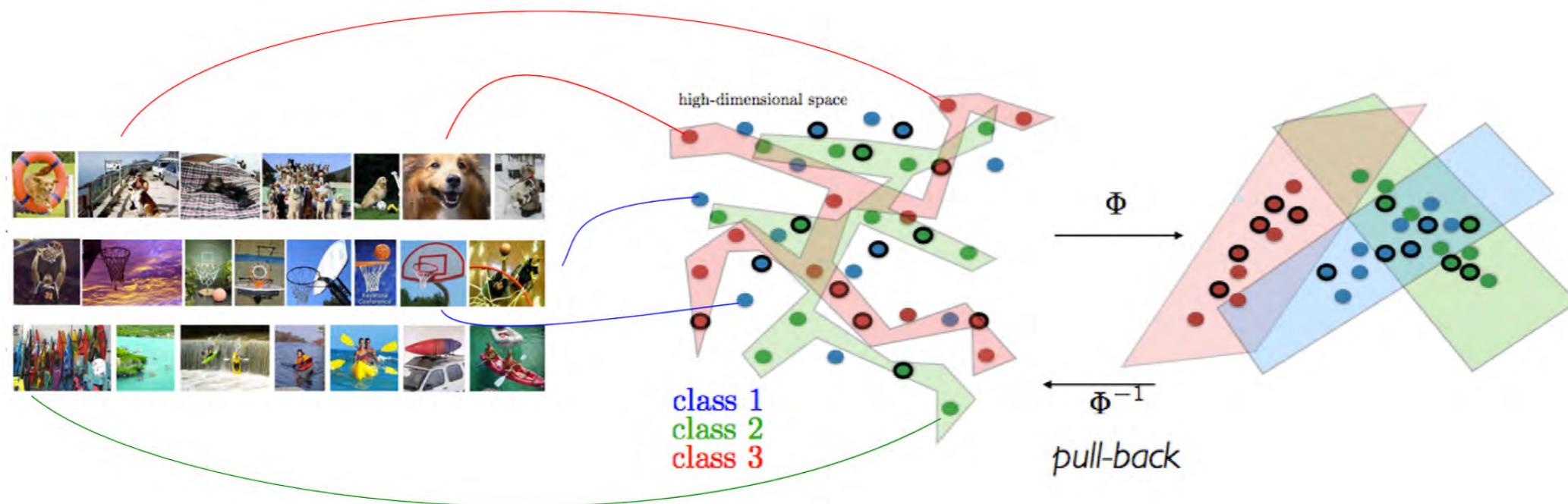


Aprendizaje Automático: Paradigma *Data Driven*

Deep Learning: Extracción y Clasificación implícitas

Deep Neural Networks:

Clase de **representaciones paramétricas no lineales** capaces de **codificar** las características del problema y de ser **optimizadas** de forma eficiente usando métodos de optimización estocástica.

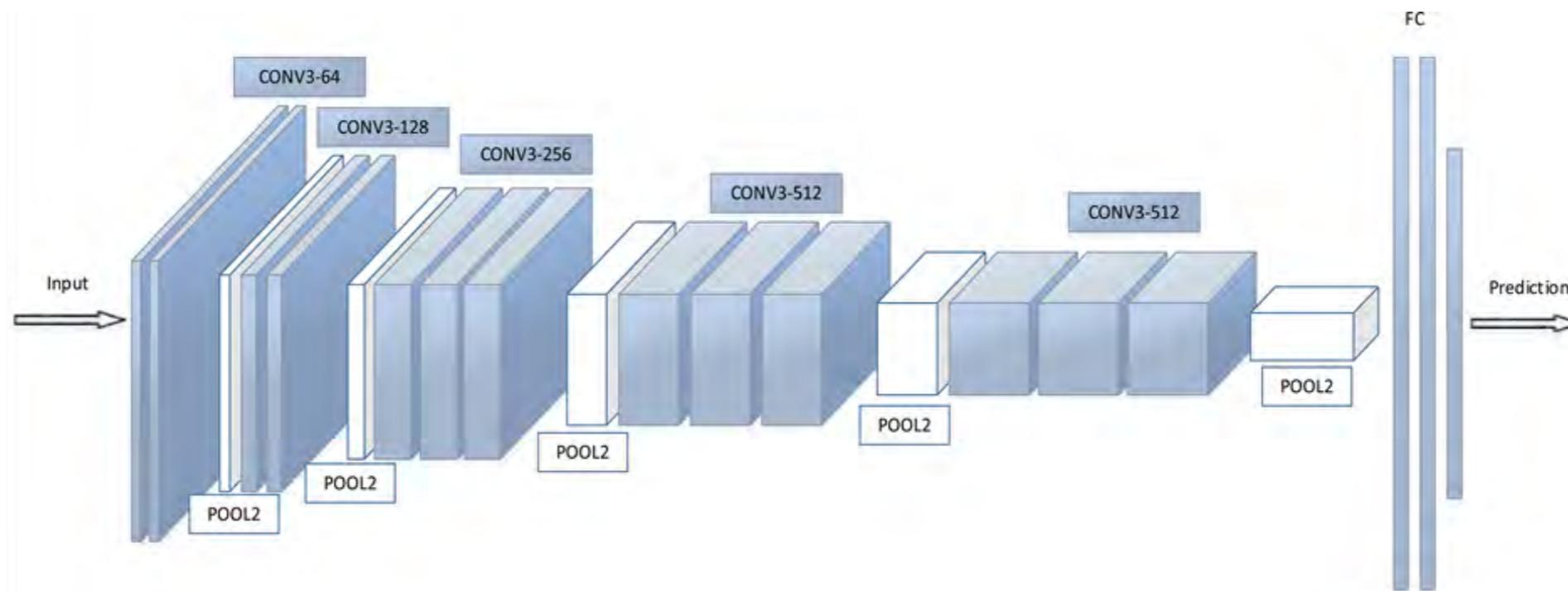


Aprendizaje Automático: Paradigma *Data Driven*

Deep Learning: Extracción y Clasificación implícitas

Deep Neural Networks:

Clase de **representaciones paramétricas no lineales** capaces de **codificar** las características del problema y de ser **optimizadas** de forma eficiente usando métodos de optimización estocástica.

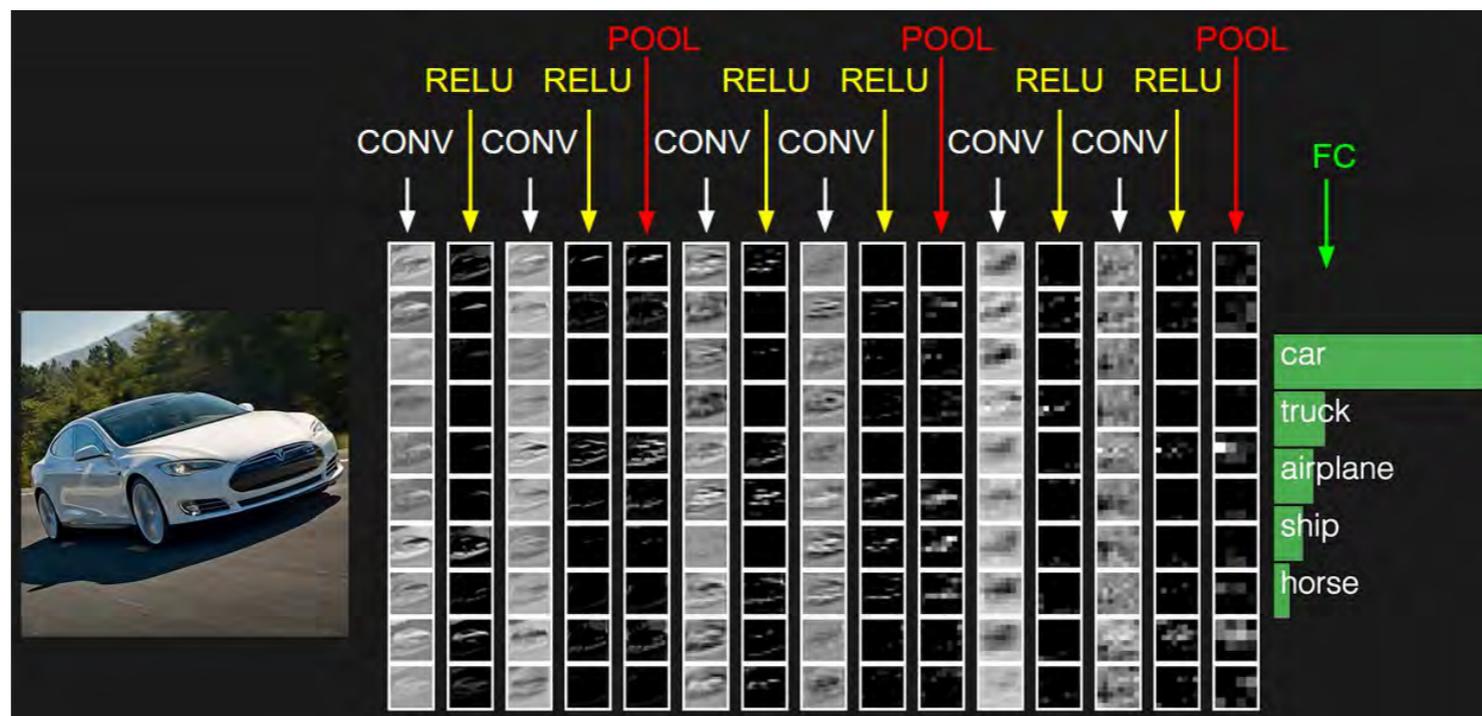


Aprendizaje Automático: Paradigma *Data Driven*

Deep Learning: Extracción y Clasificación implícitas

Deep Neural Networks:

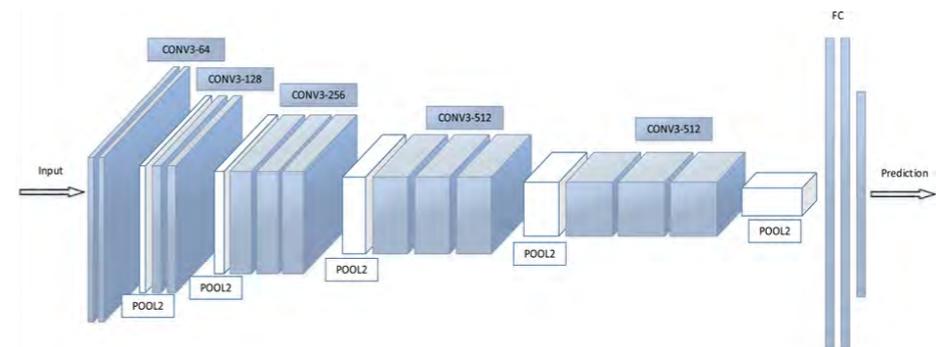
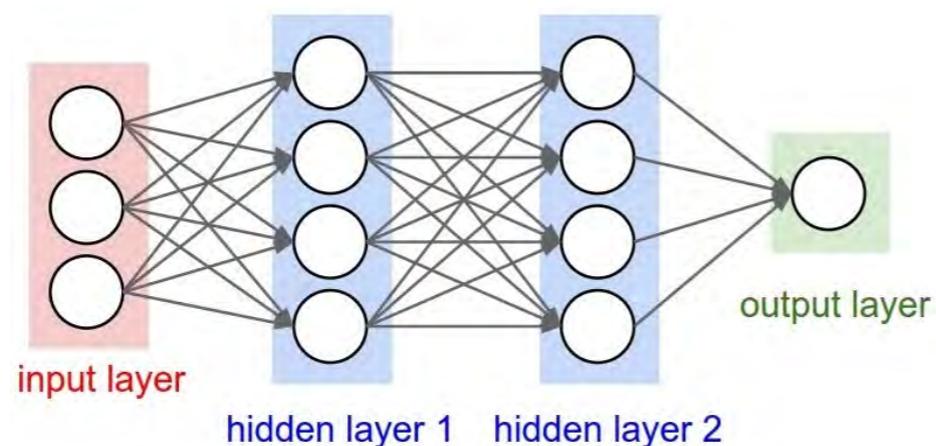
Clase de **representaciones paramétricas no lineales** capaces de **codificar** las características del problema y de ser **optimizadas** de forma eficiente usando métodos de optimización estocástica.



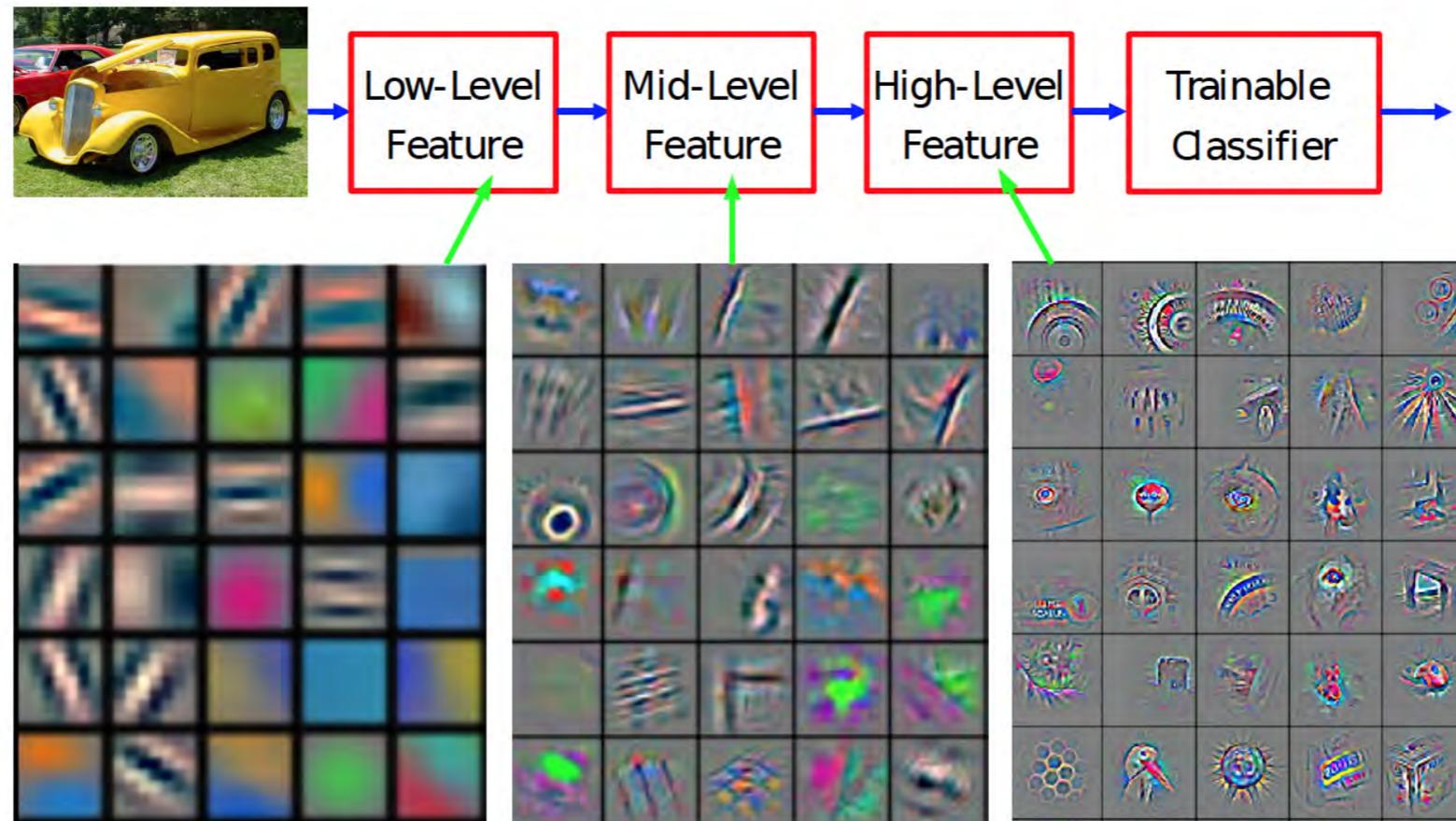
Aprendizaje Profundo (*Deep Learning*)

Representaciones jerárquicas adaptivas

- **Cascada de capas** de procesamiento **no lineal** para extraer y transformar variables.
- Cada capa usa la salida de la capa anterior como entrada: transforma la representación anterior a otra de **mayor nivel de abstracción**
- Múltiples niveles de representación se corresponden con **diferentes niveles de abstracción** (jerarquía de conceptos).

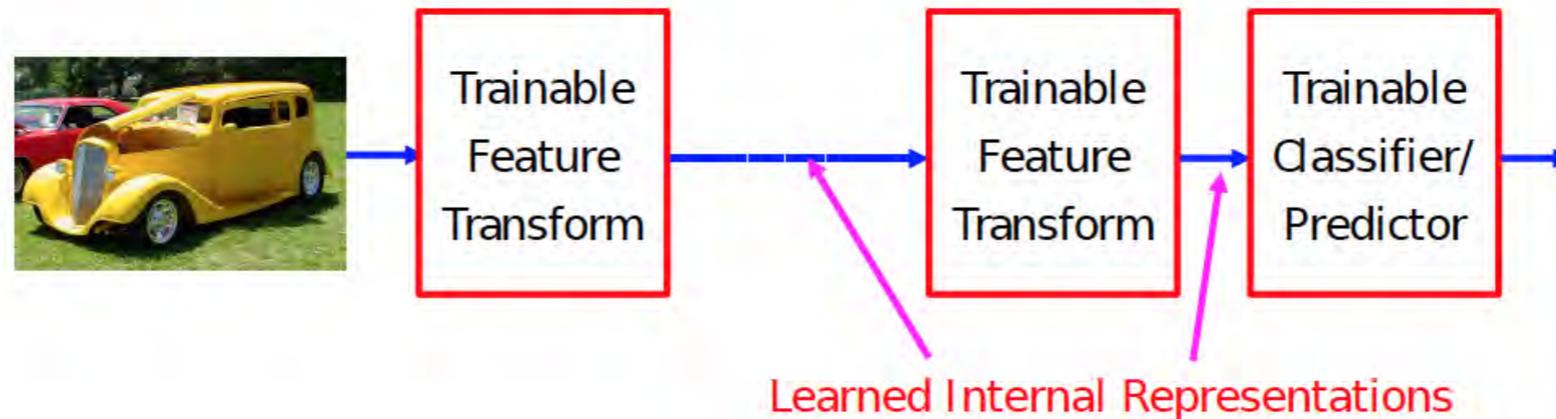


Representaciones jerárquicas adaptivas



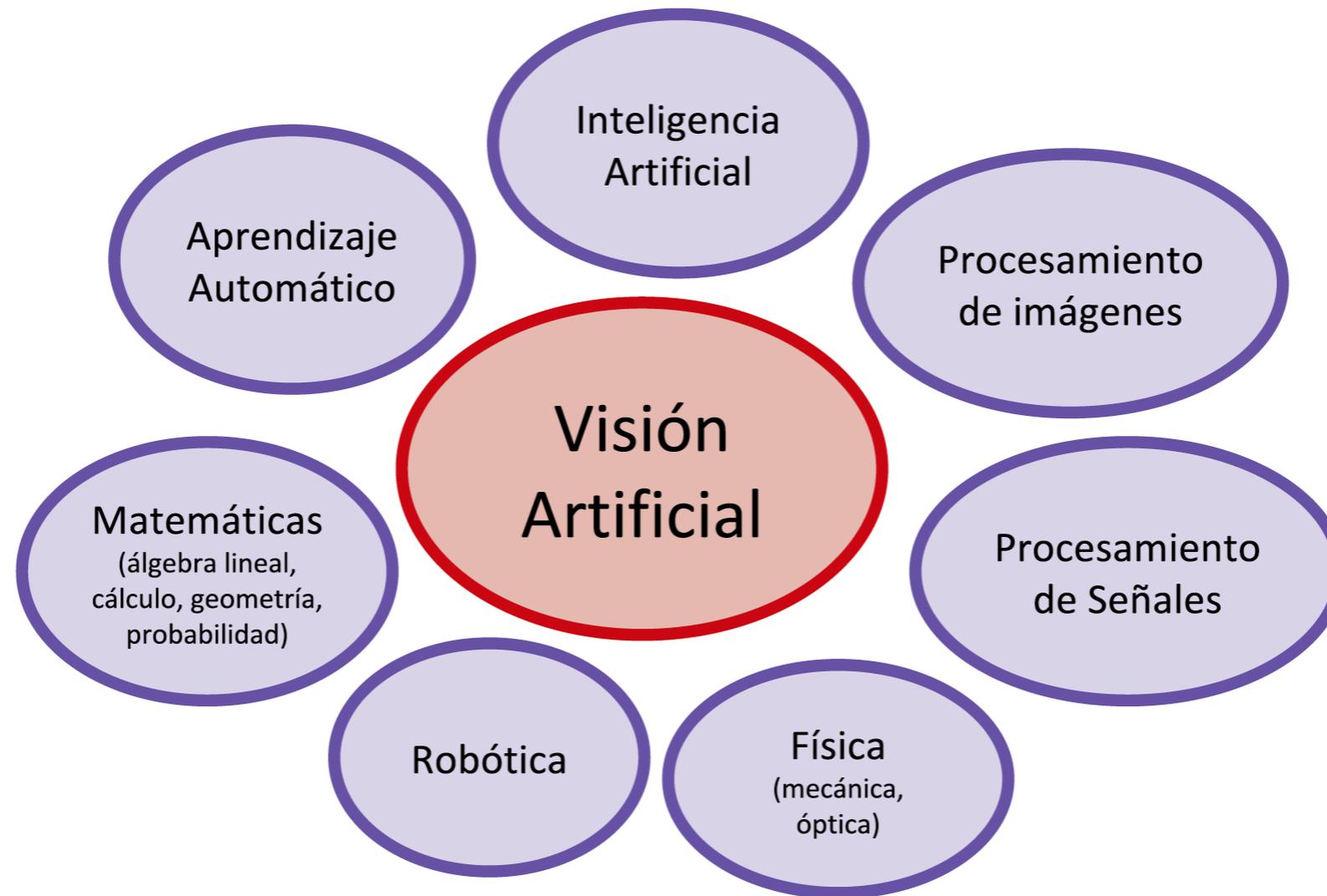
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Representaciones jerárquicas adaptivas



- Cada capa es una representación adaptiva (entrenable)
- Jerarquía de representaciones (capas): nivel creciente de abstracción:
 - **Imágenes:** Pixel → bordes → partes → objetos
 - **Texto:** Caracteres → palabras → grupos de palabras → frases → historias

Visión Artificial (*Computer Vision*)



Visión Artificial: Ejemplos



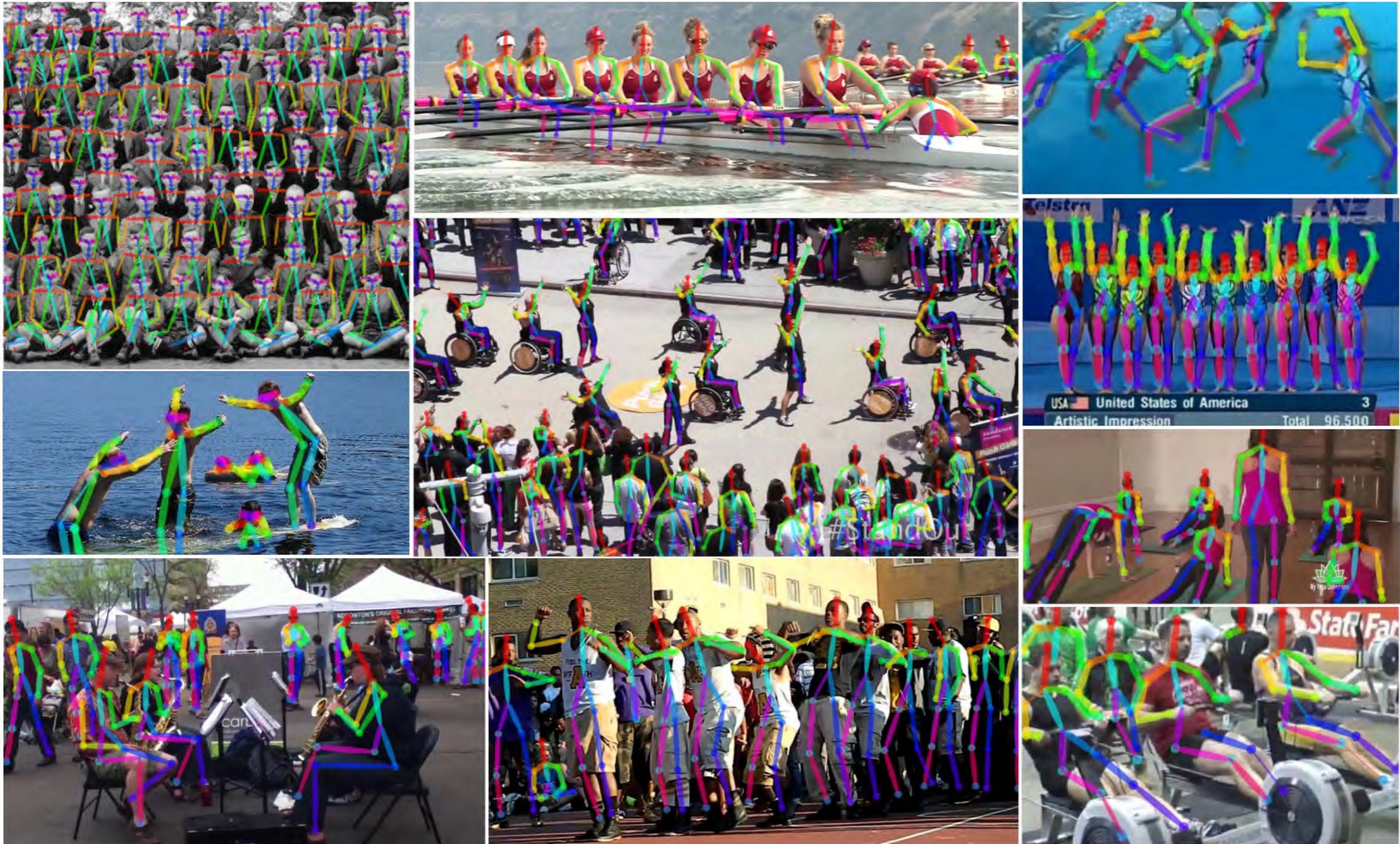
Viola and Jones. "Robust real-time face detection" IJCV 2004

Visión Artificial: Ejemplos



Farabet et al. "Learning hierarchical features for scene labeling" PAMI 2013

Visión Artificial: Ejemplos



Cao et al. "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields" arXiv 2016

Visión Artificial: Ejemplos



man in black shirt is playing guitar.



construction worker in orange safety vest is working on road.



two young girls are playing with lego toy.

Karpathy, Fei-Fe. "Deep visual-semantic alignments for generating image descriptions" CVPR 2015

Visión Artificial: Ejemplos



Gkioxari and Malik. "Finding action tubes" CVPR 2015

Departamento de Procesamiento de Señales, IIE

Procesamiento de Señales y Aprendizaje Automático

- Audio, imágenes, video, volúmenes, datos en general, ...
- Reconocimiento de patrones, machine learning, deep learning, predicción, anomalías, biometría, biomedicina...

El Departamento

- 23 docentes, 8 en régimen de dedicación total
- 8 PhD (Francia, USA, Uruguay), 4 MSc (Francia, USA, Uruguay)
- 5 PhD en curso, 6 MSc en curso
- 4 estudiantes en fase final del grado

Grupos de Investigación

- Grupo de Tratamiento de Imágenes (GTI)
- Grupo de Procesamiento de Audio (GPA)
- Núcleo de Teoría de la Información (NTI)
- Reconocimiento de Patrones (machine learning)

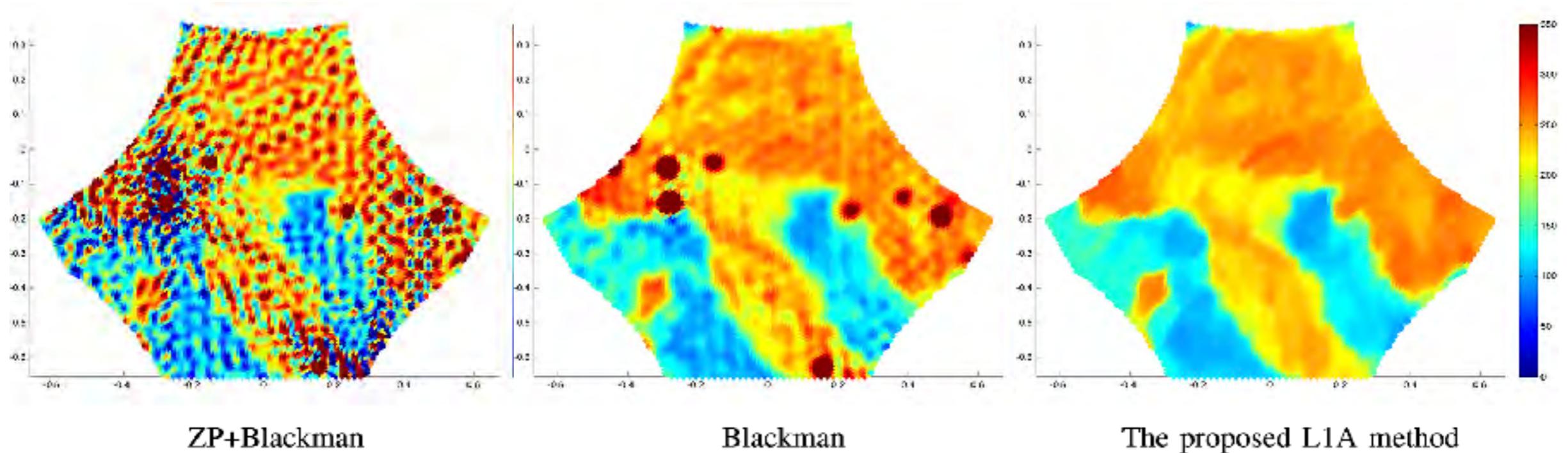
Colaboraciones nacionales e internacionales

- EEUU, Francia, España, Italia, Argentina, Brasil, Chile, ...
- IIBCE, INIA, SUL, LATU, CUDIM, FAgro, FMed, FVet, Psico, EUM, Institut Pasteur de Montevideo, ICT4V, ...

Restauración de Imágenes del Satélite SMOS

Proyecto conjunto con el Centro Nacional de Estudios Espaciales, Francia.

SMOS sensa humedad en la tierra y salinidad en el océano por interferometría en radiofrecuencia.



J. Preciozzi, A. Almansa, P. Musé, S. Durand, A. Khazaal, B. Rougé. *A sparsity-based variational approach for the restoration of SMOS images from L1A data*, IEEE Transactions on Geoscience and Remote Sensing, 2017.

Detección de floraciones algales con imágenes de satélite multiespectral

- **Objetivo:** predicción de concentración algal en el río Uruguay
- **Método:** regresión a partir de imágenes satelitales multiespectrales y datos in-situ
- **Resultado:** modelo predictivo de la concentración algal a partir de una imagen satelital multiespectral.

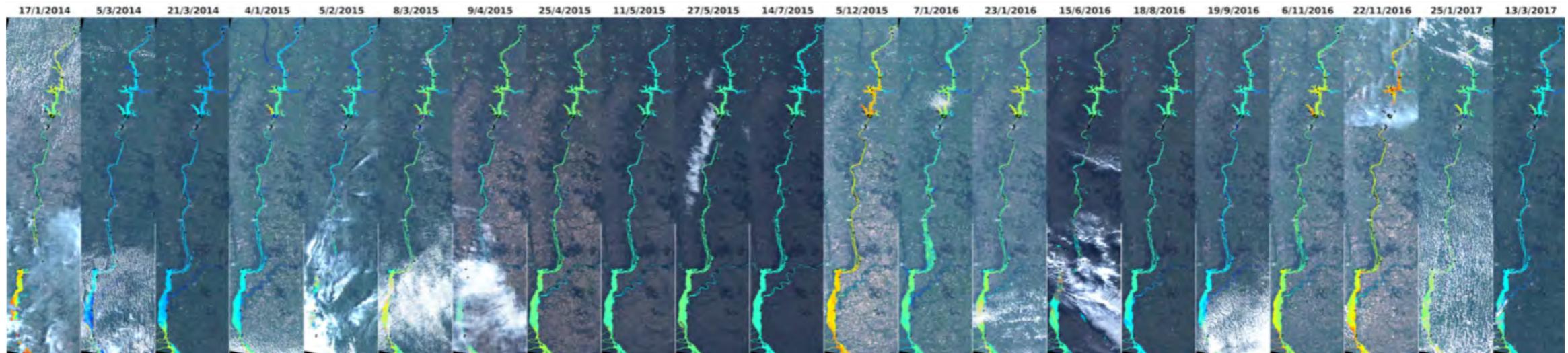
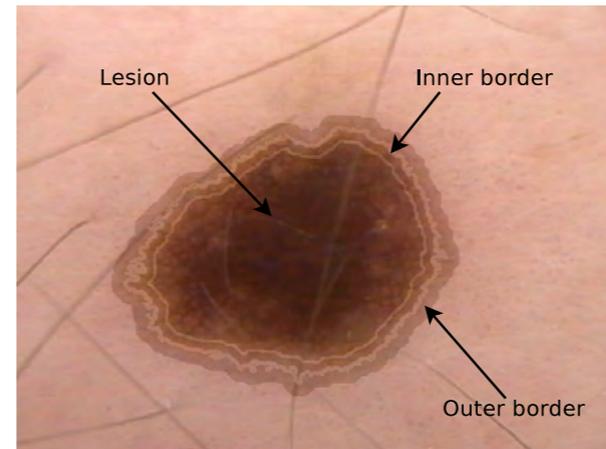
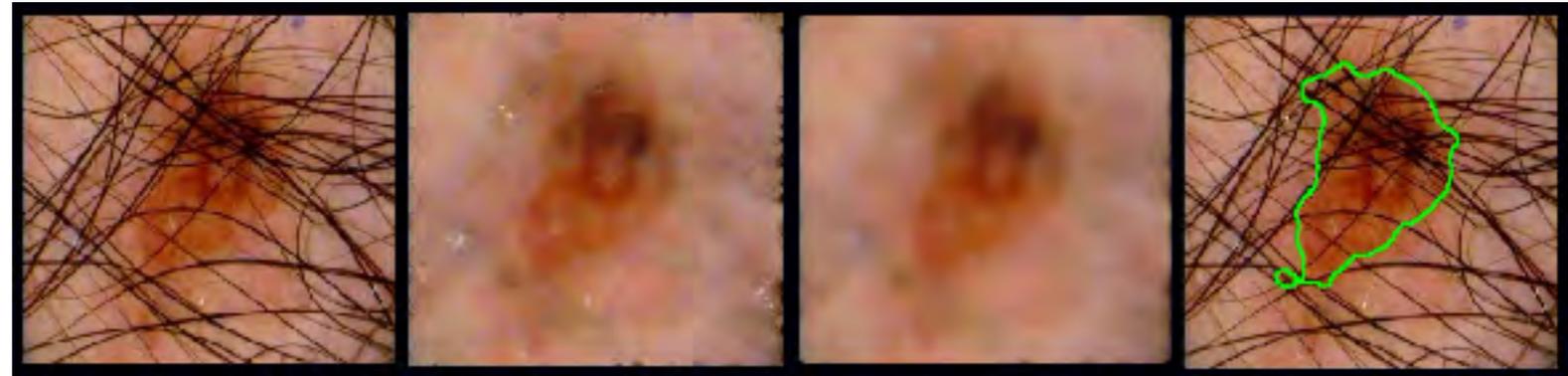
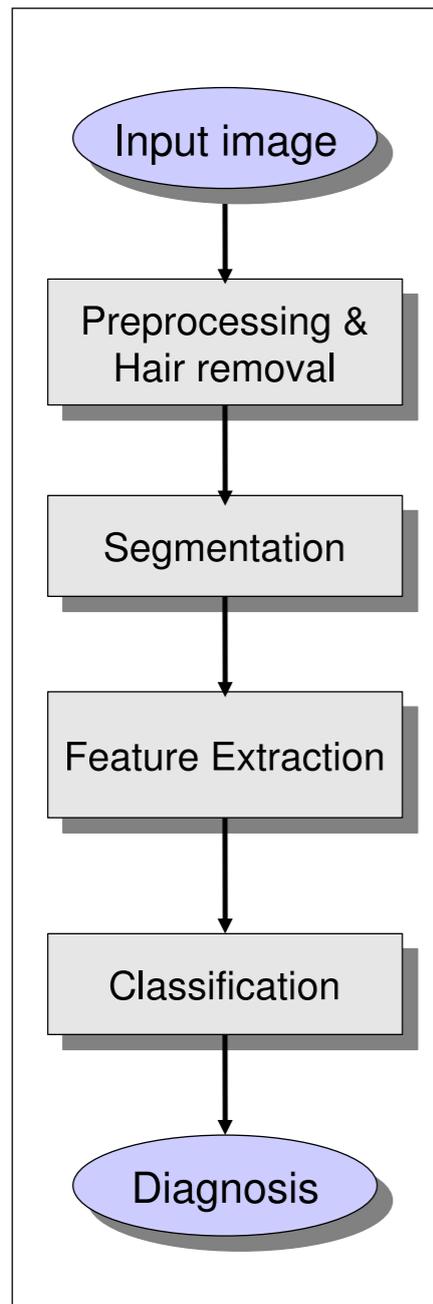


Figura: Mapa de probabilidad de floración algal

Detección de melanoma en imágenes dermatoscópicas

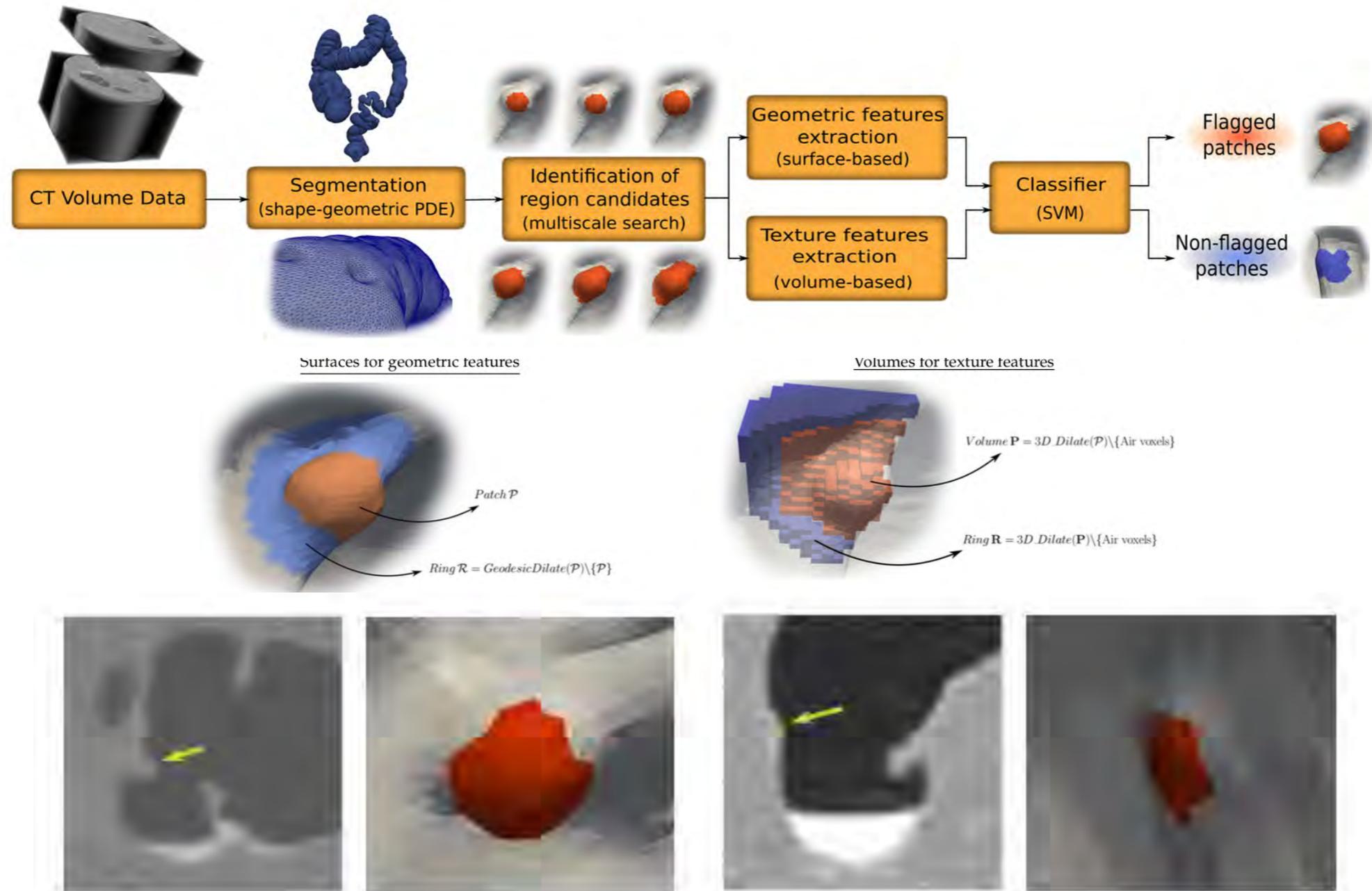


False Negatives		False Positives		
ABCD score=6 7 points=5	ABCD score=6.2 7 points=6	ABCD score=5 7 points=2	ABCD score=4.5 7 points=2	ABCD score=4.9 7 points=3

All misclassified patterns corresponding to lesion images.

G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, P. Musé. *Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions*, Pattern Recognition Letters, 2011.

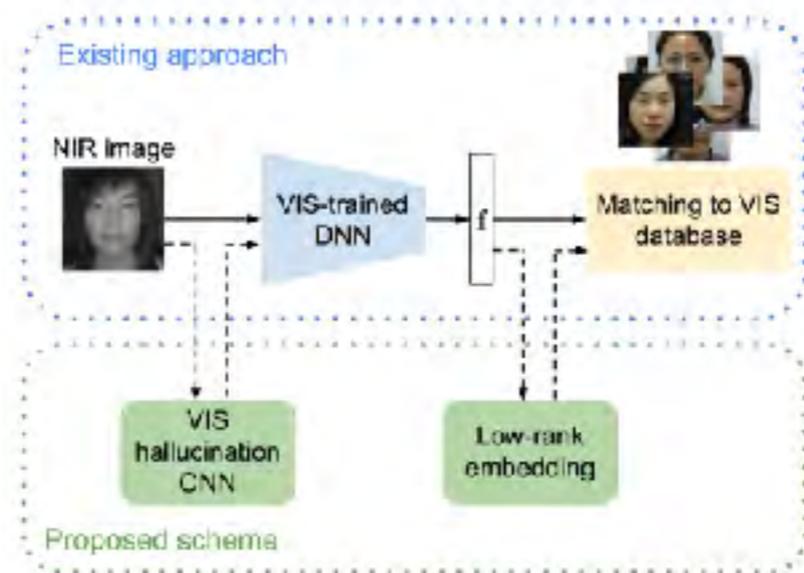
Detección de pólipos en colonoscopia virtual



M. Fiori, P. Musé, G. Sapiro. *A Complete system for candidate polyps detection in virtual colonoscopy*, Int. J. Patt. Recogn. Artif. Intell., 2014.

Reconocimiento de Caras en Infrarrojo Cercano (cámaras NIR)

- **Objetivo:** búsqueda en una base de datos color (RGB) a partir de una imagen infrarroja (NIR)
- **Método:**
 - Regresión de RGB a partir de NIR con una red neuronal
 - Post-procesamiento de las características obtenidas por una red neuronal especializada en caras
- **Resultado:** algoritmo de reconocimiento de caras a partir de una imagen infrarroja.



José Lezama, Qaing Qiu, Guillermo Sapiro. *Not afraid of the dark: NIR-VIS face recognition via cross-spectral hallucination and low-rank embedding*, CVPR 2017

Deep Video Deblurring

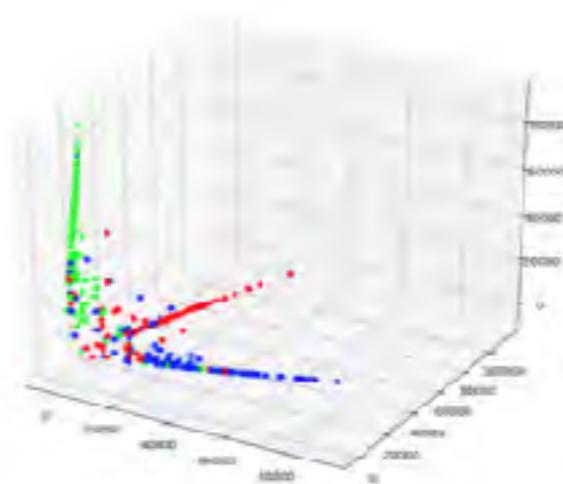
<https://www.youtube.com/watch?v=NoqRMlbqgaQ>

S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, O. Wang IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2017

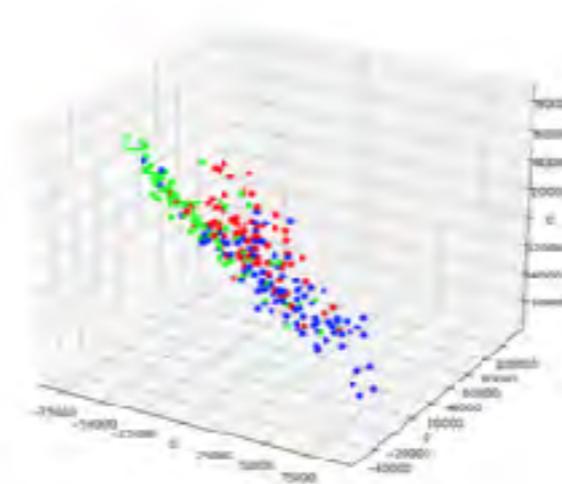
Deep Metric Learning

- **Objetivo:** aprender vectores de características con una red neuronal, que vivan en un espacio de baja dimensión
- **Método:** Restricción geométrica para la función de costo de la red
- **Resultado:** Mejora general en la clasificación para varias arquitecturas de redes neuronales profundas.

$$\sum_{i=1}^C \|X_i\|_* - \|X\|_*$$



(a) OLÉ: 78.38% accuracy



(b) Softmax: 76.69% accuracy

CIFAR 3 classes

Formación en *Machine Learning* en FING

Instituto de Ingeniería Eléctrica / Instituto de Matemática y Estadística /
Instituto de Computación

Tratamiento de imágenes por computadora

Estimación y predicción en series temporales

Procesamiento digital de señales de audio

Reconocimiento de patrones

Aprendizaje profundo para visión artificial

Introducción al procesamiento de lenguaje natural

Procesamiento semántico del lenguaje natural

Bases de datos no relacionales

Métodos de aprendizaje automático

Álgebra lineal, Probabilidad y estadística, Análisis, etc

Estadística multivariada computacional

Optimización no lineal

Optimización con aplicación a la planificación bajo incertidumbre

Computación de alta performance