Summary and More

Learning from Users

A Brief Overview

Setting

- A number of ML applications are used by many users regularly, and in the process, they also give supervisory signals.
- Examples:
  - Search Engines
  - Browsing/watching/trying products
  - Using devices (like a temperature control at home)
  - Etc.

Allied problems

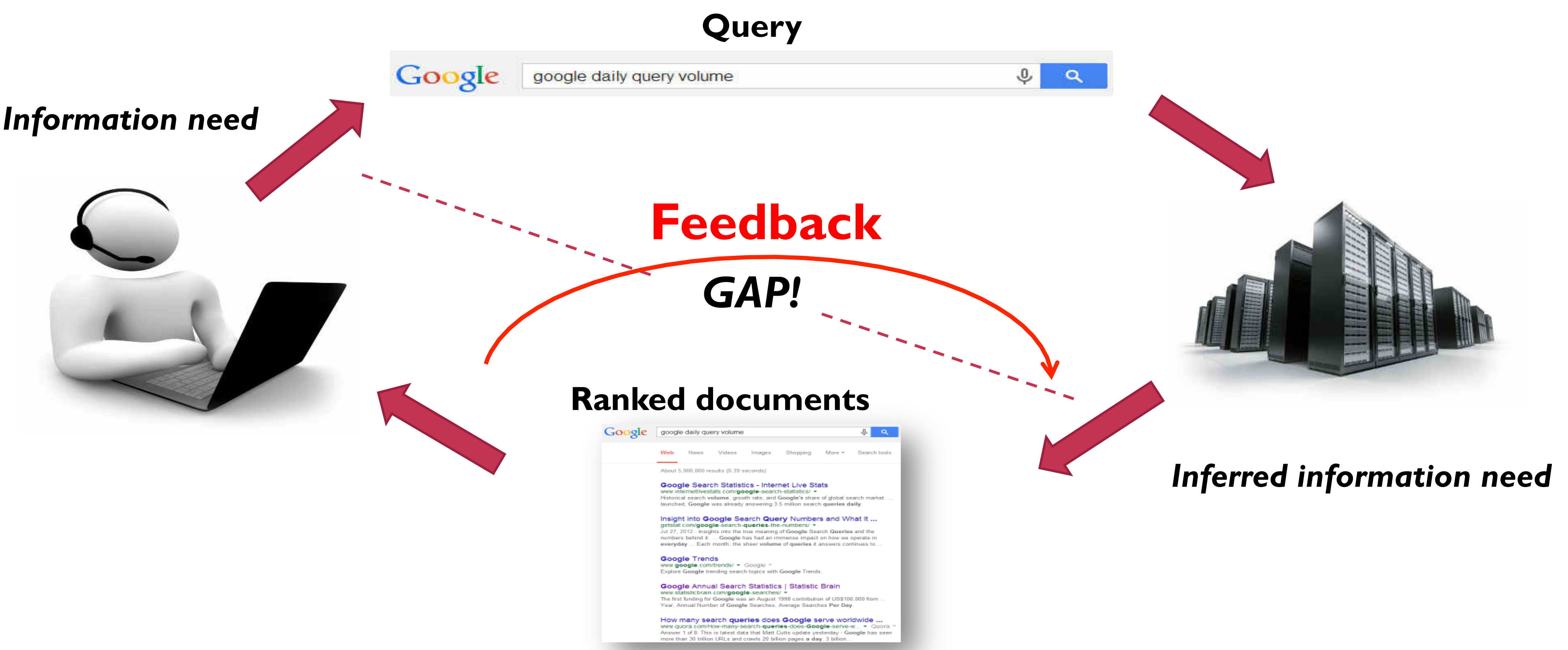
- Personalization
- Online learning/Incremental Learning
- Weak Supervision
- ..
- User feedbacks:
  - Explicit feedbacks (labels +ve and/or –ve explicitly)
  - Implicit feedbacks
- Many variations. Many special cases
- We will look at some sample scenarios.

Relevance Feedback

(Popular in IR, and Multimedia Retrieval)

User feedback

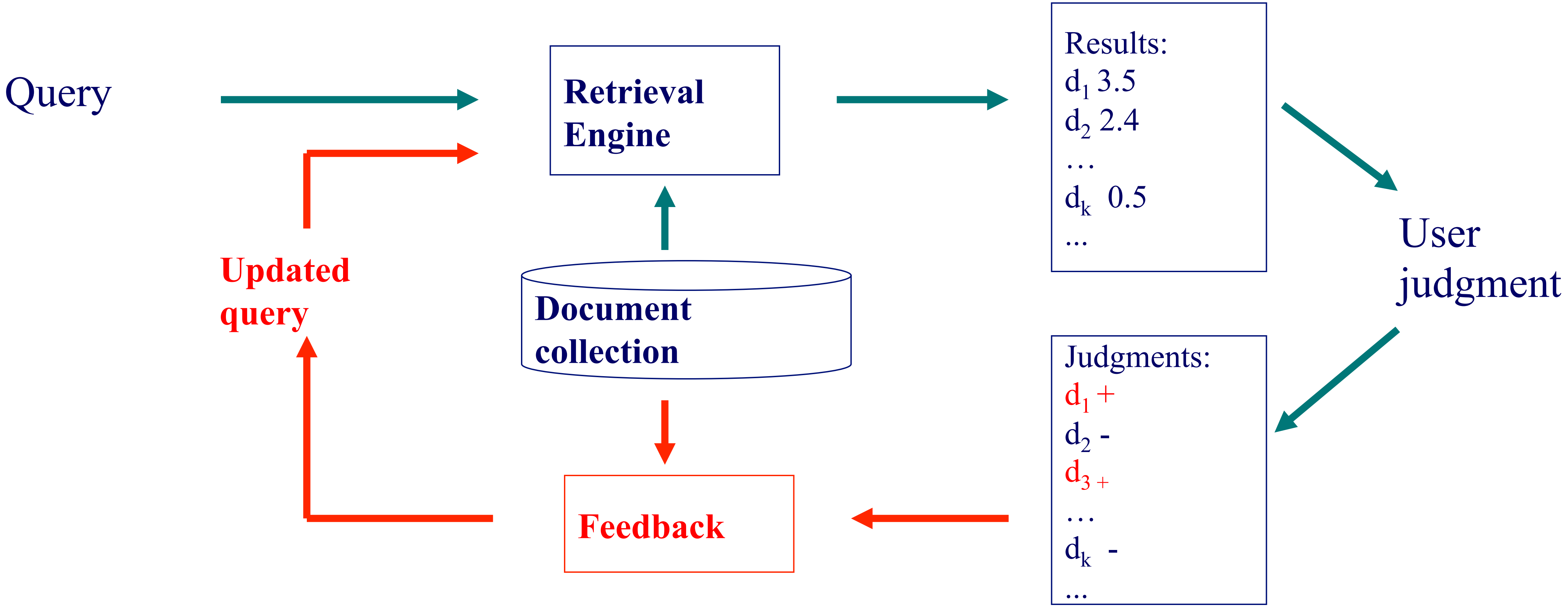
- An IR system could be an interactive system



Use Scenario

- A query **q** or a classifier **w** is given
- Search engine retrieves a set of possible answers
  - x1, x2, x3, etc.
- System guess the user intend and improve the answers
  - x7,x12,x23, etc.
- User is able to smartly navigate and get what she is looking for.
- Eg. Search for a specific fashion/design in a large database

Relevance feedback

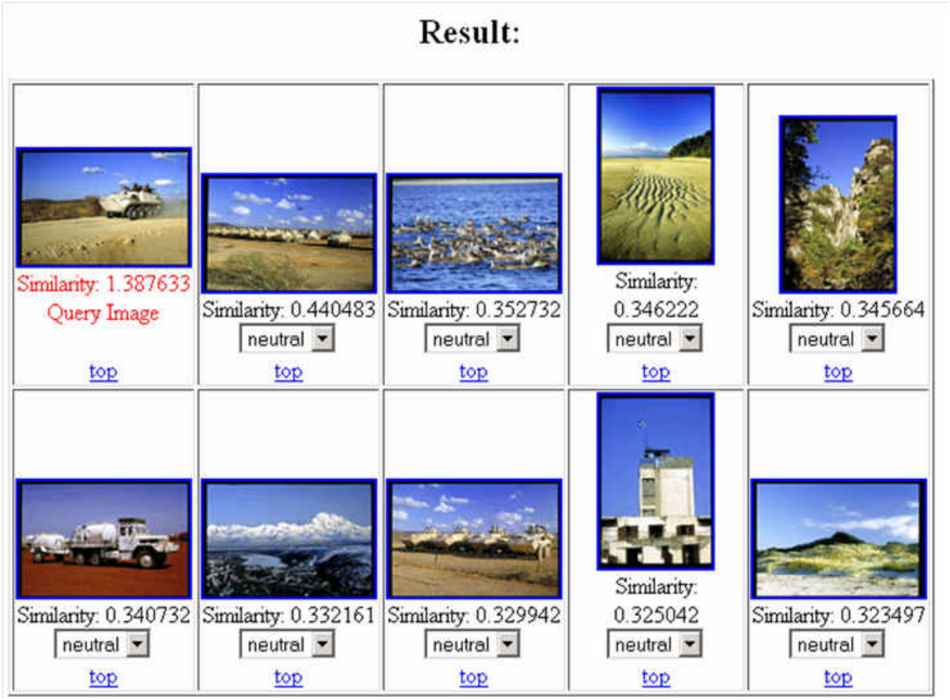


Effective and Popular(?)


Personalization - Wikipedia, the free encyclopedia
Personalization involves using technology to accommodate the differences between individuals. Once confined mainly to the Web, it is increasingly becoming a ...
en.wikipedia.org/wiki/**Personalized** - 42k - Cached - Similar pages -

Personalized Gifts from **Personalization** Mall
It shows you went out of your way to find the perfect gift and to **personalize** it to make it theirs alone! At PersonalizationMall.com, we design most of our ...
www.**personalization**mall.com/Default.aspx?&did=111028 - 47k -
Cached - Similar pages -

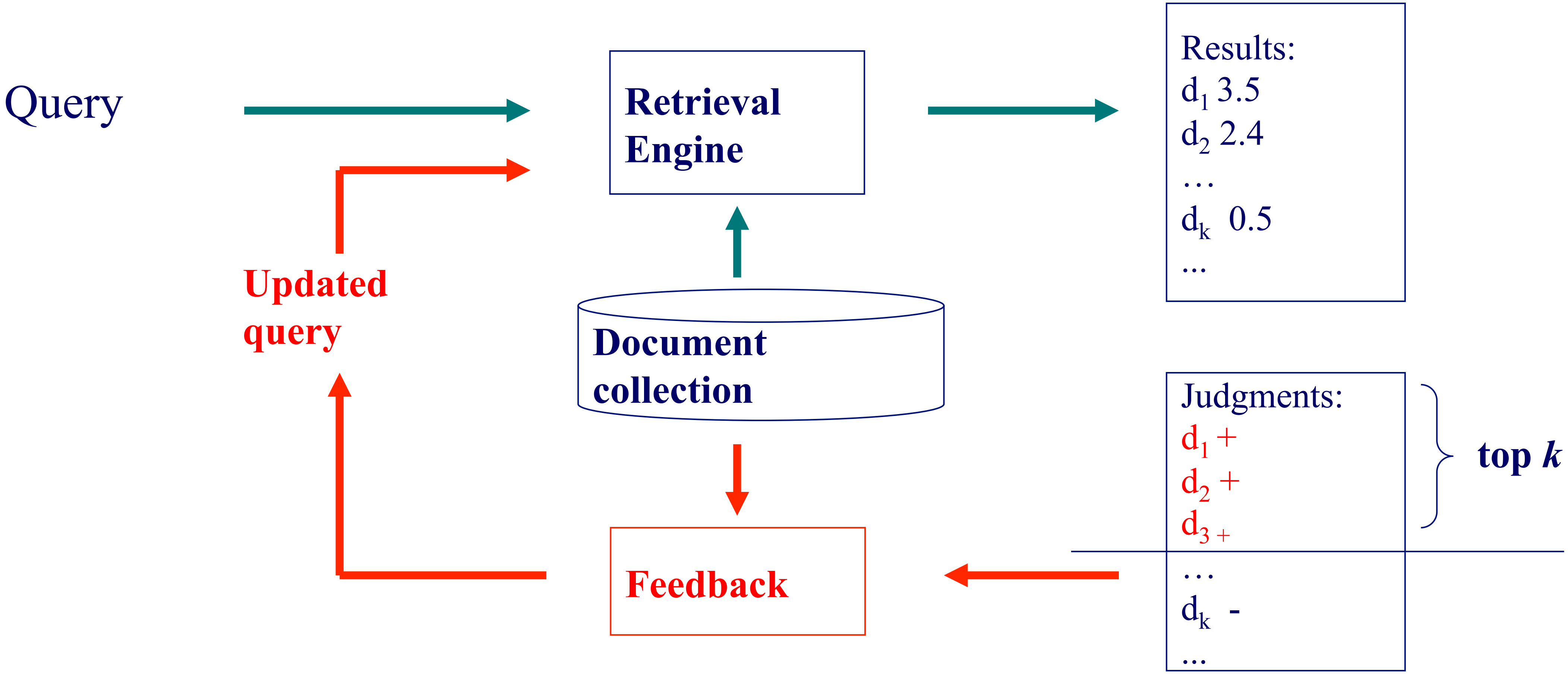
What is **personalization**? - a definition from Whatis.com
Mar 6, 2007 ... On a Web site, **personalization** is the process of tailoring pages to individual users' characteristics or preferences.
searchcrm.techtarget.com/sDefinition/0,,sid11_gci532341,00.html - 72k -
Cached - Similar pages -



Too Explicit?

Pseudo feedback and Query Expansion

- What if the users are reluctant to provide any feedback



Rocchio Model

$$Q_1 = \alpha\ Q_0 + \frac{\beta}{n_1}\sum_{i=1}^{n_1} R_i - \frac{\gamma}{n_2}\sum_{i=1}^{n_2}\ S_i$$

where

Q_0 = the vector for the initial query

R_i = the vector for the relevant document i

S_i = the vector for the non-relevant document i

n_1 = the number of relevant documents chosen

n_2 = the number of non-relevant documents chosen

α, β and γ tune the importance of relevant and nonrelevant terms (in some studies best to set β to 0.75 and γ to 0.25)

Illustration (NN View)

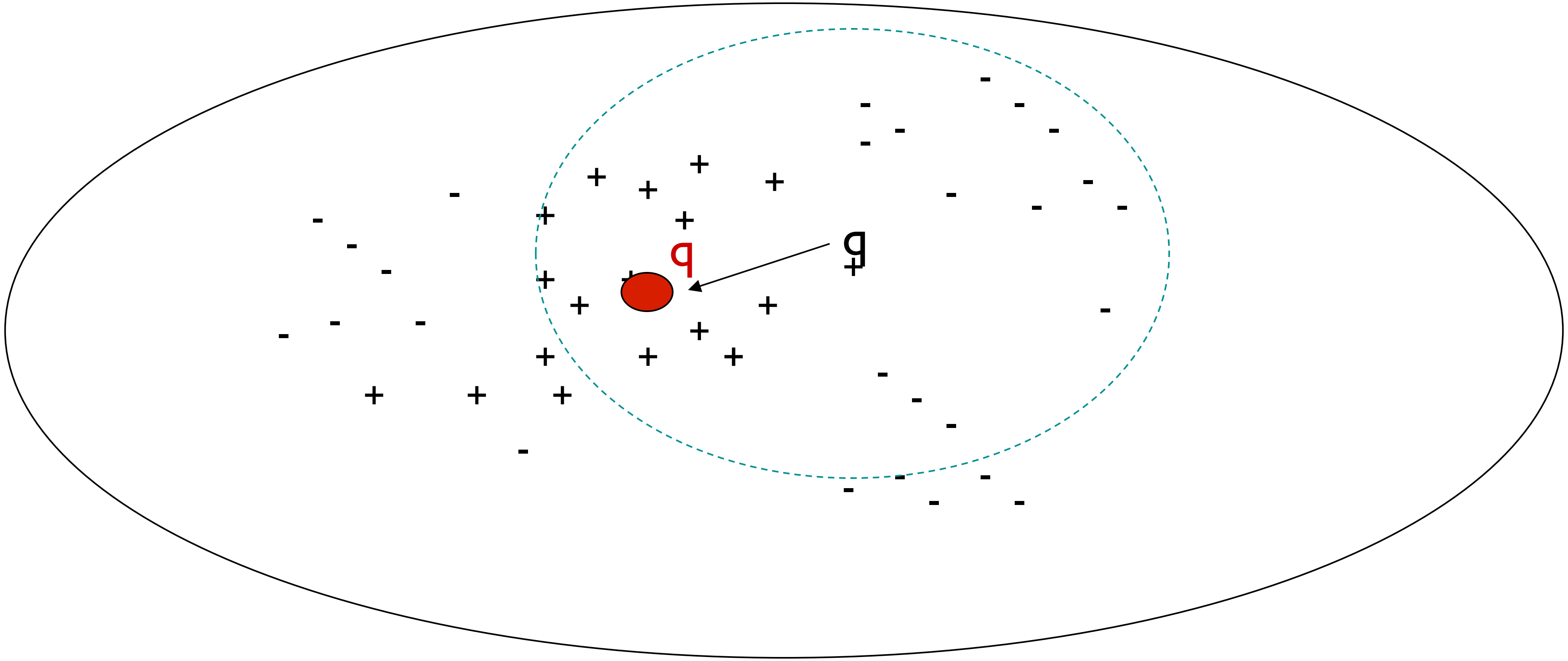
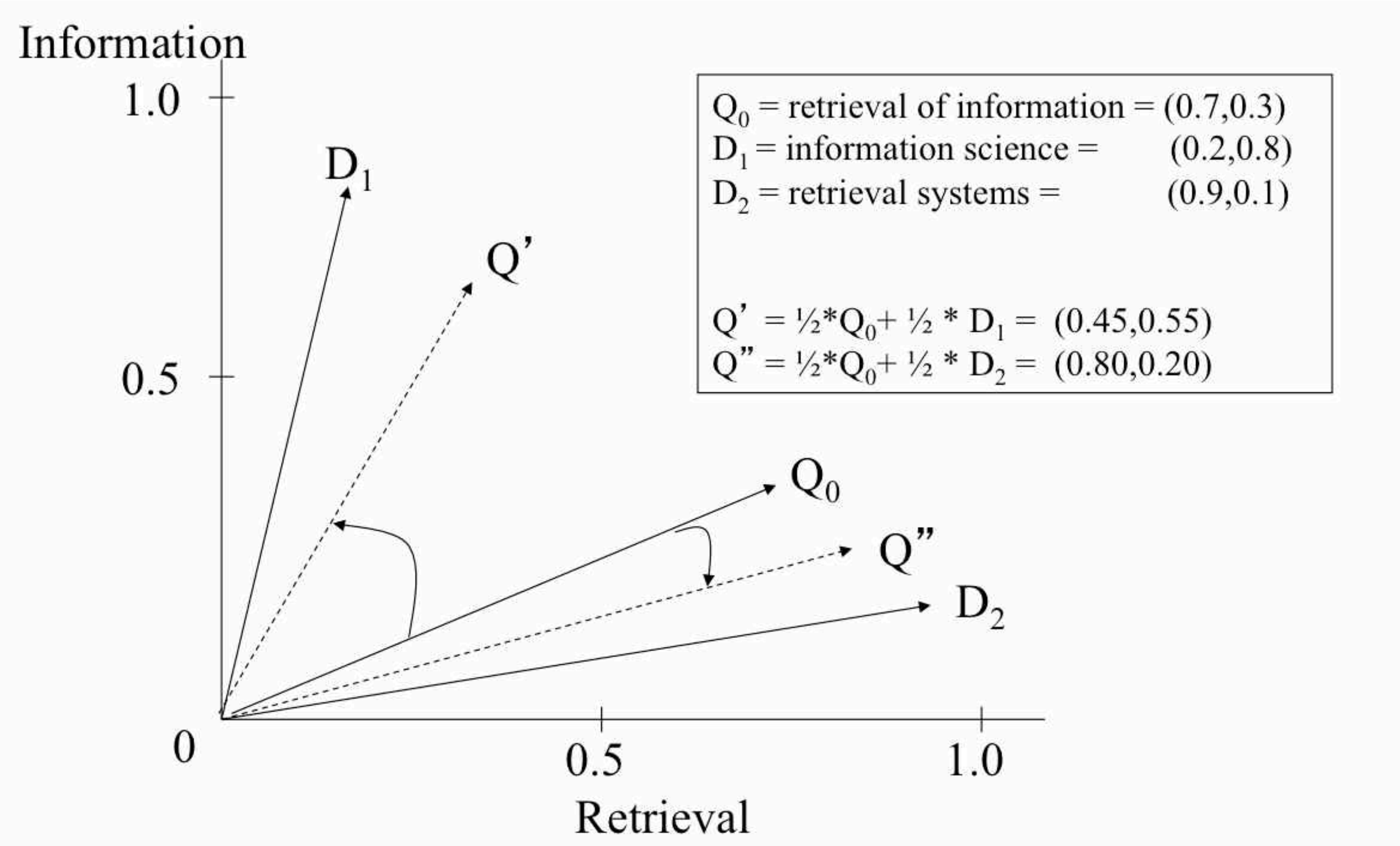


Illustration (dot product view)



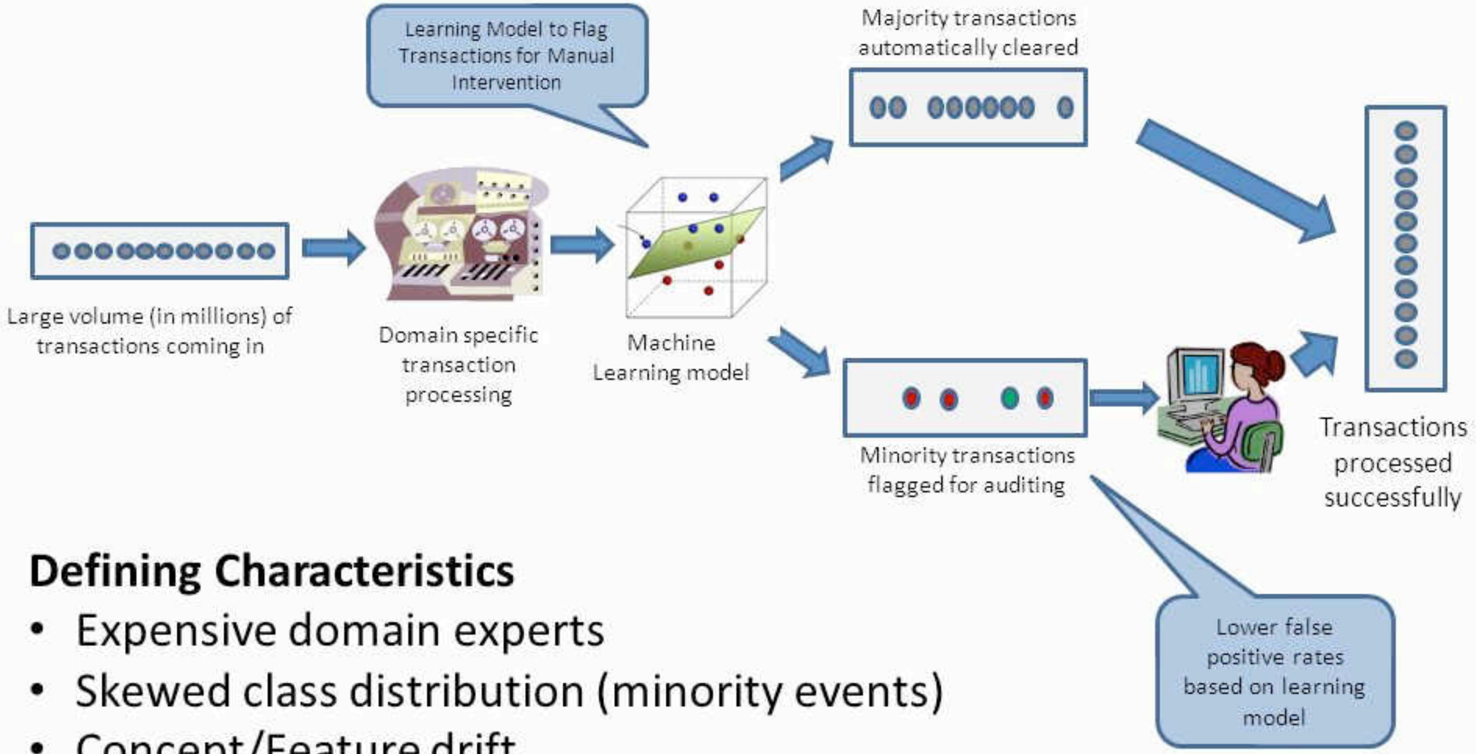
Challenges and Refinements

- Can we force user to say + and – on the answers?
  - Often + is more clear ?. But not – ve is not shared.
    - Cases when only + or Only – is available.
  - Often + is implicit (I click/browse) and not explicit.
- Examples:
  - Browsing for fashion (clothes)

Users in the ML systems

Interactive Classification

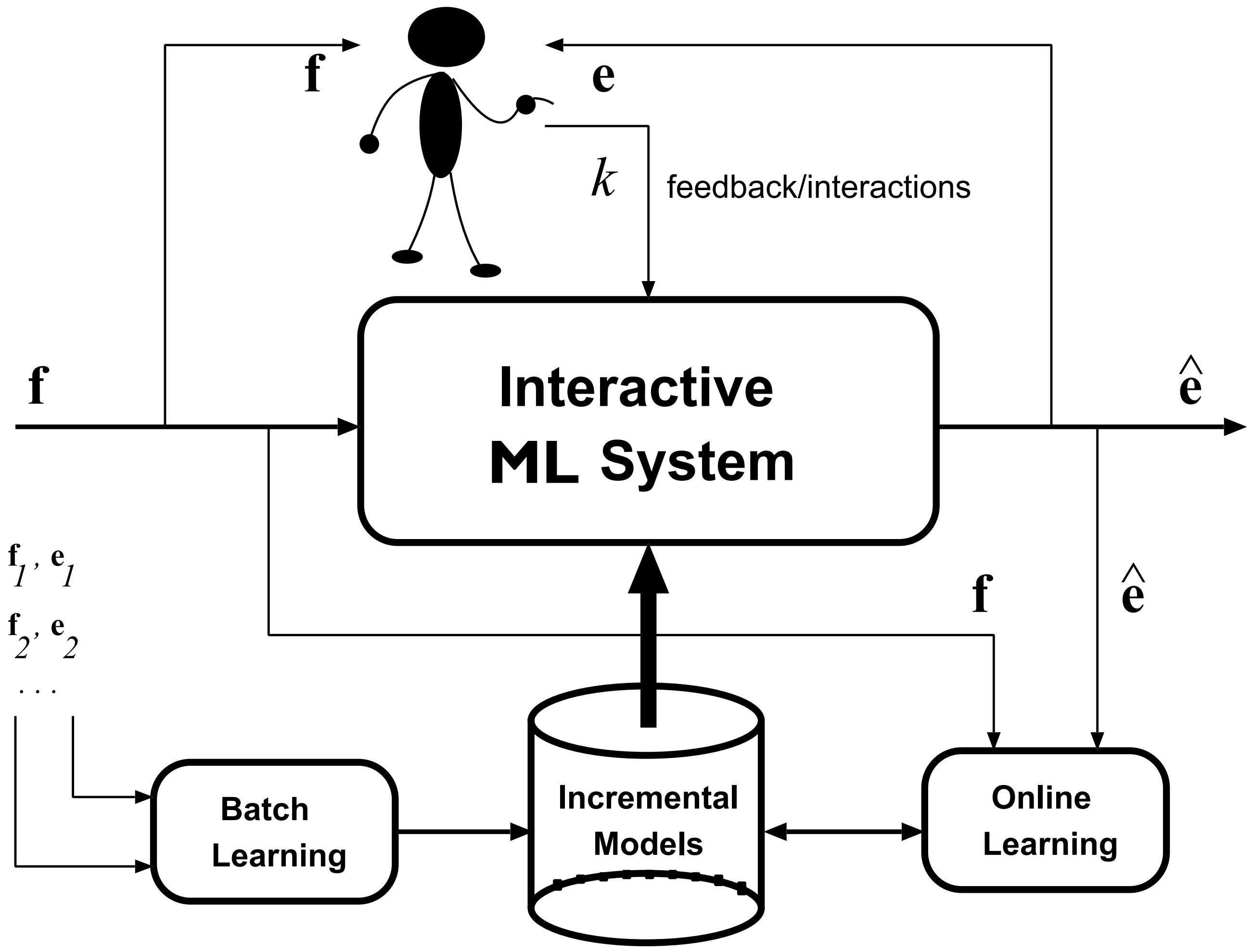
Goal: Optimize life-time Return On Investment



Defining Characteristics

- Expensive domain experts
- Skewed class distribution (minority events)
- Concept/Feature drift
- Biased sampling of labeled historical data
- Lots of unlabeled data

Another Scenario (Interactive ML)



User may adjust:

- K (control) or

- Directly give desired e

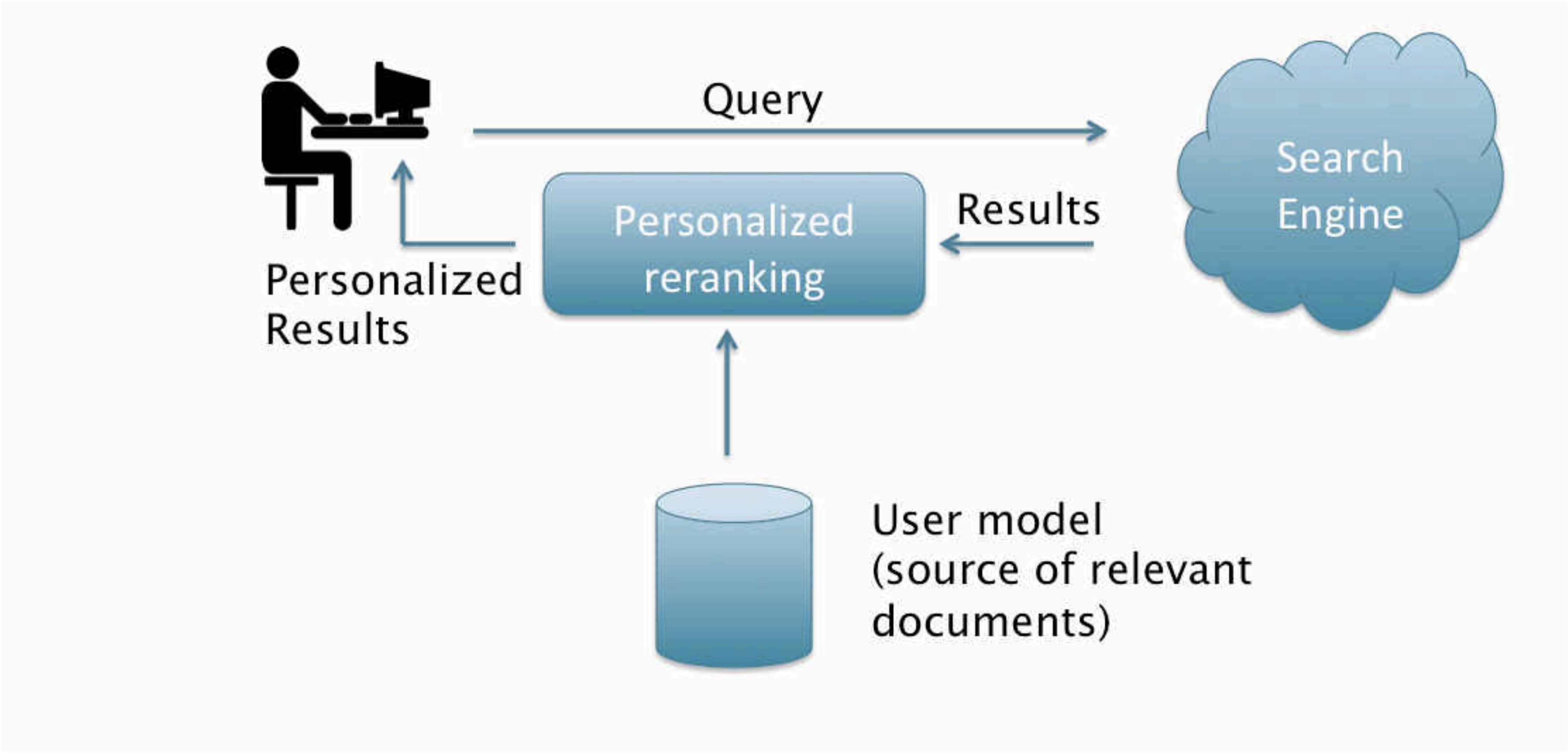
Associated Issues

- Incremental and Computational Issues
  - How do we learn, adapt and forget
  - What is the basic knowledge and what do we adapt?
- Stability
  - Am I overlearing and changing too fast?
  - Stability, convergence and other algorithmic issues.

Personalization


DON'T WORRY, IT'S ONLY MARKETERS COLLECTING OUR PERSONAL DATA SO THEY CAN CREATE MORE RELEVANT ADVERTISING FOR US.
TOM FISHBURNE

Use Case (RF style)



Many Relevant Attributes

- Consider the following pieces of information
  - Geographical Location
  - Age, gender, ethnicity, religion, etc.
  - Interests
  - Previous reviews on products
  - ……
- How could these pieces of information help?
- How to collect these information?

Approaches

- Individual Vs Collaborative
- Reactive Vs Proactive
- User Vs Item Information

Individual Vs Collaborative

- Individual approach (Eg. *Google Personalized Search*)
  - Use only individual user's data
  - Generate user profile by analyzing
    - User's browsing behavior
    - User's active feedback on the system
  - Advantage
    - Can be implemented on the client-side - no *privacy violation*
  - Disadvantage
    - Based only on past interactions.

Individual Vs Collaborative

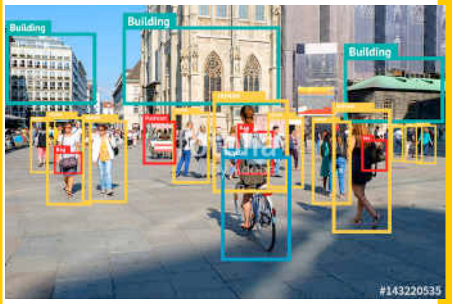
- Collaborative approach (Eg. Amazon recommendations)
  - Find the *neighborhood* of the active user
  - React according to an assumption
    - If *A* is like *B*, then *B* likes the same things as *A* likes
  - Disadvantages
    - *New item rating* problem
    - *New user* problem
  - Advantage
    - Better than individual approach - Once the two problems are addressed.

Reactive Vs Proactive

- Reactive approach
  - Explicitly ask user for preferences
    - Either in the form of query or feedback
- Proactive approach
  - Learn user preferences by user behavior
    - No explicit preference demand from the user
  - Behavior is extracted
    - *Click-through rates*
    - *Navigational pattern*

User Vs Item Information

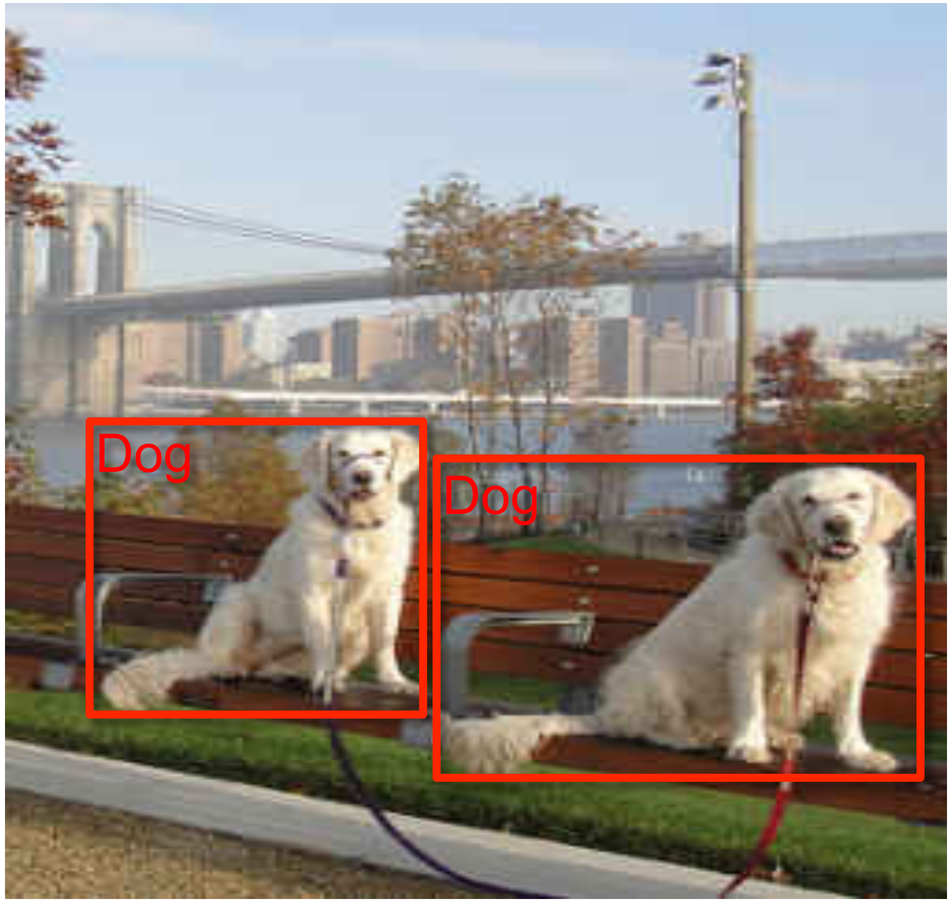
- User Information
  - Geographic location (from IP address)
  - age, gender, marital status, etc (explicit query)
  - Lifestyle, etc. (inference from past behavior)
- Item Information
  - Content of Topics – movie genre, etc.
  - Product/ domain ontology



Object Detection

A classification problem?

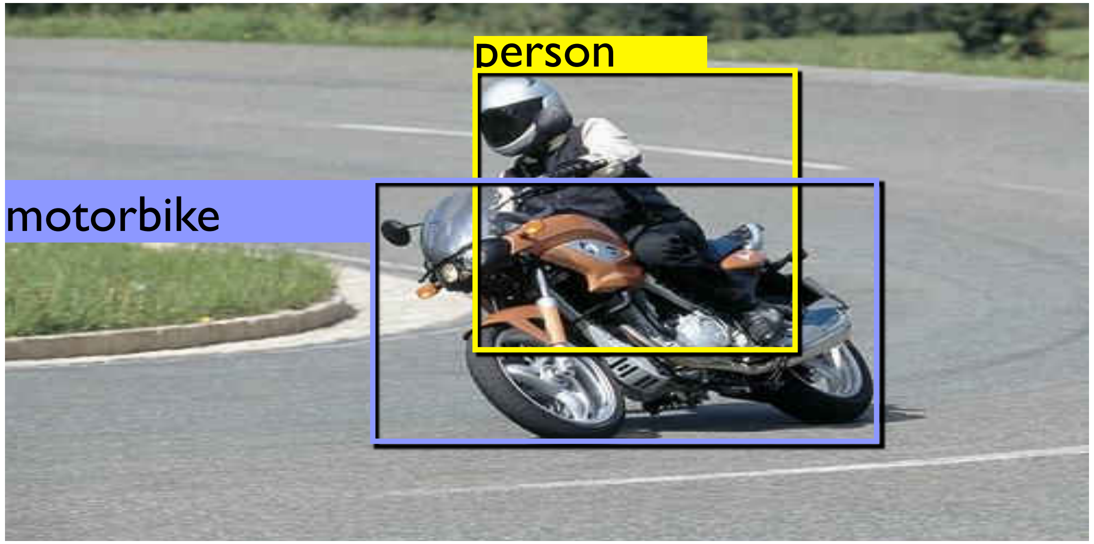
Classification vs. Detection



Problem formulation

{ airplane, bird, motorbike, person, sofa }

Input



Desired output

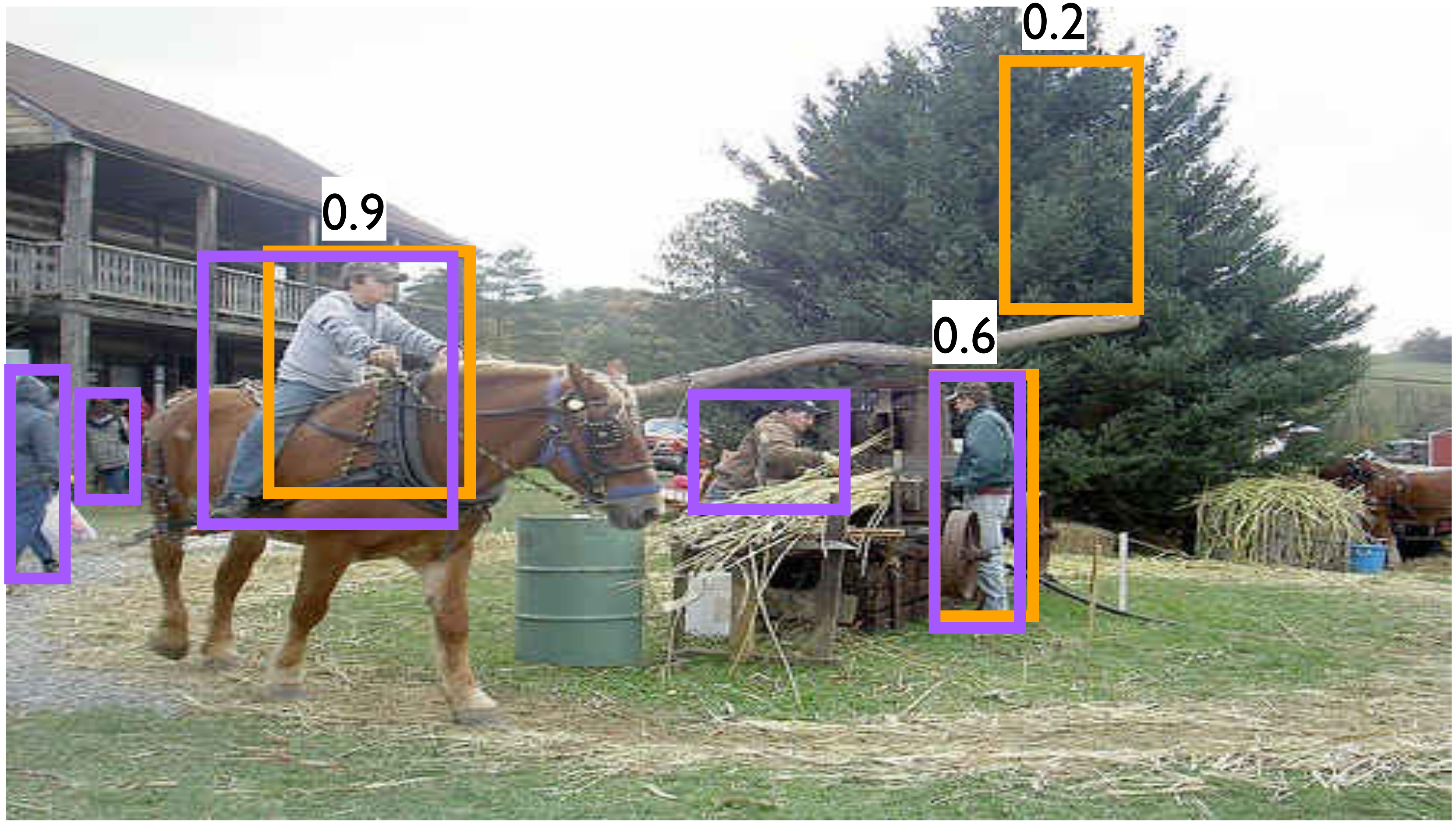
Challenges

- A single image has Millions/Billions of windows to classify
  - All locations
  - All Scales/Sizes
- Computational speed
  - If a classification takes 10 ms?
- Variability in appearance/illumination (as in the classifiers)
- Applications:
  - Face detection (cameras)
  - Obstacle detection (autonomous navigation)

Evaluating a detector

Test image (previously unseen)

Many detections with varying confidence

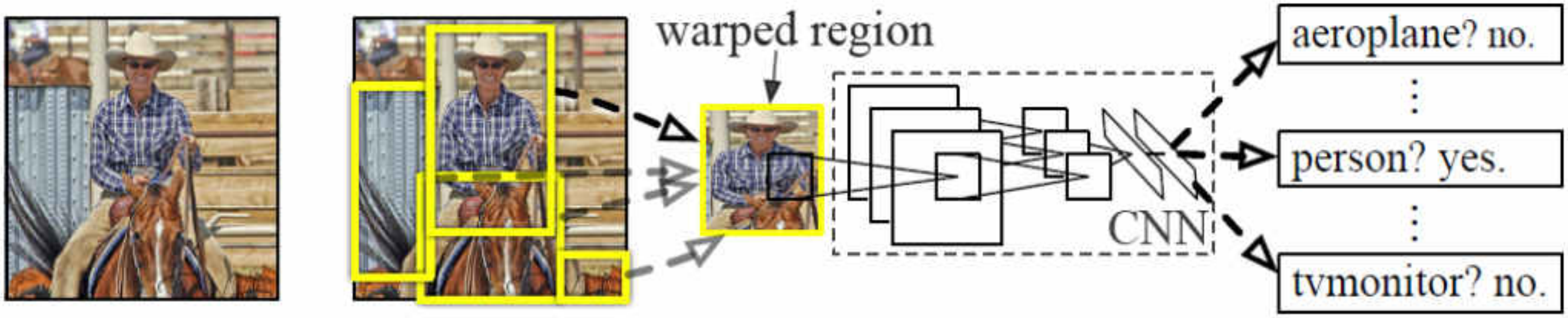


'person' detector predictions

ground truth 'person' boxes

R-CNN: Region with CNN Features

- Rich feature hierarchies for accurate object detection and semantic segmentation



Input Image

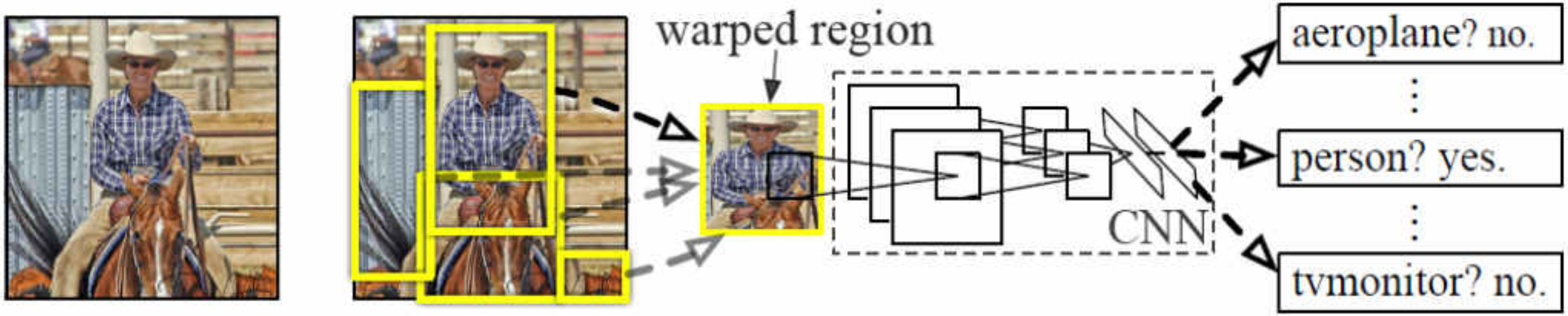
Extract region proposal (~2k/image)

Compute CNN Features

Classify Regions (linear SVM)

Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." CVPR, 2014

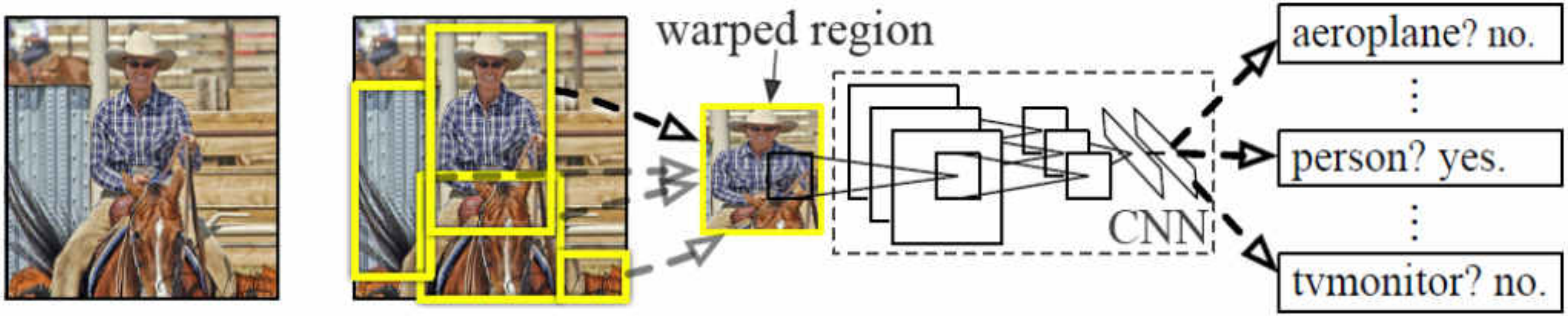
R-CNN: At test time – Step 1



Input Image → Extract region proposal (~2k/image)

- Proposal-method agnostic, many choices
  - Selective Search [van de Sande, Uijlings et al.]
  - MCG [Arbelaez et al.]
  - BING [Ming et al.]
  - CPMC [Carreira & Sminchisescu]

R-CNN: At test time – Step 2



Input Image

Extract region proposal (~2k/image)

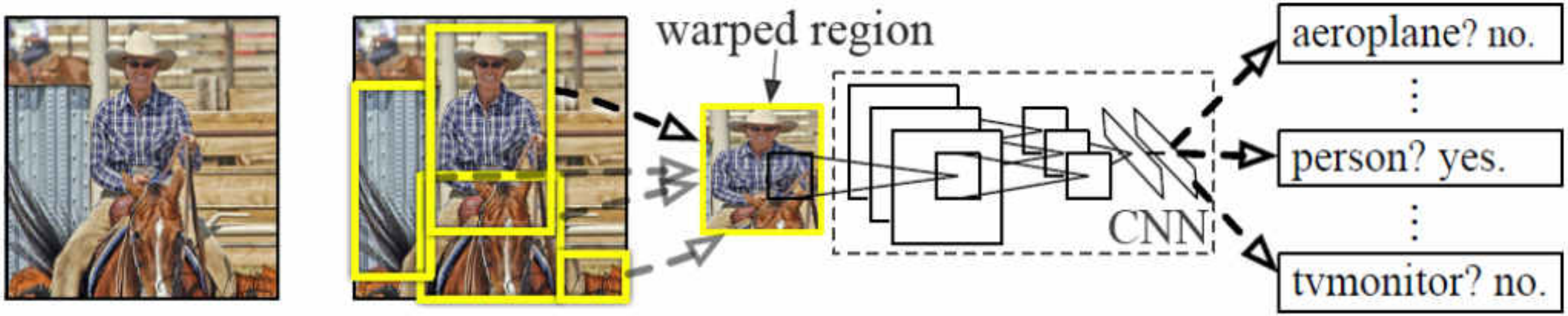
Compute CNN Features

Extract and Dilate Proposal

a. Crop

b. Scale (anisotropic) 227 x 227

R-CNN: At test time – Step 3

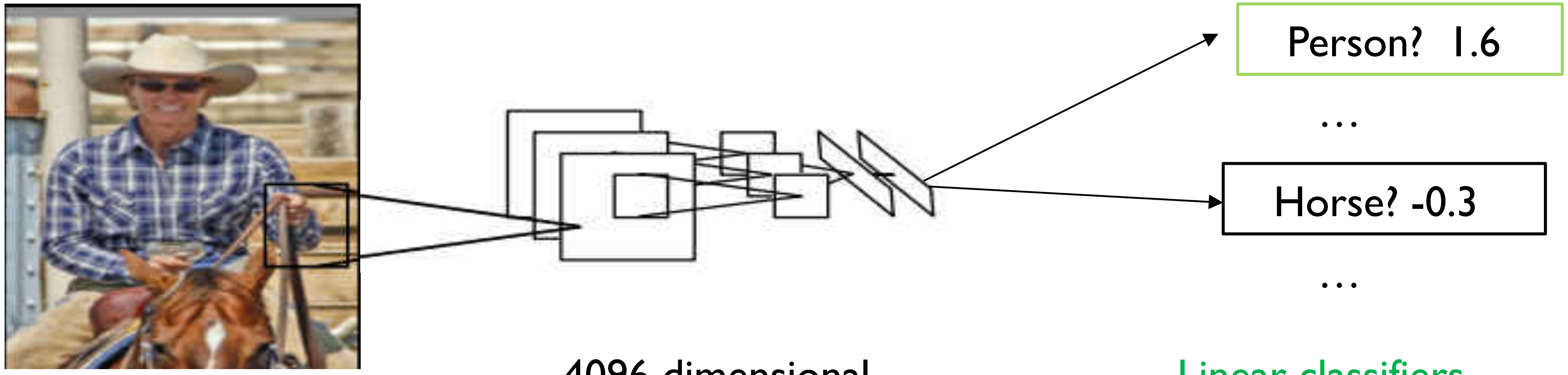


Input Image

Extract region proposal (~2k/image)

Compute CNN Features ← Classify Regions (linear SVM)

…



b. Scale (anisotropic)

4096 dimensional fc_7 feature vector

Linear classifiers (SVM or softmax)

R-CNN: At test time – Step 4

- Object Proposal Refinement (Bounding box regression)

Linear Regression

on CNN Features

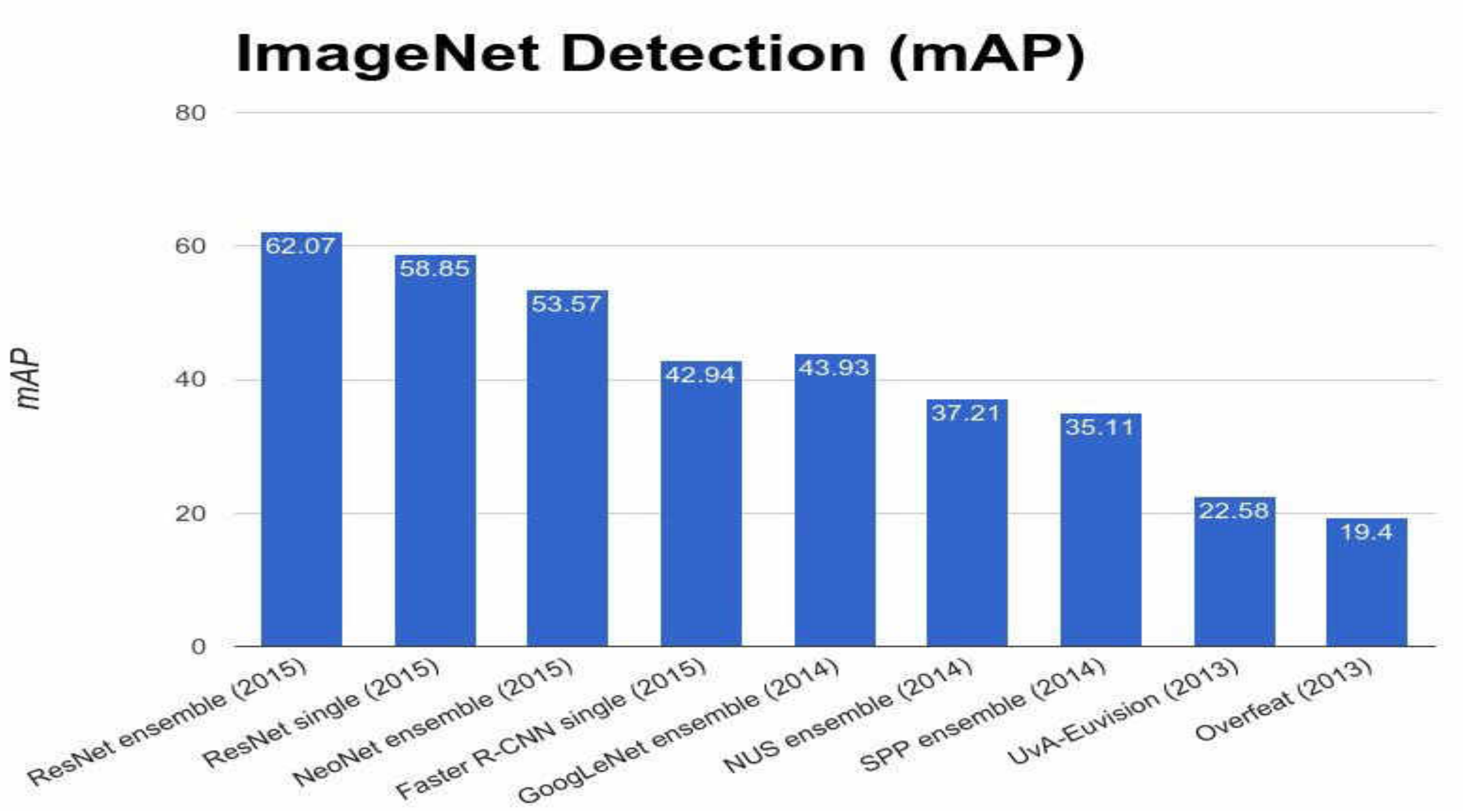
Original Image

Predicted object bounding box

Fast and Faster Versions

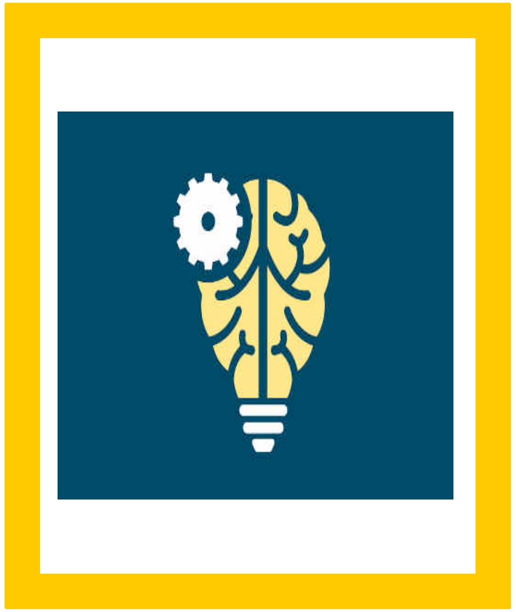
	R-CNN	**Fast R-CNN**	**Faster R-CNN**
Test time per image (with proposals)	50 seconds	2 seconds	**0.2 seconds**
(Speedup)	1x	25x	**250x**
mAP (VOC 2007)	66.0	**66.9**	**66.9**

Rapid Progress



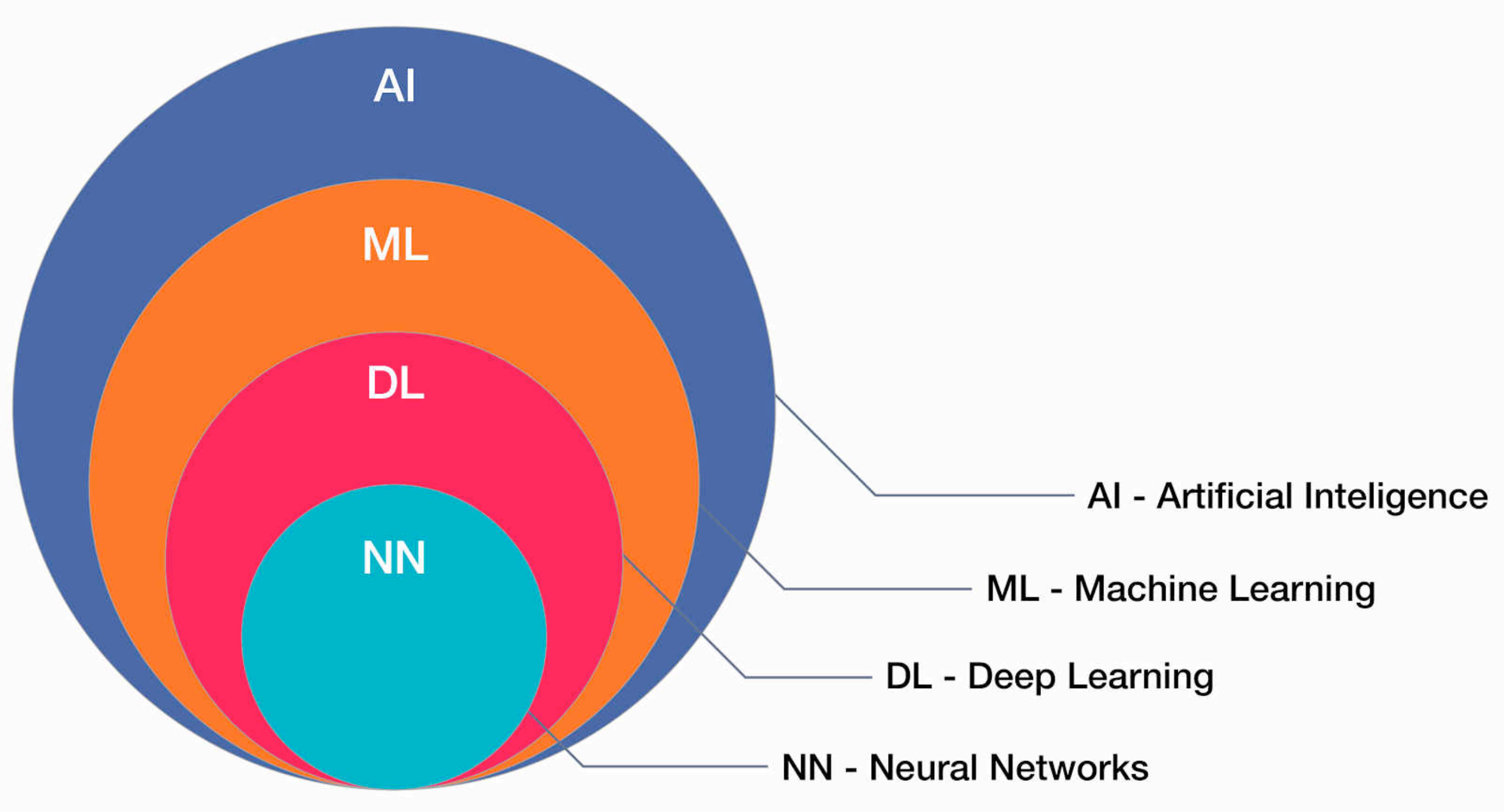
Still Recent

- SSD
- YOLO and its versions [Yolo V3?]



Summary

The broad problem space



Proliferation of Applications

Vision, Speech, Text and Beyond

ADAS

Drones

Mobile

Surveillance

Augmented Reality

Siri

Google Assistant

Hey Cortana

amazon alexa

Google Translate

Meet Jarvis.
He reminds you to do things.

Talk to him on Messenger

Deep Learning Startups

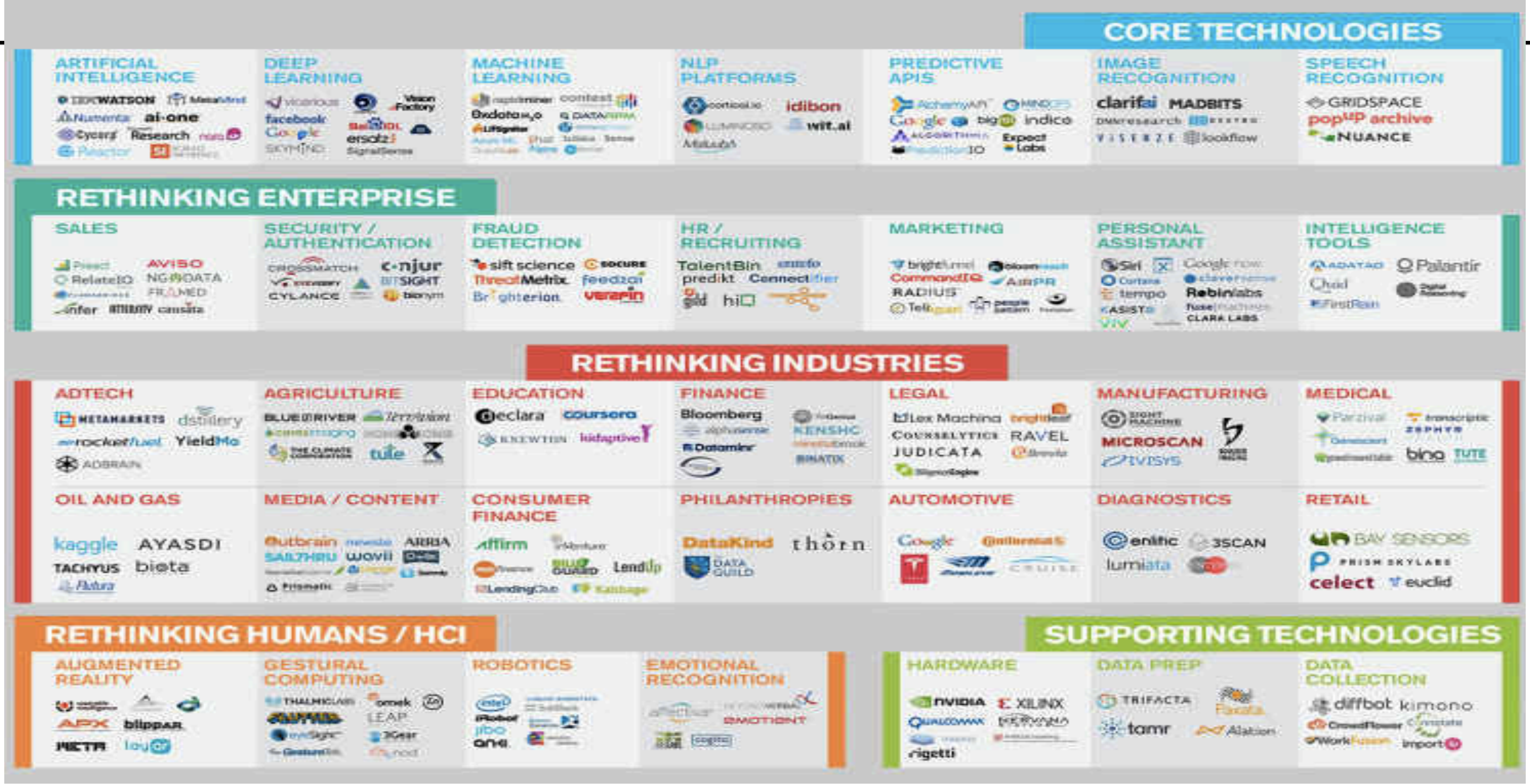
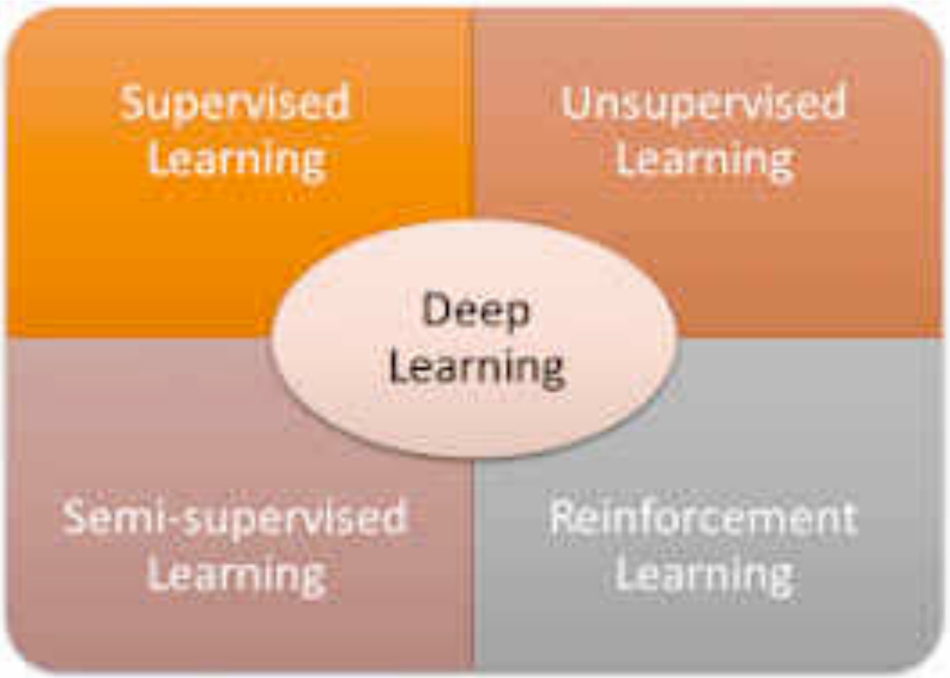


Image Credit: https://airesearch.com

What enabled this success?

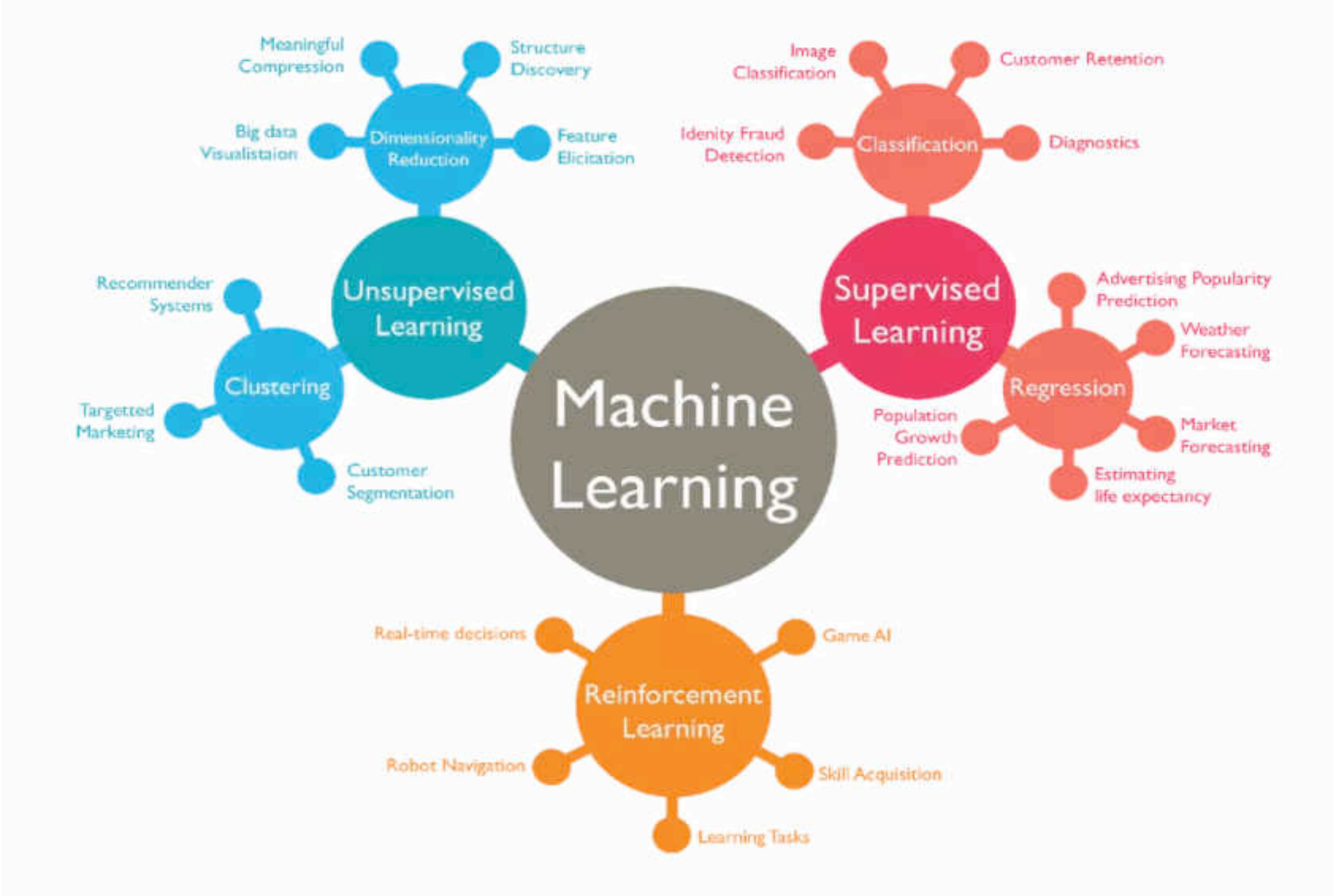
- Modern Features
  - Invariant to popular transformations
  - Capable of capturing local and global (shape, colour, texture) characteristics reliably
  - Features than can be learnt
- Machine Learning
  - Learn from examples rather than handcoding
  - New algorithms: effective, efficient
  - Efficient algorithms to solve complex optimization tasks
  - End to end learning and deep learning.
- Realistic Data
  - Huge amount; partly annotated
  - Regular competitions
  - Challenging problem statements. Evaluation Metrics
- Advances in Computational Resources
  - GPUs
  - Industrial scale clusters
  - Deep Software Stack.

One Simple Dichotomy



Different Types of Supervision

Closer Look at Problems/Algorithms



THE SPACE OF MACHINE LEARNING METHODS

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Recurrent Neural Net

Boosting

Convolutional Neural Net

Neural Net

Perceptron

SVM

Deep (sparse/denoising) Autoencoder

AutoencoderNeural Net

Sparse Coding

ΣΠ

GMM

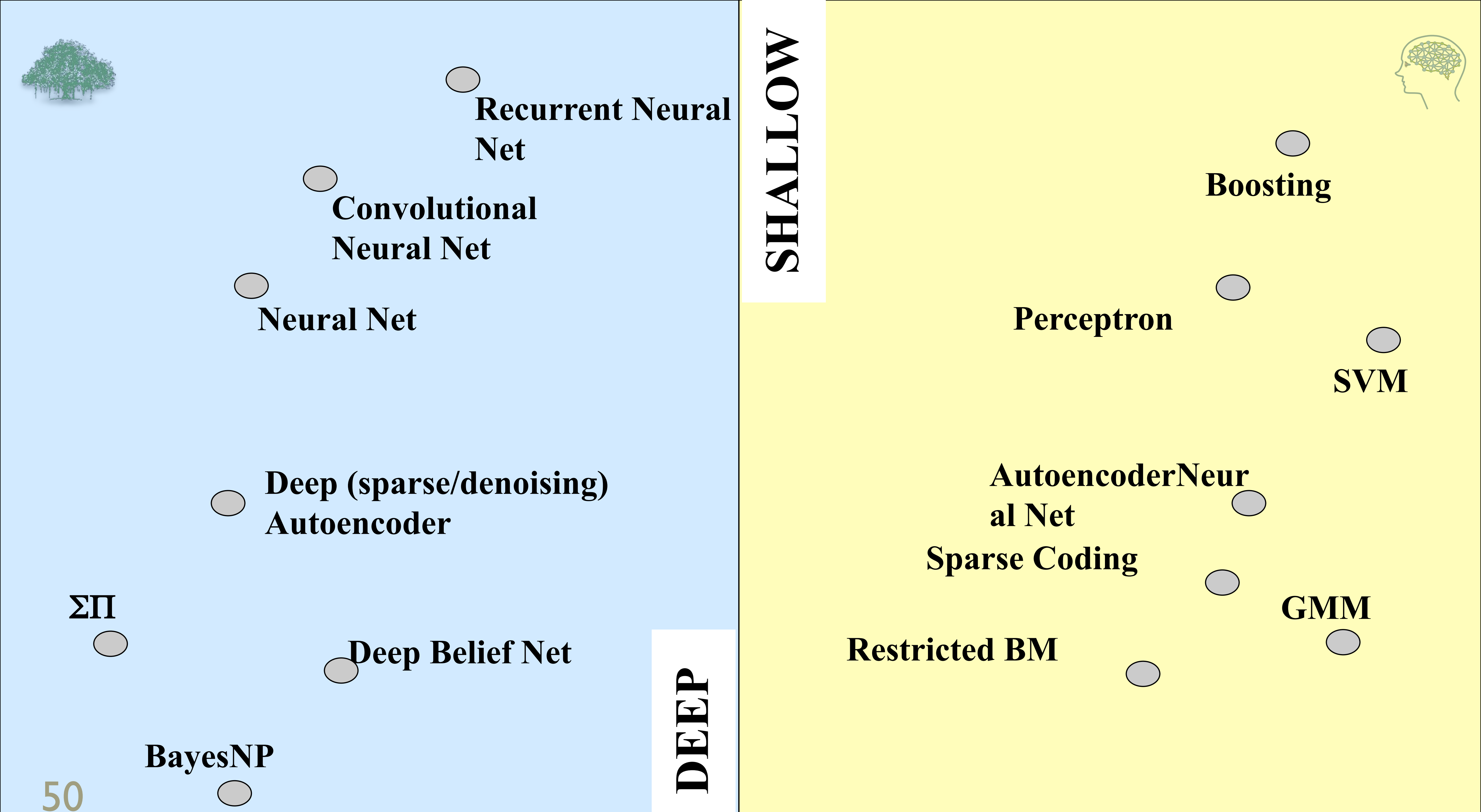
Deep Belief Net

Restricted BM

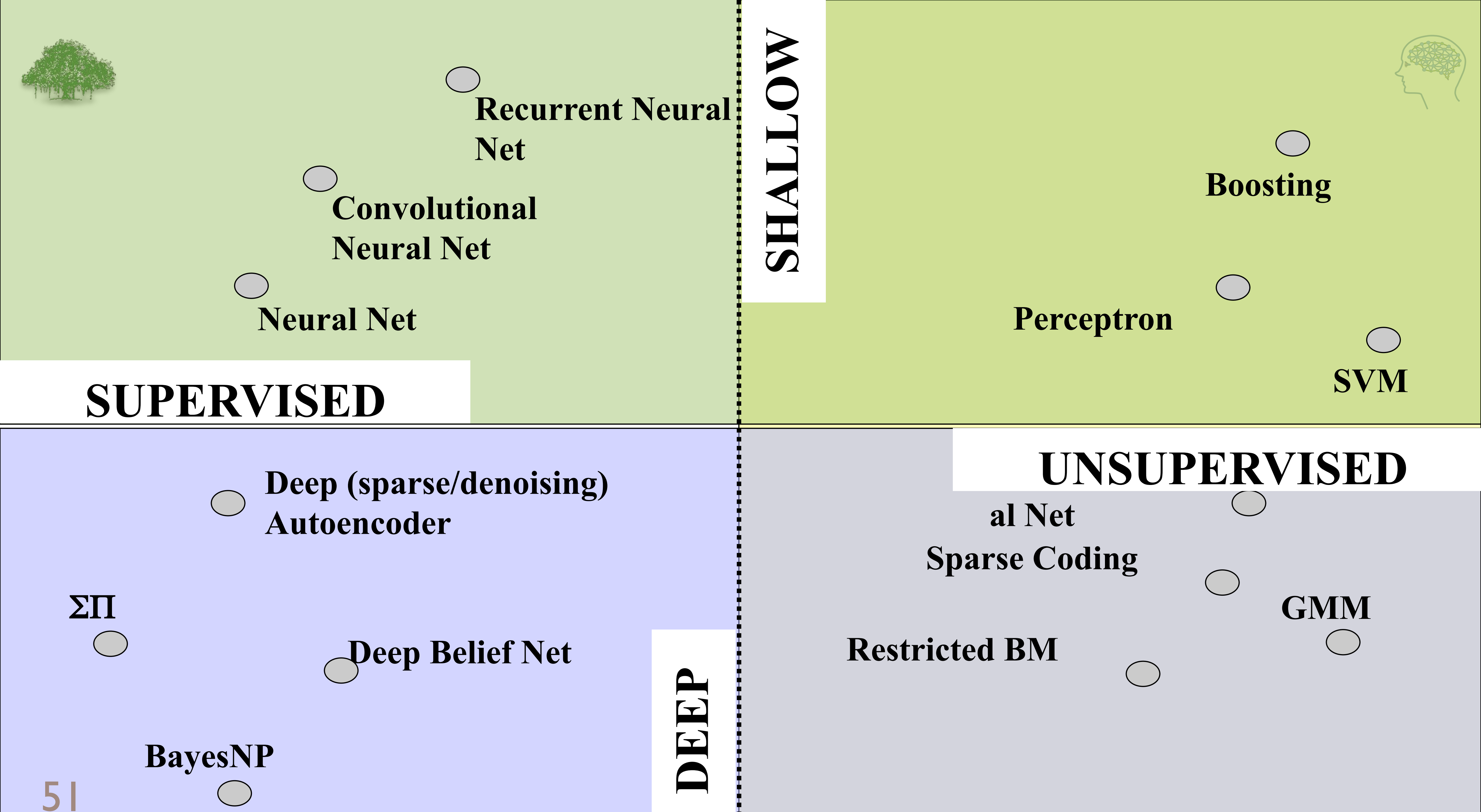
BayesNP

Disclaimer: showing only a subset of the known methods

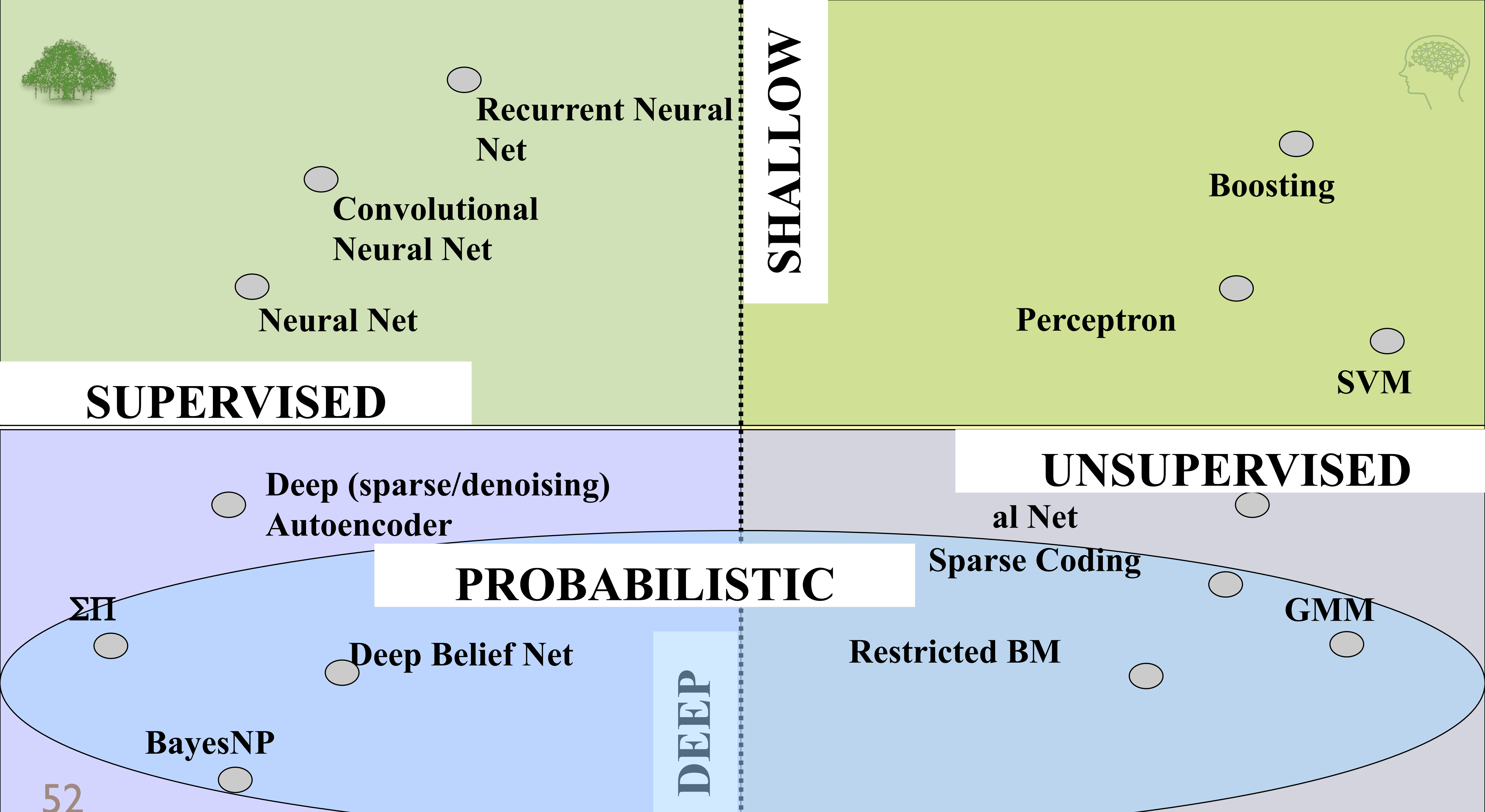
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

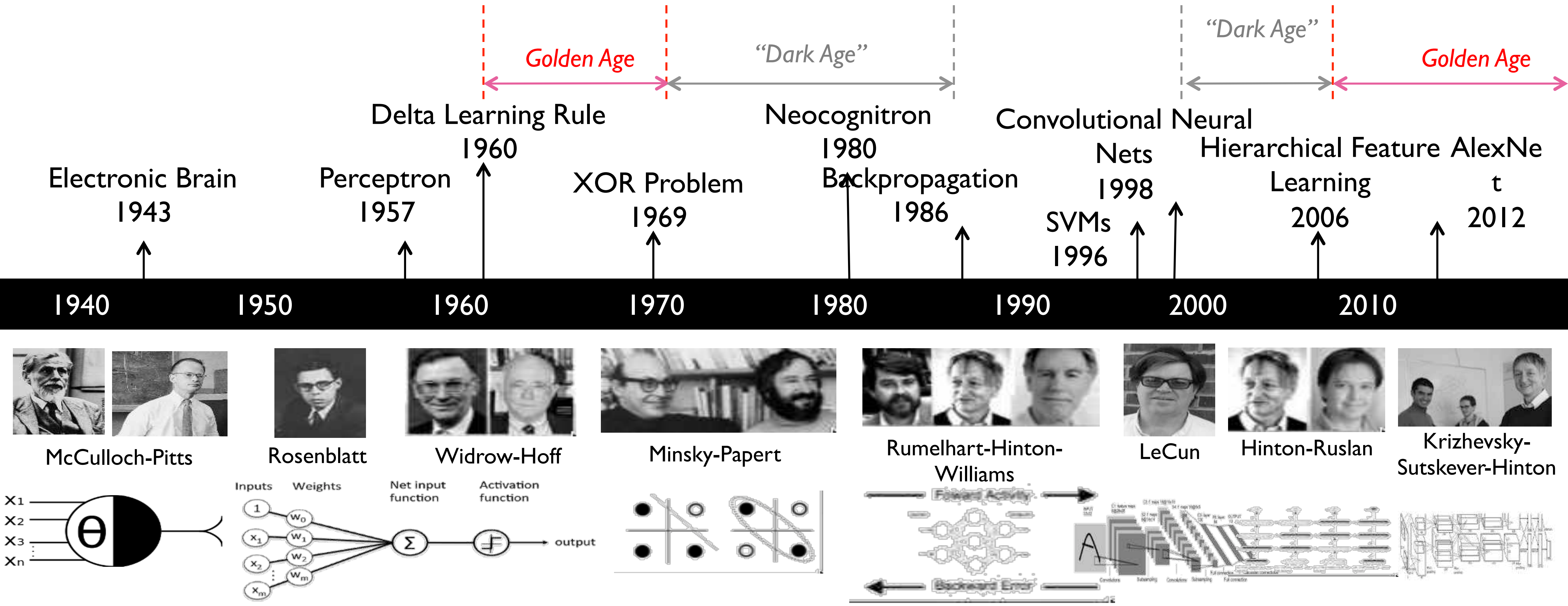


Slide Credit: Marc'Aurelio Ranzato, Yann LeCun



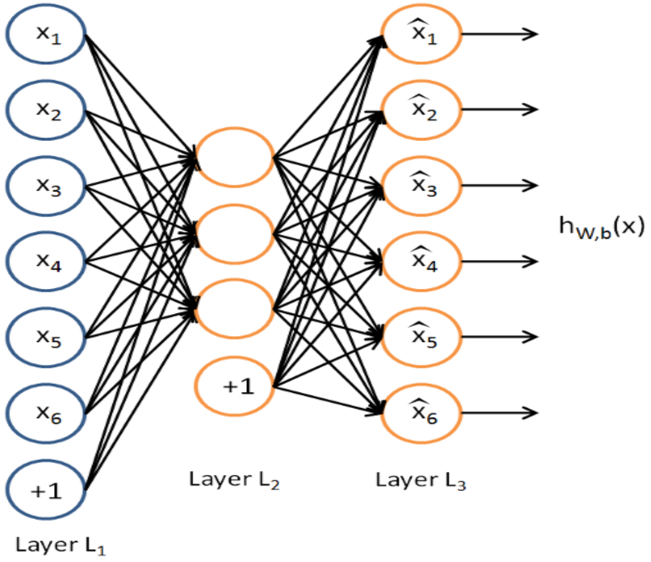
Slide Credit: Marc'Aurelio Ranzato,Yann LeCun

History of Deep Learning

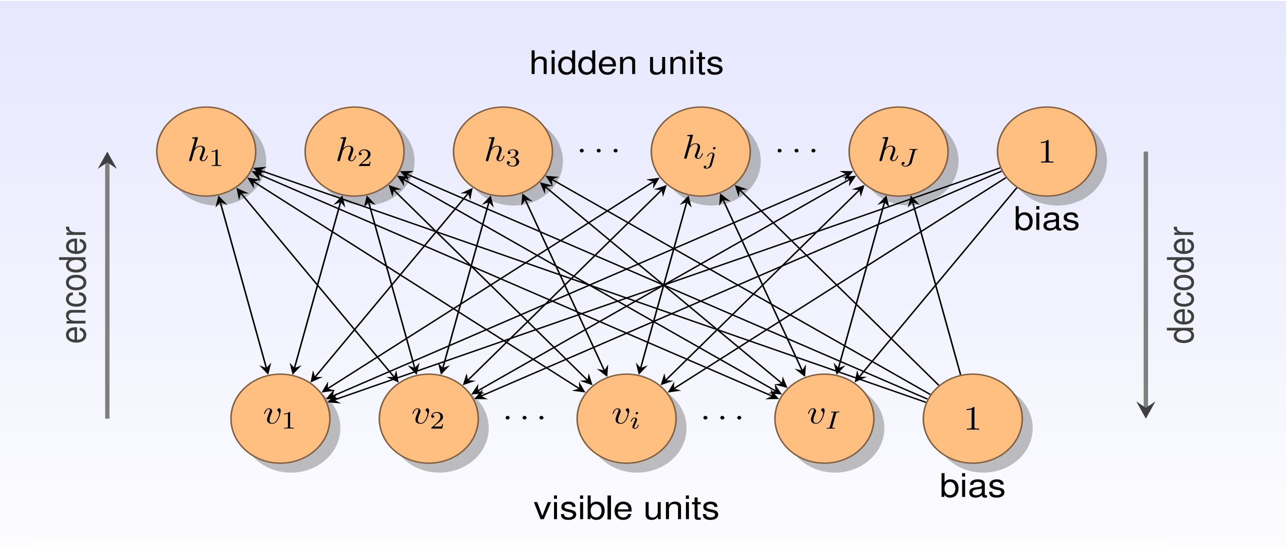


Popular DL Architectures

Auto Encoder



RBM



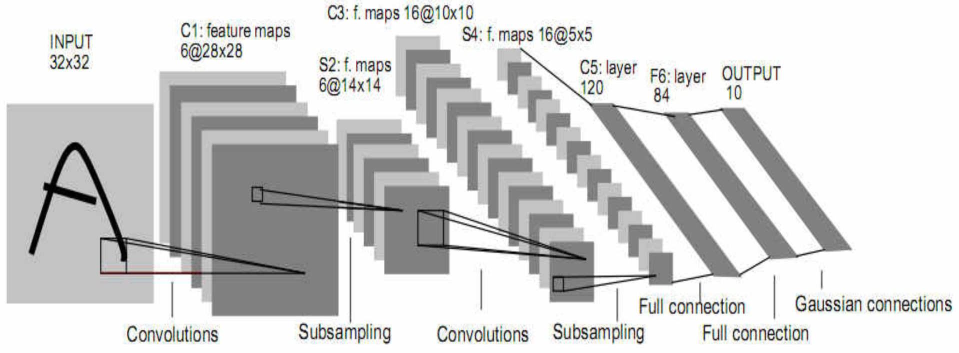
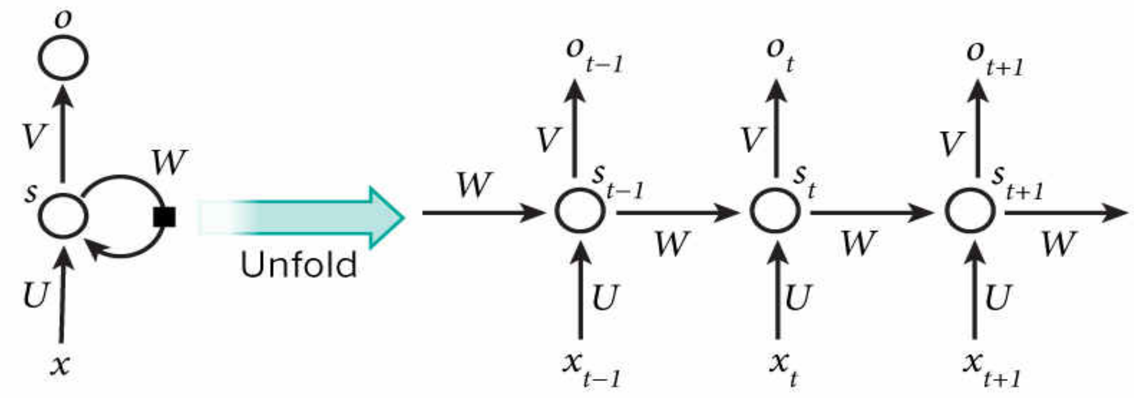


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

CNN



RNN

What is this big leap?

	LeNet(1989)	LeNet(1998)	AlexNet(2012)
Task	Digit	Digit	Objects
# Classes	10	10	1000
image size	16×16	28×28	$256 \times 256 \times 3$
# examples	7291	60,000	1.2 M
units	1256	8084	658,000
parameters	9760	60K	60 M
connections	65K	344K	652M
Operations	11 billion	412 billion	200 quadrillion

Regularization and Hyperparameter Engineering

Little Pieces that have made the Whole

Regularization

- DropOut, DropConnect, Batch Normalization, Data Augmentation, Noise in Data/Label/Gradient

Weight Initialization

- Xavier's initialization, He's initialization

Choosing Gradient Descent Parameters

- Adagrad, RMSProp, Adam, Momentum, Nesterov Momentum

Activation Functions

- ReLU, PReLU, Leaky ReLU, ELU

Loss Functions

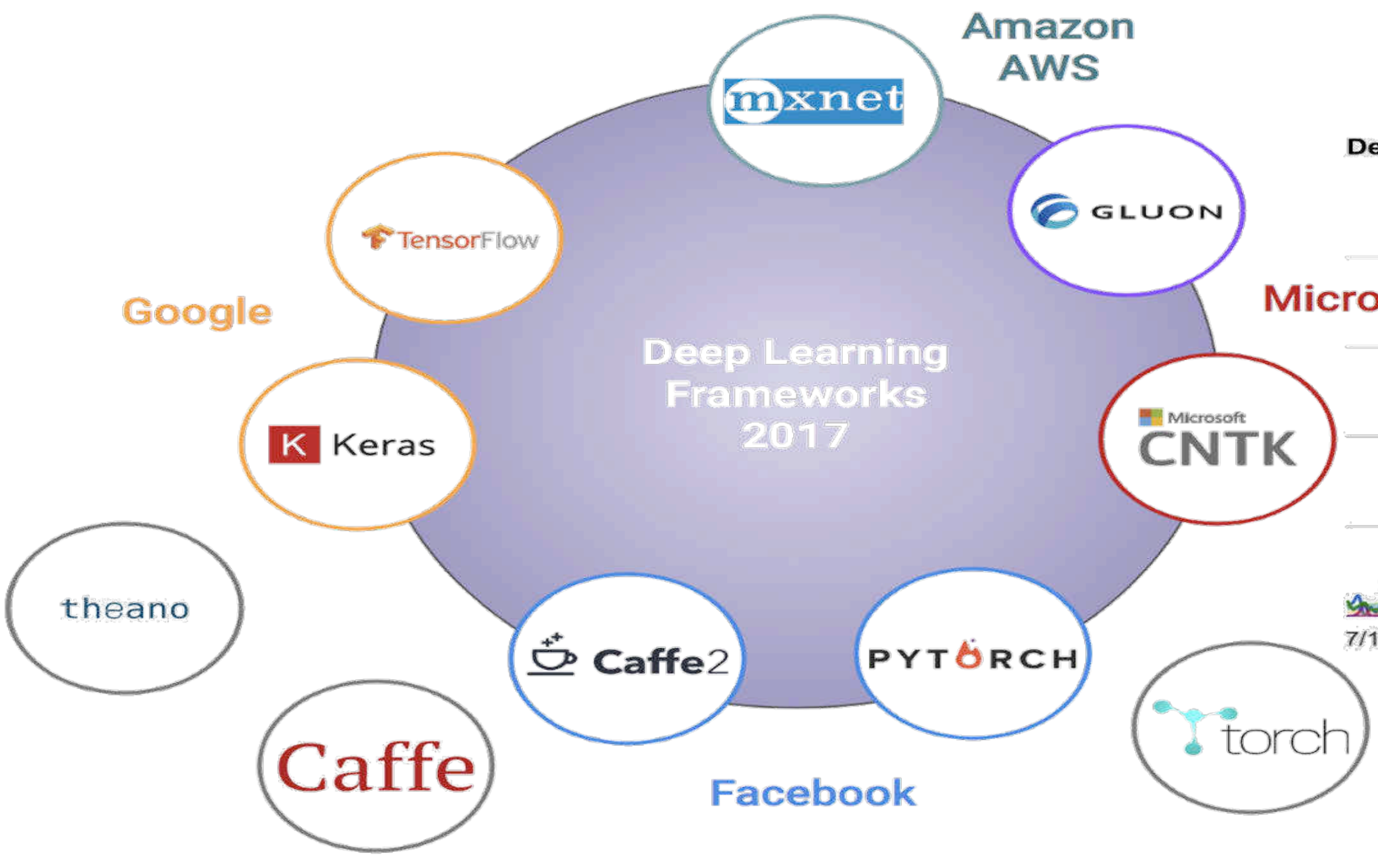
- Cross-Entropy, Embedding Loss, Mean-Squared Error, Absolute Error, KLDivergence, Max-Margin Loss

Deep Learning Libraries

Layer	Libraries
Machine Learning Libraries	•Torch •Caffe •Scikit-learn
Tensor	•Theano •TensorFlow •Numpy
Systems Programming 1	•CUDNN •CUBLAS
Systems Programming 2	•CUDA
OS	•NVIDIA Tesla drivers •NVIDIA GeForce drivers •Linux kernel
HW	•NVIDIA Tesla K40, K80 •NVIDIA GeForce GTX 950, 960, 970, 980 •CPU : Intel i7, AMD, ARM

Layer	Libraries
Machine Learning Libraries	• Caffee
Systems Programming 1	• MKL
Systems Programming 2	• OPENMP • MPI • AVX Instructions
OS	• Linux kernel
HW	• CPU : Intel i7, Xeon, Xeon Phi

Deep Learning Frameworks



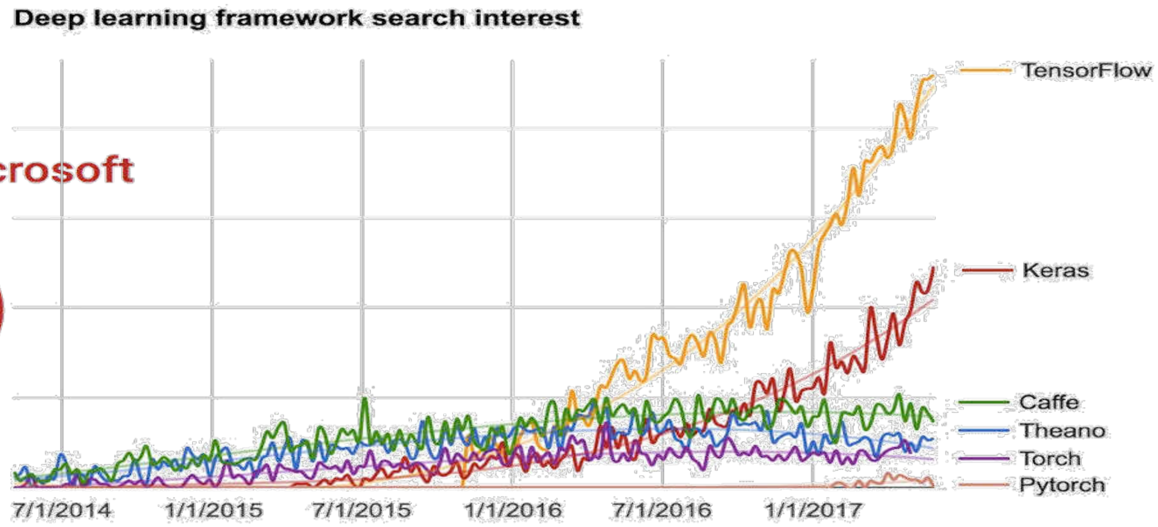


Image Credit: https://towardsdatascience.com

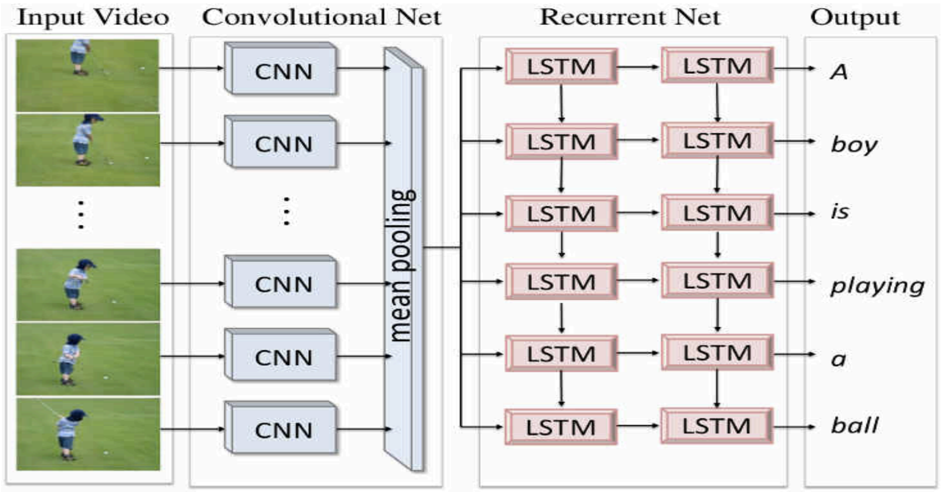
Image Credit: Francois Chollet

Many Newer Nets

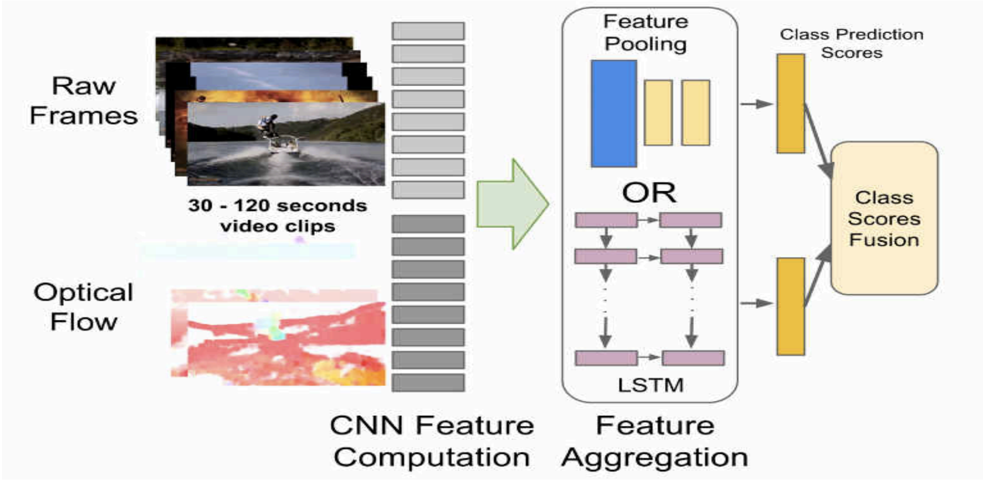
Neural Turing Machine (Graves et al., 2014)	DeepMask (Pinheiro et al., 2015)	SharpMask (Pinheiro et al., 2015)	Faster R-CNN (Ren et al, 2015)	FCN (Shelhamer , 2015)	SegNet (Badrinarayanan, 2015)
CRFasRNN (Zheng, 2015)	Ladder Network (Rasmus et al., 2015)	DenseNet (Huang et al, 2016)	DCGAN (Radford et al., 2016)	Pix2Pix (Isola et al., 2016)	Social LSTM (Alahi et al, 2016)
SketchNet (Zhang et al., 2016)	DeepFashion (Liu et al., 2016)	Pixel-RNN (Oord et al., 2016)	LocNet (Gidaris, 2016)	XNOR-Net (Rastegari et al., 2016)	UnrolledGAN (Metz et al. 2017)

And more to go …

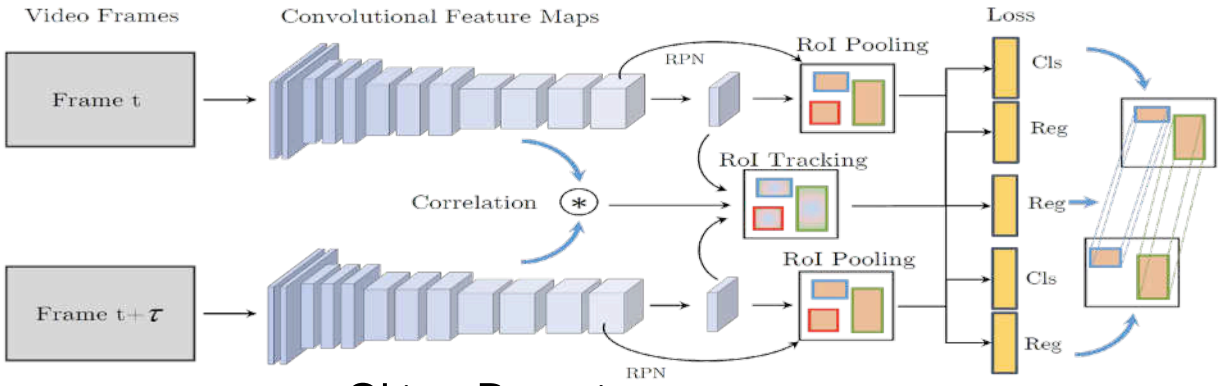
Hybrid Architectures



Video Captioning



Video Classification

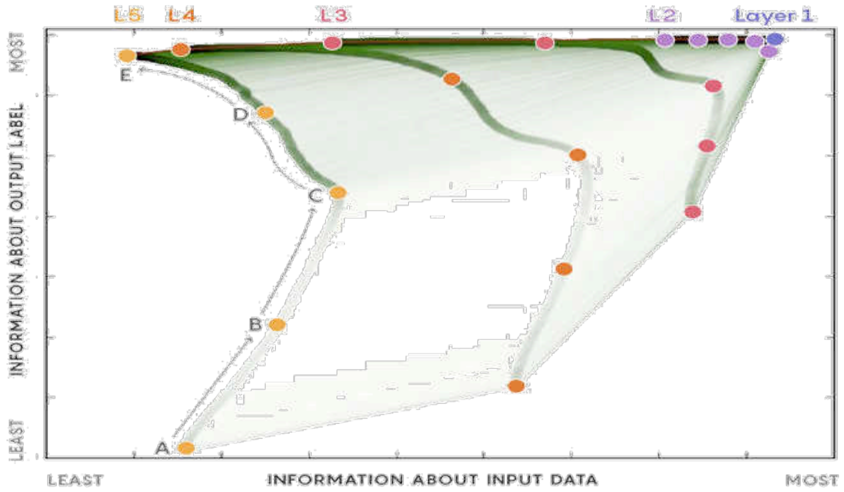


Object Detection

- End-to-end backprop on hybrid architectures
- Mix-and-match plug-and-play modules

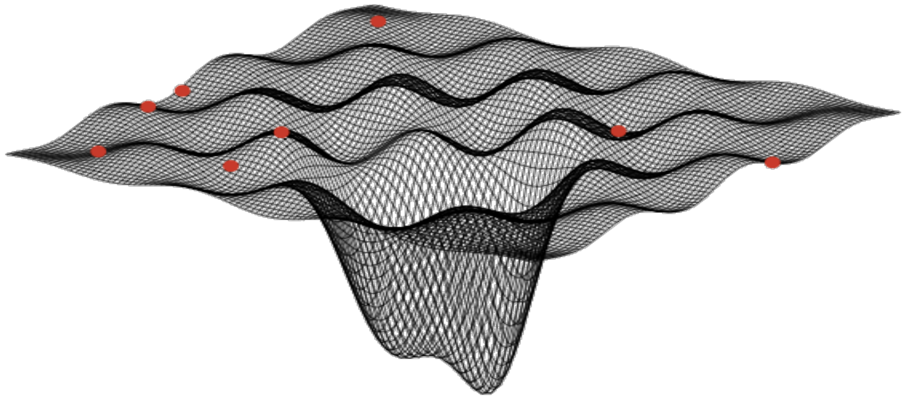
Theory and Optimization Methods

Theory of Deep Learning



- Information Bottleneck Principle, arXiv 2017
- Generalization in Deep Learning, arXiv 2017
- Random Matrix Theory, ICML 2017

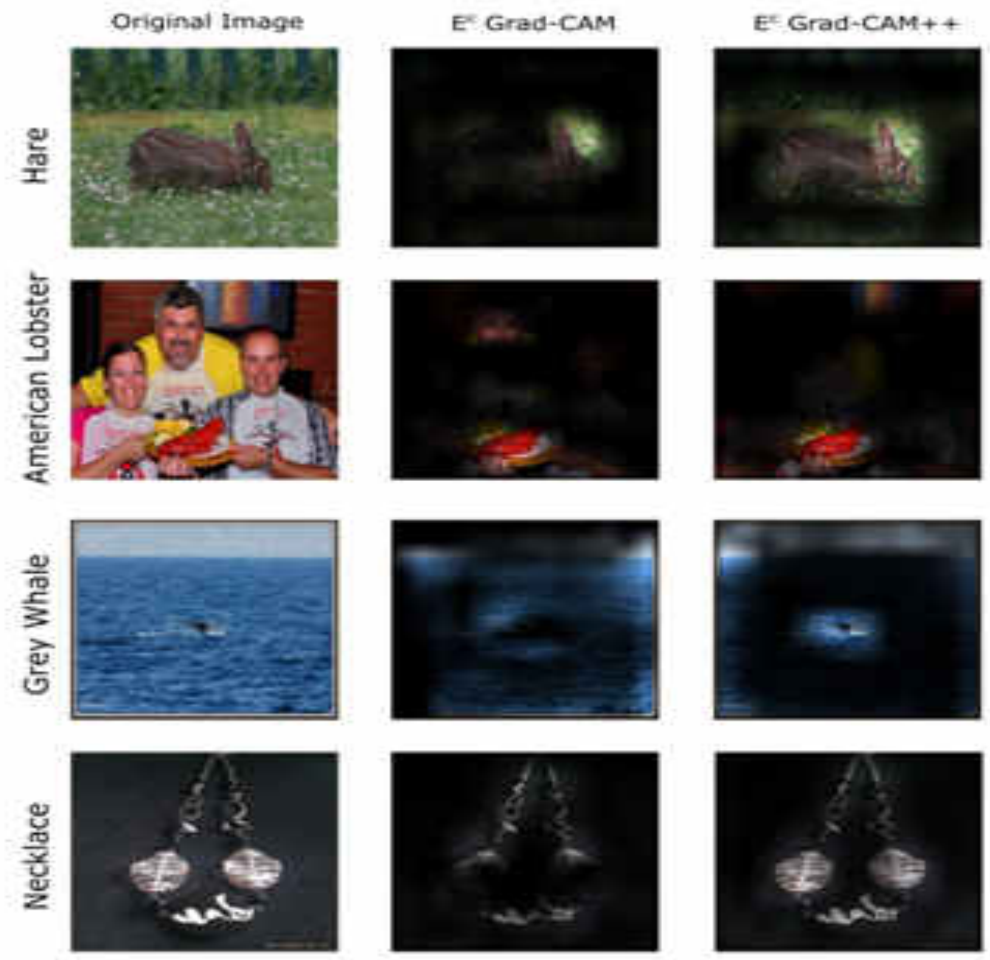
Understanding Error Surfaces



- Deep Learning without Poor Local Minima, NIPS 2016
- How to Escape Saddle Points Efficiently, arXiv 2017
- Sharp Minima can Generalize for Deep Nets, ICML 2017

Interpretability and Explainability

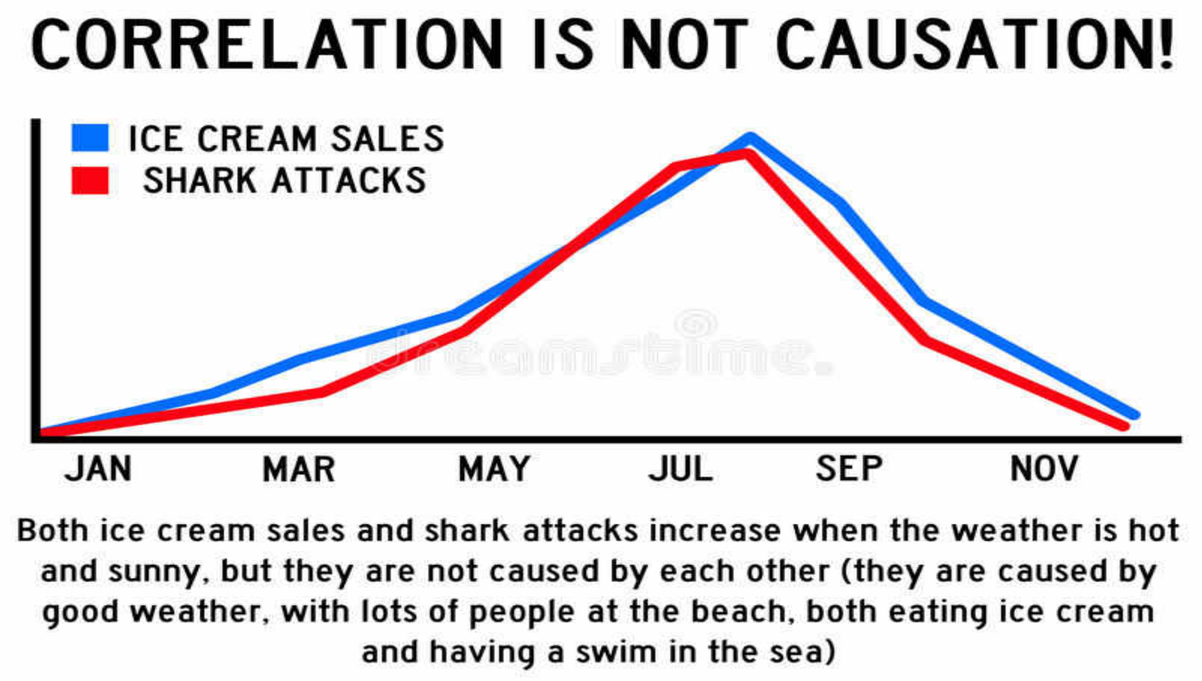
- Why deep learning models work?
  - *Theory of deep learning*
- How deep learning models work?
  - *Visualizing and Understanding CNNs, ECCV 2014*
  - *CAM, Grad-CAM, Grad-CAM++*
  - *Visualizing and Understanding RNNs, arXiv 2015*
- Long way to go!

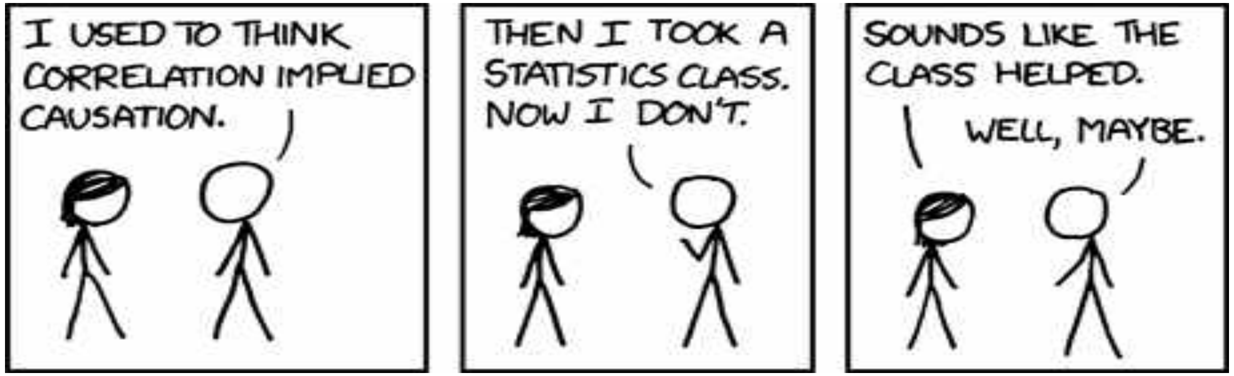


Need for Causal Inference

Causality vs Correlation

Deep learning models correlation, not causality





- Some recent efforts
  - *Discovering Causal Signals in Images, CVPR 2017*
- Long way to go!

Image Credit: http://www.statisticshowto.com/causation/

Robustness and Consistency

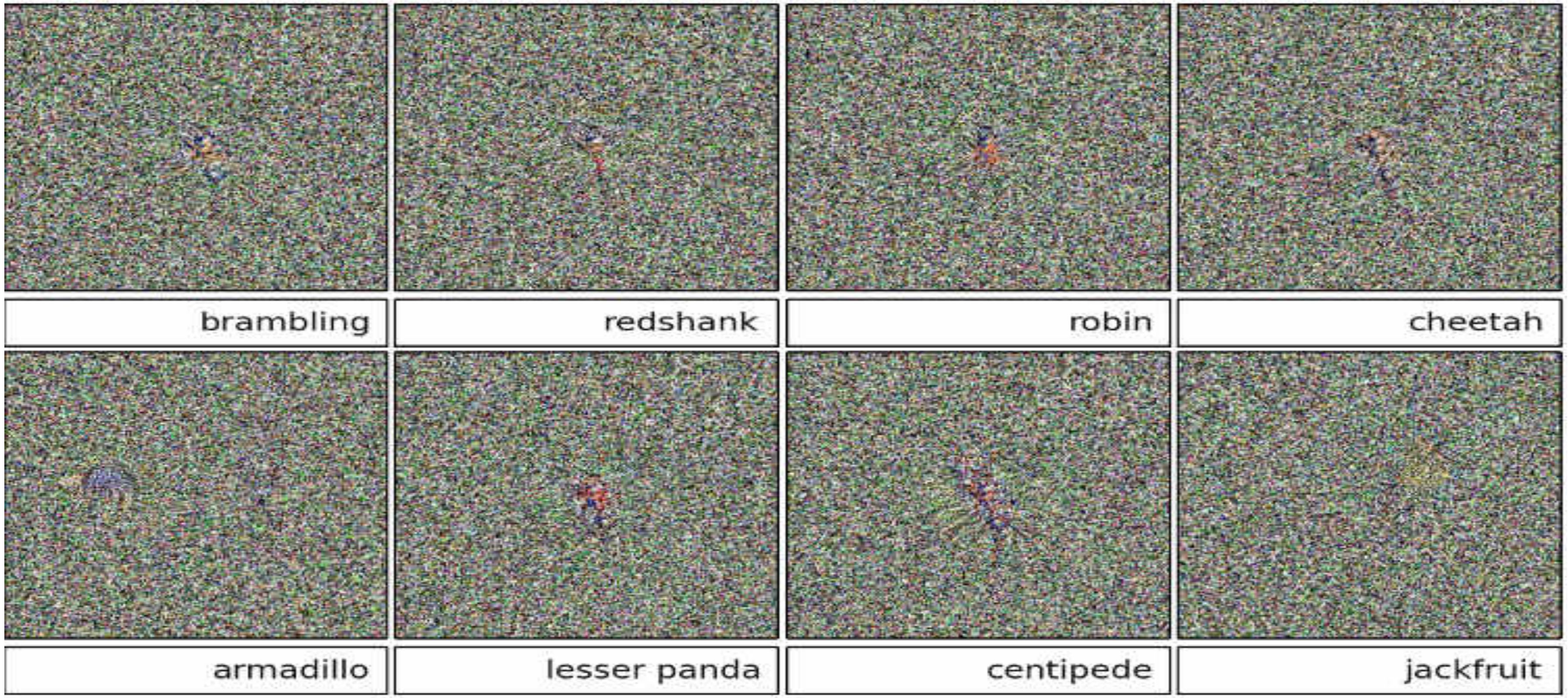
Neural networks are easily fooled, CVPR 2015

Classified with ~ 99% confidence

http://www.evolvingai.org/fooling

Robustness and Consistency

Neural networks are easily fooled, CVPR 2015



Classified with ~ 99% confidence

http://www.evolvingai.org/fooling

Our Limited Scope

- Perception
  - Understanding/recognizing Text, Speech, Image
- Tasks
  - Product rating
  - Sentiment analysis
  - Spam filter
  - Product recommendations
  - Financial time series prediction
  - Control Systems
  - Object Recognition
  - Etc.
- Applications
  - Chatbots/Dialogue Systems

Our Limited Scope

- Classification
- Regression
- Clustering
- Visualization
- Ensembling
- Feature Selection

- KNN,
- Decision Trees
- Random Forests
- Linear Classifiers
- K Means
- MLP
- SVMs, Kernels
- Deep Neural Networks
  - CNNs, RNNs, AEs

Our Limited Scope

- Word2Vec
- Bag of Words
- PCA and Eigen Face
- MFCC Features
- ISOMAP/LLE
- Perceptron Learning
- Gradient Descent
- Back Propagation
- CART
- Boosting
- Model Compression

Our Limited Scope

- Experimental Skills
  - Training, Validation, Cross validation, Hyperparameter Selection
- Harvesting Supervision
  - Google/Internet, User engagements
- Training and Testing
  - Sampling, Losses, Parameters and Tricks
- Avoiding Overfitting
  - Regularization, Jittering, Early Stop, etc.
- Evaluation Metrics
  - Accuracy, Precision, Recall, Detection, False Alarms etc.
- Problem Solving
  - Abstracting the task. Connecting to Data. Map to Algorithms/Tasks

Thanks. Questions? Comments?