Centre for
Data Analytics

Insight

Machine Learning Overview: How did we get to here?

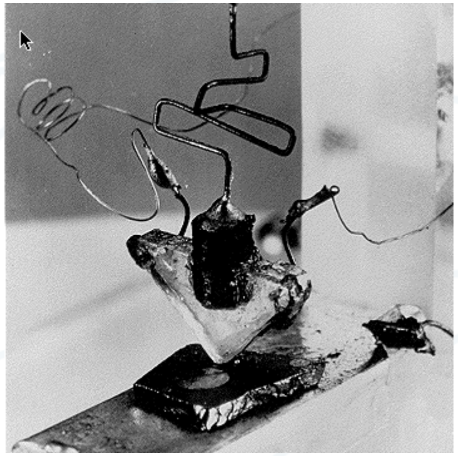
Prof. Alan F. Smeaton
Insight @ DCU

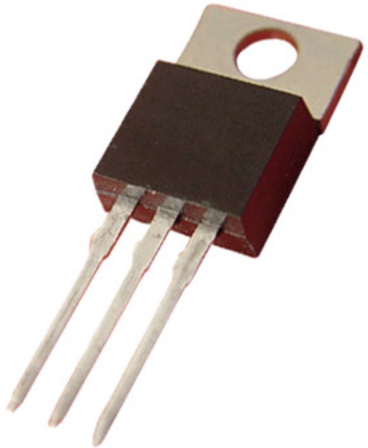
WW II, the first modern computers

- 1943 Mauchly and Eckert prepare a proposal for the US Army to build an Electronic Numerical Integrator
  – calculate a trajectory in 1 second
- 1944 early thoughts on stored program computers by members of the ENIAC team
- Spring 1945 ENIAC working well
- Alan Turing, recruited as a codebreaker, had the Automatic Computing Engine (ACE) basic design by Spring, 1946
- Pilot ACE working, May 10, 1950
- Full version of ACE at NPL, 1959

The Late 40s

- 1947 first computer bug when Grace Murray Hopper found bug killed in jaws of electromechanical relay on Mark II computer at Harvard
- 1947 –Shockley et al. invent the transistor
- 1949 –Wilkes at Cambridge developed EDSAC, the first large scale, fully operational stored program computer
- 1951 – Remington-Rand sold Univac 1 to US government for $1,000,000





The 1950s

- IBM produces series of computers with Jean Amdahl as chief architect
- Memory upgraded to magnetic core memory, magnetic tapes and disks with movable read/write heads
- Fortran was introduced in 1957, still around, used in large-scale legacy data processing
- Integrated Circuit was invented in 1958

The 1960s

- Computers begin to have business uses
- 1965 the IBM/360 Mainframe introduced using integrated circuits
- In 1965 DEC introduced PDP-8, first minicomputer
  - Bill Gates' school bought one and he used it daily
- In 1969 work began on ARPAnet (the predecessor of the internet)

The Early 1970s

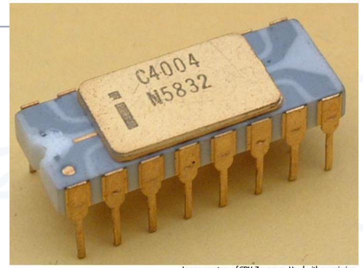


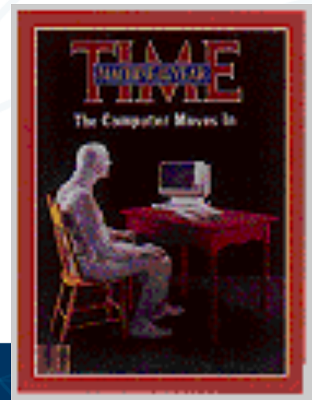
Image courtesy of CPU-Zone.com. Used with permission.

- In 1971 Intel releases the 4004, the first microprocessor
  - Also in 1971, the first floppy disk is introduced
- In 1973, Xerox invents Ethernet, the common networking standard
- In 1975, the first PC, MITS Altair 8800 (no keyboard, no display, no auxiliary storage)

- Bill Gates and Paul Allen (high school friends) wrote software for the Altair that allowed users to write their first program … became their first product

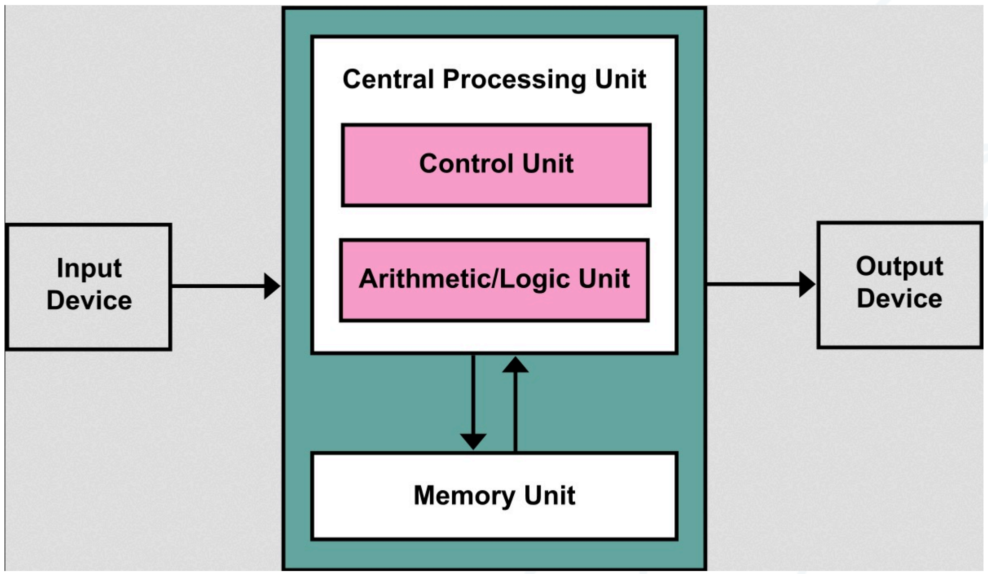
The 1980s

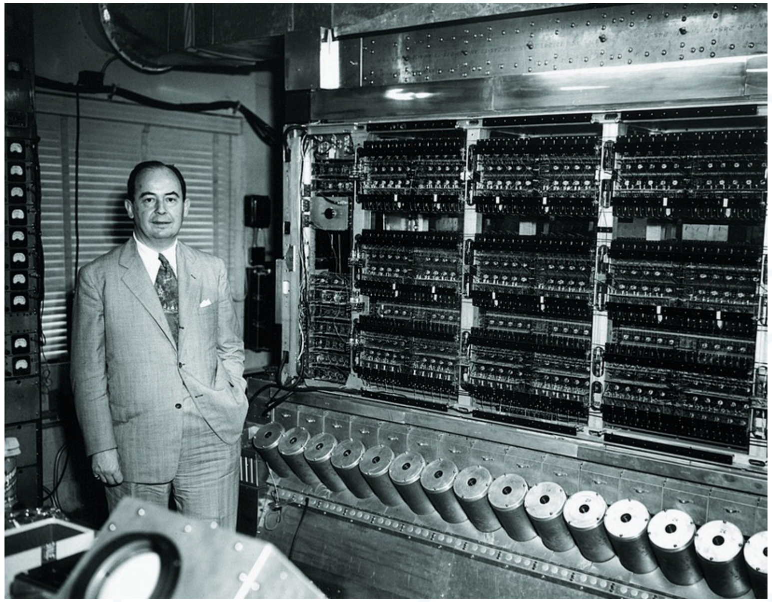
- In 1981 IBM release the IBM PC based on an Intel 8088 chip
- In the same year, Microsoft release an operating system called DOS
- In 1982, Time magazine announces the computer as the 1982 Man of the Year
- Portable computers released in 1982
- Also in 1982, the first affordable home computer, the Commodore 64
- Apple Macintosh released in 1984



What's in common ?

- John von Neumann (1952) proposed a model for computers … that's still with us





Could computers do AI ?

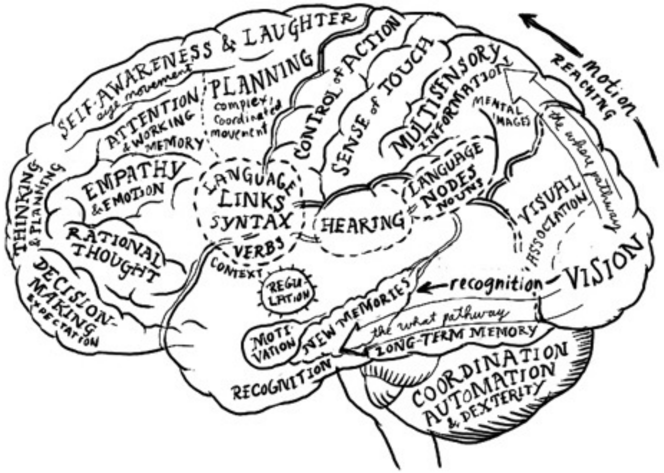
- Since the early days, always the question of could computers do artificial intelligence ?

The Human Brain

- 1.5kg, or 2% of our body weight, made of 86B neurons (grey matter) connected by trillions of connections (synapses)
- Responsible for executive functions like autonomic (heart, breathing, digestion, etc) and voluntary, in addition to executive functions like self-control, planning, reasoning, and abstract thought.
- This architecture of huge number of simple connected processors, is good for solving very complex problems, like vision, and learning

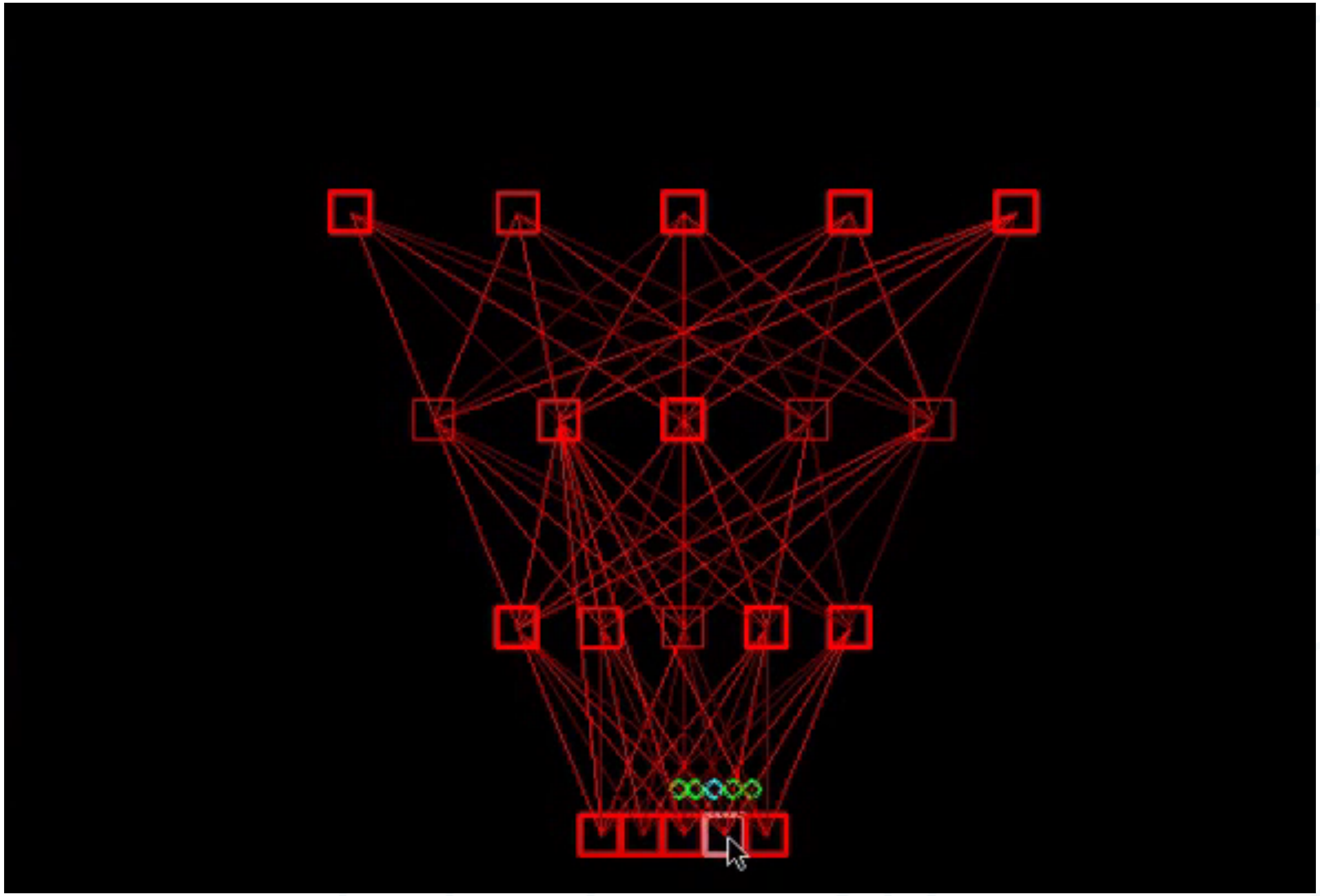
Human Memory

- The brain has been coarsely mapped



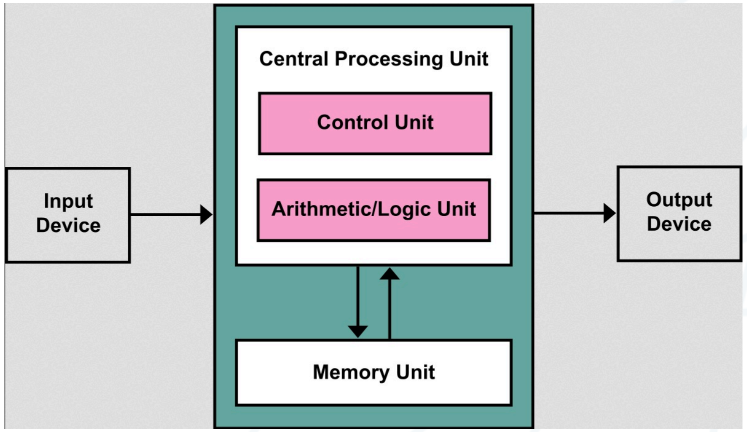
- The architecture is of simple but massively parallel processing, a form of perceptron, highly connected graph of nodes and links with signal-passing

How neurons work …



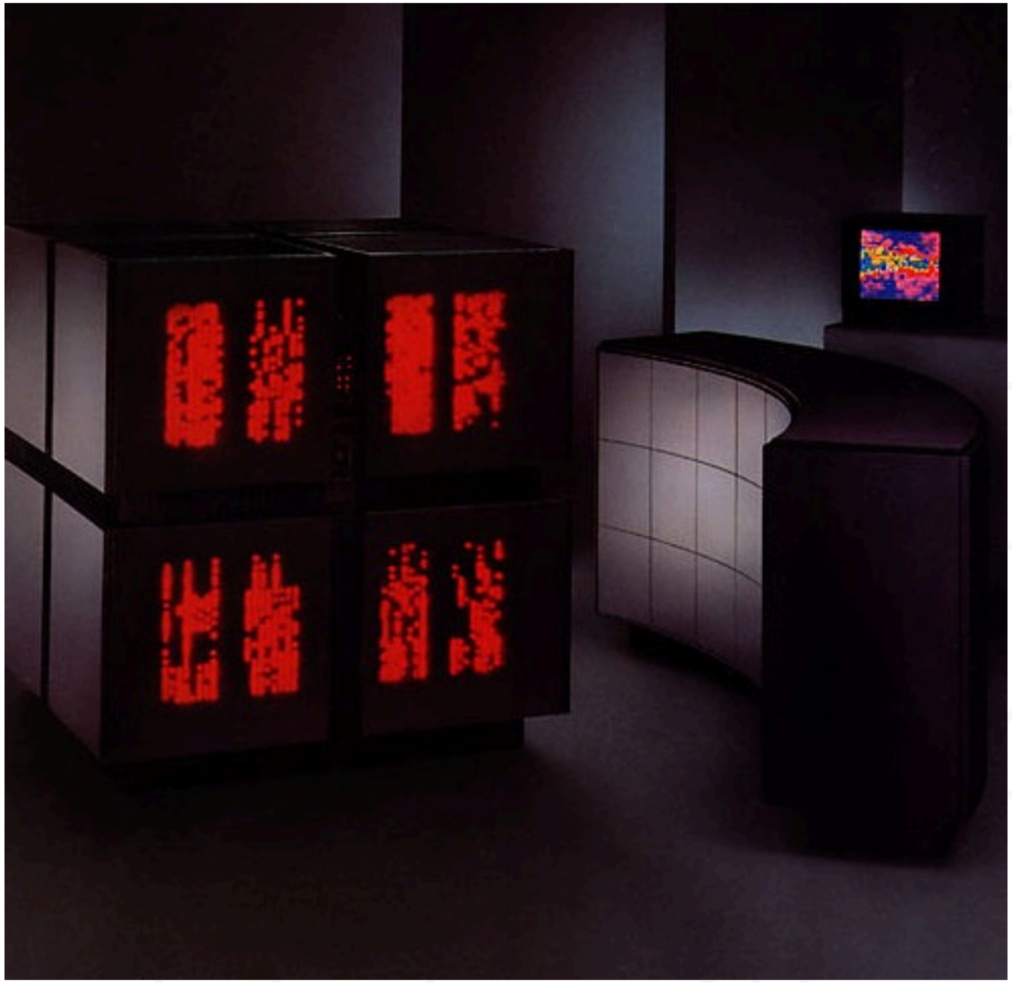
What's a BCI?

- So how do we implement Artificial Intelligence, emulating the human brain, on the vN architecture ?



- Can it be implemented directly in hardware ?

Late 1980s, Connection Machine

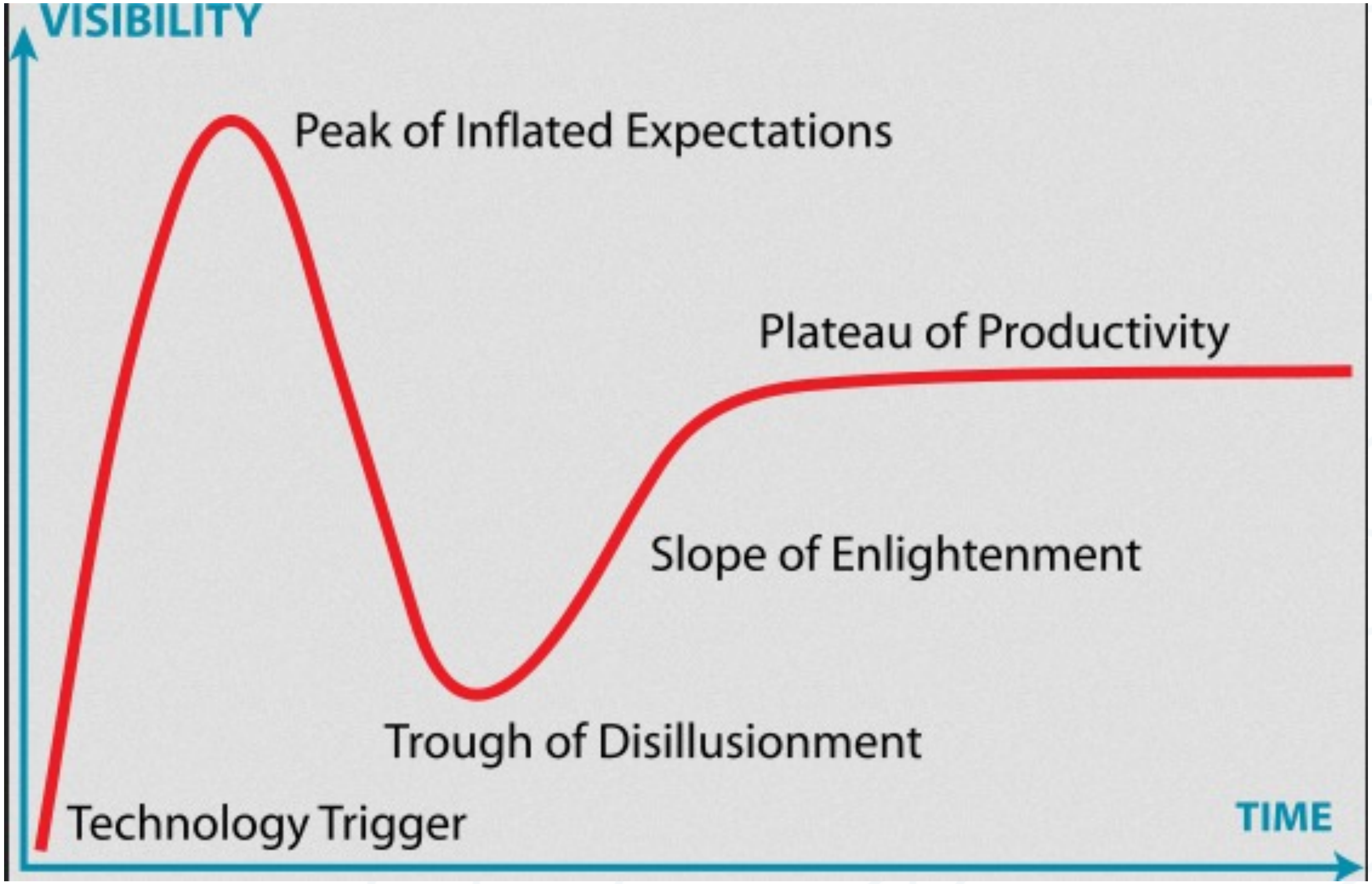


- Problem with CM was it only had 000’s nodes with connections, not enough to simulate “intelligence”
- So it was used for simple, massive parallelism apps.

Raised expectations … AI "Winter"

- Too ambitious / too big claims:
  - "The vodka is good, but the meat is rotten."
    "The spirit is willing, but the flesh is weak."
    (allusion to Mark 14:38)
  - 1966, negative report by an advisory committee, government funding of automatic translation cancelled.
- Limited knowledge of the outside world:
  - Restricted to micro-worlds (e.g. Blocks World)
  - Restricted to pattern-matching (e.g. ELIZA)
- Inherent limitations of computability:
  - Intractability, combinatorial explosion (to be discussed next week).
  - Undecidability

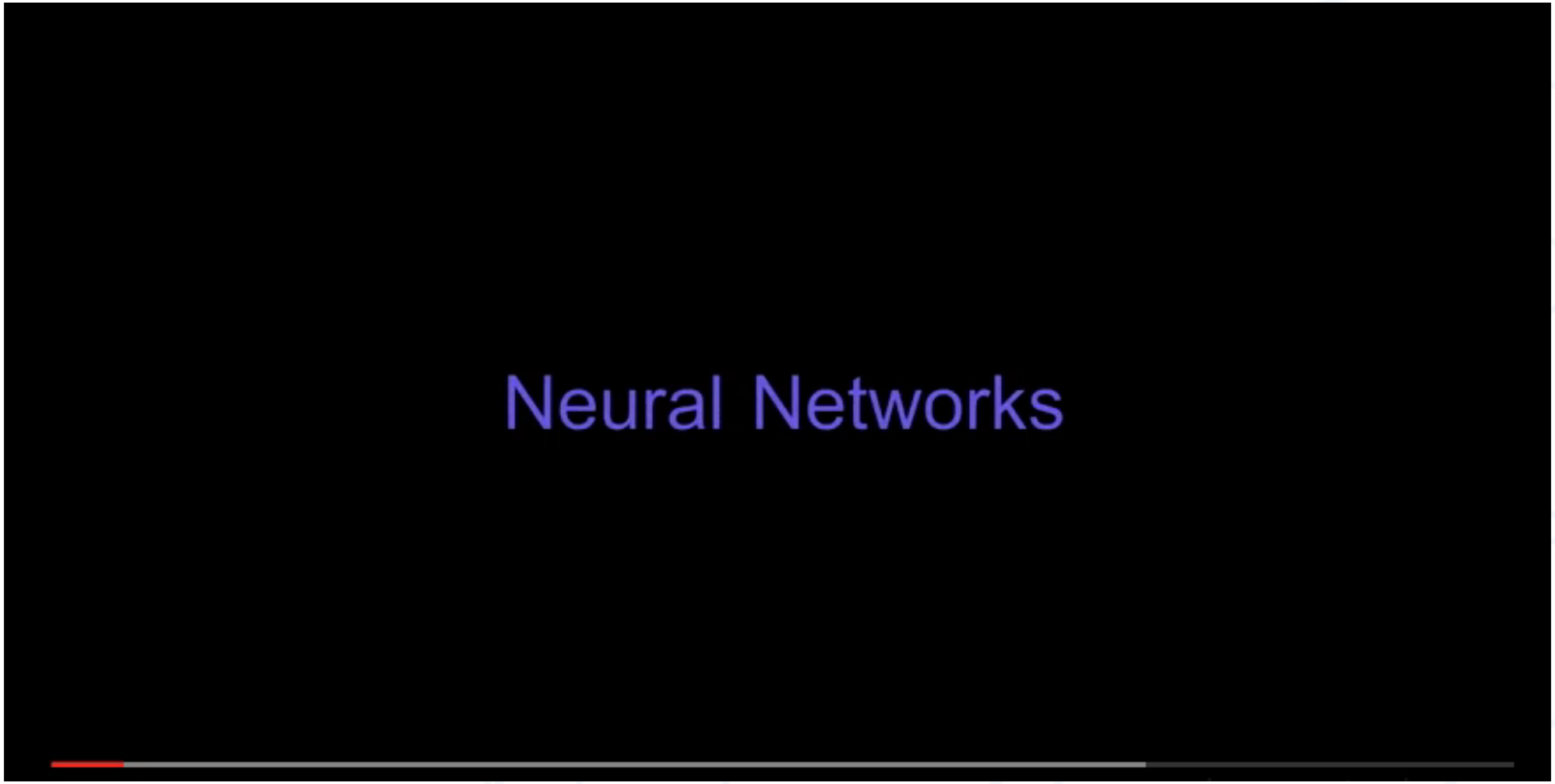
Classic Gartner Curve !



Neural Network research continued

- Research into neural networks – computational implementation of the brain's structure – continued but at a slow burn and a bit of an AI backwater !
- Meanwhile, hand-crafted, rules-driven AI research continued through 70s, 80s, 90s, even 00s
- AI applications were
  - Speech
  - Machine Translation
  - Computer Vision
  - Expert Systems
  - Etc.

Here's how Neural Networks developed ...



Insight

Meanwhile in a galaxy far far away …

AI and Machine Learning

- Machine learning evolved as an AI tool
- With mathematics and statistics input rather than any neural network connection
- Slow evolution of Machine Learning over decades
- Nourished by increasing availability of huge data volumes from … internet searching … social media …online transactions … etc.

- One application which pushed this was computer vision

What does this mean?

Machine Learning of Semantic Concepts

- Use Machine Learning to train a classifier to identify an object
  - Decision tree learning
  - Random forests
  - Genetic programming
  - **Support vector machines**
- Given some input data (e.g. SIFT features or colours or textures or all 3)
- Given a lot of + and - examples
- Let the computer figure out how to classify new examples into + or – clusters … that's the modus for machine learning

Concept Detection

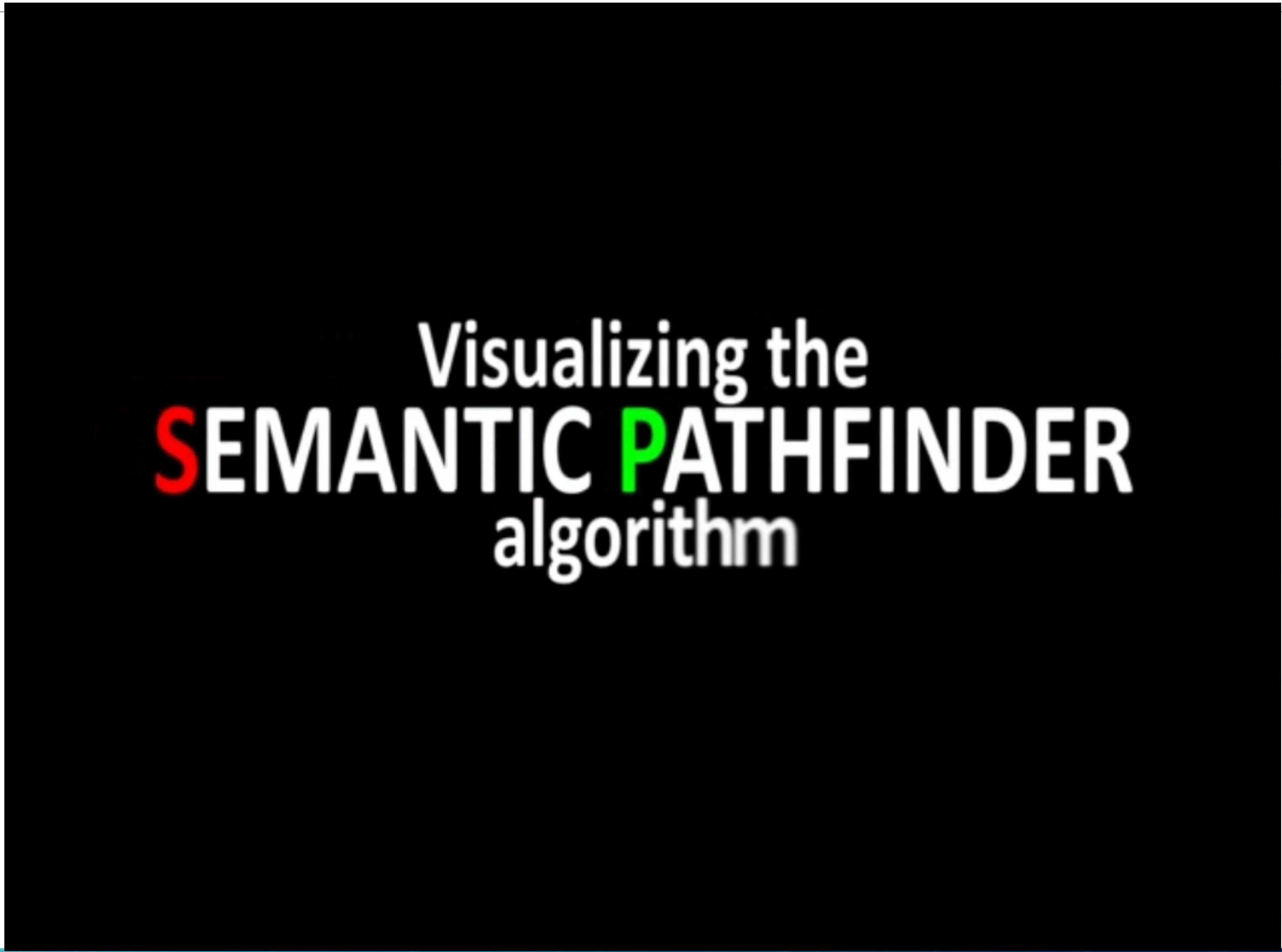
- We should develop a suite of retrieval techniques that can be used for such features …

	FEATURE DETETCION
0.2	Indoor
0.8	Outdoor
0.7	CityScape
0.3	Landscape
0.1	People
0.0	Face
0.8	Sky
0.2	Vegetation
0.7	Building

How does "standard" ML work ?

- Lots of + and – examples as training data
- Plot each data point onto an n-dimensional space
- Learn a boundary function which differentiates
- Train and test, refine, then deploy

Visualisation



So how well did it work ?

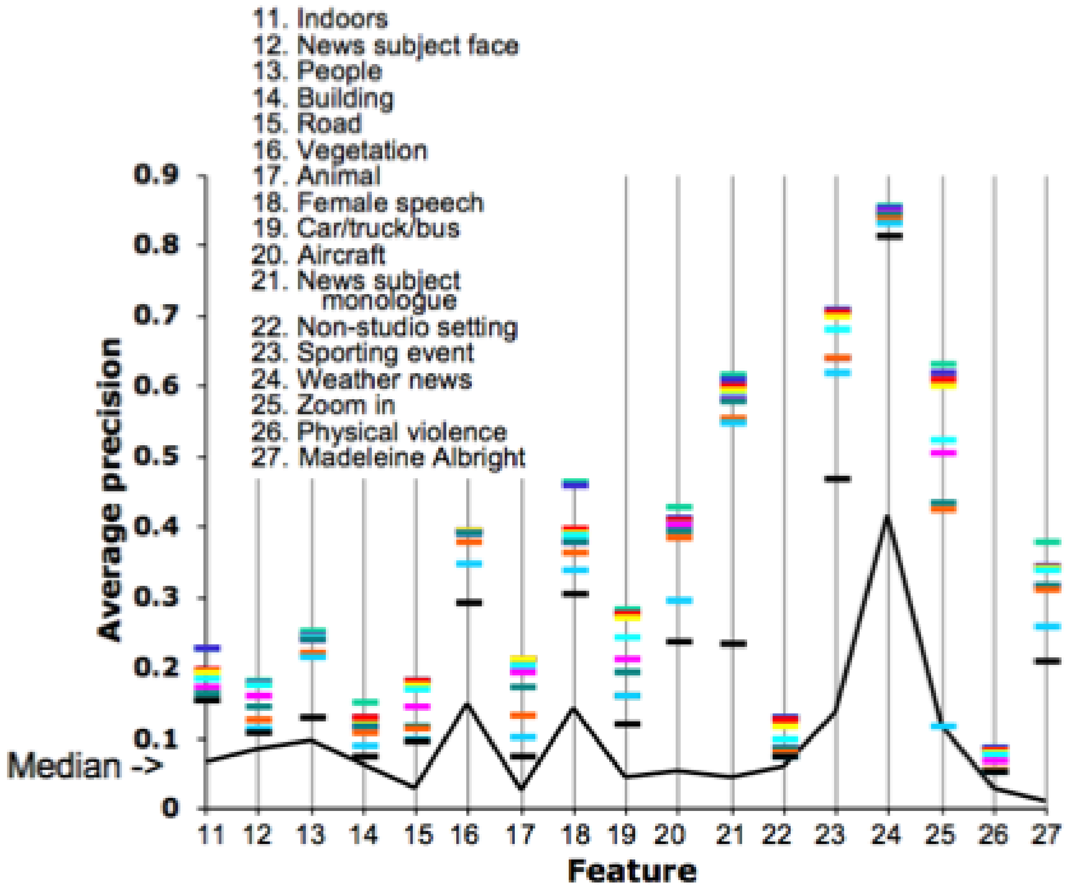


Fig. 2. Performance of Top-10 high-level Feature Detections per Feature in TRECVid 2003

So how well did it work ?

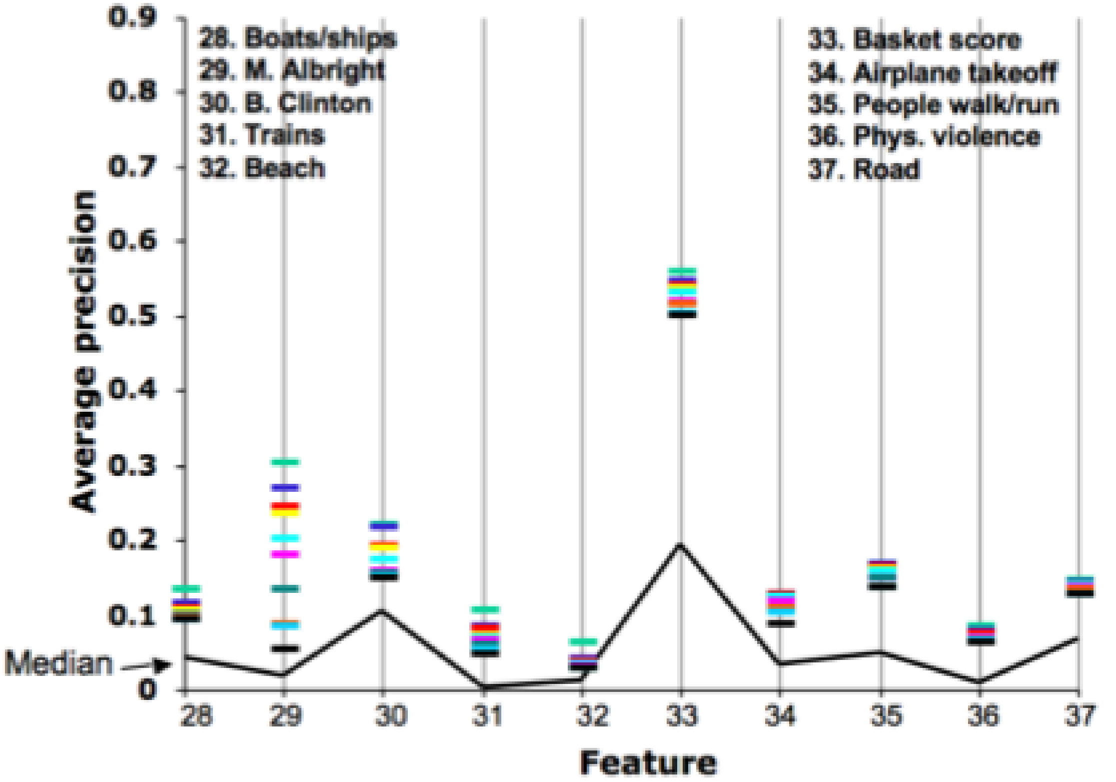


Fig. 3. Performance of Top-10 high-level Feature Detections per Feature in TRECVid 2004

So how well did it work ?

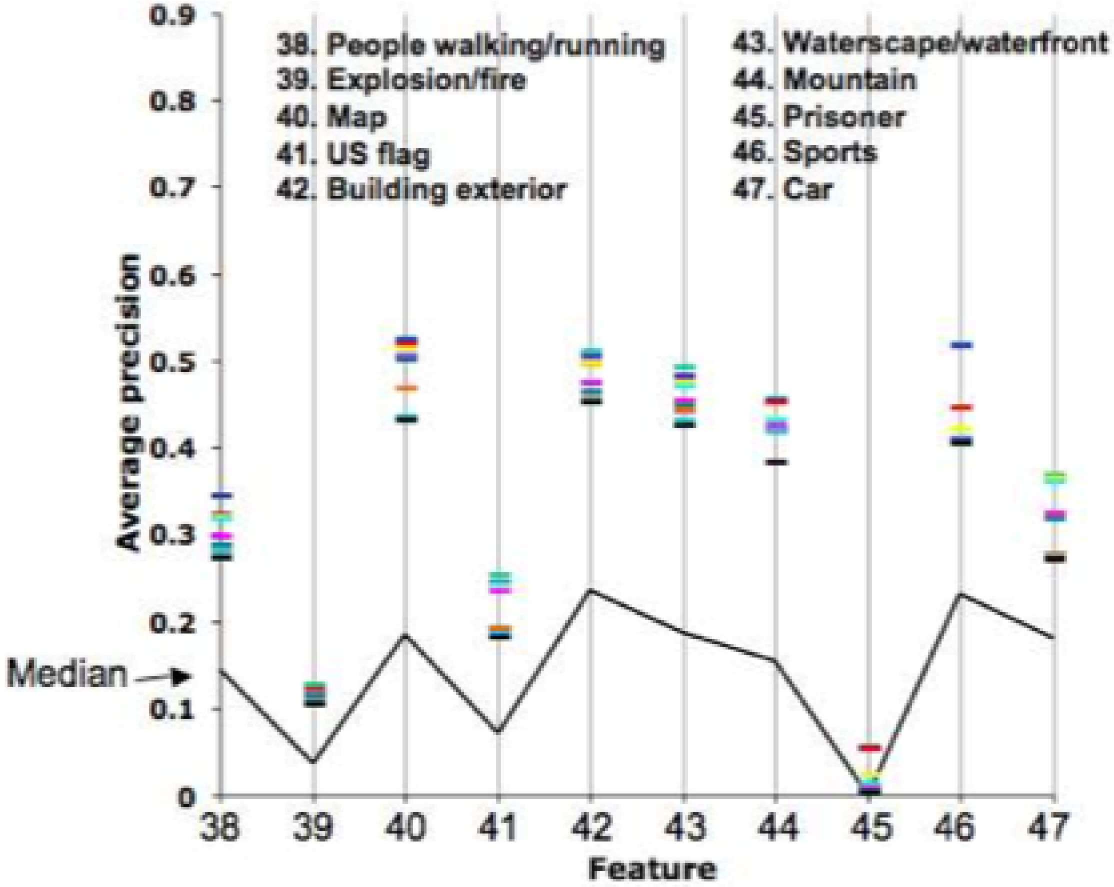
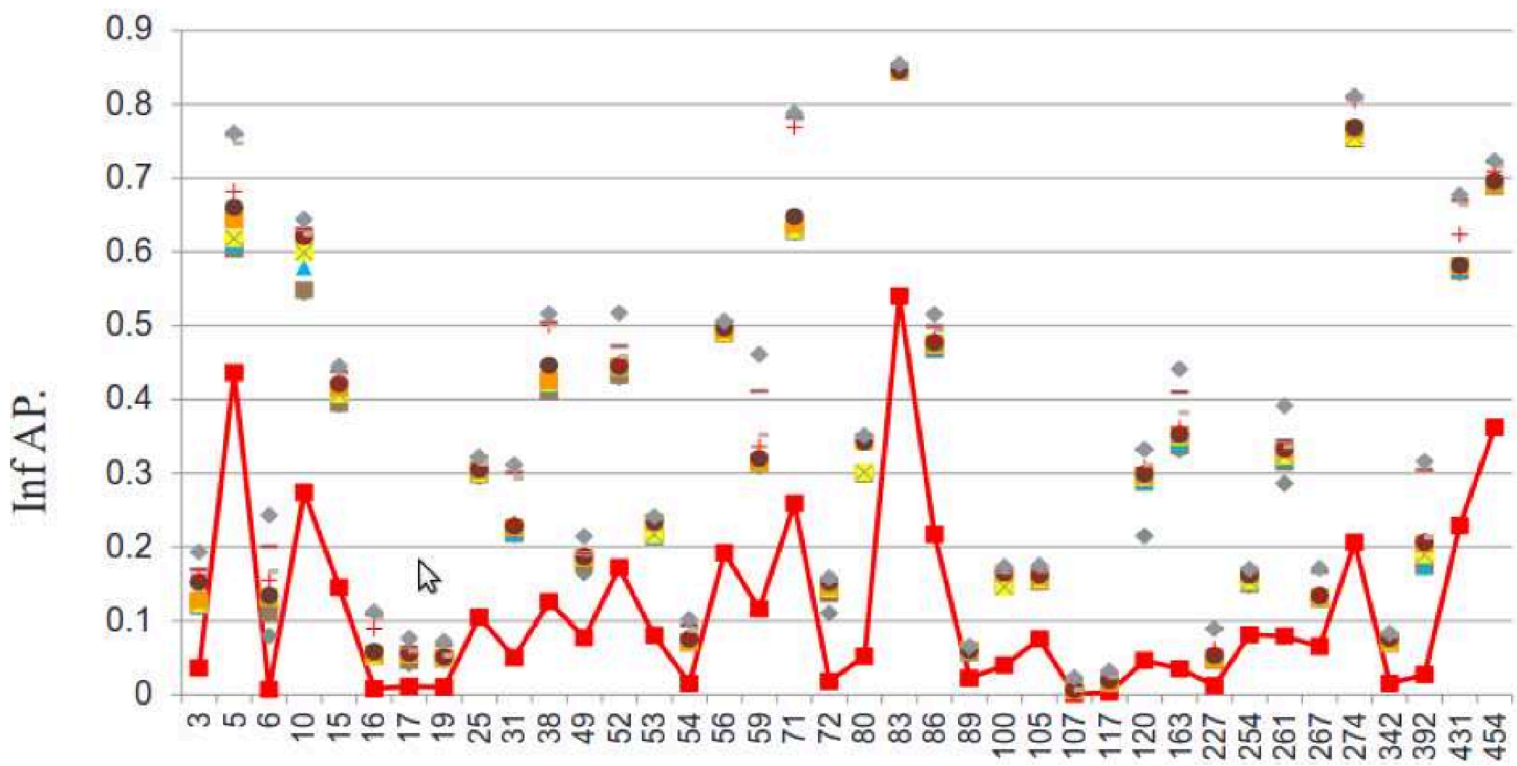


Fig. 4. Performance of Top-10 high-level Feature Detections per Feature in TRECVid 2005

And in 2013 ?

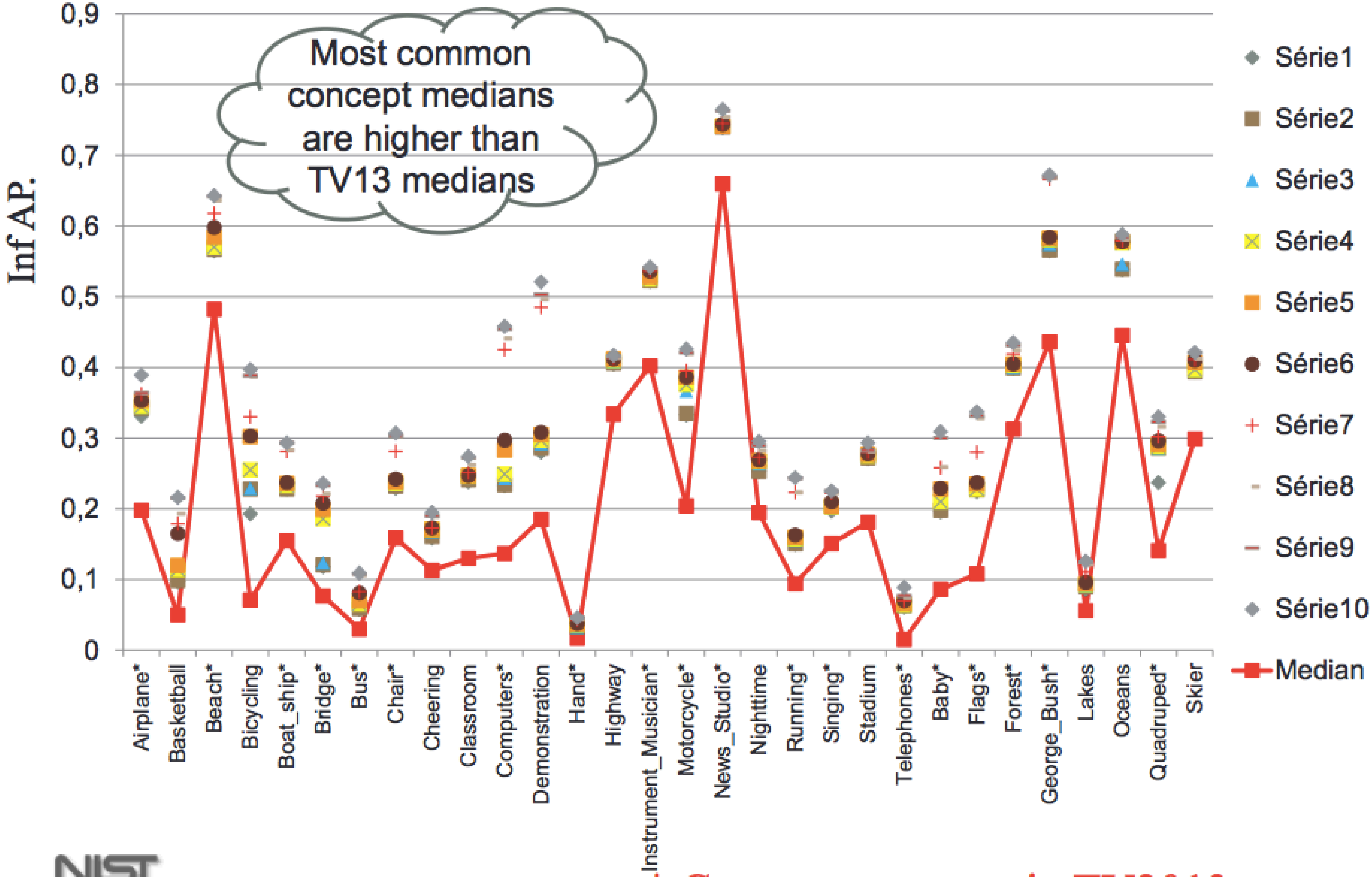
Figure 4: Top 10 runs (xinfAP) by concept



30 Single concepts evaluated(1)

3 Airplane*
9 Basketball
10 Beach*
13 Bicycling
15 Boat_Ship*
17 Bridges*
19 Bus*
25 Chair*
27 Cheering
29 Classroom
31 Computers*
41 Demonstration_Or_Protest
59 Hand*
63 Highway
71 Instrumental_Musician*
80 Motorcycle*
83 News_Studio*
84 Nighttime
100 Running*
105 Singing*
112 Stadium
117 Telephones*
163 Baby*
261 Flags*
267 Forest*
274 George_Bush*
321 Lakes
359 Oceans
392 Quadruped*
434 Skier

Top 10 InfAP scores by concept (Main runs)



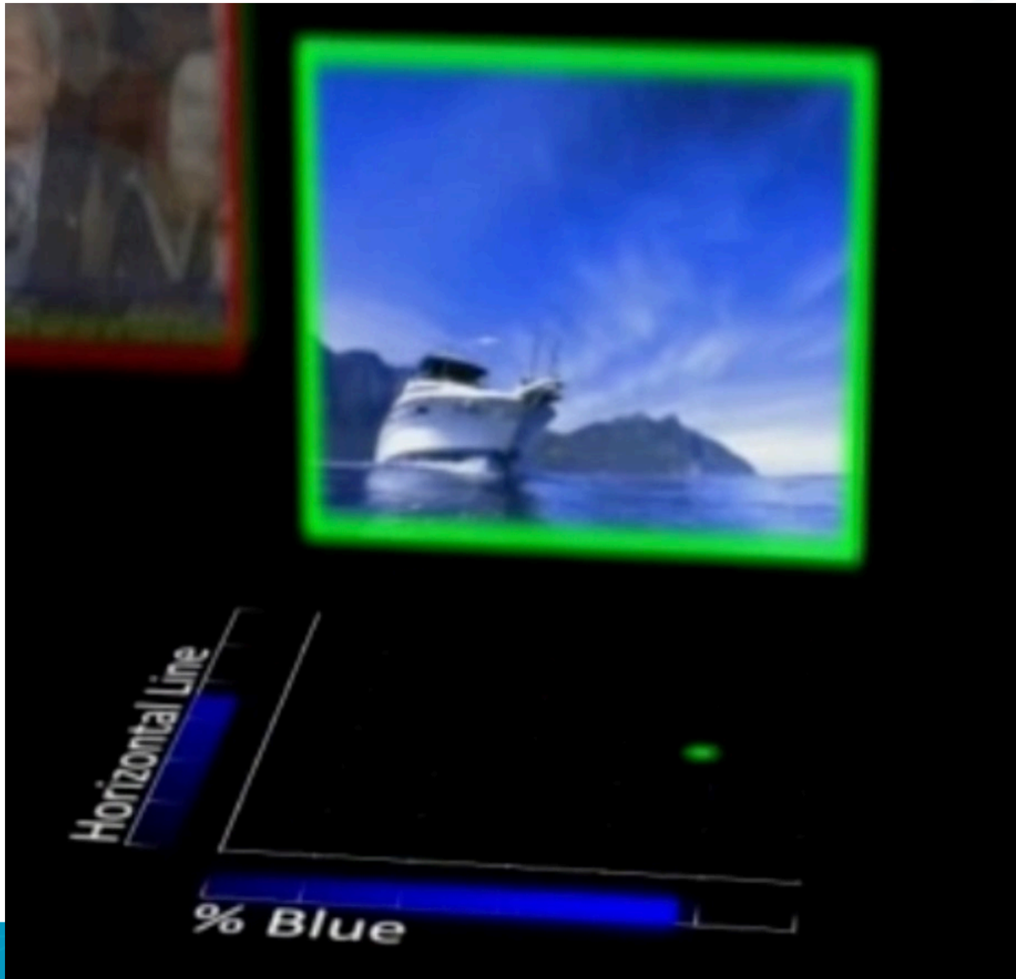
NIST National Institute of Standards and Technology

* Common concept in TV2013

Standard Machine Learning …

- So we’re bumbling along, slow and steady progress in using ML for CV, and ML is being used elsewhere also … from parole recommendations in US court cases … to recommending books from Amazon … to …

One of the pluses of "standard ML" is that we can conjure up an explanation for a result



- Another plus is that we can rate the relative importance of each of the axes .. i.e. features
- A downside is we have to do feature engineering to define the axes, and that's a black art
- Lots of applications across domains

And then this happened !

In 2012, Krizhevsky et al. “won” the ImageNet large scale visual recognition challenge with a convolutional neural network

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

How does this happen ?

In 2012, Krizhevsky et al. “won” the ImageNet large scale visual recognition challenge with a convolutional neural network

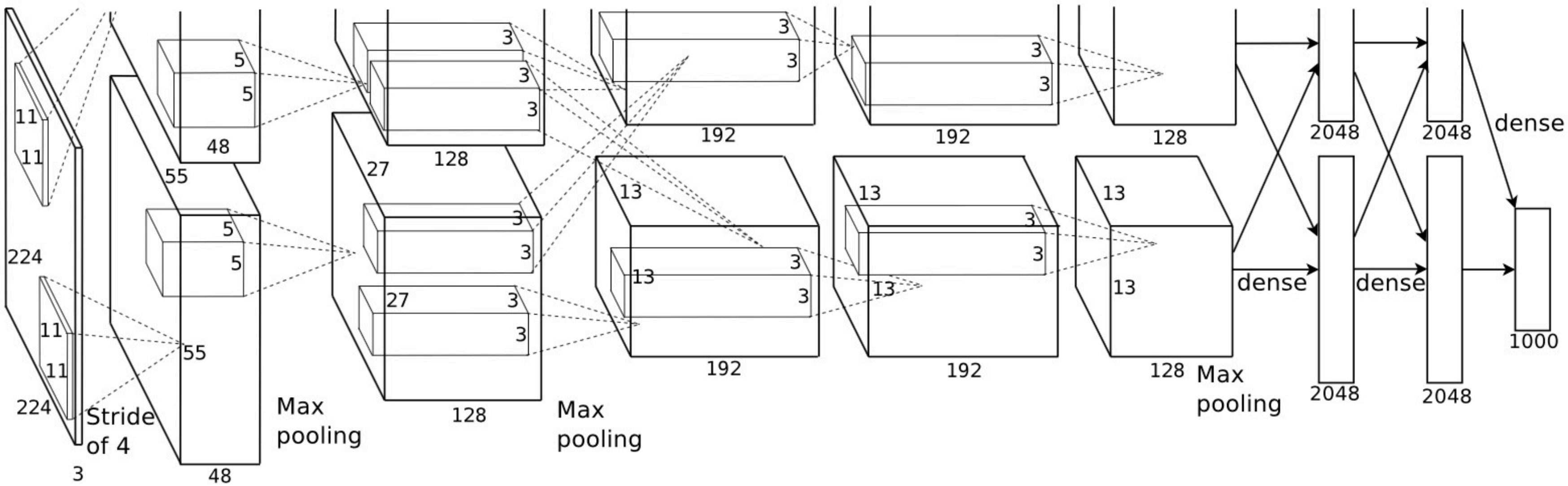
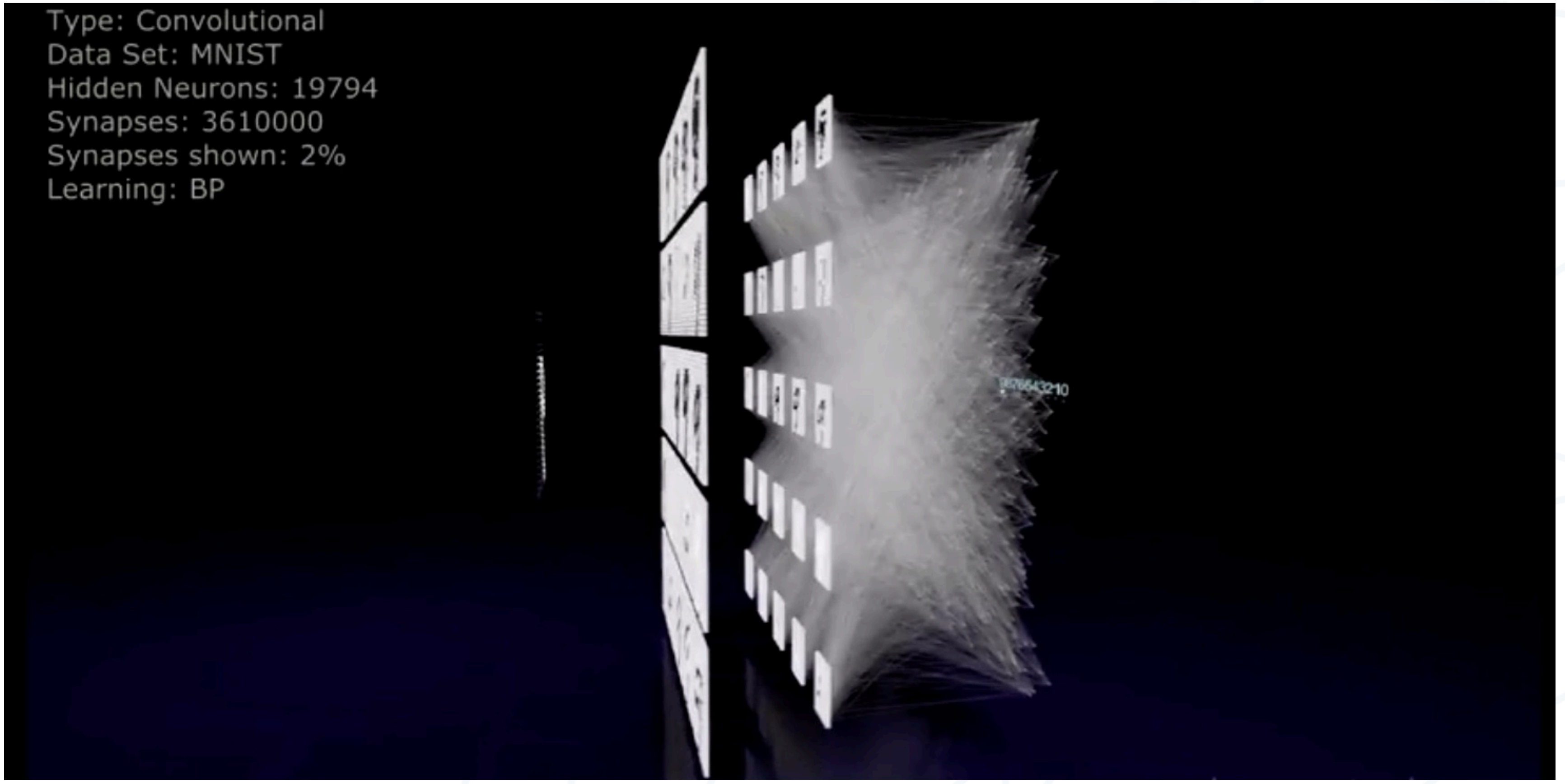


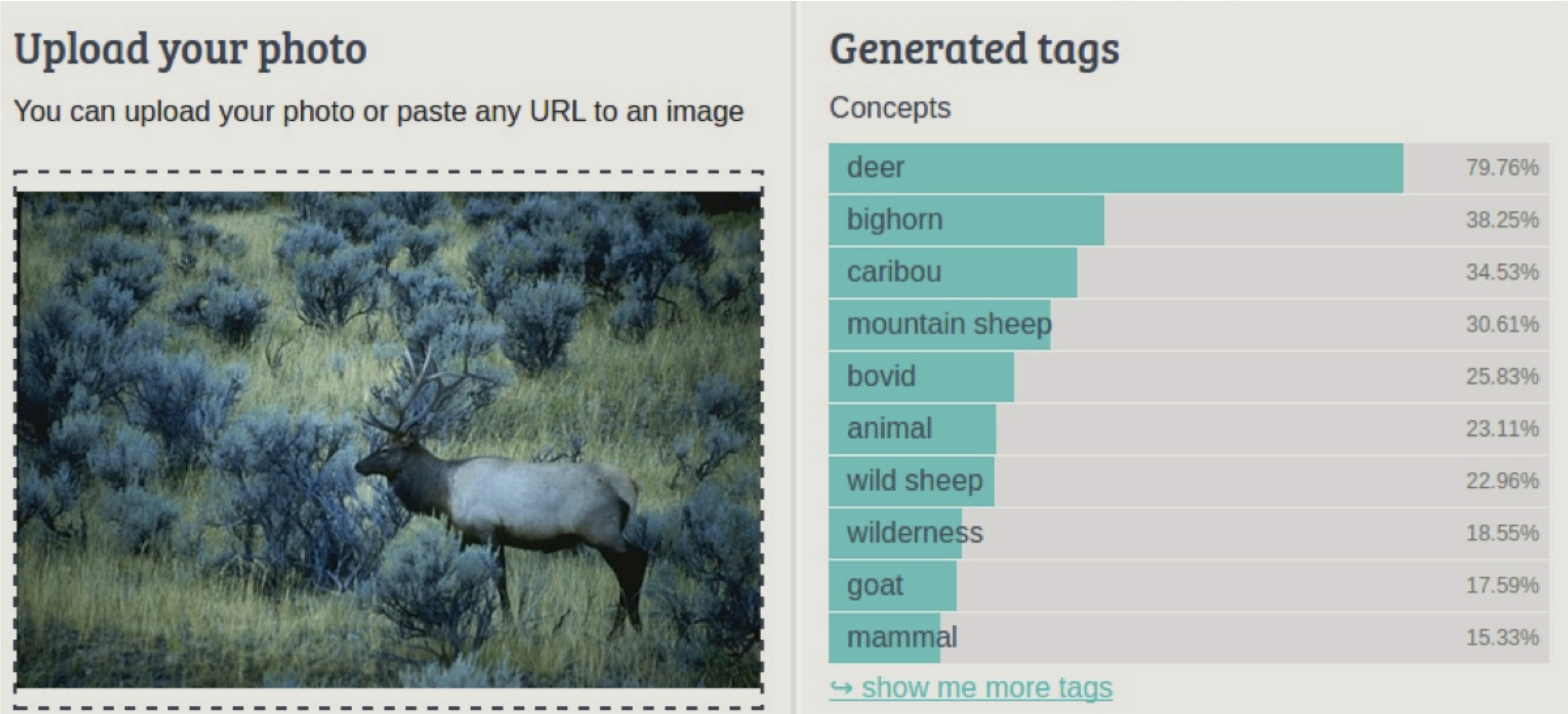
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network’s input is 150,528-dimensional, and the number of neurons in the network’s remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.


Type: Convolutional
Data Set: MNIST
Hidden Neurons: 19794
Synapses: 3610000
Synapses shown: 2%
Learning: BP

- Deep Convolutional Neural Networks are end-to-end solutions in which both the feature extraction and the classifier training are performed at once.
- The first layers extracts a type of information which is similar to the features/descriptors extracted in the classical approaches.
- These, called “deep features”, turn out to be significantly more efficient that the classical “engineered” ones, even when used with classical machine learning for classifier training
- Once a model for a concept is built, it can be packaged and released and easily run in a hosted environment

Online systems now available

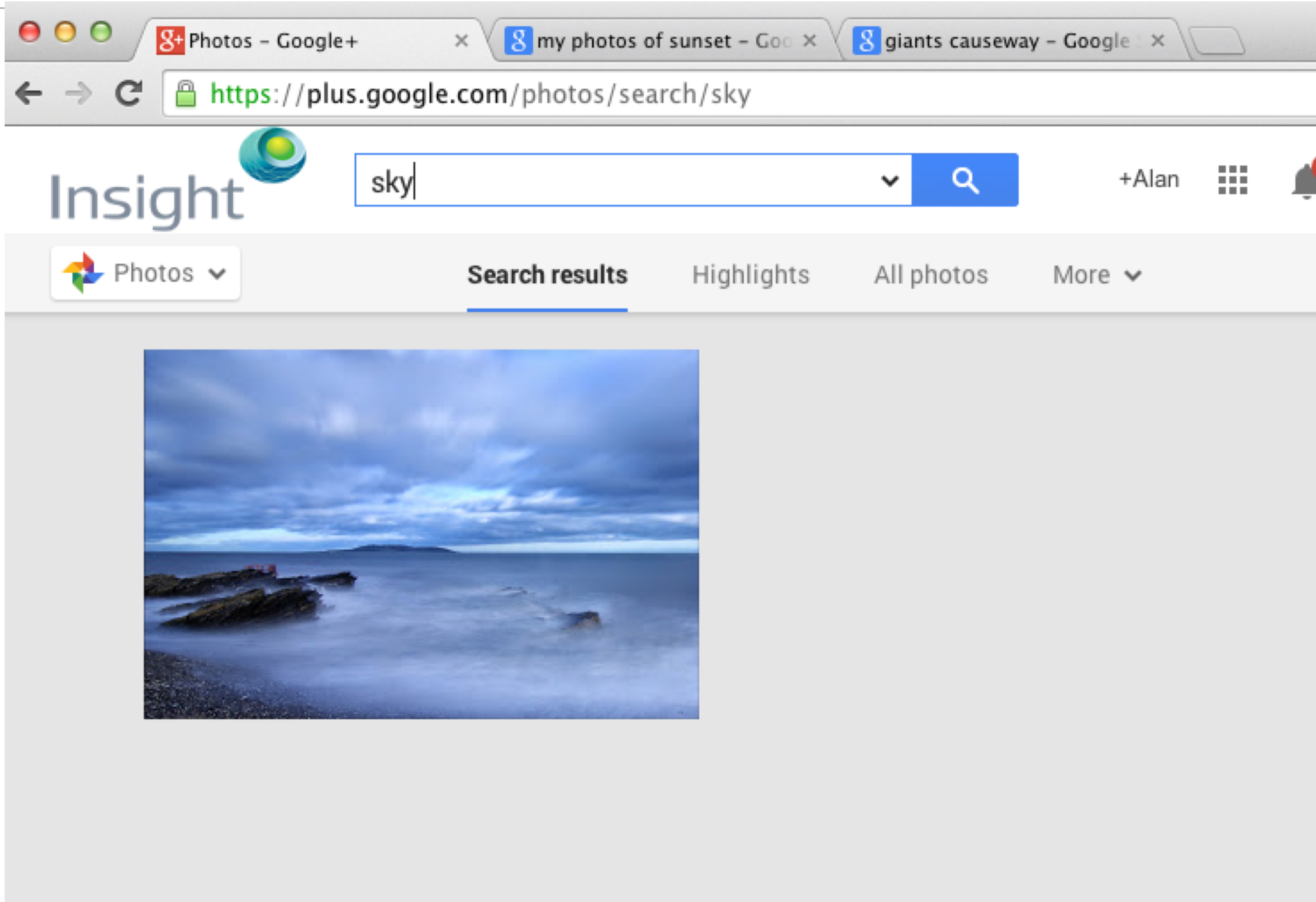
- Imagga, Bulgarian SME, offering up to 6,000 visual tags



Concepts in image search

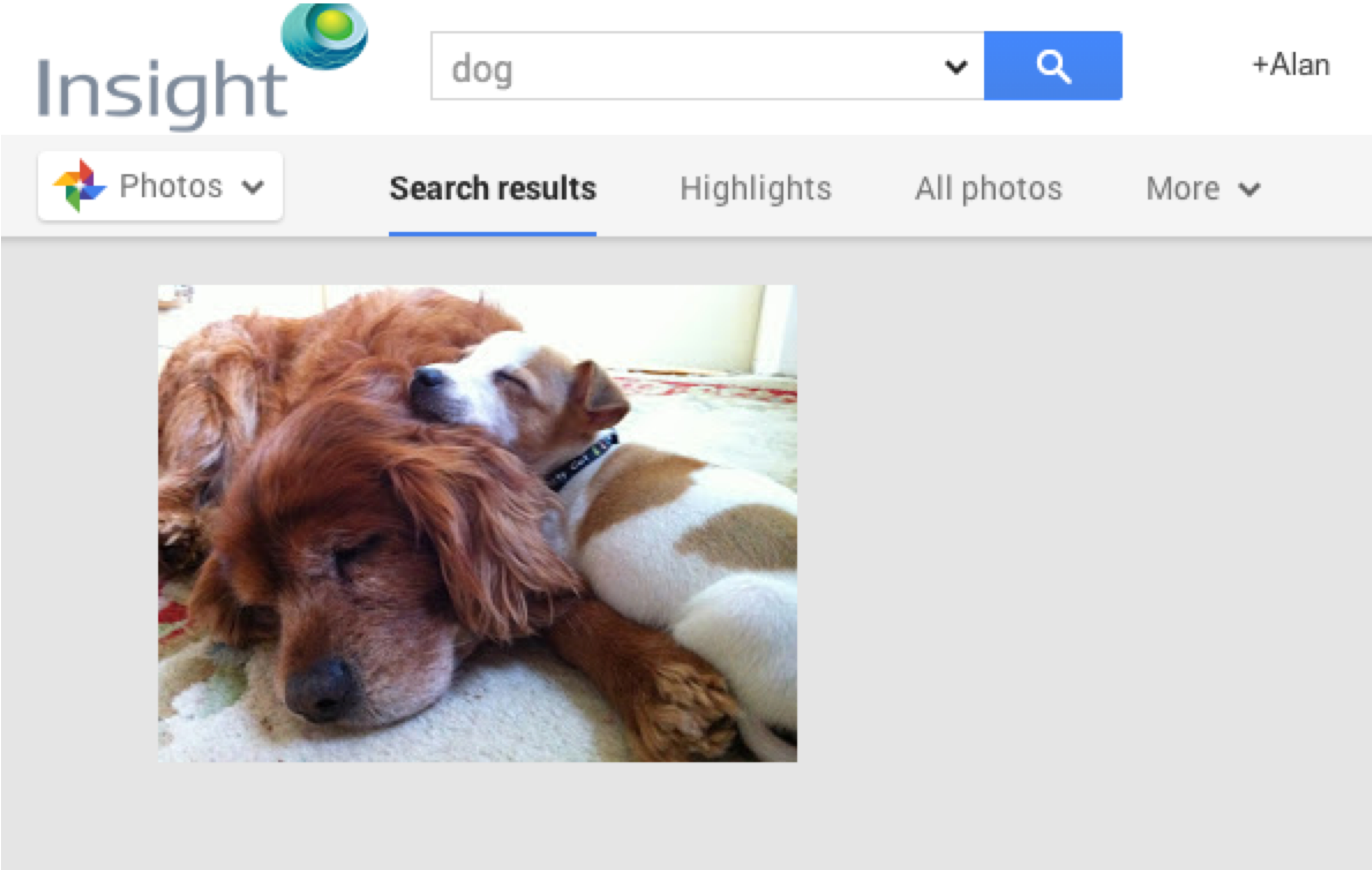
- Google+ photos now uses computer vision and machine learning to identify objects and settings in your uploaded snapshots

My uploads

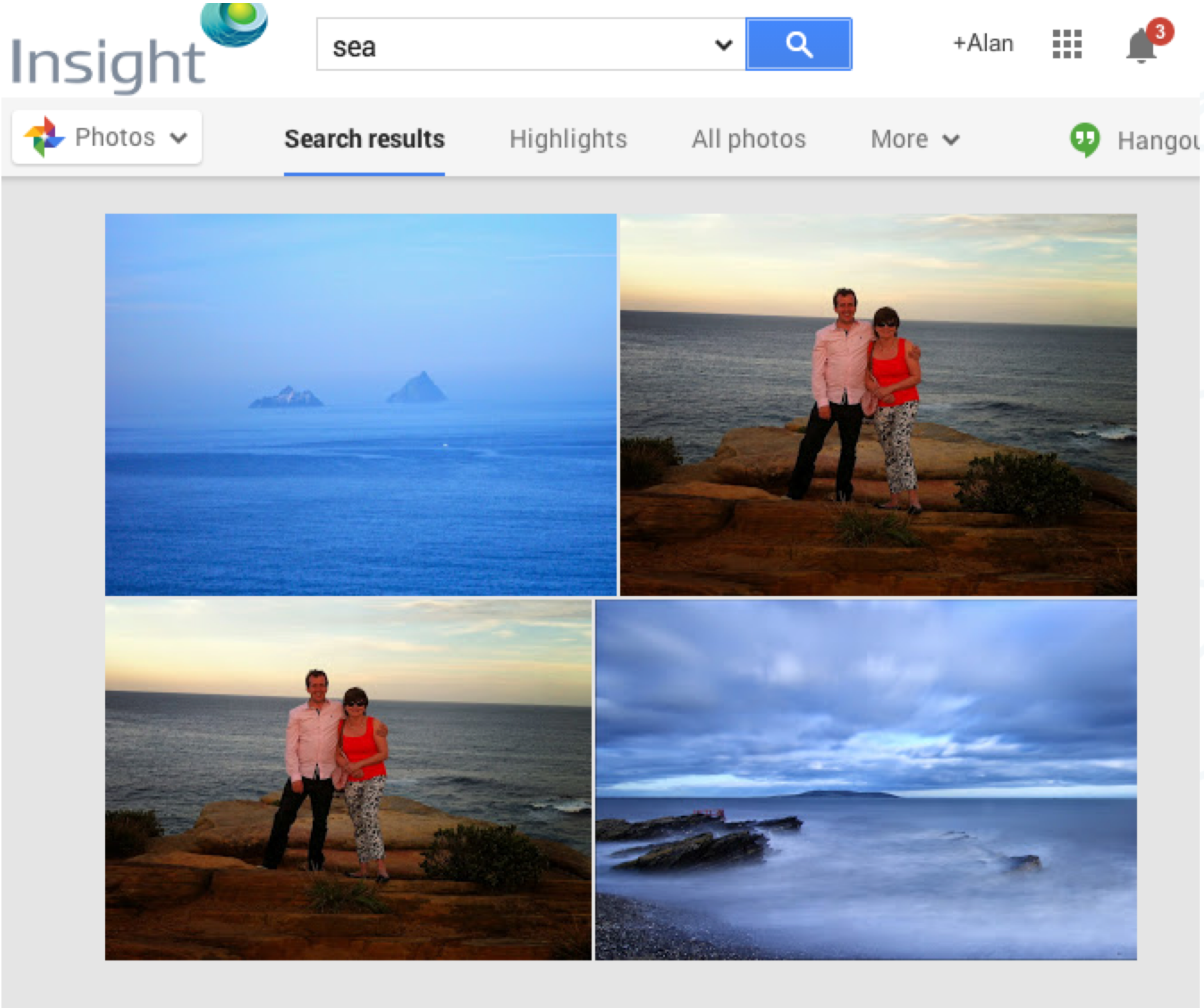


Photos - Google+
my photos of sunset - Goo
giants causeway - Google
https://plus.google.com/photos/search/sky
Insight
sky
+Alan
Photos
Search results
Highlights
All photos
More

My uploads


Insight
dog
+Alan
Photos
Search results
Highlights
All photos
More

My uploads



Insight
sea
+Alan
3
Photos
Search results
Highlights
All photos
More
Hangou

How good can a computer get at automatically annotating images without using any context at all ?

Captioned by Human and by Google's Experimental Program

Human: "A group of men playing Frisbee in the park."
Computer model: "A group of young people playing a game of Frisbee."

Captioned by Human and by Google's Experimental Program

Human: "A young hockey player playing in the ice rink."
Computer model: "Two hockey players are fighting over the puck."

2 of 6

Captioned by Human and by Google's Experimental Program

Human: "Elephants of mixed ages standing in a muddy landscape."
Model: "A herd of elephants walking across a dry grass field."

How to implement CNNs ?

- Its definitely not von Neumann architecture
- So we kludge it by throwing massive parallelism … what's the cheapest massive parallelism … GPUs
- New role for Nvidia … not necessarily fast, but many of them !

- An alternative is to design new chips …
- Intel deep learning chips code named Lake Crest followed by Knight's Crest, under development, 2017 arrival
- Samsung handset chips and Qualcomm chips, allow deep learning on devices, and Movidius (now part of Intel) specialising in computer vision using deep learning on silicon.

How to implement CNNs ?

- Google was starting to use deep networks for NLP applications like speech and also to run Panda website rating, but needed more horsepower, like x2 !
- Tensor Processing Unit (TPU) designed from scratch, very efficient
- Used for executing neural networks rather than training .. saved building x15 data centres and can run neural networks x30 times faster than conventional vN chip
- Downsides … need lots and lots and lots of training data, lots and lots of compute resources … and no facility for explaining why a decision or categorisation or output, is made

What of von Neumann architecture ?

- We used to have people, writing algorithms, encoded as programs which were stored and run
- Now we also have data, used to train networks and develop models (sets of weights and network configurations) which are stored and run