Deep Learning 과정

Introduction to Deep Learning

Gunhee Kim

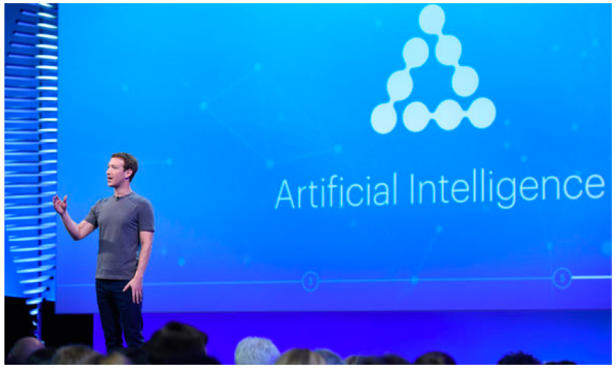
Computer Science and Engineering

서 울 대 학 교
SEOUL NATIONAL UNIVERSITY

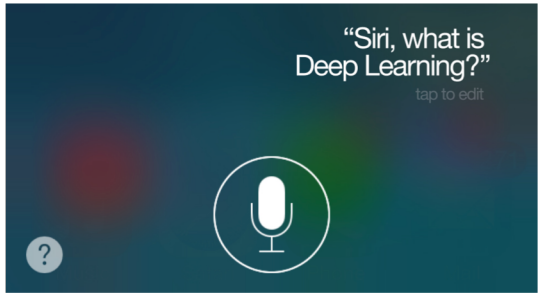
Deep Learning

One of the hottest buzzwords in both academia and industry

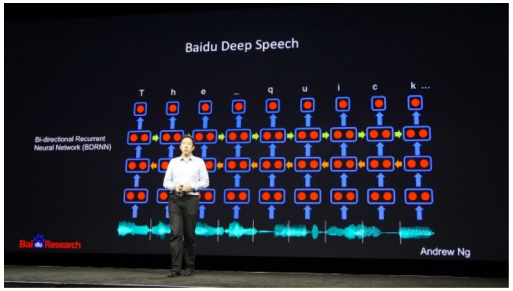
Google



Facebook

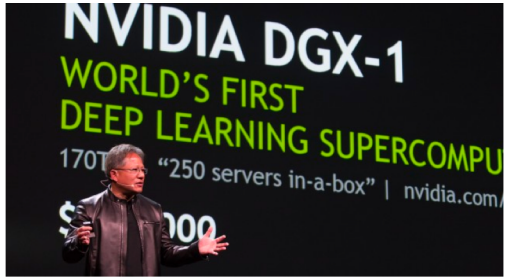


Apple



Baidu

Microsoft



NVIDIA

Deep Learning for Image Classification

Representation learning attempts to automatically learn good features or representations

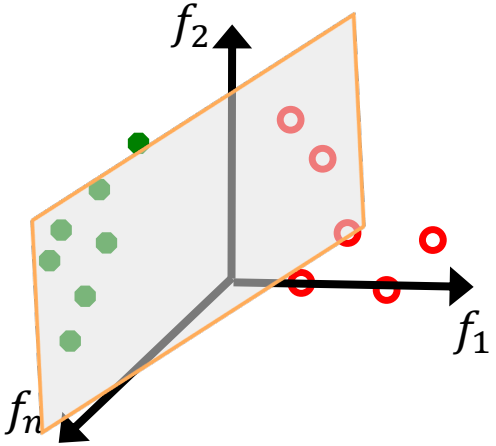
Feature learning problem

- Suppose we want to classify images whether they are faces or not

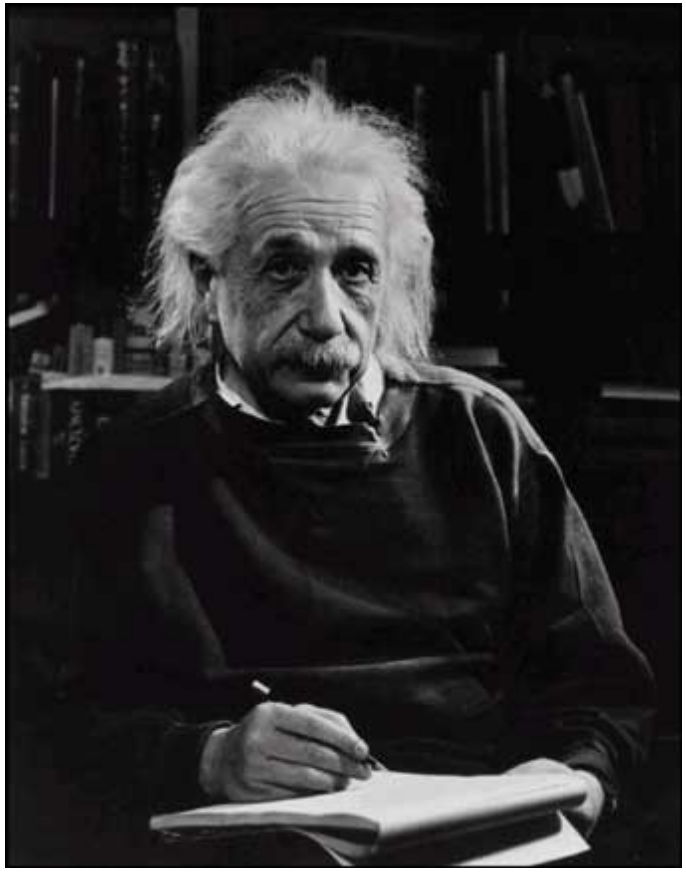
- Can we represent images using 100 real numbers?

$$\begin{pmatrix} 255 \\ 98 \\ 93 \\ 87 \\ 89 \\ 91 \\ \dots \end{pmatrix}$$

Can we find a good one?



Representation is Not Easy – An Image



What we see

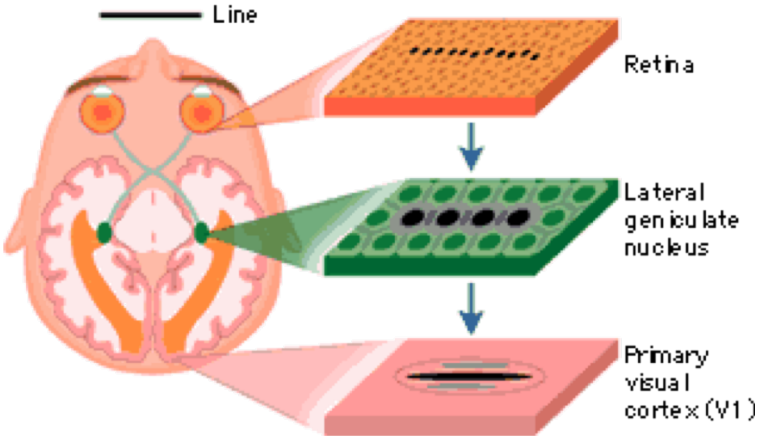
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer see

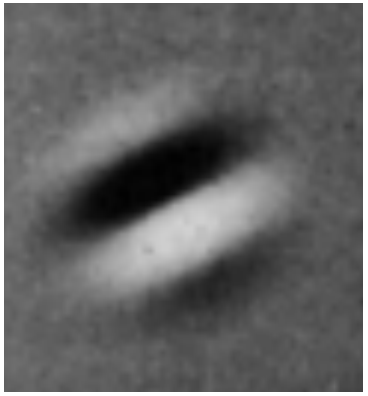
Edge Detection

Our brain first detects edges

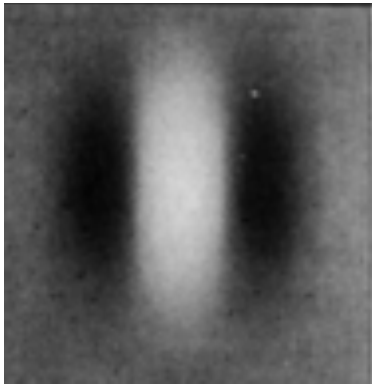
- Cells in primary visual cortex (V1) are activated by lines of a given orientation



First stage of visual processing: V1



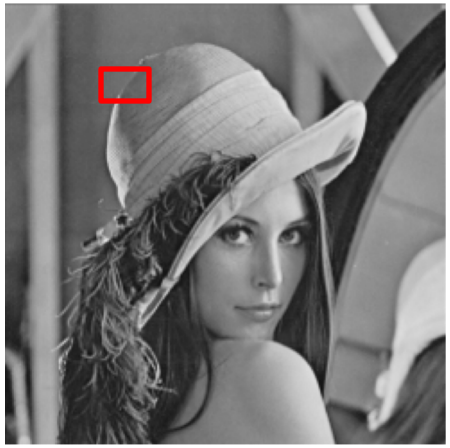
Neuron #1 of visual cortex (model)



Neuron #2 of visual cortex (model)

Edge Detection

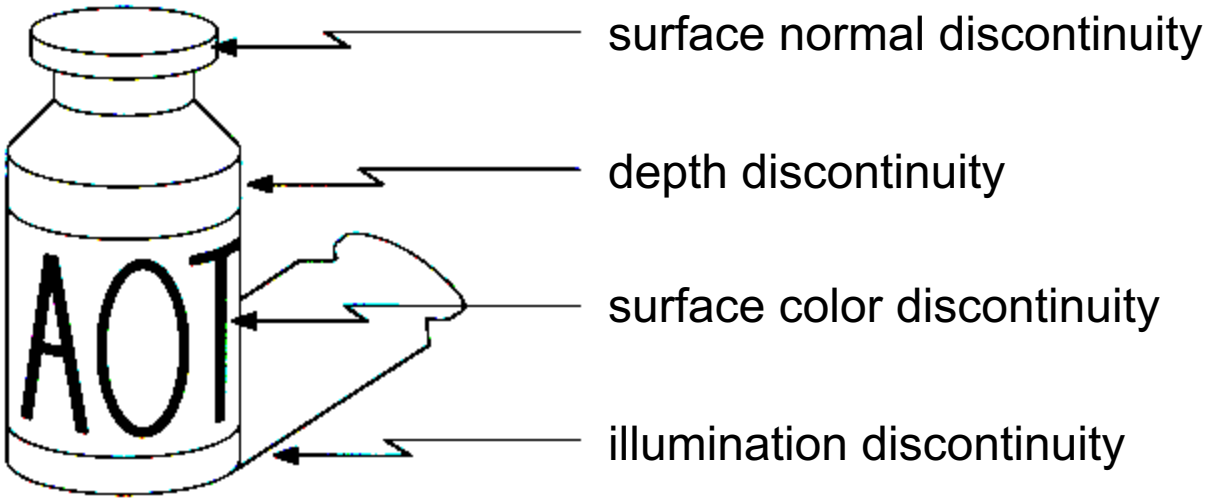
Line segments where the image brightness changes sharply (or has discontinuities)



Edge detection is not easy!

Edges

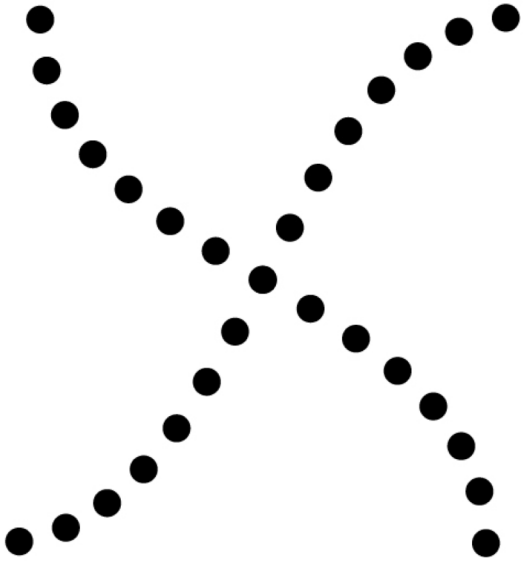
Edges are caused by a variety of factors



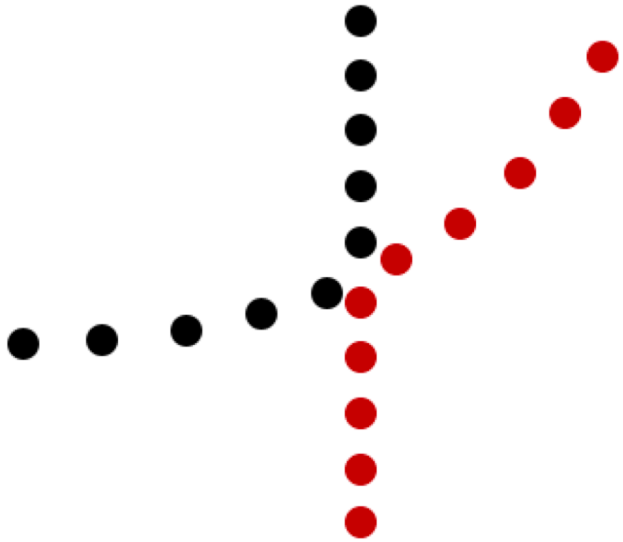
OK, edge detection is not easy… then how can we group edges?

Grouping

Grouping



People tends to mentally form a continuous line

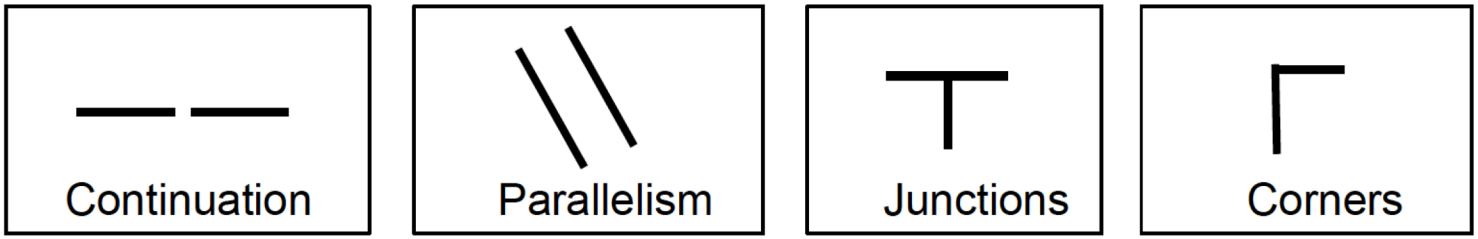


All of sudden, people use color information

People adaptively use different rules for grouping

Mid-Level Representation

Mid-level cues



"Tokens" from
Vision by D.Marr

Object parts

Difficult to hand-engineer → What about learning them?

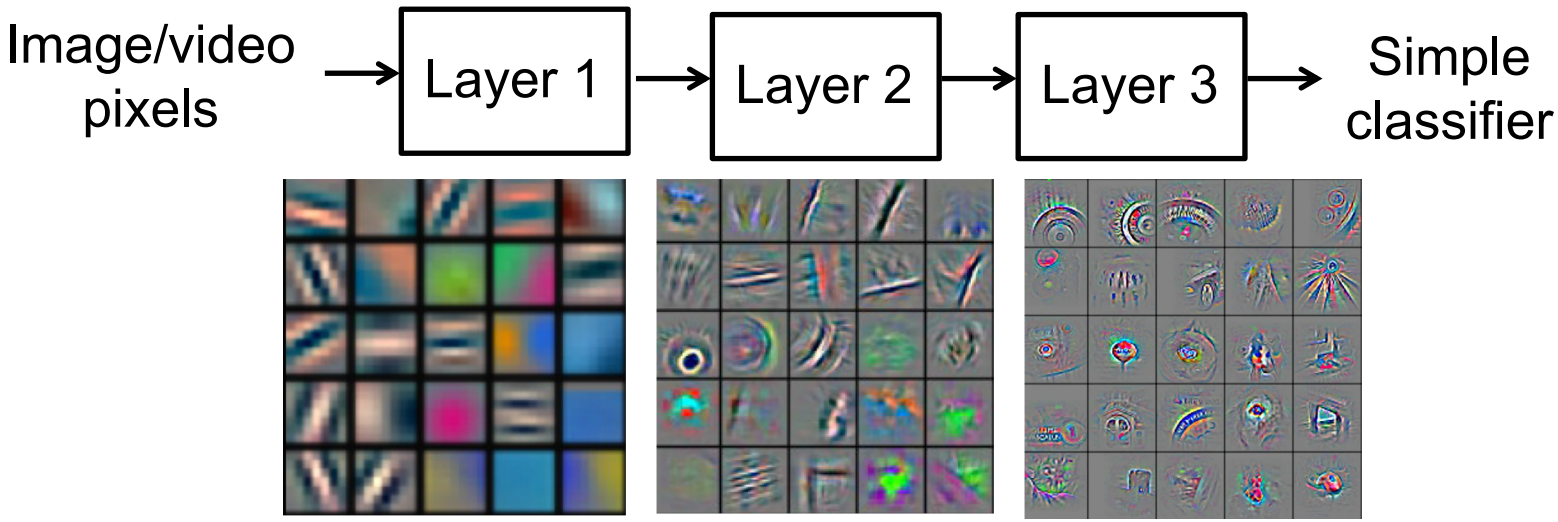
[Rob Fergus]

Deep Learning

Deep learning algorithms attempt to learn multiple levels of representation of increasing complexity abstraction

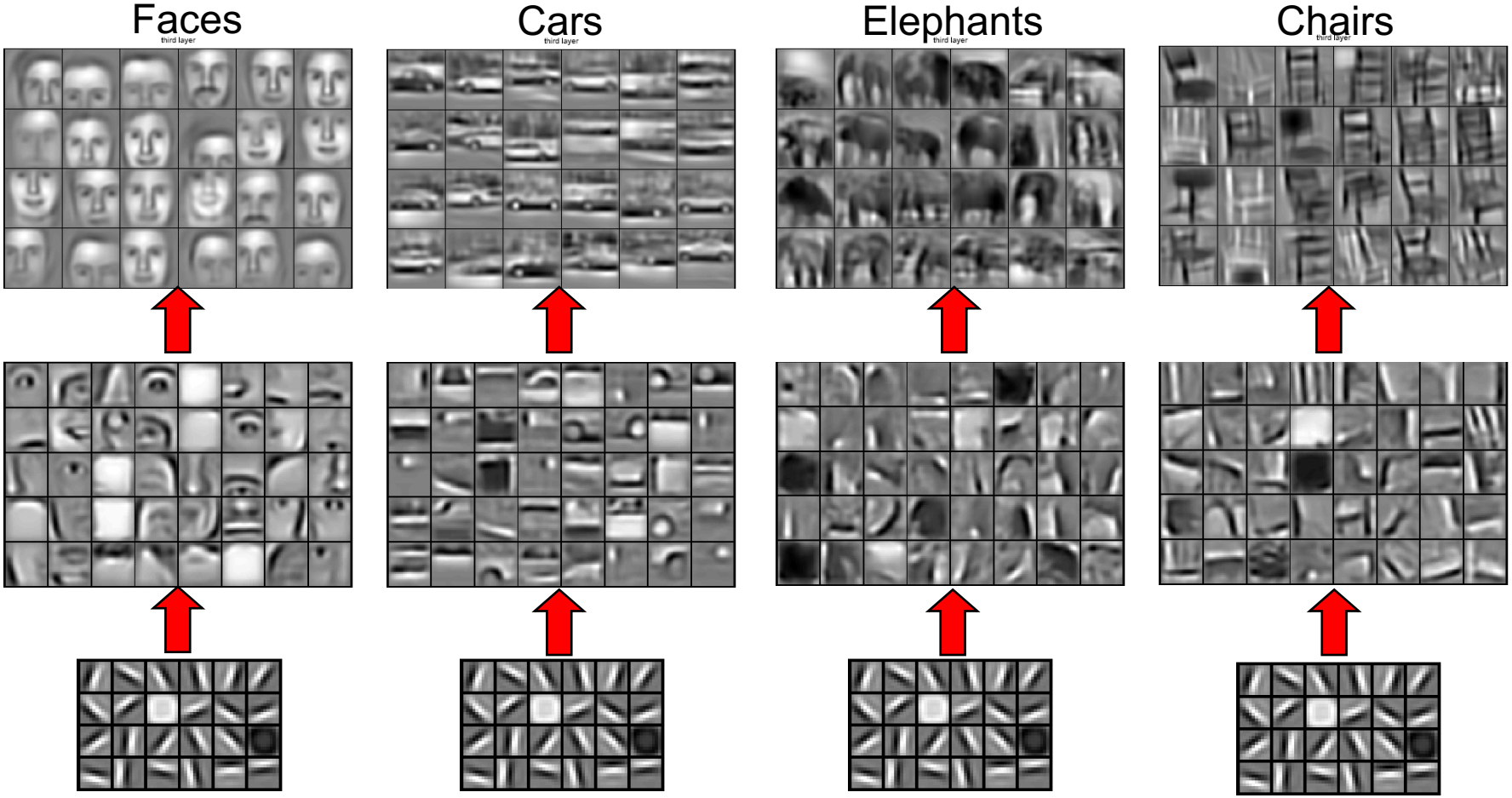
- A cascade of many layers of nonlinear processing units for feature extraction and transformation
- Each hidden layer learns different level of abstraction; the levels form a hierarchy of concepts
- End-to-end: All the way from pixels → Classifier

(Learned internal representation)



Learning of Object Parts

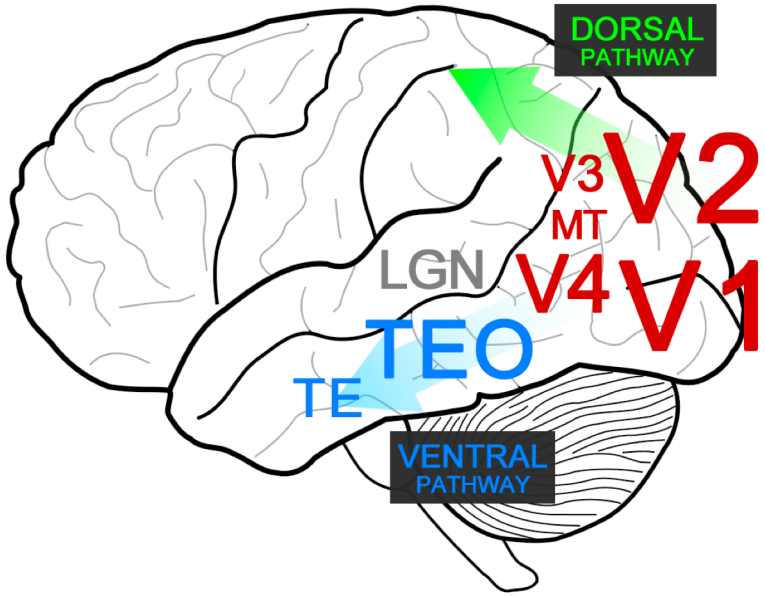
Examples of learned object parts from object categories



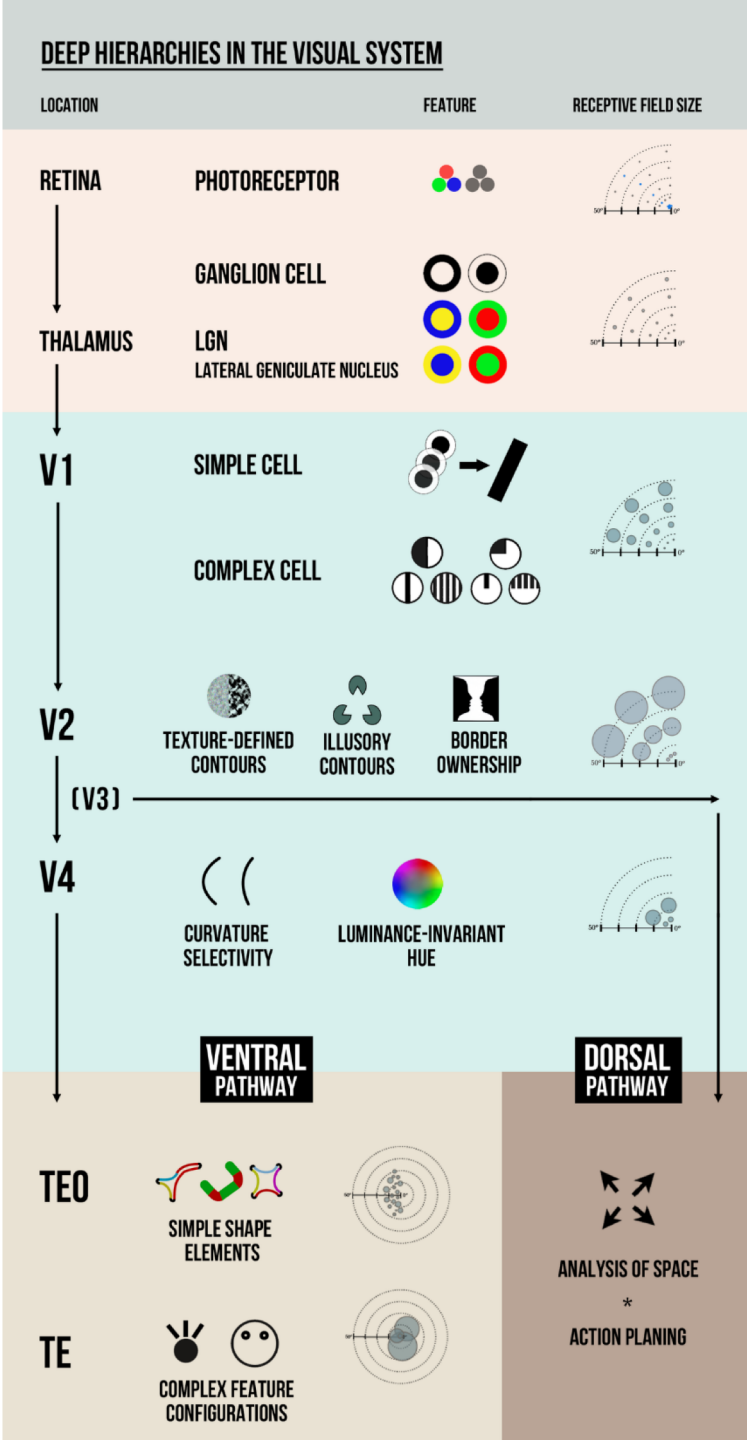
[Andrew Ng]

Deep Hierarchies in Vision

The processing of visual information happens over at least 10 functional levels



[https://en.wikibooks.org/wiki/Sensory_Systems/Visual_System]



Sub-cortical Vision

Photoreceptors on retina

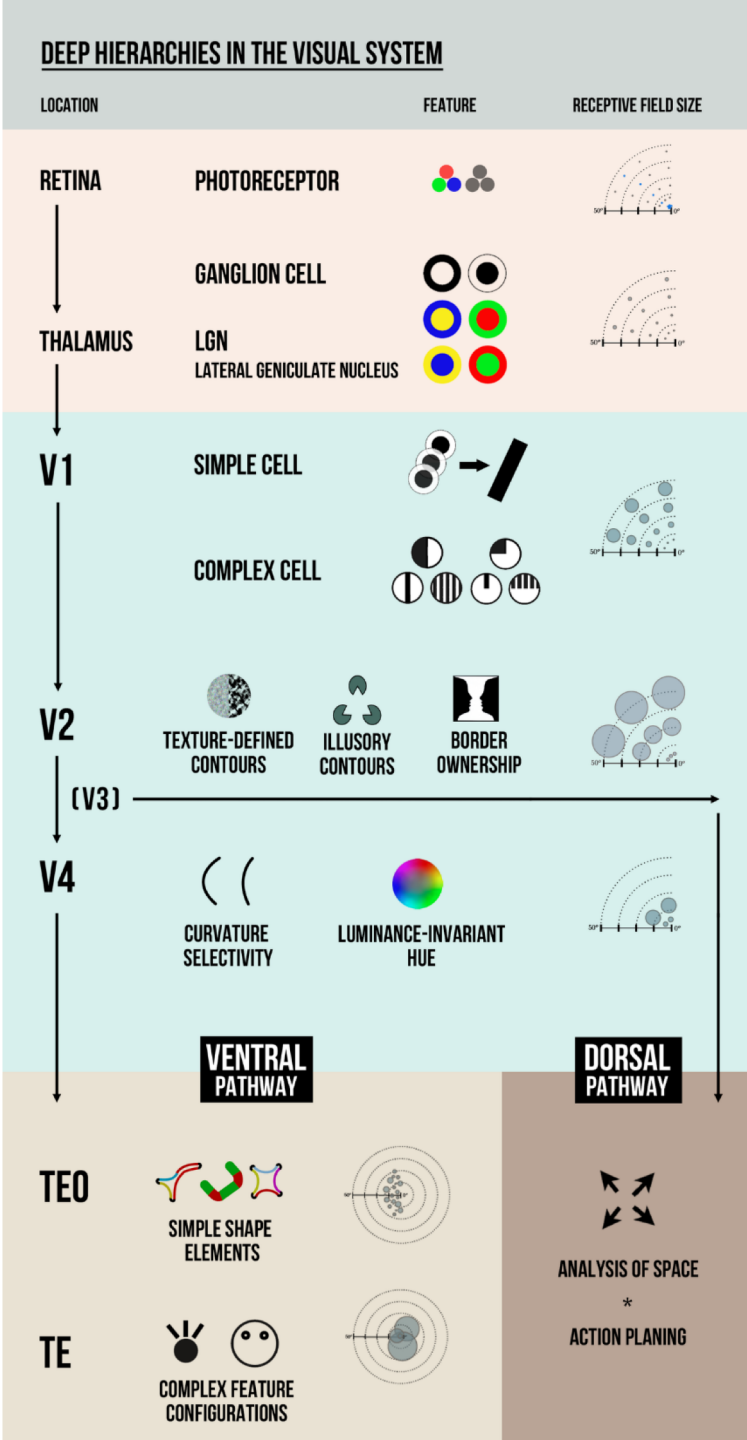
- Detect light and send signals to retinal ganglion cells
- Receptive field size: 0.01°

Ganglion cells

- Selectively tuned to detect various features (e.g. luminance/color contrast, movement direction/speed)

LGN

- Relay the signals to the cortex
- Only 5% of inputs from the retina
- Feedback like attention, expectation, imagination and filling-in the missing information



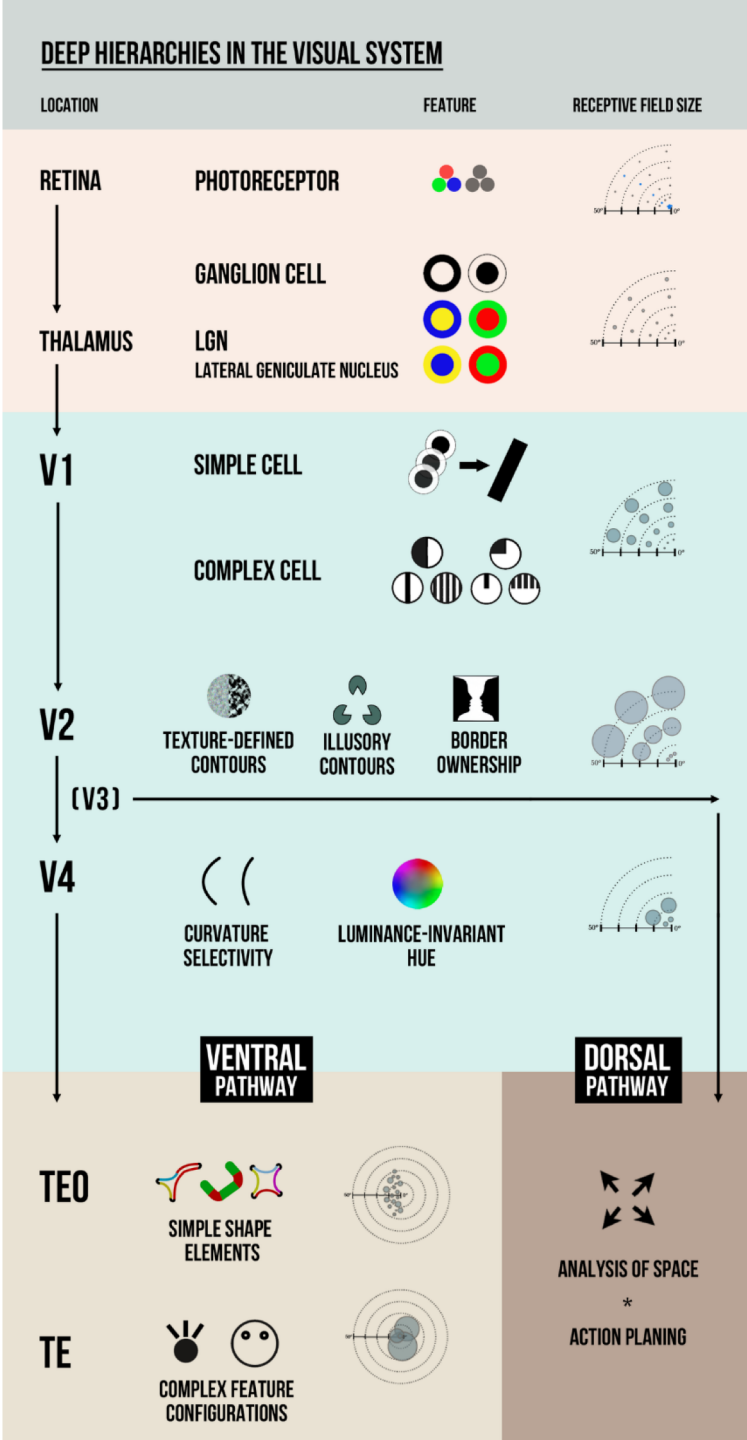
Cortical Vision

Receive input from LGN and sends outputs to dorsal and ventral streams

- Dorsal pathway: analysis of space and in action planning
- Ventral pathway: object recognition and categorization

V1

- Sensitive to edges, gratings, line-ending, motion, color and disparity
- Hierarchical bottom-up processing: linear combination of the inputs from several ganglion cells (or other V1s)



Cortical Vision

V2

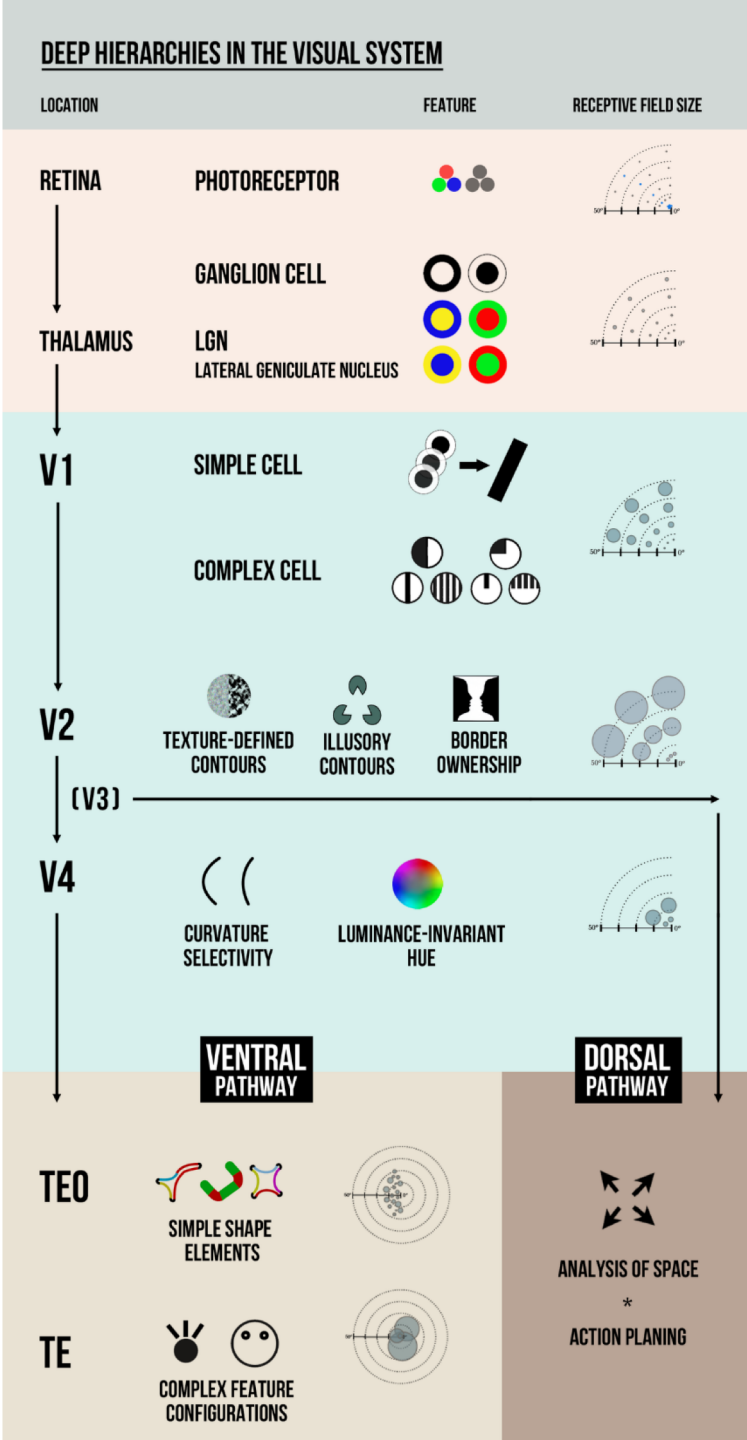
- More sophisticated contour representation including texture-defined contours, illusory contours and contours with border ownership

V3

- Very little is known about the computation taking place in V3

V4

- Sensitive to contours with different curvature and vertices with particular angles
- Coding for luminance-invariant hue



Inferior Temporal Cortex

IT (Inferior Temporal Cortex)

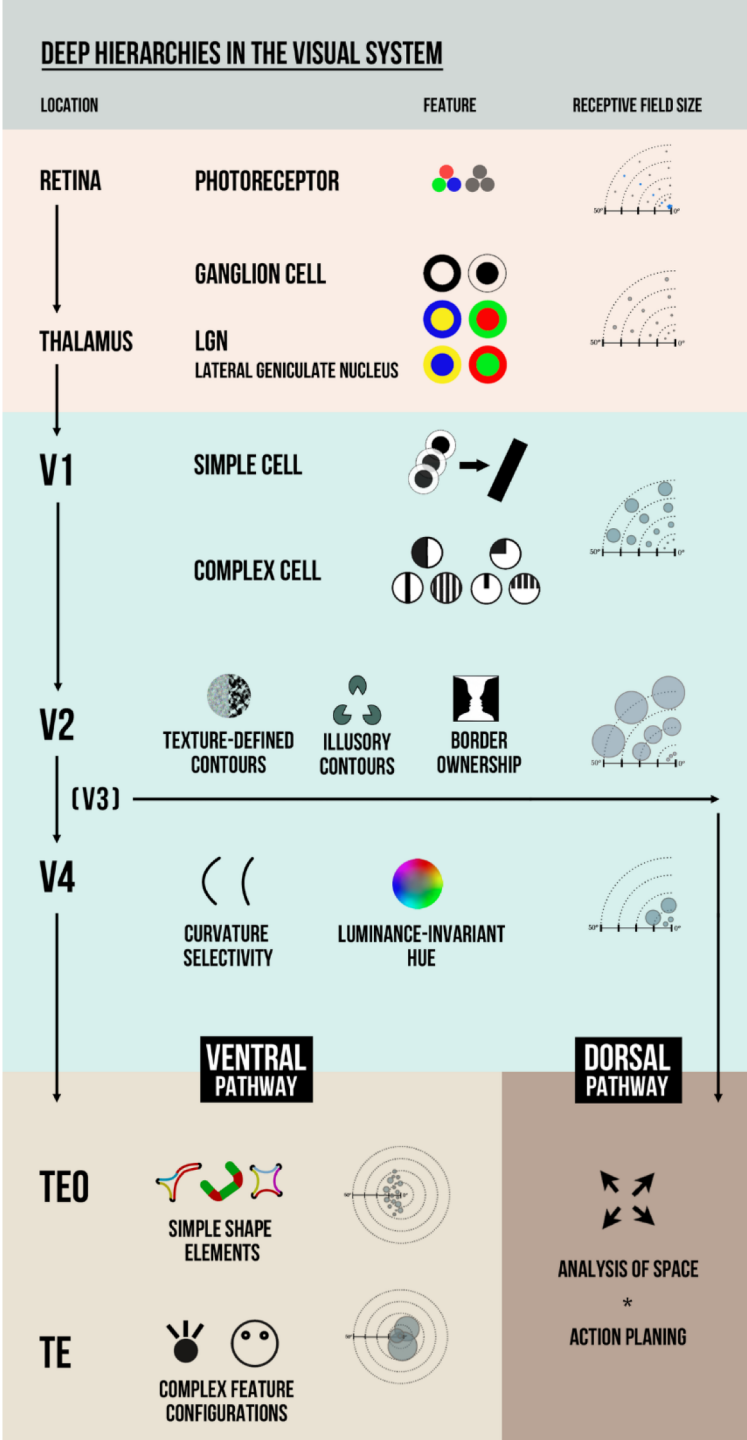
- Build models of object parts
- Pull together various features of medium complexity from lower levels in the ventral stream
- Consists of TEO and TE

TEO

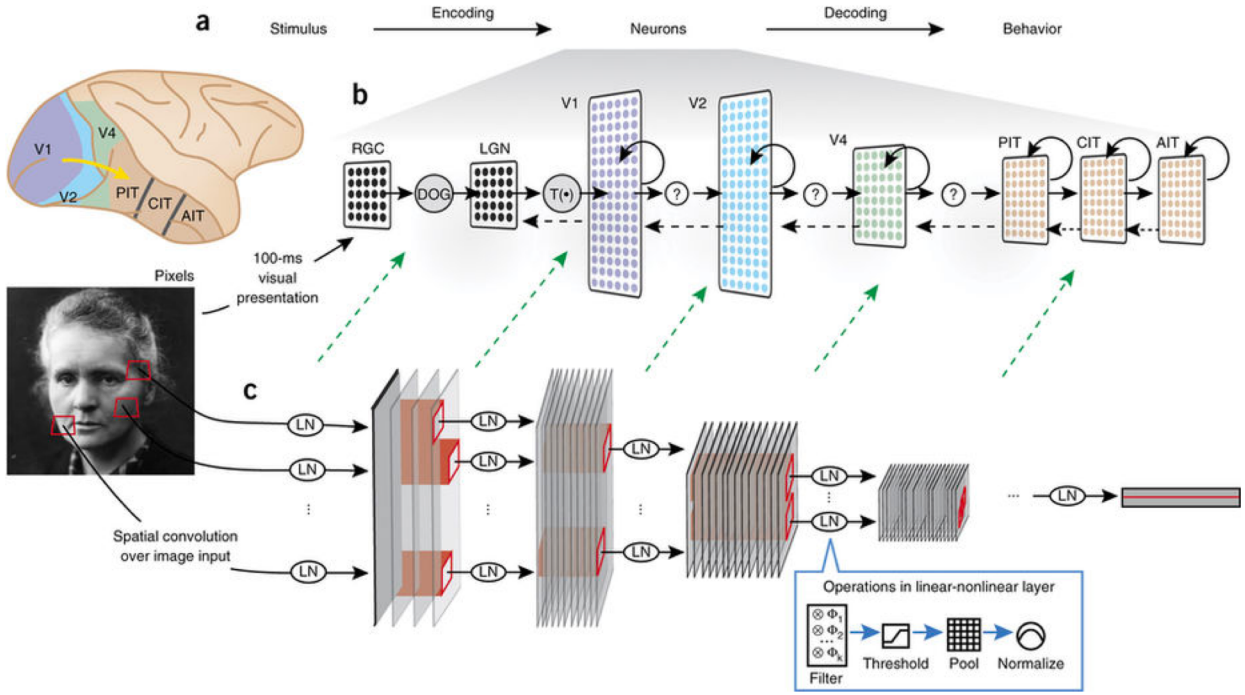
- Simple shape elements (shapes and relative positions of multiple contour elements)

TE

- Complex feature configurations (e.g. faces, hands)



Human Visual System to Convolutional Net

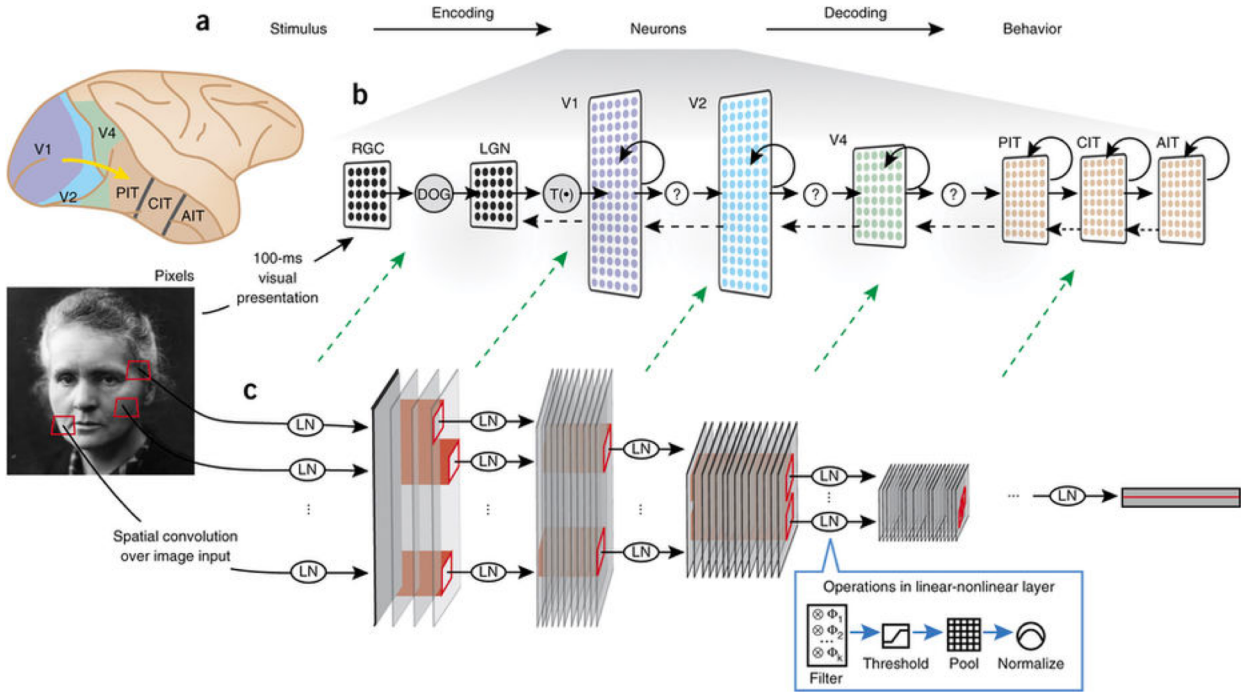


Basic framework of sensory cortex

- Encoding: Transform stimuli into neural activity patterns
- Decoding: Neural activity generates behavior
- Ventral visual pathway: the most comprehensively studied sensory cascade

[Yamins& DiCarlo. Using goal-driven deep learning models to understand sensory cortex, *Nature Neuroscience*]

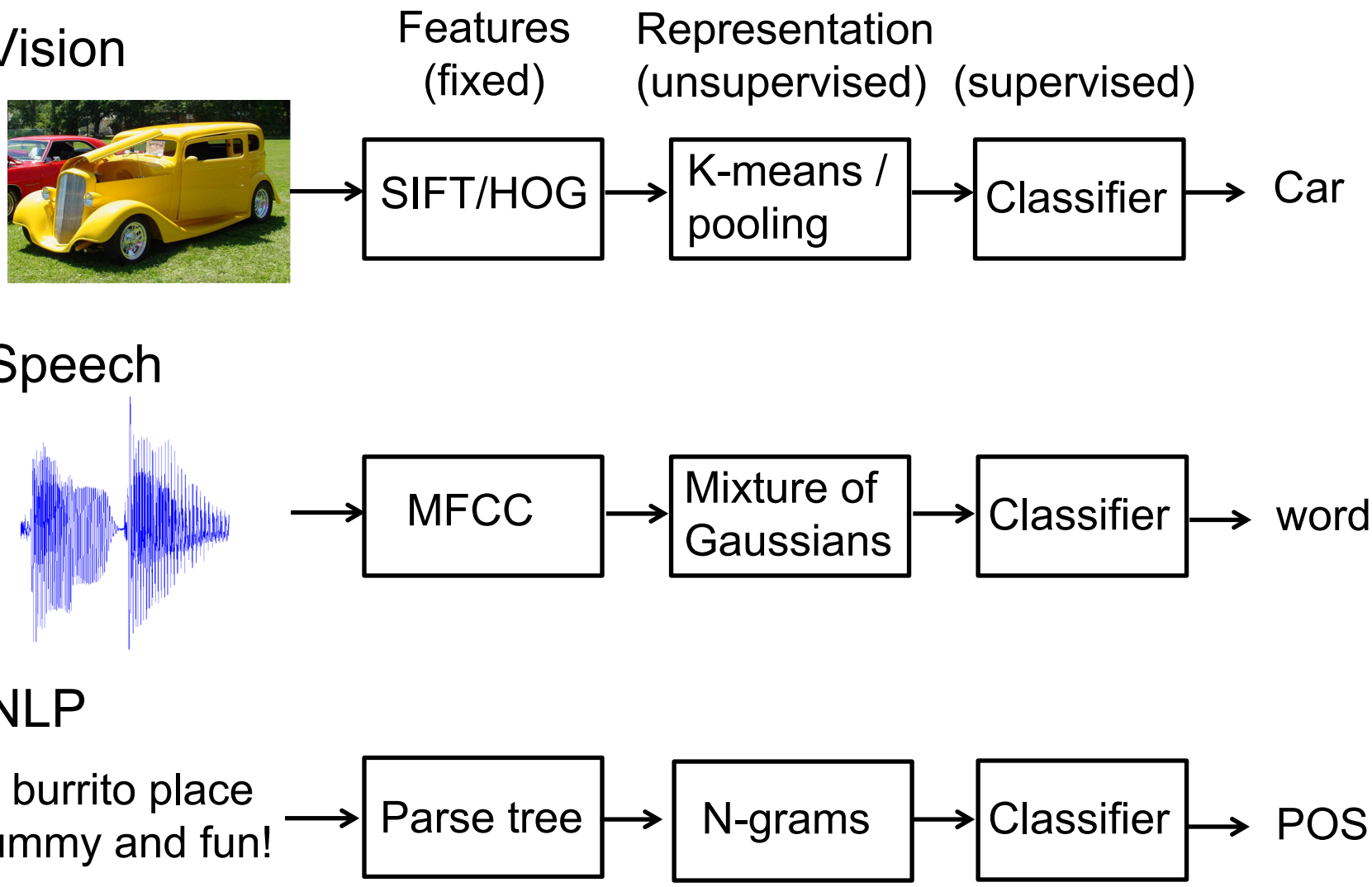
Human Visual System to Convolutional Net



Hierarchical (multilayer) CNNs

- Each layer: a linear-nonlinear (LN) combination of simple operations such as filtering, thresholding, pooling and normalization
- Filter bank (a set of weights analogous to synaptic strengths) corresponds to a distinct template
- Keep increasing sizes of receptive fields

Traditional Pattern Recognition



[Marc'Aurelio Ranzato]

(Deep) Hierarchical Compositionality

Vision

Pixels → Edges → Texton → Motif → Part → Object

Speech

Sample → Spectral band → Formant → Motif → Phone → Word

Natural language

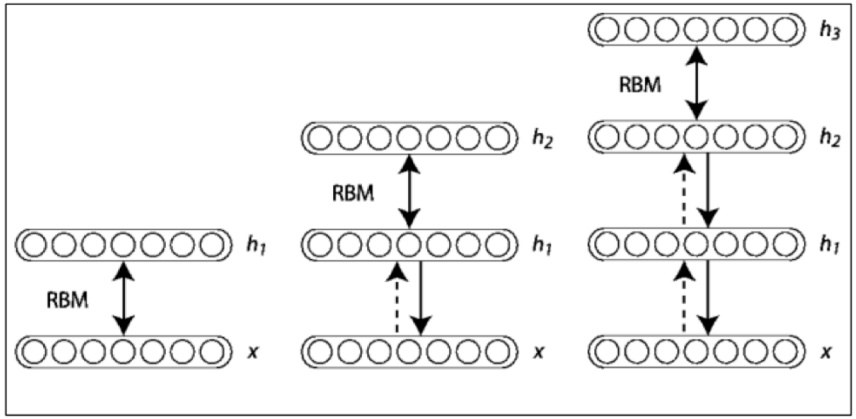
Character → Word → NP/VP/… → Clause → Sentence → Story

[Marc'Aurelio Ranzato]

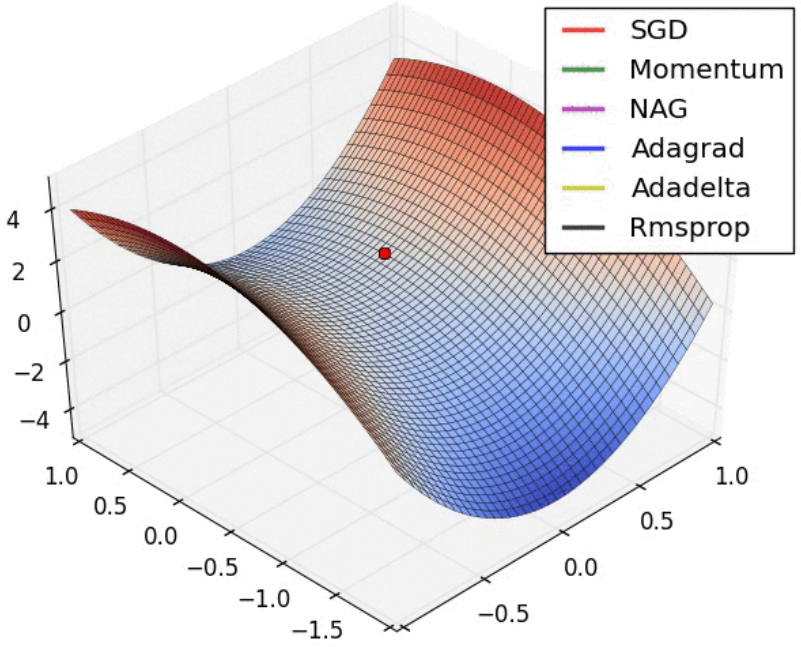
How Can DL Revive?

Progress in machine learning research

- Before 2006 training deep architectures was unsuccessful
- New methods for unsupervised pre-training have been developed (RBMs, autoencoders, contrastive estimation, etc)
- More efficient parameter estimation methods
- Better understanding of model regularization



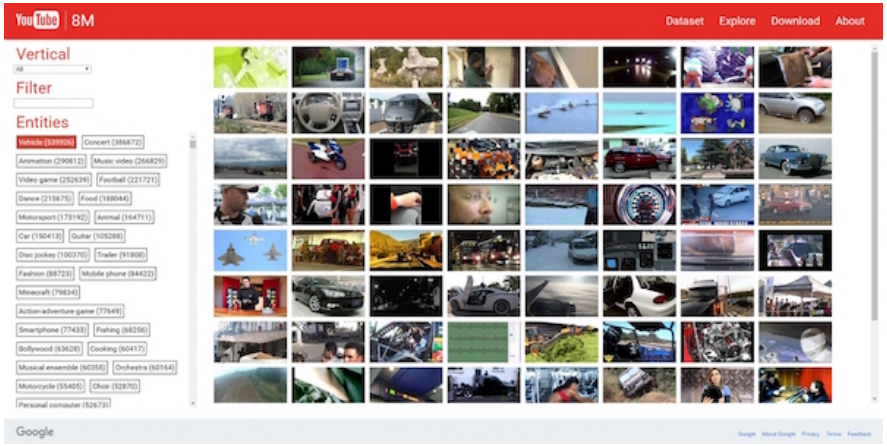
[Hinton et al. 2006]



[Honglak Lee]

How Can DL Revive?

Many training data available



Changes in computing technology favor deep learning

- Multi-core CPUs and GPUs
- Uniform parallel operations on dense vectors are faster

Open-Source Tools

- https://caffe2.ai/
- Based in C++, great Python interface

- http://www.tensorflow.org/
- Python open source software library by Google Brain team

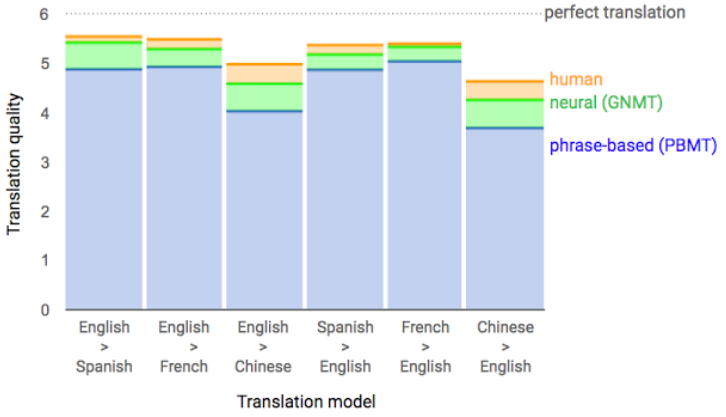
- http://torch.ch/
- Based in C/CUDA, support several script languages
- Strongly backed by Facebook

- https://mxnet.apache.org/
- Support multiple languages
- Apache open source

Limitation – 1. Need Many Clean Training Data

Machine translation is so successful. Then how are about the other NLP tasks?

- Google Neural Machine Translation system (GNMT) in 2016/09



Spell checking

- In Google News, 곱배기 (254 results) vs 곱빼기 (683)
- 외래어 표기법: 루이비통 vs 루이뷔통, 마를린 먼로 vs 메릴린 먼로

Sentiment analysis

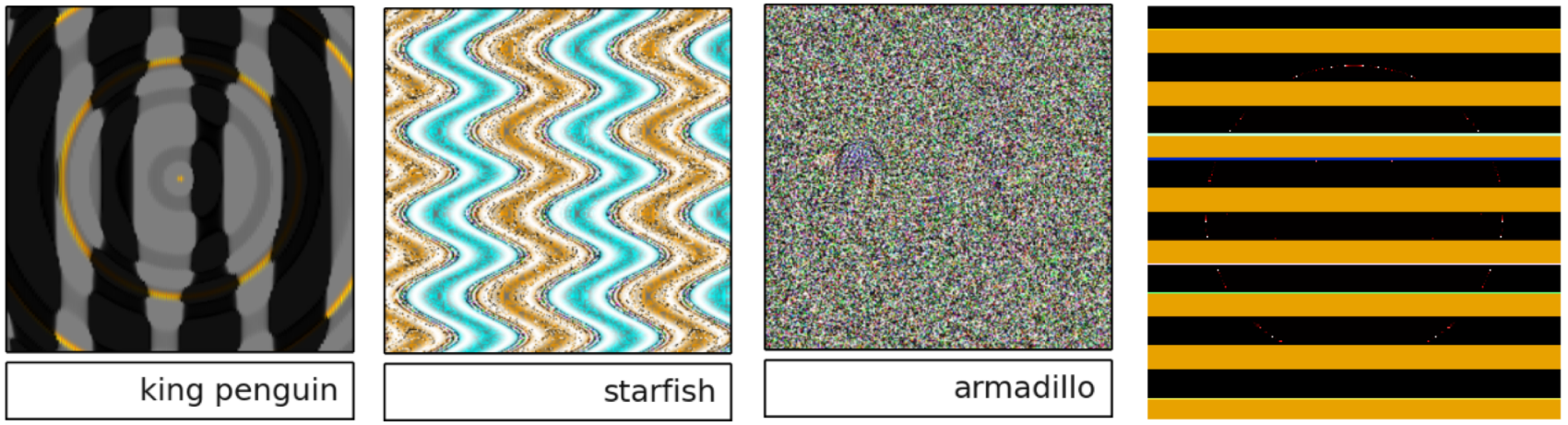
- Dorothy Parker on Katherine Hepburn:
  "*She runs the gamut of emotions from A to B*"

https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

Limitation – 2. Easily Break Down

Deep neural networks are easily fooled

- High confidence predictions for unrecognizable images
- State-of-the-art DNNs trained on ImageNet believe with ≥ 99.6% certainty to be a familiar object



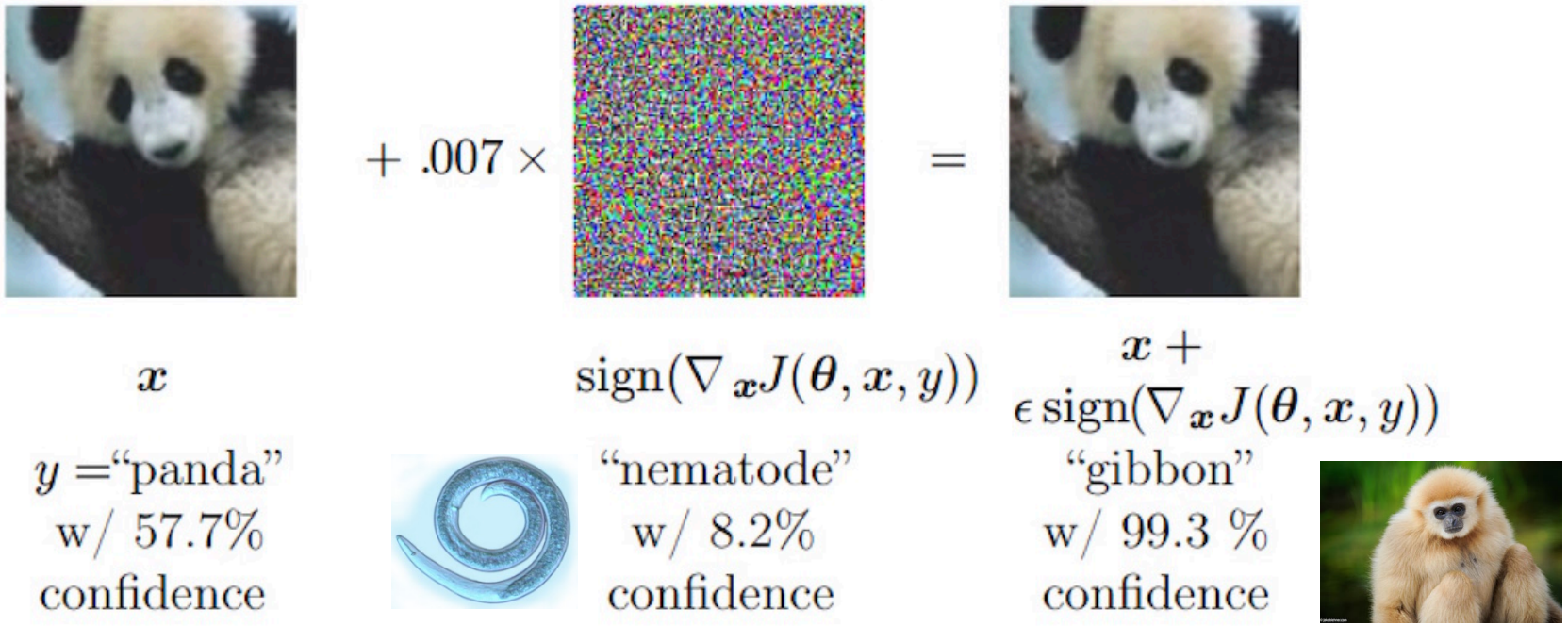
School bus!

[http://www.evolvingai.org/fooling]

Limitation – 2. Easily Break Down

Adversarial examples

- Human cannot tell the difference with the original example
- However, the network can make highly different predictions



[I. Goodfellow]

Limitation – 3. Not Energy Efficient

Not sustainable energy consumption in Nature

- Lee Sedol used about **20 Watts** of power to operate
- AlphaGo used approximately **1 MW** (200 W per CPU and 200 W per GPU)

50,000 times more!

AlphaGO	Lee Se-dol
1202 CPUs, 176 GPUs,	1 Human Brain,
100+ Scientists.	1 Coffee.

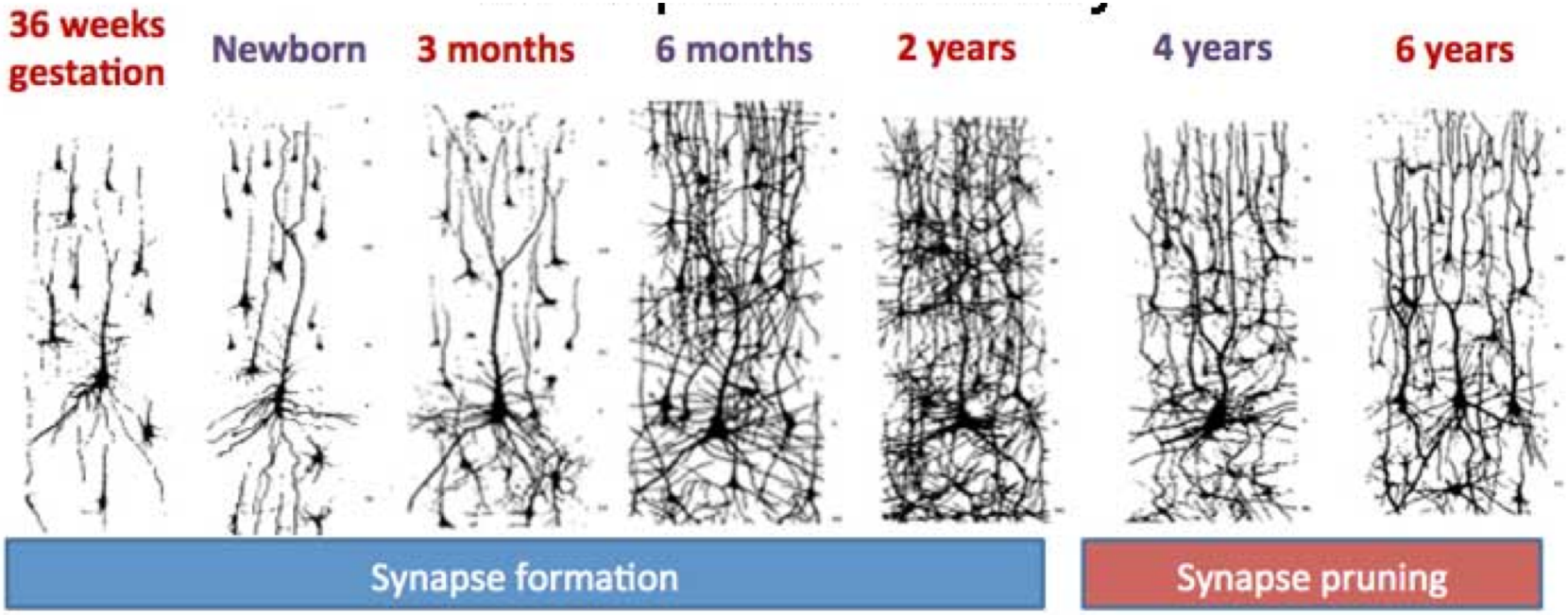
http://www.businessinsider.com/heres-how-much-computing-power-google-deepmind-needed-to-beat-lee-sedol-2016-3

Limitation – 3. Not Energy Efficient

Doing nothing is often the best action!

Developmental plasticity

- Each neuron in the cerebral cortex has approximately…
- 2,500 synapses at birth → 15,000 synapses → keep decreasing in our entire world



http://it.snhu.edu/naturalsciences/BIO320/chapter3/lectures/plasticity/printversion.htm

Limitation – 4. Lack of Semantic Information

Human can learn a new class even with a single image

- Suppose my kid knows jaguar, and leopard, and see a picture of cheetah for the first time



Does it have a **tail**?

Does it lay the **egg**?

How does its **foot** look like?

Generalization / Specialization

- First do categorization by finding commonality (it's a big cat)
- Then focus on its differences in the group (e.g. tear marks, patterns, ears, …)

Limitation – 4. Lack of Semantic Information

DL models require a large amount of training data

- Knowledge transfer is difficult
- Collect training data of a new class again…

1,000 images of Jaguar

1,000 images of Leopard

1,000 images of Cheetah

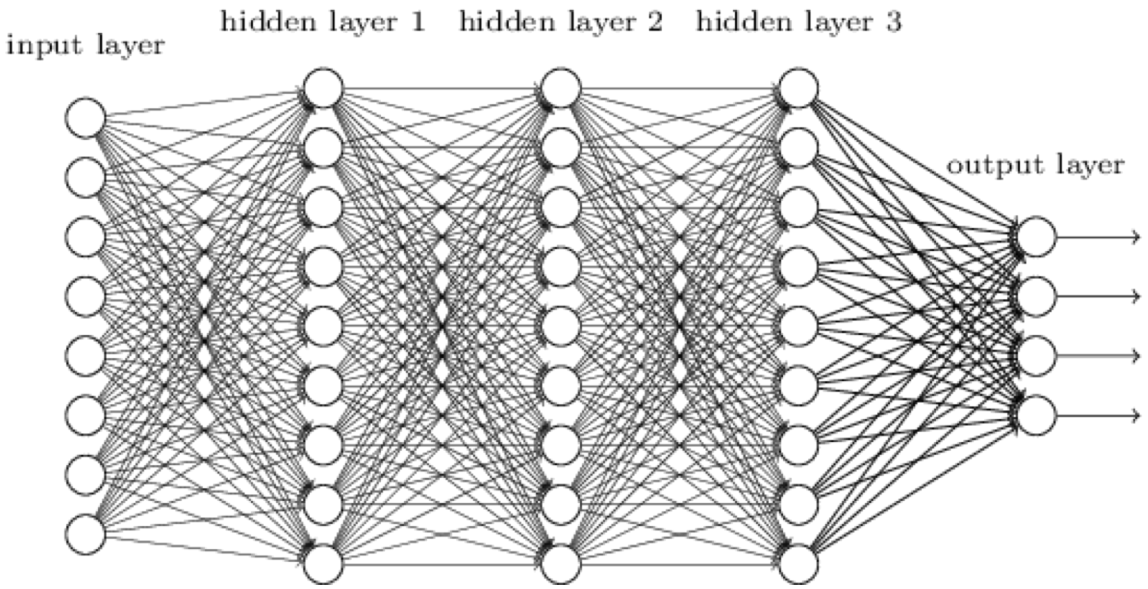
Promising research directions

- Zero-shot/one-shot learning, transfer learning, multi-task learning, semi-/unsupervised learning…

Limitation – 5. Interpretability/Explainability

Deep networks are widely regarded as black boxes but are often more accurate

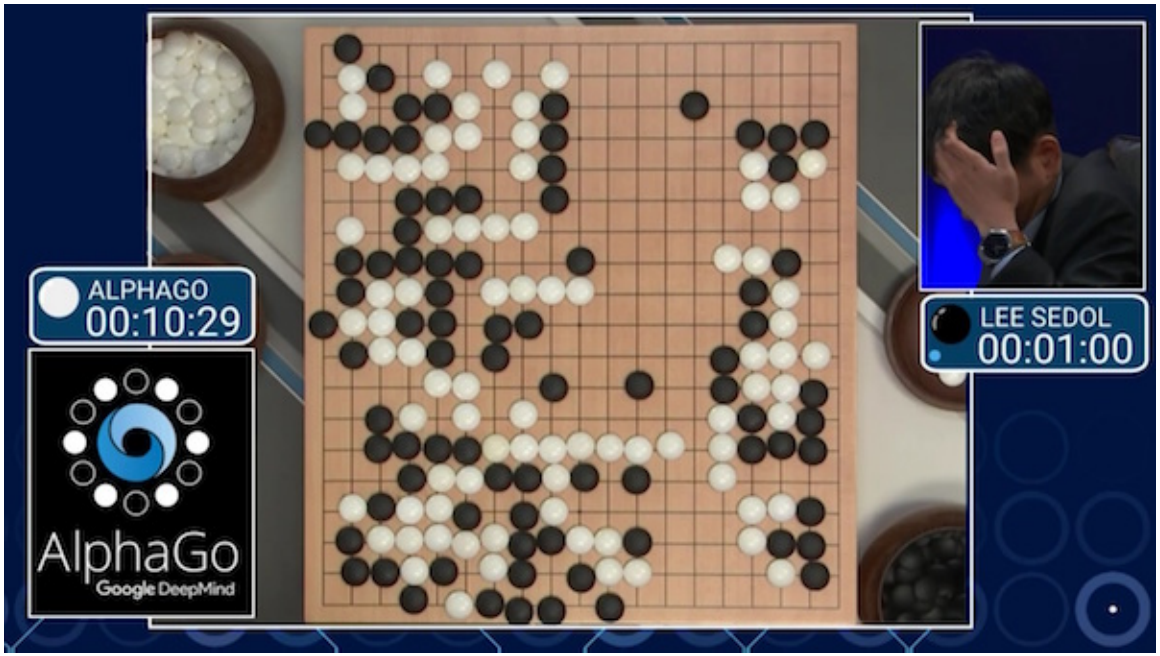
- State-of-the-art CNNs often include 10~100 millions of parameters to learn
- It is hard to know what happens inside the model



Limitation – 5. Interpretability/Explainability

Deep networks are very inferior to explain what they did

- Explainability-Accuracy trade-off
- Explainable AI should be essential; users are to understand, trust, and effectively manage



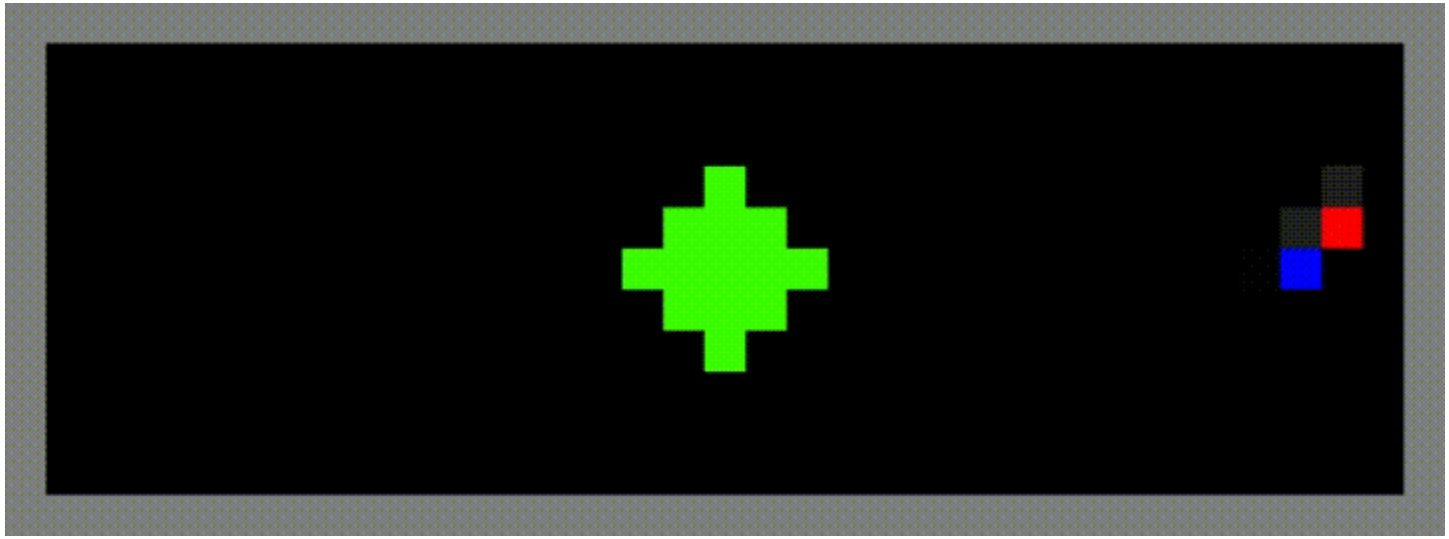
Why did you do that?
Why not something else?
When do you succeed?
When do you fail?
When can I trust you?
How do I correct an error?

Because it maximizes the winning possibility …

Limitation – 6. Emotional Intelligence for AI

Google DeepMind's Fruit gathering

- 두 agents 가 사과를 최대한 많이 모으는 게임
- 사과가 줄어들수록, 에이전트들은 서로 레이저빔을 쏘며 공격적으로 변함



Limitation – 6. Emotional Intelligence for AI

인공지능이 어떻게 정서를 처리하여 표현해야 하는지에 대한 연구가 현재 전무함